Technische Universität München

Fakultät für Mathematik
Lehrstuhl für Angewandte Numerische Analysis

Sparse Optimal Control of Multiagent Systems

Mattia Bongini

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Boris Vexler
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Massimo Fornasier
	2.	Univ.-Prof. Dr. Giuseppe Buttazzo Università di Pisa, Pisa, Italien
	3.	Prof. Dr. Benedetto Piccoli Rutgers University, Camden, USA (nur schriftliche Beurteilung)

Die Dissertation wurde am 21.04.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 11.07.2016 angenommen.

a Ilaria

Acknowledgements

This thesis could not have been possible without the guidance and support of my Ph.D. advisor, Prof. Dr. Massimo Fornasier, to whom I owe endless gratitude for introducing me into the fascinating topic of sparse control. While writing, I tried to follow the teachings of the departed Prof. Franco Montagna, who passed down to me the love for clear exposition. A special thank goes to Prof. Paolo Costantini, whose help was the first step on the path that led to this dissertation.

My research was supported by the ERC-Starting Grant "High-Dimensional Sparse Optimal Control" (HDSPCONTR – 306274) and by the Technische Universität München, which I both gratefully acknowledge. This thesis stem from various collaborations with Emiliano Cristiani, Markus Hansen, Dante Kalise, Oliver Junge, Mauro Maggioni, Francesco Rossi, and Benjamin Scharf: I am thankful to all of them for the invaluable suggestions they gave me. Francesco Solombrino and Giacomo Albi deserve a special mention for the epic moments and the endless amount of coffee we shared together.

During these years, I could not have worked so steadily and serenely without extraordinary companions at my side: my fellow doctoral colleagues of M15, wonderful mates for such a long journey; my friends in München, who always make me forget I left home; my friends in Siena, who always make me feel like I never left.

Finally, my deepest gratitude goes to my family: my grandparents Fernanda, Bruna, Dino, and Guido, who never let me feel hunger; my parents Umberto and Monica, who are my №1 supporters; my brother Tommaso, of whom I am the №1 supporter; my love Ilaria, to whom I dedicate this thesis and with whom I share my life.

Contents

Introduction

Later began the convoys of vehicles bringing food to New York [...]. *In some squares great detours had to be made because of the great pressure from the side streets, so the entire journey stopped and traveled step by step, then it all came on again, so that for a short while everything rushed by at lightning speed until as if by a single brake a slowdown took over and quieted everything again.*

— Franz Kafka, *Amerika*

In the novel *Amerika* by Franz Kafka, the main character Karl Rossmann is amazed to see how the massive traffic of vehicles flowing towards New York behaves like a single, worm-like creature, as if the synchrony of their motion were the result of divine directives to humanity. It is, indeed, a rather recent discovery, dating to the seminal work of Lorenz [138] and Mandelbrot [139], that complicated global patterns in nature may emerge from very simple reiterated local rules. Since then, numerous studies have confirmed that, at any scale of observation, complexity is often an emergent feature of large systems with elementary interactions: from the intricate interplays at cellular level [47], passing through the hierarchical structure of ants' societies [72], all the way to the myriads of religious and social customs of human beings [108]. The principles of self-organization in multiagent systems are employed in swarm robotics to produce cheap, resilient, and efficient squadrons to perform predefined tasks [16] and to render battles, swarms of animals and hair/fur textures in CGI animations [129, 154, 164]. The same concepts lie at the core of the *Game of Life*, a zero-player game whose capability to reproduce infinite self-replicating patterns from simple evolution rules makes it Turing complete [28]. The scientific literature on multiagent self-organizing systems is vast and ever-growing: the interested reader may be addressed to [19, 174] and references therein for further insights on the topic.

In all the examples above, self-organization is the result of the superimposition of binary interactions between agents amplified by a steadfast feedback loop. This reinforcement process is necessary to give momentum to the multitude of feeble local interactions and eventually let a global pattern appear. Typically, the strength of such interaction forces is a function of the "social distance" between agents: for instance, birds align with their closest neighbors [20] and people agree easier with those who already conform to their beliefs [122]. Some of the forces of the system may be of cohesive type, i.e., they tend to reduce the distance between agents: whenever cohesive forces have a comparable strength at short and long range, we call these systems *heterophilious*; if, instead, there is a long-range bias we speak of *homophilious* societies [146].

Figure 1: On the left: pilgrims walking around the Kaaba in Mecca[1]. In the center: ants coordinate to create a bridge[2]. On the right: a school of fish[3].

Heterophilious systems have a natural tendency to keep the trajectories of the agents inside a compact region, and therefore to exhibit stable asymptotic profiles, modeling the autonomous emergence of global patterns. On the other hand, self-organization in homophilious societies can be accomplished only conditionally to sufficiently high levels of initial coherence that allow the cohesive forces to keep the dynamics compact [143]. Being such systems ubiquitous in real life (e.g., see [85, 135]), it is legitimate to ask whether – in case of lost cohesion – external controls with limited strength acting on the agents of the system may restore stability and achieve pattern formation.

In the case of multiagent systems, efficient control strategies should target only few individuals of the population, instead of squandering resources on the entire group at once: taking advantage of the mutual dependencies between the agents, they should trigger a ripple effect that would spread their influence to the whole system, thus indirectly controlling the rest of the agents. The property of control strategies to target only a small fraction of the total population is known in the mathematical literature as *sparsity* [48, 96, 105]. The fundamental issue is the selection of the few agents to control: an effective criterion is to choose them as to maximize the decay rate of some Lyapunov functional associated to the stability of the desired pattern [67]. In alignment-dominated models, the sparse control targeting at each instant only the agent farthest away from the mean velocity was shown to possess this property, see [49]. The stabilizing power of sparse controls goes beyond this particular class of systems: in [32, 33] similar results for attraction-repulsion models were obtained, while in [158] sparse controls for linear systems were studied.

Alongside the above centralized approach to control (in the sense that there is an external figure implementing the control on the system), also decentralized strategies may be considered: these consist in assuming that each agent, besides being subjected to forces induced in a feedback manner by the rest of the population, follows an individual strategy to coordinate with the other agents [37]. Whenever such additional force is the result of an offline optimization among perfectly informed players, we fall into the realm of

[1] `https://commons.wikimedia.org/wiki/File:Islamm.jpg`
[2] `https://commons.wikimedia.org/wiki/File:Ant_bridge.jpg`
[3] `https://commons.wikimedia.org/wiki/File:Large_fish_school.png`

Game Theory [147, 177]. Games without an external regulator model situations where it is assumed that an automatic tendency to reach “correct” equilibria exists, like the stock market. However, this optimistic view of the dynamics is often frustrated by evidences of the convergence to suboptimal configurations [120], whence the need of an external figure controlling the evolution of the system. The action of the regulator can eventually be optimized too, choosing it as the minimizer of a given cost functional: in this case, sparse controls can be obtained when the control cost is the L^1-norm [49].

Figure 2: On the left: swarmbots performing a task[4]. In the center: Gosper's Glider Gun, a recurring pattern in the Game of Life[5]. On the right: multiagent systems employed in the videogame *Total War: Rome 2*[6].

Sometimes, it could be advantageous to control a different population (usually called the *leaders*) rather than directly interacting with the original one (the *followers*). In this case, followers are controlled indirectly through their interaction with leaders. One way to introduce sparsity in such context of *indirect control* is to require that the population of leaders is numerically inferior to the one of followers. This control setting is, actually, incredibly common in real life: for instance, to influence public opinion, the control of few key mass media is often sufficient [125]. Another example is provided by certain kind of tumors, which spawn trailblazing leader cells with the ability to lead the metastasis into surrounding tissues: if diverted, the migration process of the tumor could possibly be halted [59]. Similarly, the recent work [5] has shown the effectiveness of few undercover coordinated agents in the evacuation of a crowd from an unfamiliar environment.

Unfortunately, the computation and implementation of (direct and indirect, optimal and feedback) control strategies is affected by the so called *curse of dimensionality*: roughly speaking, algorithms to find solutions to such problems become inefficient as the number of agents N or their dimensionality d increase conspicuously [24]. In the case of N large, one solution to solve this issue is to combine mean-field approximation techniques (like the mean-field limit [31, 130], the BBGKY hierarchy [58, 116], and the binary interaction approximation [7, 151]) with the optimal control problem. *Mean-field games* stem from the application of such combination to games involving a large number of players [128, 136, 149], while its application to optimal control problems on multiagent systems leads to the *mean-field optimal control theory* [38, 103, 106]. For what concerns large values of d, recent approaches of a certain relevance to the dimensionality reduction of dynamic data are diffusion maps [68], geometric multiscale reductions [40], and Johnson-Lindenstrauss embeddings [101, 131]. Despite the large amount of work devoted to these issues and their

[4] `http://people.idsia.ch/~gianni/SwarmRobotics/swarm_navigation/start-double-chain.jpg`
[5] `https://commons.wikimedia.org/wiki/File:Gospers_glider_gun.gif`
[6] `http://forums.totalwar.org/vb/showthread.php?148830-Ostrogoth-Campaign`

relevance in many practical situations, the application of the above strategies to control problems on multiagent systems has only been investigated recently in [36].

Figure 3: On the left: fish evade predators[7]. In the center: confinement of crowds using infrastructures[8]. On the right: a cop overseeing sheep[9].

The present thesis is devoted to the study of sparse control and its dimensionality reduction for multiagent systems. Since in control problems an accurate prior knowledge of the dynamics is essential, we also address the automatic learning of the fundamental functions governing the dynamics from sampled observations of the evolutions. In some parts of this work we consider general multiagent systems, while in others we focus on specific models in order to assess quantitatively the effectiveness of sparse controls with respect to different control strategies.

The thesis is organized in seven chapters,each focusing on a specific topic mentioned above.

- Chapter 1 presents general results for the convergence at large times of alignment models to fully aligned states, as a form of consensus and more in general of self-organization. We show that the ability of such systems to converge to consensus is limited, even if we allow for feedback decentralized control strategies. We prove that consensus emergence can be guaranteed in only very few, specific cases, one of which being the situation where agents are highly heterophilious. The main results of this chapter were published in the paper [37] and are integrated with comments and observations from [49, 121].

- Motivated by the limitations of homophilious societies to self-organize, in Chapter 2 we introduce the concept of (direct) sparse control of multiagent systems. We show the effectiveness of such control strategy in the case of two prototypical self-organizing models: the Cucker-Smale system [81] and the Cucker-Dong system [80]. While sparse controls seem intuitively capable of steering multiagent dynamics, it is however surprising that, for some classes of systems, they are also the most effective control strategies: for such systems, the control of the instantaneous leaders of the dynamics is more convenient than controlling simultaneously all agents. The first significant contribution of this thesis is that the optimality of sparse controls (initially proven for the Cucker-Smale model, which is a dissipative system) can be extended also to non-dissipative systems with singular repulsive interaction forces, like the Cucker-Dong model. The existence of solutions of sparsely controlled multiagent systems is also studied. The chapter is based on the papers [32, 33], again supplemented by results from [49].

[7] `https://commons.wikimedia.org/wiki/File:Flocking.jpg`

[8] `https://pixabay.com/en/cathedral-square-ulm-human-crowds-592756/`

[9] `https://www.flickr.com/photos/31363949@N02/6707219407/sizes/l`

- In Chapter 3 we investigate the possibility of defining sparse control strategies in high dimension on the basis of the information obtained from the random projection of the original system to lower dimension. Our aim is to validate the following

  Meta-theorem. *For high dimensional multiagent systems, governed by smooth nonlinearities depending on mutual distances between the agents, one can construct low-dimensional representations of the dynamical system, which allow the computation of nearly optimal control strategies in high dimension with overwhelming confidence.*

  The above statement implies that, for a dimensionality d large, the curse of dimensionality for the implementation of efficient controls for multiagent systems can be circumvented. Since control strategies are in general highly dependent on the specific dynamical system to which they are applied, we shall prove a specific instance of the above statement, which nonetheless tackles all the relevant technical issues potentially lurking in other situations: we shall focus on the sparse control of the Cucker-Smale model, for which we have already available the tools developed in Chapter 2. Numerical tests corroborate the theoretical findings. The results shown here were first obtained in [36].

- Chapter 4 provides two complexity reduction techniques in the case that the number of agents N makes the direct numerical treatment of multiagent systems unfeasible. The results on the *mean-field limit* and the *grazing interaction limit* are personal revisions of well-known work [46, 95, 151] and are reported for the sake of completeness, being them central in the subsequent chapters. Several sections of the papers [5, 34, 38] provided the basis for this chapter.

- In Chapter 5 we show how to infer the interaction mechanism of a dynamical system from the observation of its evolution. By means of an approach that blends together the mean-field limit and a Γ-convergence argument, we illustrate that we can reconstruct with high accuracy the interaction functions of a large class of first-order multiagent systems. Such learning problem is remarkably non-standard due to the need to manipulate highly dependent dynamic data (the trajectories of the agents) to perform the reconstruction of the function, see [83]. Possible extensions to second-order models are also discussed. These results were collected in [34].

- In Chapter 6 we study dimensionality reduction techniques for direct and indirect optimal control problems for multiagent systems. Roughly speaking, when the number of agents to control N becomes very large, to compute the optimal control strategy we have to deal with a numerically intractable number of ODE constraints. One way to circumvent the curse of dimensionality in this case is to compute the mean-field limit PDE of the ODEs system and solve the optimal control problem with such PDE as unique constraint, in place of the ODEs. However, in principle, it is not clear whether the optimal control strategies we obtain with this procedure are related to the ones of the original problem or not. In this chapter we prove such connection by showing that, under certain regularity assumptions, the PDE-constrained optimal control problem is the Γ-limit of the ODE constrained problems as the number of agents N tends to ∞. First-order optimality conditions for the limit problem are also derived. The chapter's core is the paper [38] together with [103, 106].

- We conclude the thesis with Chapter 7, which presents an indirect sparse control strategy for the efficient evacuation of pedestrians from unknown environments under limited visibility. We propose a model for pedestrians composed of an exploration and an evacuation phase: being the exit visible only from specific locations, when agents do not see it, they explore the environment with a random walk coupled with *topological alignment* (in order to remain close to their nearest neighbors). Instead, whenever they see the exit, they perform a sharp motion towards it. If left free to evolve, pedestrians split into several subgroups, each determined to find the way out, but evacuation is not guaranteed. The introduction of few *invisible* leaders (in the sense that pedestrians do not recognize them as special), whose dynamics is the result of an optimization procedure, is shown to significantly improve the evacuation process. To reduce the computational effort of finding optimal strategies in the case where the number of pedestrians is very large, we derive a mean-field equation for the pedestrians using the grazing interaction limit procedure. Numerical simulations and a real experiment with people show the efficiency of the proposed control technique. This chapter is based on [5].

In accordance with the latest regulations concerning the authorship of the results obtained in Ph.D. theses, I hereby declare that the content of the following sections are original and were obtained thanks to my contribution (the symbol – means that also sections in between are considered): 1.3–1.6, 2.2–2.6, 3.2–3.7, 4.4.2, 4.4.3, 5.1.5, 5.4, 5.5, 6.2–6.4, 7.1–7.4, 7.7. The remaining sections are re-elaborated versions of results that can be found in the literature and that are necessary to make the thesis self-contained.

The pictures shown in this Introduction were licensed under a Creative Commons license, as can be verified by following the related web addresses. No changes were made to them and no authorship is claimed. The pictures in the rest of the work were either autonomously produced or the outcome of the scientific partnership with the co-authors of the papers [5, 32–34, 36, 37].

CHAPTER 1

Consensus emergence and decentralized feedback control

We start from the analysis of general properties of *alignment models*, i.e., systems where imitation is the dominant feedback mechanism. Such systems are ubiquitous in nature since several species are able to interpret and instinctively reproduce certain manoeuvres that they perceive (e.g., fleeing from a danger, searching for food, performing defense tactics, etc.), see [92]. We first derive sufficient conditions for their self-organization, which shows that only a certain balance between the strength of the alignment forces and the initial coherence of the group is able to guarantee the formation of a *consensus*, interpreted in this case as a state where all agents are aligned. We then introduce decentralized feedback controls to facilitate the emergence of such patterns. The results reported in this chapter will make clear how prone to lose their initial coherence these systems are, and how much an *external* control to help them organize is needed.

1.1 Self-organization in dynamical communication networks

In many natural and social phenomena, a group of agents faces the problem of coordinating on the basis of mutual communication. The modeling of such scenarios has to comply with the nature of protocols and customs governing the interaction among agents, limited or unreliable information transmission, and changing interaction topologies. Despite the above complications, networks of agents with oriented information flow under possible link failure or creation can be effectively represented by means of directed graphs with edges possibly switching in time.

A *directed graph* G on a set of nodes $A_1, \ldots, A_N$ is any subset of $\{A_1, \ldots, A_N\}^2$. Each pair $(A, B) \in G$ is called an *edge from A to B*, and a *directed path from A to B* in G is a sequence of edges $(A, A_{i_1}), (A_{i_1}, A_{i_2}), \ldots, (A_{i_k}, B) \in G$. The graph G is said to be *strongly* *connected* if for any pair A, B of distinct nodes there is a directed path from A to B and a directed path from B to A.

When studying under which conditions networks of agents are able to self-organize, it is usually not enough to know if two nodes are connected: the strength of the interaction between them also matters. Hence, given a system of $N \in \mathbb{N}$ agents, for each pair of agents $i, j = 1, \dots, N$ we denote by $g_{ij}(t) \in \mathbb{R}_+$ the weight of the link connecting i with j: clearly, if $g_{ij}(t) = 0$, i is not connected to j at time t. The value $g_{ij}(t)$ can be seen as the relative intensity of the information exchange flowing from agent i to agent j at time $t \geq 0$. We shall assume for the moment that each weight function $g_{ij} : \mathbb{R}_+ \to \mathbb{R}_+$ is piecewise continuous.

The weights $g_{ij}(\cdot)$ naturally induce a directed graph structure on the set of agents: we define, for any $\varepsilon \geq 0$ and $t \geq 0$, the graph $G_\varepsilon(t)$ as

$$G_\varepsilon(t) \triangleq \left\{(i,j) \in \{1,\dots,N\}^2 : g_{ij}(t) > \varepsilon\right\}.$$

The *adjacency matrix* $G_0(t)$ is the set of pairs (i,j) for which the communication channel from i to j is active at time t.

As a prototypical example of a multiagent system and to quantitatively illustrate the concept of self-organization, we introduce *alignment models*: if we denote by $\{v_1, \dots, v_N\} \subset \mathbb{R}^d$ the states of the N agents of our systems, then the instantaneous evolution of the state $v_i(t)$ of agent i at time t is given by

$$\dot{v}_i(t) = \sum_{j=1}^{N} g_{ij}(t)(v_j(t) - v_i(t)), \quad i = 1, \dots, N. \tag{1.1}$$

The meaning of the above system of differential equations is the following: at each instant $t \geq 0$, the state $v_i(t)$ of agent i tends to the state $v_j(t)$ of agent j with a speed that depends on the strength of the information exchange $g_{ij}(t)$. Since (1.1) is a system of ODEs with possibly discontinuous coefficients, we need for it a proper notion of solution.

Definition 1.1. Let $\{I_k\}_{k\in\mathbb{N}}$ denote a countable family of open intervals such that all the functions g_{ij} are continuous on every I_k and $\cup_{k\in\mathbb{N}}\overline{I}_k = \mathbb{R}_+$. Given $v^0 = (v_1^0, \dots, v_N^0) \in \mathbb{R}^{dN}$, we say that the curve $v = (v_1, \dots, v_N) : \mathbb{R}_+ \to \mathbb{R}^{dN}$ is a *solution* of (1.1) with initial datum v^0 if

(*i*) $v(0) = v^0$;

(*ii*) for every $i = 1, \dots, N$ and $k \in \mathbb{N}$, v_i satisfies (1.1) on I_k.

The notion of self-organization that we are considering for system (1.1) is that of *consensus* or *flocking*, which is the situation where the state variables of the agents asymptotically coincide.

Definition 1.2 (Consensus for system (1.1))**.** Let $v : \mathbb{R}_+ \to \mathbb{R}^{dN}$ denote a solution of (1.1) with initial datum v^0. We say that $v(\cdot)$ *converges to consensus* if there exists a $v^\infty \in \mathbb{R}^d$ such that, for every $i = 1, \dots, N$, it holds

$$\lim_{t\to+\infty} \|v_i(t) - v^\infty\|_{\ell_2^d} = 0.$$

The value v^∞ is called the *consensus state*.

In the definition above, $\|\cdot\|_{\ell_2^d}$ stands for the Euclidean norm on $\mathbb{R}^d$. The subscript ℓ_2^d shall often be omitted whenever clear from context.
Roughly speaking, a system of agents satisfying (1.1) converges to consensus regardless of the initial condition v^0 provided that the underlying communication graph is "sufficiently connected". With this we mean that each node must possess, over some dense collection of time intervals, a strong enough communication path to every other node in the network. This intuitive idea is made precise in the following result, whose proof can be found in [121]. A similar answer for discrete-time system was also provided in [145]

Theorem 1.3. *Let $v : \mathbb{R}_+ \to \mathbb{R}^{dN}$ be a solution of (1.1) with initial datum v^0. Suppose that there exists an $\varepsilon > 0$ and a strongly connected directed graph G on the set of agents on which the system spends an infinite amount of time, i.e.,*

$$\mathcal{L}^1\left(\{t \geq 0 : G_\varepsilon(t) = G\}\right) = +\infty.$$

Then $v(\cdot)$ converges to consensus with consensus state v^∞ belonging to the convex hull of $\{v_1^0, \ldots, v_N^0\}$.

The above result also says that, without further hypotheses on the interaction weights g_{ij}, the value of v^∞ is rather an emergent property of the global dynamics of system (1.1) than a mere function of the initial datum v^0. However, it is relatively simple to identify assumptions on g_{ij} for which the latter is true. For example, from a trivial computation follows

$$\frac{1}{N}\sum_{i=1}^N \dot{v}_i(t) = \frac{1}{N}\sum_{j=1}^N \left(\sum_{i=1}^N g_{ij}(t) - \sum_{i=1}^N g_{ji}(t)\right) v_j(t).$$

Hence, if for every $t \geq 0$ the weight matrix $(g_{ij}(t))_{i,j=1}^N$ has the property that $\sum_{i=1}^N g_{ij}(t) = \sum_{i=1}^N g_{ji}(t)$ for every $j = 1, \ldots, N$, then the average

$$\overline{v}(t) \triangleq \frac{1}{N}\sum_{i=1}^N v_i(t) \tag{1.2}$$

is an invariant of the dynamics. This implies that

$$v^\infty = \frac{1}{N}\sum_{i=1}^N v_i^0$$

holds, i.e., the consensus state is only a function of the initial datum v^0.

1.2 Consensus emergence in alignment models

In this section we shall see that the assumptions of Theorem 1.3 can actually be very restrictive and seldom met when dealing with specific instances of alignment models.

1.2.1 Some classic examples of alignment models

A general principle in opinion formation is the *conformity bias*, i.e., agents weight more opinions that already conform to their beliefs. This can, actually, be extended to coordination in general, since intuitively it is easier to coordinate with "near" agents than "far away" ones. Formally, this is equivalent to asking that the weights g_{ij} are a nonincreasing function of the distance between the states of the agents, i.e.,

$$g_{ij}(t) = a(\|v_i(t) - v_j(t)\|), \tag{1.3}$$

where $a : \mathbb{R}_+ \to \mathbb{R}_+$ is a nonincreasing *interaction kernel.* Notice that (1.3) trivially implies the invariance of the mean $\overline{v}$ (given by (1.2)), and that $v^\infty = \overline{v}(0)$, if it exists.

Several classic opinion formation models combine conformity bias with alignment. In the DW model, see [180], two random agents i and j update their opinions v_i and v_j to $1/2(v_i + v_j)$, provided they originally satisfy $\|v_i - v_j\| \leq R$, where $R > 0$ is fixed *a priori.* Instead, in the popular *bounded confidence* model of Hegselmann and Krause [122], opinions evolves according to the dynamics (1.1) where the weights have the form (1.3) with

$$a(r) = \chi_{[0,R]}(r) \triangleq \begin{cases} 1 & \text{if } r \in [0, R], \\ 0 & \text{otherwise,} \end{cases}$$

for some fixed *confidence radius* $R > 0$. The dynamics is thus given by the system of ODEs

$$\dot{v}_i(t) = \sum_{j=1}^{N} \chi_{[0,R]}(\|v_i(t) - v_j(t)\|)(v_j(t) - v_i(t)), \quad i = 1, \ldots, N. \tag{1.4}$$

It is straightforward to design an instance of this model not fulfilling the hypothesis of Theorem 1.3. Indeed, consider a group of $N = 2$ agents in dimension $d = 1$ with initial conditions $v_1(0) = -R$ and $v_2(0) = R$. Since $g_{12}(0) = g_{21}(0) = 0$, it follows that $G_\varepsilon(t) = \emptyset$ for all $t \geq 0$ and for all $\varepsilon \geq 0$.

Second-order models are necessary whenever we want to describe the dynamics of physical agents, like flocks of birds, herds of quadrupeds, schools of fish, and colonies of bacteria, where individuals are considered aligned whenever they move in the same direction, regardless of their position. Since in such cases it is necessary to perceive the velocities of the others in order to align, to describe the motion of the agents we need the pair position-velocity (x, v), but this time only the velocity variable v is the *consensus parameter.*

One of the first of such models, named *Vicsek's model* in honor of one of its fathers, was introduced in [173]. Very much in the spirit of (1.4), it postulates that the evolution of the spatial coordinate x_i and the orientation $\theta_i \in [0, 2\pi]$ in the plane $\mathbb{R}^2$ of the i-th agent follow the law of motion given by

$$\begin{cases} \dot{x}_i(t) = v_i(t) = \hat{v} \begin{pmatrix} \cos(\theta_i(t)) \\ \sin(\theta_i(t)) \end{pmatrix}, \\ \dot{\theta}_i(t) = \dfrac{1}{|\Lambda_R(t, i)|} \displaystyle\sum_{j=1}^{N} \chi_{[0,R]} \left(\|x_i(t) - x_j(t)\|\right) \left(\theta_j(t) - \theta_i(t)\right), \end{cases} \quad i = 1, \ldots, N, \tag{1.5}$$

where $\hat{v} > 0$ denotes the modulus of $v_i(t)$, $\Lambda_R(t,i) \triangleq \{j \in \{1,\ldots,N\} : \|x_i(t) - x_j(t)\| \leq R\}$ and $|\Lambda_R(t,i)|$ stands for its cardinality.

In this model, the orientation of the consensus parameter v_i is adjusted with respect to the other agents according to a weighted average of the differences $\theta_j - \theta_i$. The influence of the j-th agent on the dynamics of the i-th one is a function of the (physical or social) distance between the two agents: if this distance is less than R, the agents interact by appearing in the computation of the respective future orientation.

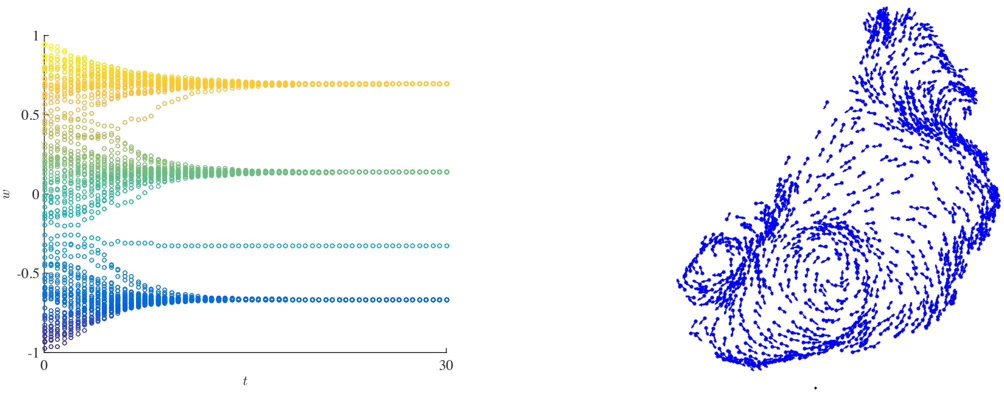


Figure 1.1: On the left: a typical evolution of the Hegselmann-Krause model. On the right: mill patterns in the Vicsek model. (Kind courtesy of G. Albi)

In [81], the authors proposed a possible extension of system (1.5) to dimensions $d > 2$ as follows

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{|\Lambda_R(t,i)|} \displaystyle\sum_{j=1}^{N} \chi_{[0,R]}\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right), \end{cases} \qquad i = 1,\ldots,N.$$

The substitution of the function $\chi_{[0,R]}$ with a strictly positive *kernel* $a : \mathbb{R}_+ \to \mathbb{R}_+$ let us drop the highly irregular and nonsymmetric normalizing factor $|\Lambda_R(t,i)|$ in favor of a simple N, and leads to the system

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N} \displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right), \end{cases} \qquad i = 1,\ldots,N. \tag{1.6}$$

Notice that the equation governing the evolution of v_i has the same form as (1.1), and since now the weights g_{ij} are symmetric (i.e., $g_{ij} = g_{ji}$ for all $i,j = 1,\ldots,N$) then $\overline{v}$ is a conserved quantity.

An example of a system of the form (1.6) is the influential model of Cucker and Smale, introduced in [81], in which the function a is

$$a(r) \triangleq \frac{H}{(\sigma^2 + r^2)^\beta}, \tag{1.7}$$

where $H > 0$, $\sigma > 0$, and $\beta \geq 0$ are constants accounting for the social properties of the group. Systems like (1.6) are usually referred to as *Cucker-Smale systems* due to the influence of their work, as can be witnessed by the wealth of literature focusing on their model, see for instance [4, 53, 87, 115, 153, 166].

In the rest of the chapter, we focus on consensus emergence for system (1.6). We shall consider a kernel $a : \mathbb{R}_+ \to \mathbb{R}_+$ which is decreasing, strictly positive, bounded and Lipschitz continuous.

As already noticed, in second-order models alignment means that all agents move with the same velocity, but *not necessarily* are in the same position. Therefore, Definition 1.2 of consensus applies here on the v_i variables only.

Definition 1.4 (Consensus for system (1.6))**.** We say that a solution

$$(x, v) = (x_1, \dots, x_N, v_1, \dots, v_N) : \mathbb{R}_+ \to \mathbb{R}^{2dN}$$

of system (1.6) *tends to consensus* if the consensus parameter vectors v_i tend to the mean $\overline{v}$, i.e.,

$$\lim_{t \to +\infty} \|v_i(t) - \overline{v}(t)\| = 0 \quad \text{for every } i = 1, \dots, N.$$

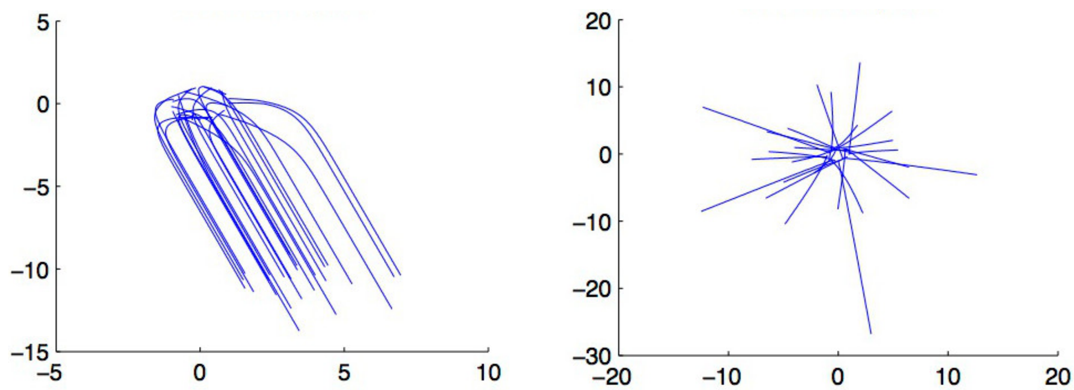


Figure 1.2: Consensus behavior of a Cucker-Smale system. On the left: agents align with the mean velocity. On the right: agents fail to reach consensus.

The following result is an easy corollary of Theorem 1.3.

Corollary 1.5. *Let* $(x(\cdot), v(\cdot))$ *be a solution of system* (1.6)*, where the interaction kernel* a *is decreasing and strictly positive. Suppose that there exists* $R > 0$ *for which it holds*

$$\mathcal{L}^1\left(\{t \geq 0 : \|x_i(t) - x_j(t)\| \leq R \textit{ for all } i, j = 1, \dots, N\}\right) = +\infty.$$

Then $(x(\cdot), v(\cdot))$ *converges to consensus.*

Proof. Since a is decreasing and strictly positive, from the initial assumptions follows

$$g_{ij}(t) = \frac{1}{N} a(\|x_i(t) - x_j(t)\|) \geq \frac{a(R)}{N} > 0,$$

for every $t \geq 0$ for which $\|x_i(t) - x_j(t)\| \leq R$ holds for every $i, j = 1, \ldots, N$. Therefore, the condition $\|x_i(t) - x_j(t)\| \leq R$ for every $i, j = 1, \ldots, N$ implies

$$G_{\frac{a(R)}{N}}(t) = \{1, \ldots, N\}^2,$$

which yields

$$\mathcal{L}^1\left(\left\{t \geq 0 : G_{\frac{a(R)}{N}}(t) = \{1, \ldots, N\}^2\right\}\right) = +\infty,$$

The statement then follows from Theorem 1.3 for the choice $\varepsilon = a(R)/N$. □

Unfortunately, the result above has the serious flaw that it cannot be invoked directly to infer convergence to consensus, since establishing a uniform bound in time for the distances of the agents is very difficult, even for smooth kernels like (1.7). Intuitively, consider the case where the interaction strength is too weak and the agents too dispersed in space to let the velocities v_i align. In this case, nothing prevents the distances $\|x_i - x_j\|$ to grow indefinitely, violating the hypothesis of Corollary 1.5. Hence, in order to obtain more satisfactory consensus results, we need to follow approaches that take into account the extra information at our disposal, which are the strength of the interaction and the initial configuration of the system.

Originally, this problem was studied in [81, 82] borrowing several tools from Spectral Graph Theory, see as a reference [61]. Indeed, system (1.6) can be rewritten in the following compact form

$$\begin{cases} \dot{x}(t) = v(t), \\ \dot{v}(t) = L(x(t))v(t), \end{cases} \tag{1.8}$$

where $L(x(t))$ is the Laplacian[1] of the matrix $\left(a\left(\|x_i(t) - x_j(t)\|\right)/N\right)_{i,j=1}^N$, which is a function of $x(t)$. Being the Laplacian of a nonnegative, symmetric matrix, $L(x(t))$ encodes plenty of information regarding the adjacency matrix $G_0(t)$ of the system, see [144]. In particular, the second smallest eigenvalue $\lambda_2(t)$ of $L(x(t))$, called the *Fiedler's number* of $G_0(t)$ is deeply linked with consensus emergence: provided that a sufficiently strong bound from below of $\lambda_2(t)$ is available, the system converges to consensus.

To establish under which conditions we have convergence to consensus, we shall follow a different approach. The advantage of it is that it can be employed also to study the issue of the controllability of several multiagent systems (see Chapter 2), as well as their dimensionality reduction (see Chapter 3).

1.2.2 The consensus region

A natural strategy to improve Corollary 1.5 would be to look for quantities which are invariant with respect to $\overline{v}$, since it is conserved in systems like (1.6).

Definition 1.6. The symmetric bilinear form $B : \mathbb{R}^{dN} \times \mathbb{R}^{dN} \to \mathbb{R}$ is defined, for any

[1] Given a real $N \times N$ matrix $A = (a_{ij})_{i,j=1}^N$ and $v \in \mathbb{R}^{dN}$ we denote by Av the action of A on $\mathbb{R}^{dN}$ by mapping v to $(a_{i1}v_1 + \cdots + a_{iN}v_N)_{i=1}^N$. Given a nonnegative symmetric $N \times N$ matrix $A = (a_{ij})_{i,j=1}^N$, the *Laplacian* L of A is defined by $L = D - A$, with $D = \mathrm{diag}(d_1, \ldots, d_N)$ and $d_k = \sum_{j=1}^N a_{kj}$.

$v, w \in \mathbb{R}^{dN}$, as

$$B(v,w) \triangleq \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}(v_i - v_j)\cdot(w_i - w_j),$$

where $\cdot$ denotes the usual scalar product on $\mathbb{R}^d$.

Remark 1.7. It is trivial to prove that

$$B(v,w) = \frac{1}{N}\sum_{i=1}^{N}(v_i \cdot w_i) - \overline{v}\cdot\overline{w}, \tag{1.9}$$

where $\overline{v}$ stands for the average of the elements of the vector $v = (v_1, \ldots, v_N)$ given by (1.2). From this representation of B follows easily that the two spaces

$$\mathcal{V}_f \triangleq \left\{v \in \mathbb{R}^{dN} : v_1 = \ldots = v_N\right\},$$
$$\mathcal{V}_\perp \triangleq \left\{v \in \mathbb{R}^{dN} : \sum_{i=1}^{N} v_i = 0\right\},$$

are perpendicular with respect to the scalar product B, i.e., $\mathbb{R}^{dN} = \mathcal{V}_f \oplus \mathcal{V}_\perp$. This means that every $v \in \mathbb{R}^{dN}$ can be written uniquely as $v = v^f + v^\perp$, where $v^f \in \mathcal{V}_f$ and $v^\perp \in \mathcal{V}_\perp$. A closer inspection reveals that for every $i = 1, \ldots, N$ it holds

$$v_i^f = \overline{v} \quad \text{and} \quad v_i^\perp = v_i - \overline{v}.$$

Notice that, since $v^\perp \in \mathcal{V}_\perp$, for any vector $w \in \mathbb{R}^d$ it holds

$$\sum_{i=1}^{N}(v_i^\perp \cdot w) = \left(\sum_{i=1}^{N} v_i^\perp\right)\cdot w = 0. \tag{1.10}$$

Since for every $v, w \in \mathbb{R}^{dN}$ we have $B(v^f, w) = 0 = B(v, w^f)$, it holds

$$B(v,w) = B(v^\perp, w) = B(v, w^\perp) = B(v^\perp, w^\perp).$$

This means that B distinguishes two vectors modulo their projection on $\mathcal{V}_f$. Moreover, from (1.9) immediately follows that B restricted to $\mathcal{V}_\perp \times \mathcal{V}_\perp$ coincides, up to a factor $1/N$, with the usual scalar product on $\mathbb{R}^{dN}$.

Remark 1.8 (Consensus manifold)**.** Notice that whenever the initial datum (x^0, v^0) belongs to the set $\mathbb{R}^{dN} \times \mathcal{V}_f$, the right-hand size of $\dot{v}_i$ in (1.6) is 0, hence the equality $v_1(t) = \ldots = v_N(t)$ is satisfied for all $t \geq 0$ and the system is already in consensus. For this reason, the set $\mathbb{R}^{dN} \times \mathcal{V}_f$ is called the *consensus manifold*.

The bilinear form B can be used to characterize consensus emergence for solutions $(x(\cdot), v(\cdot))$ of system (1.6) by setting

$$X(t) \triangleq B(x(t), x(t)) = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\|x_i(t) - x_j(t)\|^2, \tag{1.11}$$

$$V(t) \triangleq B(v(t), v(t)) = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \|v_i(t) - v_j(t)\|^2 . \tag{1.12}$$

The functionals X and V provide a description of consensus by measuring the spread, both in positions and velocities, of the trajectories of the solution $(x(\cdot), v(\cdot))$, as the following result shows. In Corollary 1.19 we shall see that actually V is a Lyapunov functional for system (1.6)

Proposition 1.9. *The following statements are equivalent:*

(i) $\lim_{t\to+\infty} \|v_i(t) - \overline{v}(t)\| = 0$ *for every* $i = 1, \ldots, N$*;*

(ii) $\lim_{t\to+\infty} v_i^{\perp}(t) = 0$ *for every* $i = 1, \ldots, N$*;*

(iii) $\lim_{t\to+\infty} V(t) = 0$.

Proof. The only implication which is not immediate is $(ii)\Rightarrow(iii)$, for which it suffices to notice that it holds

$$V(t) = \frac{1}{N} \sum_{i=1}^{N} \left\| v_i^{\perp}(t) \right\|^2 .$$

This concludes the proof. □

Remark 1.10. From Proposition 1.9 follows that, in order to establish that a solution of (1.6) tends to consensus, a possible strategy would be to prove that the Lyapunov functional V has a sufficiently strong decay. For this purpose, the following computation shall be often exploited: given a matrix $\omega \in \mathbb{R}^{N\times N}$ which is symmetric and with positive entries, i.e., for every $i, j = 1, \ldots, N$ it holds

$$\omega_{ij} = \omega_{ji} \quad \text{and} \quad \omega_{ij} > 0,$$

for any $v \in \mathbb{R}^{dN}$ we have

$$\begin{aligned} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} (v_j - v_i) \cdot v_i &= \frac{1}{2N^2} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} (v_j - v_i) \cdot v_i + \sum_{j=1}^{N} \sum_{i=1}^{N} \omega_{ji} (v_i - v_j) \cdot v_j \right) \\ &= -\frac{1}{2N^2} \sum_{j=1}^{N} \sum_{i=1}^{N} \omega_{ij} \|v_i - v_j\|^2 \\ &\leq -\min_{1\leq i,j\leq N} \omega_{ij} \frac{1}{N} \sum_{i=1}^{N} \left\| v_i^{\perp} \right\|^2 . \end{aligned} \tag{1.13}$$

A sufficient condition for consensus emergence for solutions of system (1.6) is provided by the following

Theorem 1.11 ([114, Theorem 3.1])**.** *Let* $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ *and set* $X_0 \triangleq B(x^0, x^0)$ *and* $V_0 \triangleq B(v^0, v^0)$*. If the following inequality is satisfied*

$$\int_{\sqrt{X_0}}^{+\infty} a(\sqrt{2N} r) \, dr \geq \sqrt{V_0}, \tag{1.14}$$

then the solution of (1.6) *with initial datum* (x^0, v^0) *tends to consensus.*

We shall prove this result in Section 1.6, although stated in a slightly different fashion (see Theorem 1.30). The meaning of (1.14) is that, as soon as there is a good balance between X_0, V_0 and the kernel a (i.e., the agents are neither too dispersed nor too unaligned with respect to the strength of the mutual interaction), then the system tends to consensus autonomously.

Definition 1.12 (Consensus region)**.** We call *consensus region* the set of points $(X_0, V_0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ satisfying (1.14).

The size of the consensus region gives an estimate of how large the basin of attraction of the consensus manifold $\mathbb{R}^{dN} \times \mathcal{V}_f$ is. This estimate is sharp in some cases, as the following example shows.

Example 1.13 ([81, Proposition 5])**.** Consider $N = 2$ agents in dimension $d = 1$ subject to system (1.6) with interaction kernel given by (1.7) with $H = 1/2$, $\sigma = 1$, and $\beta = 1$. If we denote by $(x_1(\cdot), v_1(\cdot))$ and $(x_2(\cdot), v_2(\cdot))$ the trajectories of the two agents, it is easy to show that the evolution of the relative main state $x(t) \triangleq x_1(t) - x_2(t)$ and of the relative consensus state $v(t) \triangleq v_1(t) - v_2(t)$ is given for every $t \geq 0$ by

$$\begin{cases} \dot{x}(t) = v(t), \\ \dot{v}(t) = -\dfrac{v(t)}{1 + x(t)^2}, \end{cases} \tag{1.15}$$

with initial condition $x(0) = x^0$ and $v(0) = v^0$ (without loss of generality, we may assume that $x^0, v^0 > 0$). An explicit solution of the above system can be easily derived by means of direct integration:

$$v(t) - v^0 = -\arctan x(t) + \arctan x^0.$$

Condition (1.14) in this case reads

$$\frac{\pi}{2} - \arctan x^0 \geq v^0.$$

Hence, suppose (1.14) is violated, i.e., $\arctan x^0 + v^0 > \pi/2$. This means $\arctan x^0 + v^0 \geq \pi/2 + \varepsilon$ for some $\varepsilon > 0$, which implies

$$|v(t)| = |-\arctan x(t) + \arctan x^0 + v^0| \geq \left| -\arctan x(t) + \frac{\pi}{2} + \varepsilon \right| > \varepsilon$$

for every $t \geq 0$. Therefore, the solution of system (1.15) with initial datum (x^0, v^0) satisfying $\arctan x^0 + v^0 > \pi/2$ does not converge to consensus, since otherwise we would have $v(t) \to 0$ for $t \to +\infty$.

Remark 1.14. Notice that, if $\int_\delta^{+\infty} a(r) dr$ diverges for every $\delta \geq 0$, then the consensus region coincides with the entire space $\mathbb{R}^{dN} \times \mathbb{R}^{dN}$. In other words, in this case the interaction force between the agents is so strong that the system will reach consensus no matter what the initial conditions are.

1.3 The effect of perturbations on consensus emergence

An immediate way to enhance the alignment capabilities of systems like (1.6) consists in adding a feedback term penalizing the distance of each agent's velocity from the average one, i.e.,

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)(v_j(t) - v_i(t)) + \gamma(\overline{v}(t) - v_i(t)), \end{cases} \tag{1.16}$$

where γ is a prescribed nonnegative constant, modeling the strength of the additional alignment term.

This approach to the enforcement of consensus is a particular instance of what in the literature is known as *decentralized control strategy*, which has been thoroughly studied especially for its application in the self-organization of *unmanned aerial vehicles* (UAVs) [99], congestion control in communication networks [150], and distributed sensor newtorks [71]. We also refer to [168] for the stability analysis of a decentralized coordination method for dynamical systems with switching underlying communication network.

As system (1.16) can be rewritten as (1.6) with the interaction kernel $a(\cdot) + \gamma$ replacing $a(\cdot)$, by Theorem 1.11 and Remark 1.14 each solution of (1.16) tends to consensus.

However, the apparently innocent fix of adding the extra term above has actually a huge impact on the interpretation of the model: as pointed out in [49], this approach requires that each agent must possess at every instant a *perfect information* of the whole system, since it has to correctly compute the mean velocity of the group $\overline{v}$ in order to compute its trajectory. This condition is seldom met in real-life situations, where it is usually only possible to ask that each agent computes an approximated mean velocity vector $\overline{v}_i$, instead of the true $\overline{v}$. These considerations lead us to the model

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)(v_j(t) - v_i(t)) + \gamma(\overline{v}_i(t) - v_i(t)). \end{cases} \tag{1.17}$$

In studying under which conditions the solutions of system (1.17) tend to consensus, it is often desirable to express the approximated feedback as a combination of a term consisting on a *true information feedback*, i.e., a feedback based on the real average $\overline{v}$, and a perturbation term, which models the deviation of $\overline{v}_i$ from $\overline{v}$. To this end, we rewrite system (1.17) in the following form:

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)(v_j(t) - v_i(t)) + \alpha(t)(\overline{v}(t) - v_i(t)) + \beta(t)\Delta_i(t), \end{cases} \tag{1.18}$$

where $\alpha(\cdot)$ and $\beta(\cdot)$ are two nonnegative, piecewise continuous functions, and $\Delta_i(\cdot)$ is the (possibly only measurable) deviation acting on the estimate of $\overline{v}$ by agent i. Therefore, solutions in this context have to be understood in terms of weak solutions in the Carathéodory sense, see [100]. For a quick reference, see Appendix B.

Remark 1.15. In what follows, we will not be interested in the well-posedness of system (1.18), but rather in finding assumptions on the functions a, α, β, and Δ_i for which we can guarantee its asymptotic convergence to consensus. To identify the additional conditions on a, α, β and Δ_i for which existence and/or uniqueness of solutions on a finite time horizon $[0, T]$ can be established for system (1.18), we need to check under which hypotheses we have that, for every $i = 1, \ldots, N$, the function

$$g_i(t, x, v) = \frac{1}{N}\sum_{j=1}^{N} a\left(\|x_i - x_j\|\right)(v_j - v_i) + \alpha(t)\left(\frac{1}{N}\sum_{j=1}^{N} v_j - v_i\right) + \beta(t)\Delta_i(t)$$

satisfies the hypotheses of Theorems B.1 and/or B.3. In case they are both satisfied for every $T \geq 0$, Remark B.4 tells us how we can obtain a unique solution of the system defined on the entire half line $\mathbb{R}_+$.

System (1.18) provides the advantage of encompassing all the previously introduced models, as can be readily seen:

- if $\alpha = \beta \equiv \gamma$ and $\Delta_i = v_i - \overline{v}$, or $\alpha = \beta \equiv 0$, then we recover system (1.6),
- the choices $\alpha \equiv \gamma$, $\Delta_i \equiv 0$ (or equivalently $\beta \equiv 0$) yield system (1.16),
- if $\alpha = \beta \equiv \gamma$ and $\Delta_i = \overline{v}_i - \overline{v}$ we obtain system (1.17).

The introduction of the perturbation term in system (1.18) may deeply modify the nature of the original model: for instance, an immediate consequence is that the mean velocity of the system is, in general, no longer a conserved quantity.

Proposition 1.16. *For system* (1.18)*, with perturbations given by the vector-valued function* $\Delta(\cdot) = (\Delta_1(\cdot), \ldots, \Delta_N(\cdot))$*, for every* $t \geq 0$ *it holds*

$$\frac{d}{dt}\overline{v}(t) = \beta(t)\overline{\Delta}(t).$$

Proof. A trivial computation shows that for every $t \geq 0$ it holds

$$\begin{aligned}
\frac{d}{dt}\overline{v}(t) &= \frac{1}{N}\sum_{i=1}^{N} \dot{v}_i(t) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)(v_j(t) - v_i(t)) + \alpha(t)(\overline{v}(t) - v_i(t)) + \beta(t)\Delta_i(t)\right) \\
&= \underbrace{\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)(v_j(t) - v_i(t))}_{=0,\text{ by simmetry.}} + \underbrace{\frac{\alpha(t)}{N}\sum_{i=1}^{N} v_i^{\perp}(t)}_{=0} + \frac{\beta(t)}{N}\sum_{i=1}^{N}\Delta_i(t)
\end{aligned}$$

$$= \beta(t)\overline{\Delta}(t).$$

This proves the statement. □

Remark 1.17. As we have already pointed out, it is possible to recover system (1.6) by setting $\Delta_i = v_i - \overline{v}$, whereas we can recover system (1.16) for the choice $\Delta_i \equiv 0$. Note that in both cases we have $\overline{\Delta}(t) = 0$ for every $t \geq 0$, therefore the mean velocity is conserved both in systems (1.6) and (1.16).

We also highlight the fact that $\overline{v}$ is not conserved even in the case that for every $t \geq 0$, and for every $i = 1, \ldots, N$ we have $\Delta_i(t) = w$, where $w \in \mathbb{R}^d \setminus \{0\}$, i.e., the case in which all agents make the same mistake in evaluating the mean velocity.

1.4 General results for consensus stabilization under perturbations

As already noticed in Remark 1.10, a possible strategy for studying under which assumptions the solutions of system (1.6) tend to consensus is to obtain an estimate of the decay of the Lyapunov functional V. We shall follow a similar approach in order to study consensus emergence for system (1.18). We begin by proving the following lemma.

Lemma 1.18. *Let* $(x(\cdot), v(\cdot))$ *be a solution of system* (1.18)*. For every* $t \geq 0$ *it holds*

$$\frac{d}{dt}V(t) \leq -2a\left(\sqrt{2NX(t)}\right)V(t) - 2\alpha(t)V(t) + \frac{2\beta(t)}{N}\sum_{i=1}^{N}\Delta_i(t)\cdot v_i^{\perp}(t). \tag{1.19}$$

Proof. Differentiating V for every $t \geq 0$, we have

$$\begin{aligned}\frac{d}{dt}V(t) &= \frac{d}{dt}\frac{1}{N}\sum_{i=1}^{N}\left\|v_i^{\perp}(t)\right\|^2 \\ &= \frac{2}{N}\sum_{i=1}^{N}\dot{v}_i^{\perp}(t)\cdot v_i^{\perp}(t) \\ &= \frac{2}{N}\sum_{i=1}^{N}\dot{v}_i(t)\cdot v_i^{\perp}(t) - \frac{2}{N}\sum_{i=1}^{N}\dot{\overline{v}}(t)\cdot v_i^{\perp}(t),\end{aligned}$$

which, inserting the expression for $\dot{v}_i(t)$, yields

$$\begin{aligned}\frac{d}{dt}V(t) = &\underbrace{\frac{2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}a\left(\|x_i(t)-x_j(t)\|\right)\left(v_j(t)-v_i(t)\right)\cdot v_i^{\perp}(t) +}_{\diamond} \\ &+\frac{2\alpha(t)}{N}\sum_{i=1}^{N}(\overline{v}(t)-v_i(t))\cdot v_i^{\perp}(t) + \frac{2\beta(t)}{N}\sum_{i=1}^{N}\Delta_i(t)\cdot v_i^{\perp}(t) - \frac{2}{N}\sum_{i=1}^{N}\dot{\overline{v}}(t)\cdot v_i^{\perp}(t).\end{aligned} \tag{1.20}$$

Since it holds

$$\begin{aligned}\|x_i(t)-x_j(t)\| &= \left\|x_i^\perp(t)-x_j^\perp(t)\right\| \\ &\le \left\|x_i^\perp(t)\right\| + \left\|x_j^\perp(t)\right\| \\ &\le \sqrt{2}\left(\sum_{k=1}^N \left\|x_k^\perp(t)\right\|\right)^{1/2} \\ &\le \sqrt{2NX(t)},\end{aligned}$$

the fact that a is nonincreasing and inequality (1.13) yield

$$\diamond \le -2a\left(\sqrt{2NX(t)}\right)V(t). \tag{1.21}$$

Using Proposition 1.16, we can rewrite the remaining term as

$$\begin{aligned}\frac{d}{dt}V(t)-\diamond &= -\frac{2\alpha(t)}{N}\sum_{i=1}^N\left\|v_i^\perp(t)\right\|^2 + \frac{2\beta(t)}{N}\sum_{i=1}^N \Delta_i(t)\cdot v_i^\perp(t) - \underbrace{\frac{2\beta(t)}{N}\sum_{i=1}^N \overline{\Delta}(t)\cdot v_i^\perp(t)}_{=0} \\ &= -2\alpha(t)V(t) + \frac{2\beta(t)}{N}\sum_{i=1}^N \Delta_i(t)\cdot v_i^\perp(t).\end{aligned} \tag{1.22}$$

Plugging (1.21) and (1.22) inside (1.20) concludes the proof. □

The following result is a trivial consequence of Lemma 1.18.

Corollary 1.19. *Let* $(x(\cdot), v(\cdot))$ *be a solution of system* (1.6)*. Then* V *is decreasing.*

Proof. The choice $\alpha = \beta \equiv 0$ in Lemma 1.18 yields the decay estimate

$$\frac{d}{dt}V(t) \le -2a\left(\sqrt{2NX(t)}\right)V(t). \tag{1.23}$$

Assume $V(0) > 0$, otherwise we are already in consensus. By the continuity of the dynamics, there exists a $T > 0$ for which $V(t) > 0$ holds for every $t \in [0, T]$. Directly integrating (1.23) gives

$$V(t) \le V(0)e^{-2\int_0^t a\left(\sqrt{2NX(s)}\right)ds}$$

for any $t \in [0, T]$. Actually this estimate holds for any $t \ge 0$ since it is trivially satisfied whenever $V(t) = 0$. Since a is strictly positive, it then follows that V is decreasing. □

In what follows, since we are interested in the case where Δ_i plays an active role in the dynamics, we assume $\beta(t) > 0$ for all $t \ge 0$. From Lemma 1.18 we obtain the following

Theorem 1.20. *Let* $(x(\cdot), v(\cdot))$ *be a solution of system* (1.18)*, and suppose that there* *exists a* $T \geq 0$ *such that for every* $t \geq T$,

$$\sum_{i=1}^{N} \Delta_i(t) \cdot v_i^{\perp}(t) \leq \phi(t) \sum_{i=1}^{N} \left\| v_i^{\perp}(t) \right\|^2 \tag{1.24}$$

for some function $\phi : [T, +\infty) \to [0, \ell]$, *where*

$$\ell < \frac{\min_{t \geq T} \alpha(t)}{\max_{t \geq T} \beta(t)}. \tag{1.25}$$

Then $(x(\cdot), v(\cdot))$ *tends to consensus.*

Proof. Under the assumption (1.24), for every $t \geq T$ the upper bound in (1.19) can be simplified to

$$\begin{aligned}
\frac{d}{dt} V(t) &\leq -2\alpha(t)V(t) + \frac{2\beta(t)}{N} \sum_{i=1}^{N} \Delta_i(t) \cdot v_i^{\perp}(t) \\
&\leq -2\alpha(t)V(t) + \frac{2\beta(t)}{N} \phi(t) \sum_{i=1}^{N} \left\| v_i^{\perp}(t) \right\|^2 \\
&= -2\alpha(t)V(t) + 2\beta(t)\phi(t)V(t) \\
&\leq 2\beta(t) \left(\ell - \frac{\alpha(t)}{\beta(t)} \right) V(t).
\end{aligned}$$

Integrating between T and t (where $t \geq T$) we get

$$V(t) \leq V(T) e^{2 \int_T^t \beta(s) \left(\ell - \frac{\alpha(s)}{\beta(s)} \right) ds},$$

and as the factor $\ell - \alpha(s)/\beta(s)$ is negative while β is nonnegative, V approaches 0 exponentially fast. □

Corollary 1.21. *If there exists* $T \geq 0$ *such that* $\Delta_i^{\perp}(t) = 0$ *for every* $t \geq T$ *and for every* $1 \leq i \leq N$, *then any solution of system* (1.18) *tends to consensus.*

Proof. Noting that $\Delta_i^{\perp} = 0$ implies $\Delta_i = \overline{\Delta}$, we have, by (1.10)

$$\sum_{i=1}^{N} \Delta_i(t) \cdot v_i^{\perp}(t) = \sum_{i=1}^{N} \overline{\Delta} \cdot v_i^{\perp}(t) = 0.$$

Hence, we can apply Theorem 1.20 with $\phi(t) = 0$ for every $t \geq T$ to obtain the result. □

Remark 1.22. A trivial implication of Corollary 1.21 is that any solution of system (1.16) tends to consensus (this was already a consequence of Theorem 1.11), but has moreover a rather nontrivial implication: also any solution of systems subjected to *deviated* uniform control, i.e., systems like (1.18) where $\Delta_i(t) = \Delta(t)$ for every $i = 1, \ldots, N$ and for every

$t \geq 0$, tends to consensus, because it holds

$$\Delta_i^{\perp}(t) = \Delta_i(t) - \frac{1}{N}\sum_{j=1}^{N} \Delta_j(t) = \Delta(t) - \Delta(t) = 0$$

for every $i = 1, \ldots, N$ and for every $t \geq 0$, therefore Corollary 1.21 applies. This means that systems of this kind converge to consensus even if the agents have an incorrect knowledge of the mean velocity, provided they all make the same mistake.

A final consequence of the previous results is the following corollary, which provides an upper bound for tolerable perturbations under which consensus emergence can be unconditionally guaranteed.

Corollary 1.23. *For every $i = 1, \ldots, N$, let $\varepsilon_i : \mathbb{R}_+ \to [0, \ell]$ for $\ell > 0$ as in* (1.25)*. If there exists $T \geq 0$ such that $\|\Delta_i(t)\| \leq \varepsilon_i(t) \left\|v_i^{\perp}(t)\right\|$ for every $t \geq T$ and for every $i = 1, \ldots, N$, then any solution of system* (1.18) *tends to consensus.*

Proof. By using the Cauchy-Schwarz inequality we have

$$\begin{aligned} \sum_{i=1}^{N} \Delta_i(t) \cdot v_i^{\perp}(t) &\leq \sum_{i=1}^{N} \varepsilon_i(t) \left\|v_i^{\perp}(t)\right\|^2 \\ &\leq \ell \sum_{i=1}^{N} \left\|v_i^{\perp}(t)\right\|^2 . \end{aligned}$$

The conclusion follows by taking $\phi(t) = \ell$ for every $t \geq T$ in Theorem 1.20. □

The result above shows that, provided that the magnitude of the perturbation is smaller than the one of the deviation of the agent's velocity from the mean, then convergence to consensus is obtained unconditionally with respect to the initial condition. This is the case of local estimates of the average, as the largest error that an agent can make when estimating the group average upon a subset of agents is precisely its own deviation from the mean, $v_i^{\perp}$.

1.5 Perturbations as linear combinations of velocity deviations

We begin this section by considering a simple case study, which is nevertheless relevant as it addresses consensus stabilization based on a leader-following feedback.

Example 1.24. Let us use Lemma 1.18 to study the convergence to consensus of a system like (1.17), where each agent computes its local mean velocity $\overline{v}_i$ by taking into account itself plus a single common agent (x_1, v_1), which in turn takes into account only itself by computing $\overline{v}_1 = v_1$. Formally, given two finite conjugate exponents p, q (i.e., two positive real numbers satisfying $1/p + 1/q = 1$), we assume that for any $i = 1, \ldots, N$ it holds

$$\overline{v}_i(t) = \frac{1}{p} v_i(t) + \frac{1}{q} v_1(t) \quad \text{for every } t \geq 0.$$

We shall prove that any solution of this system tends to consensus, no matter how small the positive weight $1/q$ of v_1 in $\overline{v}_i$ is. We start by writing the system under the form (1.18), with $\alpha(t) = \beta(t) = \gamma > 0$ and

$$\Delta_i(t) = \frac{1}{p} v_i^{\perp}(t) + \frac{1}{q} v_1^{\perp}(t) \quad \text{for every } t \geq 0.$$

Hence, the perturbation term in the estimate (1.19) on the decay of V becomes

$$\begin{aligned} \frac{2\gamma}{N} \sum_{i=1}^{N} \Delta_i(t) \cdot v_i^{\perp}(t) &= \frac{2\gamma}{N} \sum_{i=1}^{N} \frac{1}{p} (v_i^{\perp}(t) + \frac{1}{q} v_1^{\perp}(t)) \cdot v_i^{\perp}(t) \\ &= \frac{1}{p} \frac{2\gamma}{N} \sum_{i=1}^{N} \left\| v_i^{\perp}(t) \right\|^2 + \frac{1}{q} \frac{2\gamma}{N} v_1^{\perp}(t) \cdot \underbrace{\sum_{i=1}^{N} v_i^{\perp}(t)}_{=0} \\ &= \frac{2\gamma}{p} V(t). \end{aligned}$$

Lemma 1.18 let us bound the growth of V as

$$\frac{d}{dt} V(t) \leq 2\gamma \left(-1 + \frac{1}{p} \right) V(t) = -\frac{2\gamma}{q} V(t),$$

which ensures the exponential decay of the functional V for any $q > 0$.

Motivated by the latter example, we turn our attention to the study of systems like (1.18) where the perturbation of the mean of the i-th agent has the specific form

$$\Delta_i(t) = \sum_{j=1}^{N} \omega_{ij}(t) v_j^{\perp}(t) \quad \text{for every } t \geq 0, \tag{1.26}$$

for some measurable mapping $\omega : \mathbb{R}_+ \to \mathbb{R}^{N \times N}$ with the property $\omega_{ij}(t) > 0$ for all $i, j = 1, \ldots, N$ and $t \geq 0$.
We shall see that the results obtained in Section 1.3 help us identify under which assumptions on the coefficients ω_{ij} we can infer unconditional convergence to consensus.

Theorem 1.25. *Consider a system of the form* (1.18)*, where* Δ_i *is given as in* (1.26)*. Then, if for every* $t \geq 0$ *and every* $i, j = 1, \ldots, N$ *we have* $\omega_{ij}(t) = \omega_{ji}(t)$*, and we set*

$$I(t) \triangleq \min_{1 \leq i,j \leq N} \omega_{ij}(t) \quad \text{and} \quad S(t) \triangleq \max_{1 \leq i \leq N} \sum_{j=1}^{N} \omega_{ij}(t),$$

then the following estimate holds:

$$\frac{d}{dt} V(t) \leq -2a \left(\sqrt{2NX(t)} \right) V(t) + 2\beta(t) \left(S(t) - NI(t) - \frac{\alpha(t)}{\beta(t)} \right) V(t). \tag{1.27}$$

Therefore, if there exists a $T \geq 0$ *such that the quantity* $S(t) - NI(t) - \frac{\alpha(t)}{\beta(t)}$ *is bounded from above by a constant* $C < 0$ *in* $[T, +\infty)$*, then any solution of the system tends to* *consensus.*

Proof. Standard computations yield

$$\begin{aligned}
\sum_{i=1}^{N} \Delta_i(t) \cdot v_i^{\perp}(t) &= \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_{ij}(t) v_j^{\perp}(t) \cdot v_i^{\perp}(t) \\
&= \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_{ij}(t)(v_j^{\perp}(t) - v_i^{\perp}(t)) \cdot v_i^{\perp}(t) + \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \omega_{ij}(t) \right) \left\| v_i^{\perp}(t) \right\|^2 \\
&= N^2 \left(-\frac{1}{2N^2} \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_{ij}(t) \left\| v_j(t) - v_i(t) \right\|^2 \right) + \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \omega_{ij}(t) \right) \left\| v_i^{\perp}(t) \right\|^2 \\
&\leq N \left(-NI(t) + S(t) \right) V(t),
\end{aligned}$$

having used (1.13). Applying this into (1.19) and collecting β, we get (1.27). □

In the following results we shall assume that for every $t \geq 0$ the matrix $\omega(t)$ is *stochastic*, i.e., $\sum_{j=1}^{N} \omega_{ij}(t) = 1$ for every $i = 1, \ldots, N$. This implies, by Proposition 1.16, that $\overline{\Delta}(t) = 0$ for every $t \geq 0$ and that $\overline{v}$ is conserved. In particular, for the choice $\alpha = \beta$ the system can eventually be rewritten in the form of system (1.6), and in this case the following results can be seen as consequences of Theorem 1.11.

Corollary 1.26. *Let $\omega(t) \in \mathbb{R}_+^{N \times N}$ be a symmetric stochastic matrix for every $t \geq 0$. If there exists a constant $\vartheta > 0$ such that for every $t \geq 0$ it holds*

$$I(t) = \min_{1 \leq i,j \leq N} \omega_{ij}(t) \geq \vartheta > \frac{\beta(t) - \alpha(t)}{N\beta(t)},$$

then any solution of the system tends to consensus.

Proof. Under the above hypotheses, the quantity $S(t) - NI(t) - \alpha(t)/\beta(t)$ of Theorem 1.25 is bounded from above by $1 - N\vartheta - \alpha(t)/\beta(t)$, which, by assumption, is negative. □

The next result focuses on a specific form of the weight matrix ω which shall be studied further in the upcoming section.

Corollary 1.27. *Suppose that for all $i, j = 1, \ldots, N$ the function $\omega_{ij} : \mathbb{R}_+ \to \mathbb{R}_+$ satisfies*

$$\omega_{ij}(t) = \frac{\phi(\|x_i(t) - x_j(t)\|)}{\eta(t)} \qquad \textit{for every } t \geq 0,$$

where $\phi : \mathbb{R}_+ \to (0, 1]$ is a non increasing, positive, bounded function, and $\eta : \mathbb{R}_+ \to \mathbb{R}_+$ is a nonnegative bounded function satisfying

$$1 \leq N\frac{\beta(t)}{\alpha(t)} \leq \eta(t) \quad \textit{for every } t \geq 0. \tag{1.28}$$

Then, any solution of system (1.18) with Δ_i as in (1.26) tends to consensus.

Proof. If we consider the quantity X, there are at most two cases: either $X(t)$ is bounded from above by a constant $\overline{X}$ for all $t \geq 0$, or $X(t)$ is unbounded.
In the first case, we can bound $S(t)$ from above by N. Since $\|x_i(t) - x_j(t)\| \leq \sqrt{2NX(t)} \leq \sqrt{2N\overline{X}}$ and ϕ is non increasing, we have that

$$I(t) = \min_{1 \leq i,j \leq N} \omega_{ij}(t) \geq \phi\left(\sqrt{2N\overline{X}}\right),$$

and thus, from (1.28), it follows that for every $t \geq 0$ it holds

$$S(t) - NI(t) - \frac{\alpha(t)}{\beta(t)}\eta(t) \leq N - N\phi\left(\sqrt{2N\overline{X}}\right) - \frac{\alpha(t)}{\beta(t)}\eta(t) \leq -N\phi\left(\sqrt{2N\overline{X}}\right). \quad (1.29)$$

From the positivity of ϕ we can therefore conclude the proof by invoking Theorem 1.25.
Suppose now, instead, that $X(t)$ is unbounded: in this case the term $I(t)$ is bounded from below by a term going to 0 (and hence not giving us any useful information) and we have to take advantage of $S(t)$ as shall be shown now. By definition, $X(t)$ is unbounded if and only if there exist two agents with indexes h and k such that $\|x_h(t) - x_k(t)\|$ is unbounded. By the triangle inequality, it follows that for any index i, there is an index $j(i)$ for which $\|x_i(t) - x_{j(i)}(t)\|$ is unbounded. Thus, we fix $\rho > 0$ and let $T > 0$ be the maximum time t such that $\phi(\|x_i(t) - x_{j(i)}(t)\|) < 1 - \rho$ for every $i = 1, \ldots, N$. Then for every $t \geq T$ we can bound $S(t)$ with

$$S(t) \leq \frac{1}{\eta(t)}(N - 1 + 1 - \rho) \leq \frac{\alpha(t)}{N\beta(t)}(N - \rho),$$

and therefore, again from (1.28), we have the estimate

$$\begin{aligned}
S(t) - NI(t) - \frac{\alpha(t)}{\beta(t)}\eta(t) &\leq \frac{\alpha(t)}{N\beta(t)}(N - \rho) - \frac{\alpha(t)}{\beta(t)}\eta(t) \\
&\leq -\frac{\alpha(t)}{N\beta(t)}\rho + \frac{\alpha(t)}{\beta(t)} - N \\
&\leq -\frac{\alpha(t)}{N\beta(t)}\rho \\
&\leq -\frac{\rho}{\eta(t)}
\end{aligned}$$

which is negative for every $t \geq T$ (since η is bounded). Again an application of Theorem 1.25 gives us the result. □

Remark 1.28. A concrete example of a system for which we can apply Corollary 1.27 is obtained by considering the functions

$$\phi(r) \triangleq \frac{1}{(1 + r^2)^\epsilon},$$

(which is the Cucker-Smale function (1.7) with $H = 1, \sigma = 1$ and $\beta = \epsilon$) and

$$\eta(t) \triangleq \max_{1 \leq i \leq N} \left\{ \sum_{j=1}^{N} \phi(\|x_i(t) - x_j(t)\|) \right\}.$$

In this case, $\epsilon \in \mathbb{R}_+$ can be thought of as a parameter tuning the ability of each agent to gather information about the speed of the other agents. Indeed, consider the case $\|x_i(t) - x_j(t)\| > 0$, if $i \neq j$: then, if $\epsilon = 0$, it holds $\overline{v}_i = \overline{v}$ for every $i = 1, \dots, N$, and each agent communicates at the same rate with both near and far away agents; if instead $\epsilon \to +\infty$ then $\overline{v}_i$ approaches v_i, hence each agent is unable to gain sufficient knowledge about the speed of the others in order to compute the correct mean. The function η serves to the purpose of being a common normalizing factor: naturally one would choose for every agent i the normalizing factor given by

$$\sum_{j=1}^{N} \phi(\|x_i(t) - x_j(t)\|), \tag{1.30}$$

but that would produce a non symmetric matrix ω, for which the above results are not valid. In this context, the function η is a suitable replacement, being also coherent with the asymptotic behavior of (1.30) for $\epsilon \to 0$ and $\epsilon \to +\infty$.

Remark 1.29. The request of positivity of the function ϕ cannot be removed from Corollary 1.27, as the function $\phi = \chi_{[0,R]}$ shows. Indeed, what fails in the argument of the proof is the case in which we suppose that the functional X is bounded from above by $\overline{X}$: if the quantity $\sqrt{2N\overline{X}}$ is not less or equal to R, then $\phi(\sqrt{2N\overline{X}}) = 0$ in the inequality (1.29), and we cannot invoke Theorem 1.25 in order to infer consensus.

1.6 Perturbations due to local averaging

An interesting case of a system like (1.17) is the one where the local mean is given by

$$\overline{v}_i(t) = \frac{1}{|\Lambda_R(t,i)|} \sum_{j \in \Lambda_R(t,i)} v_j(t) \quad \text{for every } t \geq 0, \tag{1.31}$$

where $\Lambda_R(i) \triangleq \{j \in \{1, \dots, N\} : \|x_i(t) - x_j(t)\| \leq R\}$ and $|\Lambda_R(t,i)|$ stands for its cardinality. In this case, we model the situation in which each agent estimates the average velocity of the group in the extra feedback term by only counting those agents inside a ball of radius R centered on him. We want to address the issue of characterizing the behavior of system (1.17) with the above choice for $\overline{v}_i$ when the radius R of each ball is either reduced to 0 or set to grow to $+\infty$. We shall see that we can reformulate this decentralized system again as a Cucker-Smale model for a different interaction function for which we can apply Theorem 1.11. We shall show how tuning the radius R affects the convergence to consensus, from the case $R \geq 0$ where only conditional convergence is ensured, to the unconditional convergence result given for $R = +\infty$.

1.6.1 Preserving the asymptotics

First of all, by means of $\chi_{[0,R]}$ we can rewrite $\overline{v}_i(t)$ as

$$\overline{v}_i(t) = \frac{1}{\sum_{k=1}^{N} \chi_{[0,R]}(\|x_i(t) - x_k(t)\|)} \sum_{j=1}^{N} \chi_{[0,R]}(\|x_i(t) - x_j(t)\|) v_j(t). \tag{1.32}$$

As already noted in Remark 1.28, the normalizing terms $\sum_{k=1}^N \chi_{[0,R]}(\|x_i(t)-x_k(t)\|)$ give rise to a matrix of weights which is not symmetric. Since this will be an issue also in the present section, we take η_R to be a function approximating the above normalizing terms and which also preserves its asymptotics for $R \to 0$ and $R \to +\infty$, as for instance,

$$\eta_R(t) = \max_{1\le i\le N} \left\{ \sum_{k=1}^N \chi_{[0,R]}(\|x_i(t)-x_k(t)\|) \right\}. \tag{1.33}$$

Therefore, we replace the vector $\overline{v}_i(t)$ by

$$\frac{1}{\eta_R(t)} \sum_{j=1}^N \chi_{[0,R]}(\|x_i(t)-x_j(t)\|) v_j(t) \,.$$

On top of this, notice that the vector

$$\left(\frac{1}{\eta_R(t)} \sum_{j=1}^N \chi_{[0,R]}(\|x_i(t)-x_j(t)\|) \right) v_i(t)$$

is an approximation of $v_i(t)$ for $R \to 0$ and $R \to +\infty$. This motivates the replacement of the term $\overline{v}_i - v_i$ where $\overline{v}_i$ is as in (1.32) with

$$\overline{v}_i - v_i \approx \frac{1}{\eta_R} \sum_{j=1}^N \chi_{[0,R]}(\|x_i-x_j\|) v_j - \left(\frac{1}{\eta_R} \sum_{j=1}^N \chi_{[0,R]}(\|x_i-x_j\|) \right) v_i. \tag{1.34}$$

The term (1.34) can be rewritten as $1/\eta_R \sum_{j=1}^N \chi_{[0,R]}(\|x_i-x_j\|)(v_j-v_i)$, which can be further simplified as follows

$$\begin{aligned} \frac{1}{\eta_R} \sum_{j=1}^N \chi_{[0,R]}(r_{ij})(v_j - v_i) &= \frac{1}{\eta_R} \sum_{i=1}^N (v_j - v_i) + \frac{1}{\eta_R} \sum_{j=1}^N (1-\chi_{[0,R]}(r_{ij}))(v_i - v_j) \\ &= \frac{N}{\eta_R}(\overline{v} - v_i) + \frac{1}{\eta_R} \sum_{j=1}^N (1-\chi_{[0,R]}(r_{ij}))(v_i - v_j), \end{aligned} \tag{1.35}$$

where we have written r_{ij} in place of $\|x_i - x_j\|$ and removed the time dependencies for the sake of compactness.

It is clear that the choice of the function $\chi_{[0,R]}$ is arbitrary and other alternatives can be selected, provided they give a coherent approximation of the local average (1.31). For instance, instead of $\chi_{[0,R]}$ and η_R, we can consider two generic functions ψ_ε and η_ε, where ε is a parameter ranging in a nonempty set Ω, satisfying the following properties:

(i) $\psi_\varepsilon : \mathbb{R}_+ \to [0,1]$ is a nonincreasing measurable function for every $\varepsilon \in \Omega$;

(ii) $\eta_\varepsilon \in L^\infty(\mathbb{R}_+)$ for every $\varepsilon \in \Omega$;

(iii) there are two disjoint subsets Ω_{CS} and Ω_U of Ω such that

- if $\varepsilon \in \Omega_{CS}$ then $\psi_\varepsilon = \chi_{\{0\}}$ and $\eta_\varepsilon \equiv 1$;

- if $\varepsilon \in \Omega_U$ then $\psi_\varepsilon = \chi_{\mathbb{R}_+}$ and $\eta_\varepsilon \equiv N$.

Under the above hypotheses, we consider the perturbation given for every $t \geq 0$ by

$$\Delta_i^\varepsilon(t) \triangleq \frac{1}{\eta_\varepsilon(t)} \sum_{j=1}^{N} (1 - \psi_\varepsilon(\|x_i(t) - x_j(t)\|))(v_i(t) - v_j(t)). \tag{1.36}$$

With requirement (iii) we impose that whenever $\varepsilon \in \Omega_{CS}$ then it holds

$$\Delta_i^\varepsilon(t) = -\frac{N}{\eta_\varepsilon(t)}(\overline{v}(t) - v_i(t)),$$

therefore recovering the Cucker-Smale system (1.6) from (1.37), while whenever $\varepsilon \in \Omega_U$ then $\Delta_i^\varepsilon(t) = 0$ holds, and we obtain a particular instance of system (1.16).
By means of (1.36), we can rewrite our system with the local average (1.31) in the form of system (1.18)

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N} \displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right) + \gamma \dfrac{N}{\eta_\varepsilon(t)}(\overline{v}(t) - v_i(t)) + \gamma \Delta_i^\varepsilon(t). \end{cases} \tag{1.37}$$

1.6.2 The enlarged consensus region

In order to study under which conditions the solutions of system (1.37) for $\psi_\varepsilon = \chi_{[0,R]}$ and η_ε given by (1.33) converge to consensus, we cannot use the results of Section 1.5. This follows from Remark 1.29 since, if we compare (1.34) and (1.35) to the following computations

$$\frac{1}{\eta_R} \sum_{j=1}^{N} \chi_{[0,R]}(\|x_i - x_j\|) v_j - \left(\frac{1}{\eta_R} \sum_{j=1}^{N} \chi_{[0,R]}(\|x_i - x_j\|) \right) v_i =$$
$$= \left(\frac{1}{\eta_R} \sum_{j=1}^{N} \chi_{[0,R]}(\|x_i - x_j\|) \right) (\overline{v} - v_i) + \frac{1}{\eta_R} \sum_{j=1}^{N} \chi_{[0,R]}(\|x_i - x_j\|) v_j^\perp,$$

we obtain that the system we are considering is precisely the version of system (1.18) where

$$\alpha(t) = \frac{1}{\eta_R} \sum_{j=1}^{N} \chi_{[0,R]}(\|x_i(t) - x_j(t)\|) \quad \text{and} \quad \beta(t) = 1 \quad \text{for every } t \geq 0$$

and Δ_i is of the same kind as the one mentioned in Remark 1.29.
Therefore, we present the following result which gives a sufficient condition on the initial data for which the solutions of system (1.37) converge to consensus. We point out that

system (1.37) can be rewritten into the shape of a Cucker-Smale type system as follows

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N}\left(a\left(\|x_i(t)-x_j(t)\|\right) + \gamma\frac{N}{\eta_\varepsilon(t)}\psi_\varepsilon(\|x_i(t)-x_j(t)\|)\right)(v_j(t)-v_i(t)), \end{cases}$$

and therefore the following result can be also obtained from Theorem 1.11.

Theorem 1.30. *Fix $\gamma \geq 0$, consider system* (1.37) *where Δ_i^ε is as in* (1.36) *and let $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$. If $X_0 \triangleq B(x^0, x^0)$ and $V_0 \triangleq B(v^0, v^0)$ satisfy*

$$\int_{\sqrt{X_0}}^{+\infty} a\left(\sqrt{2N}r\right)\, dr + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}}\int_{\sqrt{X_0}}^{+\infty}\psi_\varepsilon\left(\sqrt{2N}r\right)\, dr \geq \sqrt{V_0}, \tag{1.38}$$

then the solution of system (1.37) *with initial datum* (x^0, v^0) *tends to consensus.*

Proof. Notice that we can assume $V(t) > 0$ for every $t \geq 0$, otherwise the system is already in consensus at time t. From requirement (i) we have

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}\Delta_i(t)\cdot v_i^\perp(t) &= \frac{1}{N\eta_\varepsilon(t)}\sum_{i=1}^{N}\sum_{j=1}^{N}(1-\psi_\varepsilon(\|x_i(t)-x_j(t)\|))(v_i(t)-v_j(t))\cdot v_i^\perp(t) \\ &\leq \frac{N}{\eta_\varepsilon(t)}\left(1-\psi_\varepsilon\left(\sqrt{2NX(t)}\right)\right)V(t), \end{aligned}$$

and inequality (1.19) reads

$$\frac{d}{dt}V(t) \leq -2\left(a\left(\sqrt{2NX(t)}\right) + \frac{\gamma N}{\eta_\varepsilon(t)}\psi_\varepsilon\left(\sqrt{2NX(t)}\right)\right)V(t).$$

Moreover, since

$$\frac{d}{dt}\sqrt{V(t)} = \frac{1}{2\sqrt{V(t)}}\frac{d}{dt}V(t) \quad \text{ for every } t \geq 0,$$

from (ii) we have

$$\begin{aligned} \frac{d}{dt}\sqrt{V(t)} &\leq -2\left(a\left(\sqrt{2NX(t)}\right) + \frac{\gamma N}{\eta_\varepsilon(t)}\psi_\varepsilon\left(\sqrt{2NX(t)}\right)\right)\sqrt{V(t)} \\ &\leq -2\left(a\left(\sqrt{2NX(t)}\right) + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}}\psi_\varepsilon\left(\sqrt{2NX(t)}\right)\right)\sqrt{V(t)}, \end{aligned} \tag{1.39}$$

and integrating between 0 and t we obtain

$$\sqrt{V(t)} - \sqrt{V(0)} \leq -\int_0^t\left(a\left(\sqrt{2NX(s)}\right) + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}}\psi_\varepsilon\left(\sqrt{2NX(s)}\right)\right)\sqrt{V(s)}\, ds. \tag{1.40}$$

We now work on changing the variable inside the integral. We can actually claim that

$$\frac{d}{dt}X(t) \leq \sqrt{V(t)} \quad \text{for every } t \geq 0, \tag{1.41}$$

since, indeed, the following computation

$$\begin{aligned}
\frac{d}{dt}X(t) &= \frac{1}{N}\sum_{i=1}^{N}\frac{d}{dt}\left\|x_i^{\perp}(t)\right\|^2 \\
&= \frac{2}{N}\sum_{i=1}^{N} x_i^{\perp}(t)\cdot \dot{x}_i^{\perp}(t) \\
&\leq \frac{2}{N}\sum_{i=1}^{N}\left\|x_i^{\perp}(t)\right\|\left\|v_i^{\perp}(t)\right\| \\
&\leq \frac{2}{N}\left(\sum_{i=1}^{N}\left\|x_i^{\perp}(t)\right\|\right)^{1/2}\left(\sum_{i=1}^{N}\left\|v_i^{\perp}(t)\right\|^2\right)^{1/2} \\
&= 2\sqrt{X(t)}\sqrt{V(t)},
\end{aligned}$$

and $\frac{d}{dt}X(t) = \frac{d}{dt}(\sqrt{X(t)}\sqrt{X(t)}) = 2\sqrt{X(t)}\frac{d}{dt}\sqrt{X(t)}$, together yield (1.41). Notice that in the above computation we have used the following fact

$$\frac{d}{dt}x_i^{\perp}(t) = \frac{d}{dt}\left(x_i(t) - \frac{1}{N}\sum_{j=1}^{N}x_j(t)\right) = v_i(t) - \frac{1}{N}\sum_{j=1}^{N}v_j(t) = v_i^{\perp}(t).$$

Setting $r = \sqrt{X(s)}$ and using (1.41), we can change variable in (1.40) as follows:

$$\sqrt{V(t)} - \sqrt{V(0)} \leq -\int_{\sqrt{X(0)}}^{\sqrt{X(t)}}\left(a\left(\sqrt{2N}r\right) + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}}\psi_\varepsilon\left(\sqrt{2N}r\right)\right)\, dr. \tag{1.42}$$

Let us suppose that (1.38) is true. Since $X(0) = X_0$ and $V(0) = V_0$, if $V(0) = 0$ there is nothing to prove because we are already in consensus. If, instead, it holds

$$0 < \sqrt{V(0)} \leq \int_{\sqrt{X(0)}}^{+\infty}\left(a\left(\sqrt{2N}r\right) + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}}\psi_\varepsilon\left(\sqrt{2N}r\right)\right)\, dr, \tag{1.43}$$

then there is a $\overline{X} > X(0)$ such that

$$\sqrt{V(0)} = \int_{\sqrt{X(0)}}^{\sqrt{\overline{X}}}\left(a\left(\sqrt{2N}r\right) + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}}\psi_\varepsilon\left(\sqrt{2N}r\right)\right)\, dr$$

(having used the fact that, from (i), the integrand is a non increasing function). Now, either equality holds in (1.43), and $\lim_{t\to+\infty} V(t) = 0$ follows by passing to the limit in (1.42), or we have a strict inequality. But in this case $\overline{X} \geq X(t)$ must hold for every $t \geq 0$, since otherwise there would be a $T>0$ for which we have

$$\begin{aligned}\sqrt{V(0)} &\geq \sqrt{V(T)} + \int_{\sqrt{X(0)}}^{\sqrt{X(T)}} \left(a\left(\sqrt{2N}r\right) + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}} \psi_\varepsilon\left(\sqrt{2N}r\right)\right) dr \\ &> \int_{\sqrt{X(0)}}^{\sqrt{\overline{X}}} \left(a\left(\sqrt{2N}r\right) + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}} \psi_\varepsilon\left(\sqrt{2N}r\right)\right) dr \\ &= \sqrt{V(0)},\end{aligned}$$

which is obviously a contradiction. Thus, we have that the inequality $\overline{X} \geq X(t)$ is true for every $t \geq 0$, and from (1.39) we have

$$\frac{d}{dt}V(t) \leq -2\left(a\left(\sqrt{2N\overline{X}}\right) + \frac{\gamma N}{\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)}} \psi_\varepsilon\left(\sqrt{2N\overline{X}}\right)\right) V(t).$$

The fact that $\lim_{t\to+\infty} V(t) = 0$ follows from the above inequality. □

An example of a family of functions $(\psi_\varepsilon, \eta_\varepsilon)_{\varepsilon\in\Omega}$ satisfying the hypothesis (i)–(iii) is given by

$$\psi_\varepsilon(r) = \frac{1}{(1+r^2)^\varepsilon} \qquad \text{where } \varepsilon \in \Omega = [0,\infty],$$

for which we set $\psi_\infty \triangleq \chi_{\{0\}}$, and

$$\eta_\varepsilon(t) \triangleq \max_{1\leq i\leq N} \left\{ \sum_{k=1}^N \psi_\varepsilon(\|x_(t) - x_k(t)\|) \right\}.$$

In this case, $\Omega_{CS} = \{0\}$ and $\Omega_U = \{+\infty\}$, and (1.38) is satisfied as soon as it holds

$$\int_{\sqrt{X_0}}^{+\infty} a\left(\sqrt{2N}r\right) dr + \gamma \int_{\sqrt{X_0}}^{+\infty} \psi_\varepsilon\left(\sqrt{2N}r\right) dr \geq \sqrt{V_0},$$

since in this case $\|\eta_\varepsilon\|_{L^\infty(\mathbb{R}_+)} \leq N$.
The most interesting example of such a family is the one which has introduced this section: we consider $\Omega = [0,+\infty]$, the sequence of functions $(\chi_{[0,R]})_{R\in\Omega}$ and η_R as in (1.33) (notice that, as before, we have $\Omega_{CS} = \{0\}$ and $\Omega_U = \{\infty\}$). Since, again, it holds $\|\eta\|_{L^\infty(\mathbb{R}_+)} \leq N$, if R is sufficiently large to satisfy $\sqrt{2NX_0} \leq R$, condition (1.38) is satisfied as soon as

$$\int_{\sqrt{X_0}}^{+\infty} a\left(\sqrt{2N}r\right) dr + \gamma\left(\frac{R}{\sqrt{2N}} - \sqrt{X_0}\right) \geq \sqrt{V_0},$$

by means of a trivial integration. If, instead, R is so small that $\sqrt{2NX_0} > R$ holds, condition (1.38) is satisfied as soon as

$$\int_{\sqrt{X_0}}^{+\infty} a\left(\sqrt{2N}r\right) dr \geq \sqrt{V_0},$$

recovering Theorem 1.11. As can be seen, we have enlarged the original consensus region provided by Theorem 1.11 by a term whose size is linearly increasing in R. This implies that, in the case $R = +\infty$, the consensus region coincides with the entire space $\mathbb{R}^{dN} \times \mathbb{R}^{dN}$, hence the system converges to consensus regardless of the initial datum.

The above results can be seen as the asymptotic outcome of the following more general approach: consider the set $\Omega = [0, \infty] \times (1, \infty]$, write ε as the pair of parameters (R, θ) and set

$$\psi_{R,\theta}(r) \triangleq \begin{cases} 1 & \text{if } r \leq R, \\ \dfrac{1}{(r - R + 1)^\theta} & \text{if } r > R. \end{cases}$$

Take $\eta_{R,\theta}(t)$ to be either

$$\max_{1 \leq i \leq N} \left\{ \sum_{k=1}^{N} \psi_{R,\theta}(\|x_i(t) - x_k(t)\|) \right\} \quad \text{or} \quad \min_{1 \leq i \leq N} \left\{ \sum_{k=1}^{N} \psi_{R,\theta}(\|x_i(t) - x_k(t)\|) \right\}$$

(notice that in both cases we have $\|\eta\|_{L^\infty(\mathbb{R}_+)} \leq N$). This time we have

$$\Omega_{CS} = \{0\} \times \{+\infty\} \quad \text{and} \quad \Omega_U = \{+\infty\} \times (1, +\infty].$$

If we suppose that R is sufficiently large to satisfy $\sqrt{2NX_0} \leq R$, then it holds

$$\begin{aligned} \int_{\sqrt{X_0}}^{+\infty} \psi_{R,\theta}\left(\sqrt{2N}r\right)\, dr &= \int_{\sqrt{X_0}}^{\frac{R}{\sqrt{2N}}} dr + \int_{\frac{R}{\sqrt{2N}}}^{+\infty} \frac{11}{(\sqrt{2N}r - R + 1)^\theta}\, dr \\ &= \frac{R}{\sqrt{2N}} - \sqrt{X_0} + \frac{1}{\theta - 1}, \end{aligned}$$

hence condition (1.38) is satisfied as soon as

$$\int_{\sqrt{X_0}}^{+\infty} a\left(\sqrt{2N}r\right)\, dr + \gamma\left(\frac{R}{\sqrt{2N}} - \sqrt{X_0} + \frac{1}{\theta - 1}\right) \geq \sqrt{V_0},$$

which shows that the consensus region grows linearly with the radius R while it is inversely proportional to the growth of θ.

Otherwise, if R is so small that $\sqrt{2NX_0} > R$ holds, since in this case

$$\begin{aligned} \int_{\sqrt{X_0}}^{+\infty} \psi_{R,\theta}\left(\sqrt{2N}r\right)\, dr &= \int_{\sqrt{X_0}}^{+\infty} \frac{1}{(\sqrt{2N}r - R + 1)^\theta}\, dr \\ &= \frac{1}{(\theta - 1)(\sqrt{2NX_0} - R + 1)^{\theta - 1}}, \end{aligned}$$

we have that condition (1.38) is true whenever

$$\int_{\sqrt{X_0}}^{+\infty} a\left(\sqrt{2N}r\right)\, dr + \frac{\gamma}{\theta - 1} \frac{1}{(\sqrt{2NX_0} - R + 1)^{\theta - 1}} \geq \sqrt{V_0}\,.$$

In this case, the consensus region is in practice not modified by R, but only by the decay of the far-away interaction θ (the faster the decay, the smaller the consensus region). In both cases, if θ tends to $+\infty$, we recover the result we obtained for the family of functions $(\chi_{[0,R]})_{R\in[0,+\infty]}$.

1.7 Numerical simulations

We present a series of numerical tests illustrating the main results developed throughout this chapter. We begin by describing the generic setting upon which the initial configurations of agents are determined; we follow similar ideas as those presented in [54]. We consider a system of N agents in dimension $d = 2$ with a randomly generated initial configuration of positions and velocities

$$(x^0, v^0) \in [-1,1]^{2N} \times [-1,1]^{2N} ,$$

interacting by means of the kernel (1.7) with $H = 1$, $\sigma = 1$ and $\beta = 1$. We recall that relevant quantities for the analysis of our results are given by (here we stress the dependance on x and v)

$$X[x](t) \triangleq \frac{1}{2N^2}\sum_{i,j=1}^{N}\|x_i(t)-x_j(t)\|^2 , \quad \text{and} \quad V[v](t) \triangleq \frac{1}{2N^2}\sum_{i,j=1}^{N}\|v_i(t)-v_j(t)\|^2 .$$

Notice that, once a random initial configuration has been generated, it is possible to rescale it to a desired (X_0, V_0) parametric pair, by means of

$$(x, v) = \left(\sqrt{\frac{X_0}{X[\tilde{x}]}}\tilde{x}, \sqrt{\frac{V_0}{V[\tilde{v}]}}\tilde{v}\right) ,$$

such that $(X[x], V[v]) = (X_0, V_0)$. As simulations of the trajectories have been generated by prescribing a value for the pair (X_0, V_0), which is used to rescale randomly generated initial conditions, there are slight variations on the initial positions and velocities in every model run, which can affect the final consensus direction. However our results are stated in terms of X, V, and independently of the specific initial configuration. For simulation purposes the system is integrated in time with the specific feedback control by means of a Runge-Kutta 4th-order scheme.

1.7.1 Leader-based feedback

The first case that we address is the one presented in Example 1.24, where we consider a system like (1.17) where the local mean is computed upon local information and a single leader, i.e.,

$$\overline{v}_i(t) = (1-q)v_i(t) + qv_1(t) \qquad \text{for every } i = 1, \dots, N \text{ and } t \geq 0,$$

where for convenience we have selected the first agent as the leader of the group. Figure 1.3 shows the behavior of the group of agents depending on the parameter $q \in (0, 1]$, which represents the influence of the leader in the local average. Our result asserts that for every such q, the system will converge to consensus independently of the initial configuration, which is illustrated by our numerical experiments, as shown in Figures 1.3 and 1.4. It can be observed that, the weaker the influence of the leader, the longer the group of agents takes to reach consensus.

1.7.2 Feedback under perturbed information

Next, we deal with the setting presented again in Section 1.5, by considering a system of the form (1.18) where α and β are constant functions and the feedback Δ_i is a structured perturbation written as

$$\Delta_i(t) = \frac{1}{\eta_i(t)} \sum_{j=1}^{N} \omega_{ij}(t)(v_j(t) - \overline{v}(t)) \quad \text{for every } t \geq 0.$$

In particular, we address the case where the weighting function ω_{ij} corresponds to a Cucker-Smale kernel as for the dynamics, i.e.,

$$\omega_{ij}(t) = \frac{1}{(1 + \|x_i(t) - x_j(t)\|^2)^\epsilon} \quad \text{and} \quad \eta_i(t) = \frac{1}{N} \sum_{j=1}^{N} \omega_{ij}(t).$$

In this test, we fix a large value of $\beta = 10$, representing a strong perturbation of the feedback, and a small value of $\epsilon = 1e - 5$, related to a disturbance which is distributed among all the agents. In Figures 1.5 and 1.6 , it is shown how increasing the value of α in system (1.18), representing the energy of the *correct information feedback*, induces faster consensus emergence.

1.7.3 Local feedback control

The last numerical case study presented concerns the dynamical system considered in Section 1.6, where the feedback is computed according to the local average (1.32). Simulations in Figure 1.7 illustrate the setting. From an uncontrolled system, represented by a local feedback radius $R = 0$, by increasing this quantity, partial alignment is consistently achieved, until full consensus is observed for large radii mimicking a total information feedback control.

From a theoretical perspective, this result is presented in Theorem 1.30, which describes a consensus region for a feedback control based on local averages. As it was shown in Example 1.13, that estimates for consensus regions such as the one provided by Theorem 1.11, are not sharp in many situations. In this direction, we proceed to contrast the theoretical consensus estimates with the numerical evidence. For this purpose, for a fixed number of agents, we span a large set of possible initial configurations determined by different values of (X_0, V_0). For every pair (X_0, V_0) we randomly generate a set of 20 initial conditions, and we simulate for a sufficiently large time frame. We measure consensus according to a threshold established on the final value of V; we consider that consensus has been achieved if the final value of V is lower or equal to $1e - 5$. We proceed by computing empirical probabilities of consensus for every point of our state space (X_0, V_0); results in this direction are presented in Figures 1.8 and 1.10. We first consider the simplified case of 2 agents; according to Example 1.13, for this particular case, the consensus region estimate provided by Theorem 1.11 is sharp, as illustrated by the results presented in Figure 1.8.

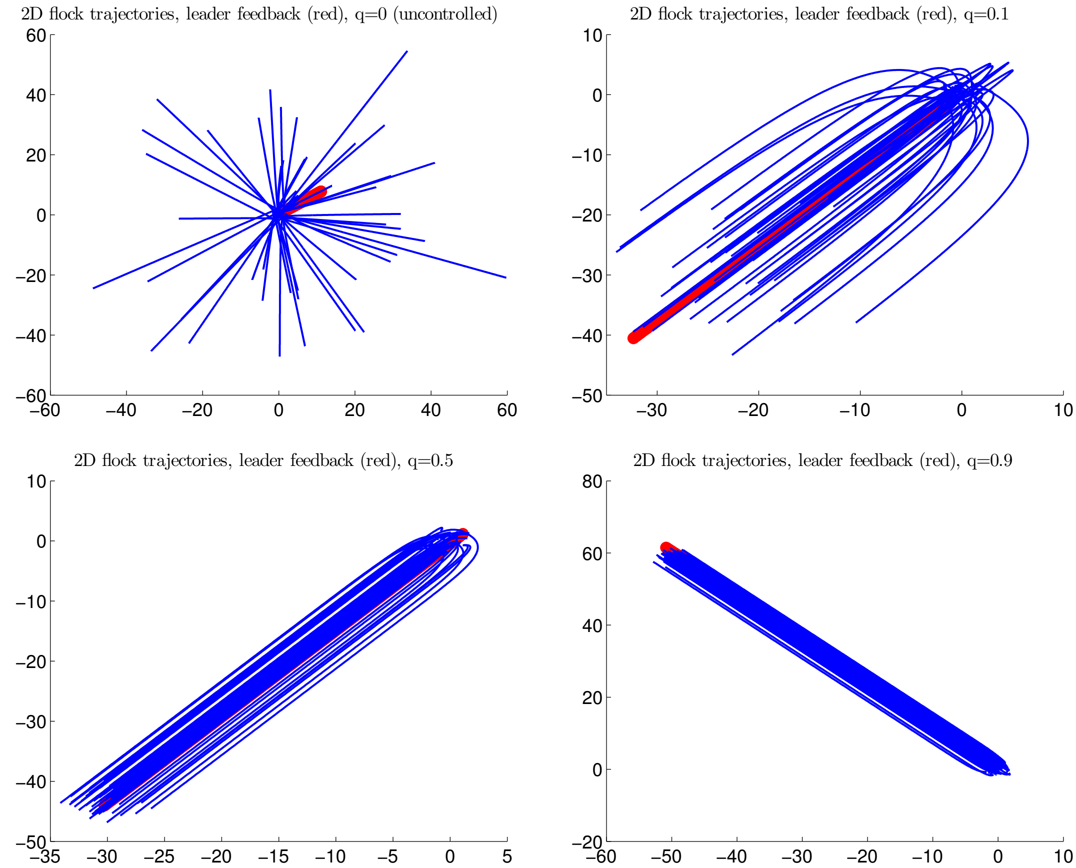


Figure 1.3: Leader-based feedback control. Simulations with 100 agents, the value q indicates the strength of the of the leader in the partial average. It can be observed how, as the strength of the leader is increased, convergent behavior is improved.

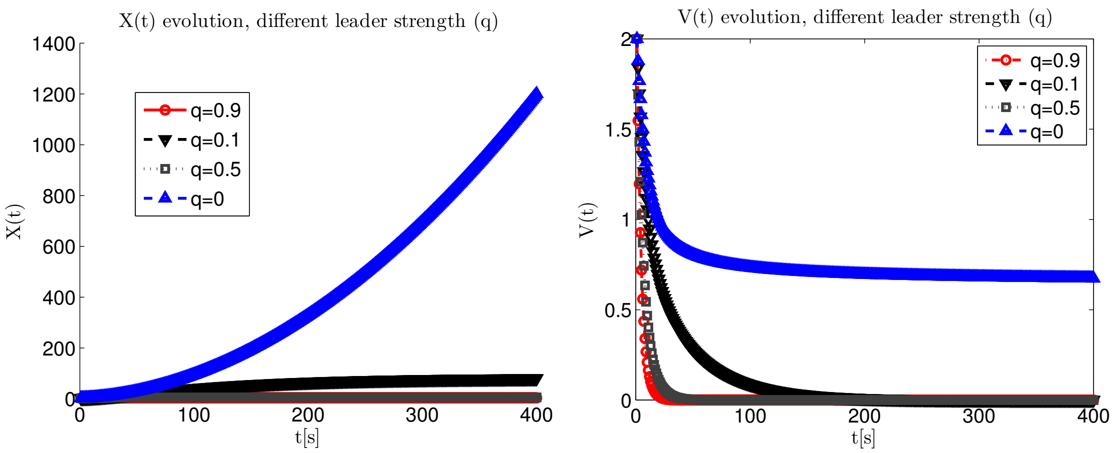


Figure 1.4: Leader-based feedback control. Simulations with 100 agents, the value q indicates the strength of the of the leader in the partial average. Evolution of X and V for the simulations in Figure 1.3.

Furthermore, it is also the case for Theorem 1.30; for $R > 0$, the predicted consensus region coincides with the numerically observed ones.

Figures 1.9 and 1.10 illustrate the case when a larger number of agents is considered. In a similar way as for Theorem 1.11, the consensus region estimate is conservative if compared with the region where numerical experiments exhibit convergent behavior. Nevertheless, Theorem 1.30 is consistent in the sense that the theoretical consensus region increases gradually as R grows, eventually covering any initial configuration, which is the case of the total information feedback control, as presented in [49, Proposition 2]. The numerical experiments also confirm this phenomena, as shown in Figure 1.9 , where contour lines showing the 80% probability of consensus for different radii locate farther from the origin as R increases.

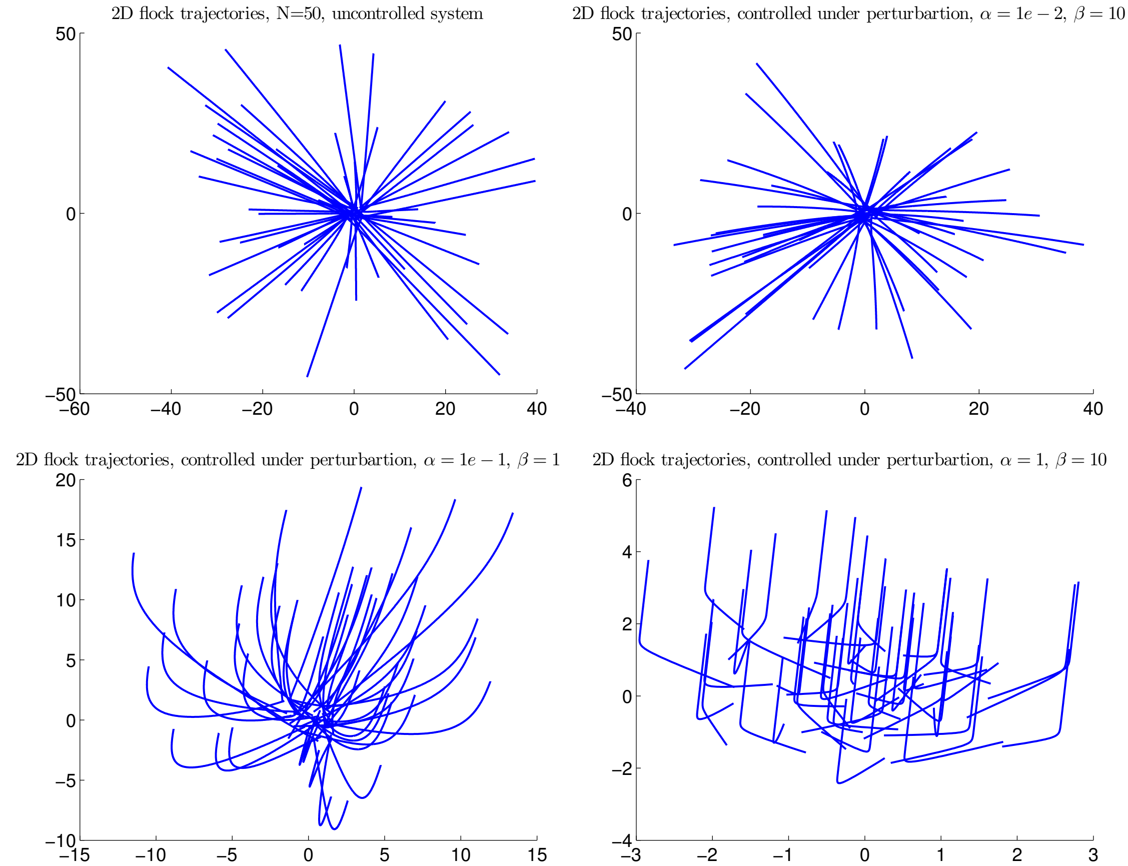


Figure 1.5: Total feedback control under structured perturbations. For a fixed strong structured perturbation term ($\beta = 10$), different energies for the unperturbed control term α generate different consensus behavior; the stronger the correct information term is, the faster consensus is achieved.

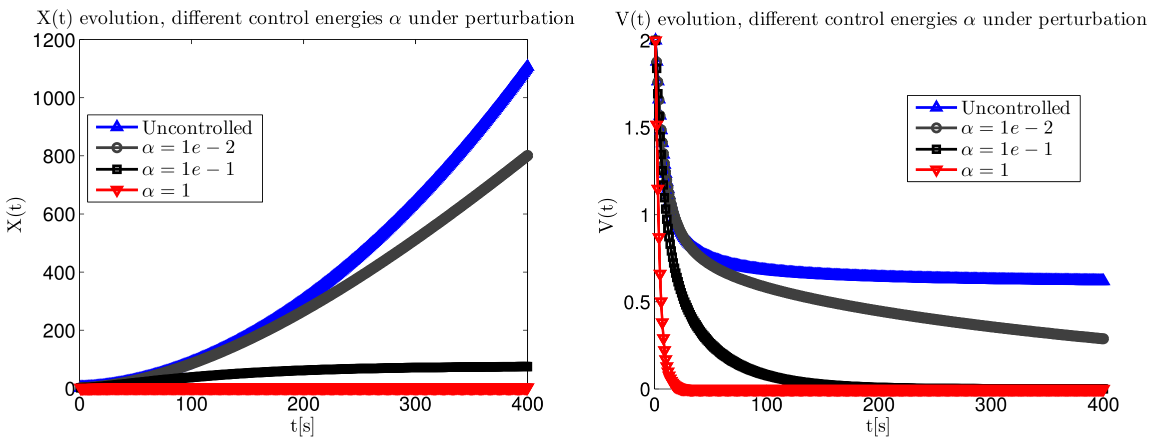


Figure 1.6: Total feedback control under structured perturbations. Evolution of X and V for the simulations in Figure 1.5.

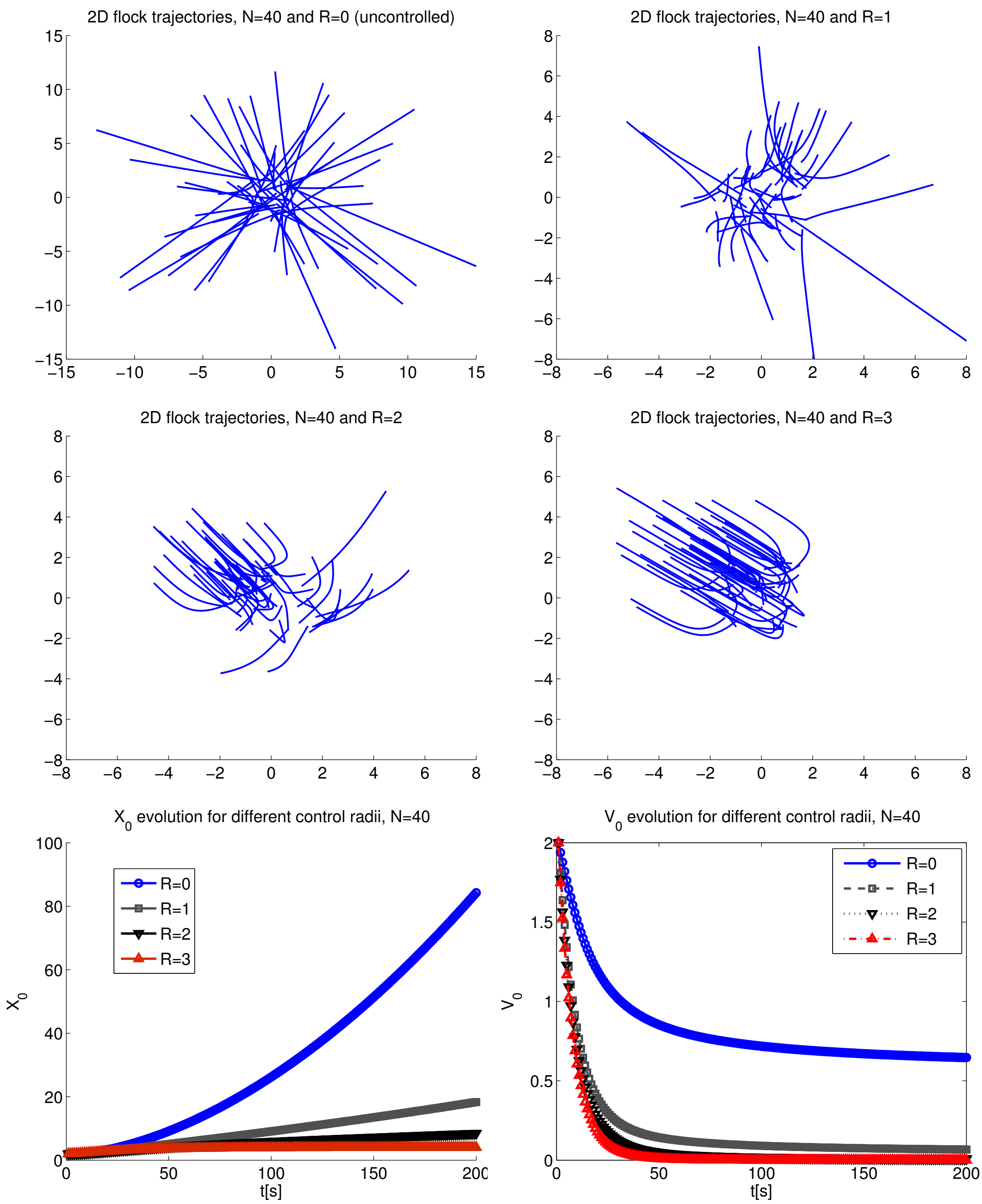


Figure 1.7: Local feedback control. Simulations with $N = 40$ agents, and different control radii R. By increasing the value of R the systems transits from uncontrolled behavior, to partial alignment, up to total, fast alignment.

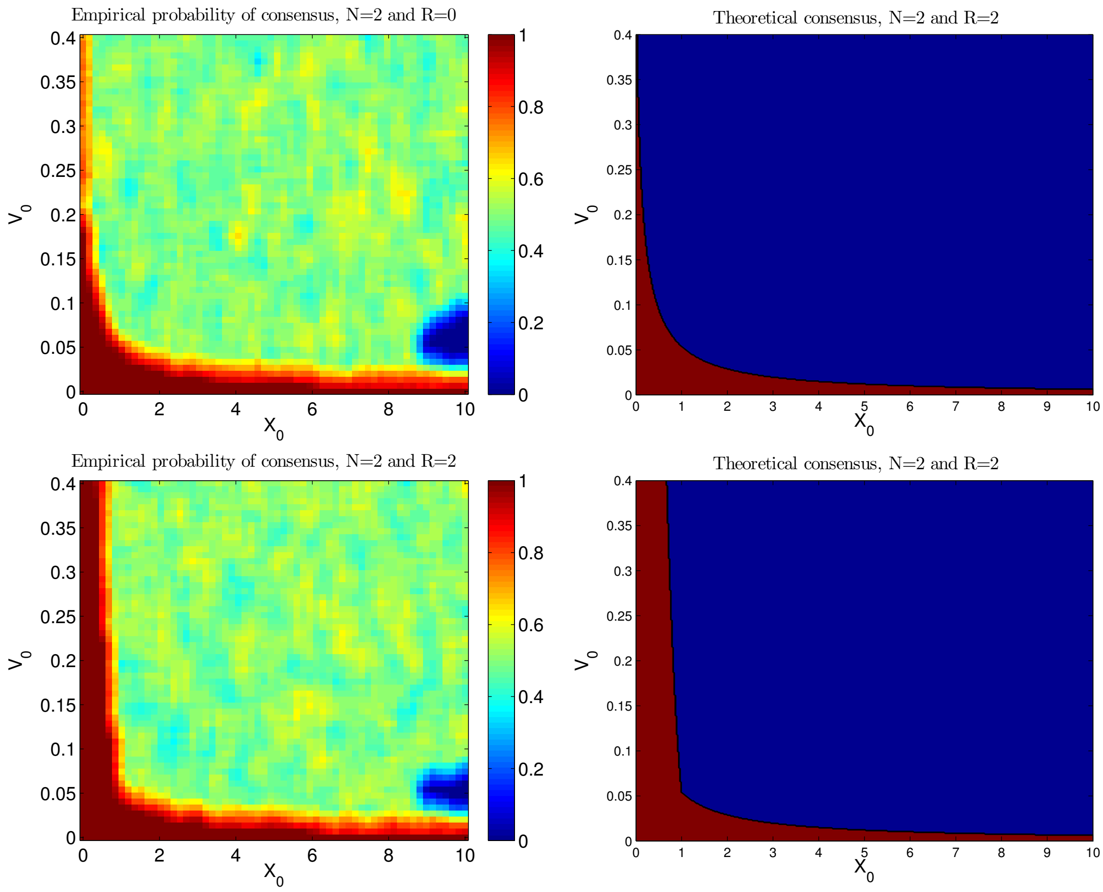


Figure 1.8: Local feedback control. Empirical consensus regions and theoretical estimates for two-agent systems.

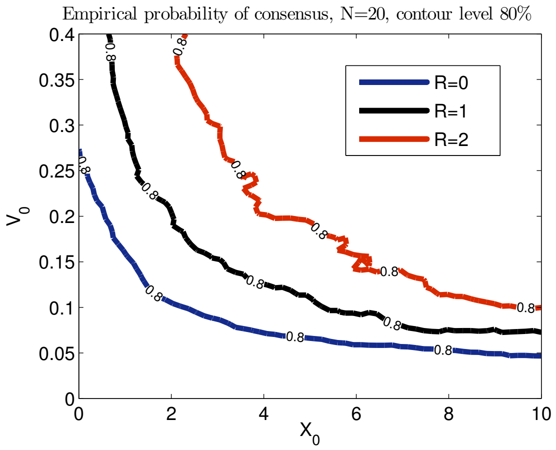


Figure 1.9: Local feedback control. Empirical contour lines for the 80% probability of consensus with different control radii.

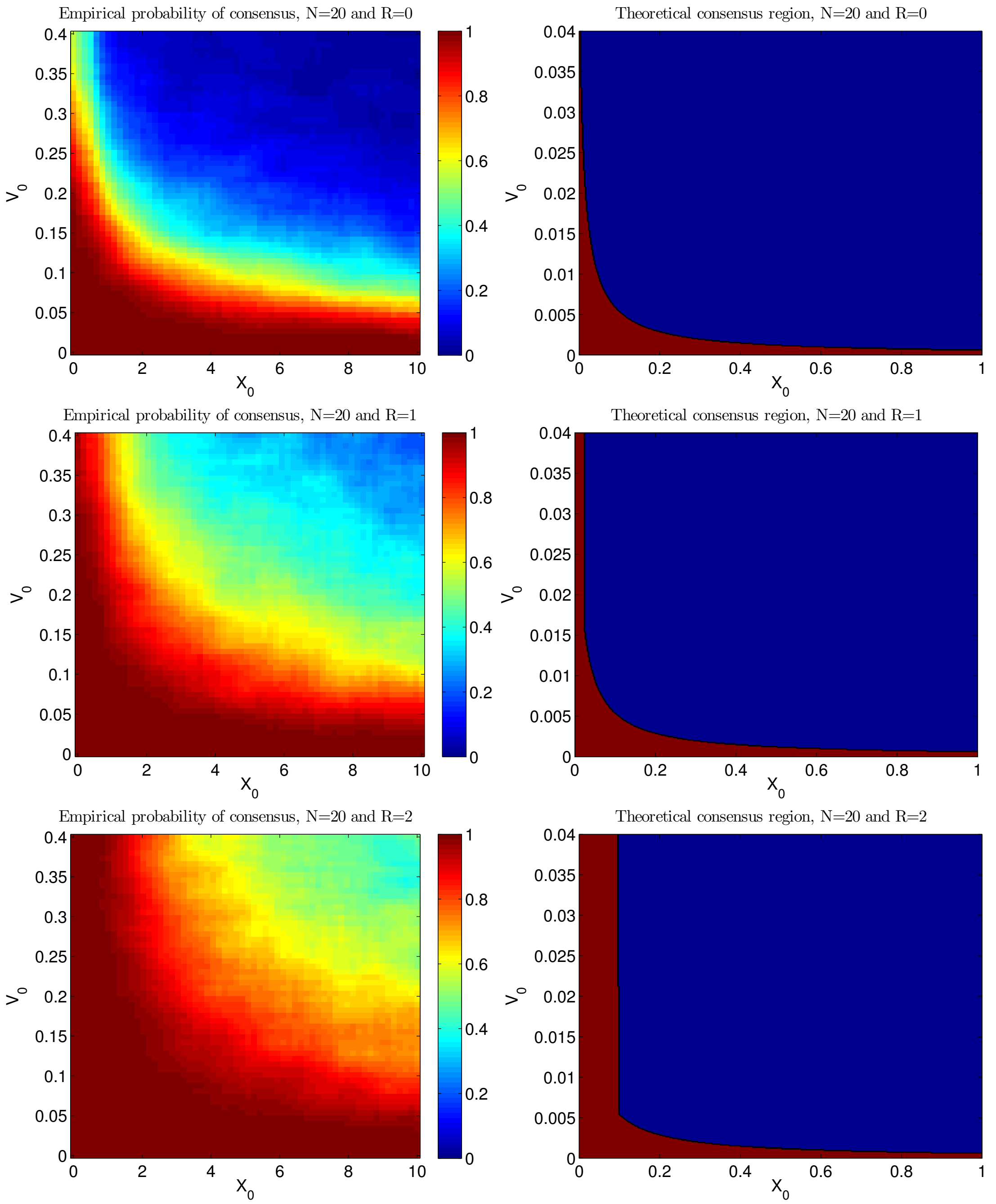


Figure 1.10: Local feedback control. Empirical consensus regions and theoretical estimates for $N = 20$ agents and different control radii R.

CHAPTER 2

Sparse control of dynamical systems

In Chapter 1 we have seen the limitations of decentralized controls in guaranteeing consensus emergence. To overcome them, in this chapter we shall introduce an external policy maker overseeing the evolution of the dynamics with the task to help the agents coordinate. The *centralized* control strategy we will design is parsimonious, optimal in terms of maximizing the decay rate of a certain Lyapunov functional related to the pattern we want to enforce, and numerically easy to implement.

2.1 Sparse control of the Cucker-Smale model

We have seen throughout Chapter 1 how difficult it is to ensure unconditional convergence to consensus for alignment models. In particular, in Section 1.6 we have proven that the addition of a local feedback does not always help: Theorem 1.30 shows that we can guarantee unconditional convergence to consensus with respect to the initial datum for dynamical systems of the form

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right) + \gamma\left(\frac{1}{|\Lambda_R(t,i)|}\sum_{j\in\Lambda_R(t,i)} v_j(t) - v_i(t)\right). \end{cases}$$

only in the case $R = +\infty$, for which the identity

$$\frac{1}{|\Lambda_R(t,i)|}\sum_{j\in\Lambda_R(t,i)} v_j(t) = \overline{v}(t)$$

holds. This means that either the agents have perfect information of the state of the entire system (so that the local mean $\overline{v}_i$ is equal to the true mean $\overline{v}$) or, as the numerical simulations in Section 1.7.3 show, there are situations where the agents are not able to converge to consensus. As already pointed out in Section 1.3, this is a very strong requirement to ask for, and not many real-life scenarios are able to support it. Consider, for instance, the case of an assembly of people trying to reach an unanimous decision, like the European Union Council: since the extra term can be interpreted as an additional desire of each agent to agree with people whose goal is near to his, the requirement $R = +\infty$ corresponds to asking that all the individual goals are close, i.e., all agents pursue the same end. A truly imaginative world indeed! We are thus facing an inherent, severe limitation of the decentralized approach.
To overcome this apparent dead-end, let us write $u_i(t) = \gamma(\overline{v}(t) - v_i(t))$, i.e.,

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right) + u_i(t). \end{cases} \tag{2.1}$$

Instead of interpreting u_i as a *decentralized* force, let us consider it as an *external* force from an outside source acting on the system to help it coordinate. This new approach casts a completely different light on the problem: with respect to the example considered before, is like introducing a moderator heading the discussion, who can make pressure on the participants to the council to speed up the process of fruitful decision-making. Adding an external figure implementing intervention policies broadens further the expressive power of the problem: indeed, since we are in principle no more tied to specific interventions of the form $u_i(t) = \gamma(\overline{v}(t) - v_i(t))$, this setting enables us to ask ourselves the following question

(Q) *given a set of constraints, which control u is the best to reach a specific goal?*

In this section, we shall study a specific instance of this very general issue in the case of system (2.1). In our setting, the constraints shall be

(*i*) we are not able to forecast the evolution of the system (for computational reasons or for lack of knowledge of the interaction mechanism, for instance). This means that, since we do not know how the system may react in the future to our intervention, we focus our attention on *locally optimal* control strategies.

(*ii*) the control should act on the least amount of agents possible at any time;

(*iii*) there is a maximal amount of resources $M > 0$ that the central policy maker can spend at any given time for the intervention.

For the time being, our goal is again alignment, hence we seek for a control u for which the associated solution to system (2.1) tends to consensus in the sense of Definition 1.4. We have already seen in Chapter 1 that an effective criterion for consensus emergence is the minimization of the Lyapunov functional V: if we are able to prove that our control strategy is able to drive V below the threshold level given by Theorem 1.11, we have automatically consensus emergence (see Figure 2.1). The maximization of the decay rate of V is a locally optimal criterion, and hence compatible with point (i); a globally optimal one, available when predictions on the possible evolution of the system are possible, shall be discussed in Chapter 6.

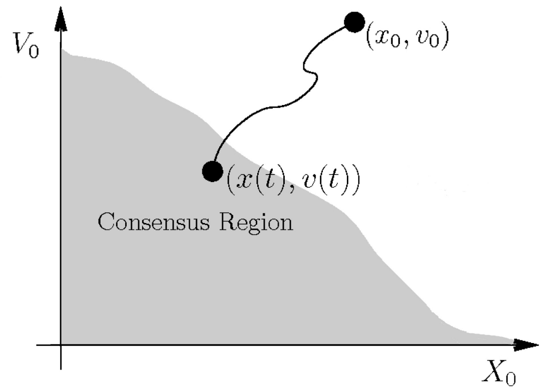


Figure 2.1: Steering the system to a point fulfilling the conditions of Theorem 1.11.

However this is not enough. Even if controls can be successful in enforcing alignment, they may be unfeasible in practice. Undesirable controls are, for example, those that force the central coordinator to interact at every instant with *almost all* the agents in the system, since keeping such a large communication network always active can be extremely expensive and highly dispersive. This motivates point (ii) and is the reason why we look for interventions that target the fewest number of agents at any given time.

To translate these words into a mathematical statement we borrow a leaf from *compressed sensing* (see as a reference [48, 96, 105]), a branch of mathematical signal processing which provides optimal results on the ability of the ℓ_1-norm minimization of robustly recovering *sparse* solutions, i.e., with very few non-zero entries. Therefore, one can obtain control strategies $u = (u_1, \ldots, u_N) \in \mathbb{R}^{dN}$ acting on very few agents by requiring the following mixed $\ell_1^N - \ell_2^d$-norm of u

$$\sum_{i=1}^{N} \|u_i\|_{\ell_2^d}$$

to be as small as possible: the ℓ_1-norm constraint on the number of components relative to the agents will select controls that at every instant interact with very few of them. The above mixed norm has been already used, for instance, in [98, 104]. The use of ℓ_1-norms to penalize controls was first introduced in the seminal paper [74] to model linear fuel consumption, while lately the use of L^1 minimization in optimal control problems with partial differential equation has become very popular, for instance in the modeling of optimal placing of sensors [56, 66, 126, 167, 178].

By incorporating also the constraint on the maximal amount of available resources M given by point (iii), we obtain the following definition of *admissible controls.*

Definition 2.1 (Admissible controls)**.** A measurable function $u = (u_1, \ldots, u_N) : \mathbb{R}_+ \to$

$\mathbb{R}^{dN}$ is an *admissible control* if it satisfies

$$\sum_{i=1}^{N} \|u_i(t)\| \leq M \quad \text{for every } t \geq 0. \tag{2.2}$$

The following result shows that the problem of finding admissible controls steering the system to consensus is well-posed.

Proposition 2.2 ([49, Proposition 2]). *Fix $M > 0$, an initial condition $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, and $0 < \alpha \leq M/(N\sqrt{V_0})$. Then, the feedback control defined pointwise in time as*

$$u(t) = -\alpha v^{\perp}(t) \quad \text{for every } t \geq 0,$$

is admissible. Moreover, system (2.1) admits a unique solution associated to u which tends to consensus.

Proof. Let $(x, v) : \mathbb{R}_+ \to \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ be a solution of system (2.1) with u as in the statement. Since with this choice of u we retrieve system (1.17), such solution exists and is unique by Remark B.4 and tends to consensus by Theorem 1.11. Moreover, since V is decreasing by Corollary 1.19, it holds

$$\sum_{i=1}^{N} \|u_i(t)\| \leq \sqrt{N}\sqrt{\sum_{i=1}^{N} \|u_i(t)\|^2} = \alpha\sqrt{N}\sqrt{\sum_{i=1}^{N} \|v_i^{\perp}(t)\|^2} = \alpha N\sqrt{V(t)} \leq \alpha N\sqrt{V_0} \leq M,$$

which implies the admissibility of the control. □

2.1.1 Sparse feedback controls

As already noted in [49], Proposition 2.2 has merely a theoretical value: the feedback control $u = -\alpha v^{\perp}$ is not convenient for practical purposes, since it requires the external policy maker to interact at every instant with all the agents in order to steer the system to consensus, a procedure that requires a large amount of instantaneous communications. For this reason, we choose controls according to a specific variational principle leading to a componentwise sparse stabilizing feedback law.

Definition 2.3. For every $M > 0$ and every $(x, v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, let $U(x, v)$ be the set of solutions of the variational problem

$$\min_{u \in \mathbb{R}^{dN}} \left(B(v, u) + \gamma(B(x, x)) \frac{1}{N} \sum_{i=1}^{N} \|u_i\| \right) \quad \text{subject to } \sum_{i=1}^{N} \|u_i\| \leq M,$$

where the *threshold functional* γ is defined as

$$\gamma(X) \triangleq \int_{\sqrt{X}}^{\infty} a(\sqrt{2N}r)dr. \tag{2.3}$$

Each value of $\gamma(B(x, x))$ yields a partition of $\mathbb{R}^{dN} \times \mathbb{R}^{dN}$ into four disjoint sets:

$\mathcal{P}_1 \triangleq \{(x, v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN} : \max_{1 \leq i \leq N} \left\|v_i^{\perp}\right\| < \gamma(B(x, x))^2\}$,

$\mathcal{P}_2 \triangleq \{(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN} : \max_{1\le i\le N} \|v_i^\perp\| = \gamma(B(x,x))^2$ and $\exists k \ge 1$ and $i_1,\dots,i_k \in \{1,\dots N\}$ such that $\|v_{i_1}^\perp\| = \dots = \|v_{i_k}^\perp\|$ and $\|v_{i_1}^\perp\| > \|v_j^\perp\|$ for every $j \notin \{i_1,\dots,i_k\}\}$,

$\mathcal{P}_3 \triangleq \{(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN} : \max_{1\le i\le N} \|v_i^\perp\| > \gamma(B(x,x))^2$ and $\exists! i \in \{1,\dots N\}$ such that $\|v_i^\perp\| > \|v_j^\perp\|$ for every $j \neq i\}$,

$\mathcal{P}_4 \triangleq \{(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN} : \max_{1\le i\le N} \|v_i^\perp\| > \gamma(B(x,x))^2$ and $\exists k > 1$ and $i_1,\dots,i_k \in \{1,\dots N\}$ such that $\|v_{i_1}^\perp\| = \dots = \|v_{i_k}^\perp\|$ and $\|v_{i_1}^\perp\| > \|v_j^\perp\|$ for every $j \notin \{i_1,\dots,i_k\}\}$,

Moreover, since we are minimizing $B(u,v) = B(u,v^\perp)$, it is easy to see that, for every $(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ and every element $u(x,v) = (u_1(x,v),\dots,u_N(x,v))^T \in U(x,v)$ there exist nonnegative real numbers $\varepsilon_i \ge 0$ such that, for every $i = 1,\dots,N$, it holds

$$u_i(x,v) = \begin{cases} -\varepsilon_i \dfrac{v_i^\perp}{\|v_i^\perp\|} & \text{if } \|v_i^\perp\| \neq 0, \\ 0 & \text{if } \|v_i^\perp\| = 0, \end{cases} \tag{2.4}$$

where $0 \le \sum_{i=1}^N \varepsilon_i \le M$. The values of the ε_i's can be determined on the basis of which partition (x,v) belongs to:

- if $(x,v) \in \mathcal{P}_1$ then $\varepsilon_i = 0$ for every $i = 1,\dots,N$;
- if $(x,v) \in \mathcal{P}_2$ then indicating with $i_1,\dots,i_k$ the indexes such that $\|v_{i_1}^\perp\| = \dots = \|v_{i_k}^\perp\| = \gamma(B(x,x))$ and $\|v_{i_1}^\perp\| > \|v_j^\perp\|$ for every $j \notin \{i_1,\dots,i_k\}$, we have $\varepsilon_j = 0$ for every $j \notin \{i_1,\dots,i_k\}$;
- if $(x,v) \in \mathcal{P}_3$ then, indicating with i the only index such that $\|v_i^\perp\| > \|v_j^\perp\|$ for every $j \neq i$, we have $\varepsilon_i = M$ and $\varepsilon_j = 0$ for every $j \neq i$;
- if $(x,v) \in \mathcal{P}_4$ then, indicating with $i_1,\dots,i_k$ the indexes such that $\|v_{i_1}^\perp\| = \dots = \|v_{i_k}^\perp\|$ and $\|v_{i_1}^\perp\| > \|v_j^\perp\|$ for every $j \notin \{i_1,\dots,i_k\}$, we have $\varepsilon_j = 0$ for every $j \notin \{i_1,\dots,i_k\}$ and $\sum_{\ell=1}^k \varepsilon_{i_\ell} = M$.

Notice that any control $u(x,v) \in U(x,v)$ acts as an additional force which pulls agents towards having the same mean consensus parameter. The imposition of the $\ell_1^N - \ell_2^d$-norm constraint has the function of enforcing *sparsity*: from the observation above clearly follows that

$$U|_{\mathcal{P}_1} = \{0\} \quad \text{and} \quad U|_{\mathcal{P}_3} = \{(0,\dots,0,-Mv_i^\perp/\|v_i^\perp\|,0,\dots,0)^T\},$$

for some unique $i \in \{1,\dots,N\}$, i.e., the restrictions of U to $\mathcal{P}_1$ and to $\mathcal{P}_3$ are single-valued. However, even if not all controls belonging to U are sparse, there exist selections with maximal sparsity.

Definition 2.4 ([49, Definition 4])**.** We select the *sparse feedback control* $u(x,v) \in U(x,v)$ according to the following criterion:

- if $\max_{1\le i\le N}\left\|v_i^{\perp}\right\| \le \gamma(B(x,x))^2$, then $u(x,v)=0$;
- if $\max_{1\le i\le N}\left\|v_i^{\perp}\right\| > \gamma(B(x,x))^2$, denote with $\hat{\imath}(x,v)\in\{1,\dots,N\}$ the smallest index such that

$$\left\|v_{\hat{\imath}(x,v)}^{\perp}\right\| = \max_{1\le i\le N}\left\|v_i^{\perp}\right\|.$$

Then

$$u_j(x,v) \triangleq \begin{cases} -M\dfrac{v_{\hat{\imath}(x,v)}^{\perp}}{\|v_{\hat{\imath}(x,v)}^{\perp}\|} & \text{if } j=\hat{\imath}(x,v),\\ 0 & \text{otherwise.}\end{cases}$$

Unfortunately, the existence of solutions of system (2.1) with the control u as in Definition 2.4 is still an open problem. We shall discuss the reasons in the next section.

2.1.2 Properties of the sparse feedback selection

The feedback control of Definition 2.4 has several other good properties, apart from being sparse. The next result shows that, if a solution of system (2.1) with control u from Definition 2.4 exists, then u is the instantaneous maximizer of the decay rate of the Lyapunov functional V in the set U.

Proposition 2.5. *The sparse feedback control of Definition 2.4 is for every $t\ge 0$ an instantaneous minimizer of*

$$\mathcal{D}(t,u)=\frac{d}{dt}V(t)$$

over all possible feedback controls in $U(x(t),v(t))$. The Lyapunov functional V of the solution associated to it satisfies

$$\sqrt{V(t)}\le\sqrt{V(0)}-\frac{M}{N}t \quad \textit{for every } t\ge 0.$$

Proof. Let $u(t)\triangleq u(x(t),v(t))\in U(x(t),v(t))$. Similar computations to (1.20) for system (2.1) yield

$$\begin{aligned}\frac{d}{dt}V(t)=&\frac{2}{N^2}\sum_{i=1}^N\sum_{j=1}^N a\left(\|x_i(t)-x_j(t)\|\right)(v_j(t)-v_i(t))\cdot v_i^{\perp}(t)+\\&+\frac{2}{N}\sum_{i=1}^N u_i(t)\cdot v_i^{\perp}(t)-\frac{2}{N}\sum_{i=1}^N\dot{\overline{v}}(t)\cdot v_i^{\perp}(t).\end{aligned}\tag{2.5}$$

Since $\sum_{i=1}^N\dot{\overline{v}}(t)\cdot v_i^{\perp}(t)=0$ by (1.10), in order to minimize $\mathcal{D}(t,u)$ with respect to u we have to work on the term $\sum_{i=1}^N u_i(t)\cdot v_i^{\perp}(t)$. Now, since from (2.4) it follows

$$\frac{2}{N}\sum_{i=1}^N u_i(t)\cdot v_i^{\perp}(t)=-\sum_{\left\{i\in\{1,\dots,N\}:v_i^{\perp}\neq 0\right\}}\varepsilon_i\left\|v_i^{\perp}\right\|,$$

minimizing $\mathcal{D}(t,u)$ (i.e., maximizing the decay rate of V) is equivalent to solve

$$\max \sum_{\{i\in\{1,\ldots,N\}:v_i^\perp\neq 0\}} \varepsilon_i \left\| v_i^\perp(t) \right\| \quad \text{subject to} \quad \sum_{i=1}^N \varepsilon_i \leq M. \tag{2.6}$$

But notice that, if the index j is such that $\left\|v_j^\perp(t)\right\| \geq \|v_i^\perp(t)\|$ for all i, then the following upper bound holds

$$\begin{aligned}\sum_{i=1}^N \varepsilon_i \left\| v_i^\perp(t)\right\| &\leq \left\|v_j^\perp(t)\right\| \sum_{i=1}^N \varepsilon_i \\ &\leq M \left\|v_j^\perp(t)\right\|.\end{aligned}$$

This clearly implies that the sparse feedback control of Definition 2.4 is a maximizer of (2.6), since it attains this maximum. Moreover, notice that the solution of (2.6) is unique whenever for any $t \geq 0$ there exists a unique j such that $\|v_j^\perp(t)\| > \|v_i^\perp(t)\|$ for all $i \neq j$. Consider now the sparse feedback control u and plug it into (2.5). What we get is

$$\begin{aligned}\frac{d}{dt}V(t) &\leq -\frac{2M}{N}\left\| v^\perp_{\hat{\imath}(x(t),v(t))}(t)\right\| \\ &= -\frac{2M}{N}\sqrt{V(t)}.\end{aligned} \tag{2.7}$$

Integrating between 0 and t yields the desired linear decay rate estimate of $\sqrt{V}$. □

An immediate consequence of the decay estimate given by (2.7) is that we can estimate a priori an upper bound for X, which in practice implies consensus, as seen in Corollary 1.5. This technical lemma can be found also in [49], but with a slightly different argument we could improve the inequalities presented there and get rid of an N^2 factor. This shall play a major role in Chapter 3, when sharp estimates of the time of entrance in the consensus region of controlled Cucker-Smale systems are crucial.

Lemma 2.6. *If there exists $\eta > 0$ and $T > 0$ such that*

$$\frac{d}{dt}V(t) \leq -\eta\sqrt{V(t)}$$

for almost every $t \in [0,T]$, then

$$V(t) \leq \left(\sqrt{V(0)} - \frac{\eta}{2}t\right)^2 \qquad \text{and} \qquad X(t) \leq 2X(0) + \frac{2}{\eta^2}V(0)^2$$

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

Proof. We can assume $V(t) > 0$ for every $t \in [0,T]$, otherwise we are already in consensus. Integrating the assumption one has

$$\int_0^t \frac{d}{ds}V(s)\frac{1}{\sqrt{V(s)}}\, ds \leq -\eta t,$$

and hence

$$\sqrt{V(t)} - \sqrt{V(0)} = \frac{1}{2}\int_0^t \frac{d}{ds}V(s)\frac{1}{\sqrt{V(s)}}\ ds \leq -\frac{\eta}{2}t.$$

Furthermore, to prove the second statement we observe

$$\int_0^t \sqrt{V(s)}\ ds \leq -\frac{1}{\eta}\int_0^t \frac{d}{ds}V(s)\ ds = -\frac{1}{\eta}(V(t)-V(0)) \leq \frac{1}{\eta}V(0).$$

On the other hand, using the (vector-valued) Minkowski inequality in the second step

$$\begin{aligned}
\sqrt{X(t)} &= \left(\frac{1}{2N^2}\sum_{i,j=1}^N \|x_i(t)-x_j(t)\|^2\right)^{1/2} \\
&\leq \left(\frac{1}{2N^2}\sum_{i,j=1}^N \|x_i(0)-x_j(0)\|^2\right)^{1/2} + \left(\frac{1}{2N^2}\sum_{i,j=1}^N \left(\int_0^t \|v_i(s)-v_j(s)\|\ ds\right)^2\right)^{1/2} \\
&\leq \sqrt{X(0)} + \int_0^t \left(\frac{1}{2N^2}\sum_{i,j=1}^N \|v_i(s)-v_j(s)\|^2\right)^{1/2} ds \\
&= \sqrt{X(0)} + \int_0^t \sqrt{V(s)}\ ds \\
&\leq \sqrt{X(0)} - \frac{1}{\eta}\int_0^t \frac{d}{ds}V(s)ds \\
&\leq \sqrt{X(0)} + \frac{1}{\eta}V(0)
\end{aligned}$$

and furthermore by $(x+y)^2 \leq 2x^2+2y^2$ it follows

$$X(t) \leq 2X(0) + \frac{2}{\eta^2}V^2(0),$$

which concludes the proof. □

With the estimate of the decay rate of the functional V given by (2.7), it is now very easy to see that the sparse feedback control forces the system to satisfy the condition

$$V(t) \leq \gamma(X(t)) = \int_{\sqrt{X(t)}}^{+\infty} a(\sqrt{2N}r)dr$$

in finite time. Since, by Definition 2.4, after this event happens the control shuts off, it follows from Theorem 1.11 that the system has entered the consensus region, and thus tends to consensus. More precisely, we can estimate *a priori* that the system will enter the consensus region before the time horizon

$$T_0 \triangleq \frac{N}{M}\left(\sqrt{V(0)} - \gamma(\overline{X})\right),$$

where $\overline{X} \triangleq 2X(0) + \frac{N^2}{2M^2}V(0)^2$. Indeed, by Lemma 2.6 and the decreasing monotonicity of

V and γ, for every $t \geq T_0$ it holds

$$\sqrt{V(t)} \leq \sqrt{V(T_0)} \leq \sqrt{V(0)} - \frac{M}{N} T_0 = \gamma(\overline{X}) \leq \gamma(X(t)).$$

Another direct consequence of Lemma 2.5 is that, for Cucker-Smale systems, a feedback stabilization is most effective if all the attention of the controller is focused on very few (actually in this case only one) agents at each time. This also means that, despite the fact that the external policy maker may have few resources at disposal and can allocate them at each time only on very few key players in the system, it is always possible to effectively stabilize the dynamics to return to energy levels where the system tends autonomously to consensus. This result is perhaps surprising if confronted with the more intuitive strategy of controlling more, or even all, agents at the same time. This let us answer to the question (Q) raised at the beginning of this section as follows:

(A) *under the constraints* $(i) - (iii)$, *sparse is better.*

However, the sparse feedback control can be in general highly irregular in time, and this is the reason why it is so difficult to prove the well-posedness of the sparse control problem. For instance, if there are only three agents and their consensus parameters form an equiangular and equinormal set of vectors at a given time t, then u is not pointwise computable after t because of chattering effects. A method to avoid chattering in such trajectories is the use of *sampling solutions*, first introduced in [64].

Definition 2.7. Let $U \subseteq \mathbb{R}^m$ and consider $f : \mathbb{R}^m \times U \to \mathbb{R}^m$ such that $f(\xi, u)$ is continuous in ξ and u as well as locally Lipschitz in ξ uniformly on every compact subset of $\mathbb{R}^m \times U$. Given a feedback control function $u : \mathbb{R}^m \to U$, a sampling time $\tau > 0$, and an initial datum $\xi^0 \in \mathbb{R}^m$ we define the *sampling solution associated with the sampling time* τ of the differential system

$$\begin{cases} \dot{\xi}(t) = f(\xi(t), u(\xi(t))), \\ \xi(0) = \xi^0, \end{cases}$$

as the piecewise $\mathcal{C}^1$-function $\xi : [0, T] \to \mathbb{R}^m$ solving

$$\dot{\xi}(t) = f(\xi(t), \widetilde{u}(t))$$

in the interval $t \in [n\tau, (n+1)\tau]$ recursively for $n \in \mathbb{N}$, where $\widetilde{u}(t) \triangleq u(\xi(n\tau))$ is constant for $t \in [n\tau, (n+1)\tau]$. As the initial value $\xi(n\tau)$ we use the endpoint of the solution of the preceding interval, and start with $\xi(0) = \xi^0$.

This notion of solution has the clear advantage of being very down-to-earth: in practice, controls cannot be updated instantaneously but require a minimal interval of time between two switchings. Moreover, there is no need to worry about existence of sampling solutions, since in this case controls are piecewise constant. The following result shows that, if the sampling time τ is sufficiently small (i.e., the controller is sufficiently reactive) then sampled sparse feedback controls are effective in enforcing consensus emergence.

Theorem 2.8 ([49, Theorem 3]). *Fix $M > 0$, $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, and let u be the sparse feedback control introduced in Definition 2.4. Denote by $(x_\tau(\cdot), v_\tau(\cdot))$ the sampling solution of system* (2.1) *associated to the control u, the sampling time τ and the initial condition (x^0, v^0). Then there exists $\tau_0 > 0$ small enough such that for all $\tau \in [0, \tau_0]$, $(x_\tau(\cdot), v_\tau(\cdot))$ reaches the consensus region in finite time.*

Unfortunately, it is not true that, as $\tau \to 0$, the above sampled control strategy converges to the sparse feedback control of Definition 2.4: nonetheless, it is not difficult to show that for every $t \geq 0$ the limit control belongs to the set $U(x(t), v(t))$ introduced in Definition 2.3. By exploiting this fact, it is possible to prove Theorem 2.9 below, which shows that, as the sampling time becomes smaller and smaller, the sampling solutions of (2.1) associated with the sparse feedback controls approach uniformly a Filippov solution (that is, an absolutely continuous function, see [100]) of the differential inclusion

$$(\dot{x}(t), \dot{v}(t)) \in F(x(t), v(t)) \quad \text{for every } t \geq 0, \tag{2.8}$$

where the set-valued functional F is defined as

$$F(x, v) \triangleq \{(v, L(x)v + u) : u \in U(x, v)\},$$

and $L(x)$ is the Laplacian of the matrix $(a(\|x_i - x_j\|)/N)_{i,j=1}^N$.

Theorem 2.9 ([49, Theorem 2]). *Fix $M > 0$, $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, let u be the sparse feedback control introduced in Definition 2.4, and let $(x_\tau(\cdot), v_\tau(\cdot))$ be as in Theorem 2.8. Then, as τ tends to 0, $(x_\tau(\cdot), v_\tau(\cdot))$ tends uniformly to a Filippov solution of the differential inclusion* (2.8)*. Moreover, this solution tends to consensus.*

Hence, we can establish the well-posedness of the control problem of system (2.1) for a control u belonging to the set U which, as already noticed, contains the sparse feedback control of Definition 2.4. In addition, the above results show that the stabilization of a Cucker-Smale system by means of sparse controls is *unconditional*, i.e., it does not depend on the initial conditions of the dynamics.

2.2 Sparse control of the Cucker-Dong model

In the next sections we shall see how the sparse feedback control strategy previously developed has far more reaching potential, as it can address also situations which do not match the structure (1.8), like the Cucker and Dong model of cohesion and avoidance introduced in [80]. This system has the form

$$\begin{cases} \dot{x}(t) = v(t), \\ \dot{v}(t) = -(L^a(x(t)) - L^f(x(t)))x(t) - v(t)b(t), \end{cases} \tag{2.9}$$

where $L^a(x(t))$ and $L^f(x(t))$ are graph-Laplacians associated to competing avoidance and cohesion forces respectively, and $b(t)$ is a vector-valued function modeling dampening. Notice that, in contrast to the Cucker-Smale system (1.8), now the Laplacians are acting on the variable x and not anymore on v, mixing the dynamics of the two components of the state. Similar models considering attraction, repulsion and other effects, such as alignment or self-drive, appear in the recent literature and they seem effectively describing realistic situations of conditional pattern formation, see, e.g., some of the most related contributions [52, 60, 79, 97].

We shall introduce this model in more detail below, but we can anticipate that, under certain conditions on the attraction and repulsion forces, if the total energy of the system (2.9), composed of the sum of its kinetic and potential parts, is below a certain critical threshold, then such systems are known to converge autonomously to the stable configuration of keeping confined and collision avoiding in space uniformly in time, see [80, Theorem 2.1]. This is the pattern we want to enforce when dealing with this model, i.e., *cohesiveness and collision avoidance*, opposed to the desired alignment pattern of the Cucker-Smale system (1.6)

If the energy is above such a critical level, then the space coherence can be lost, and this introduces us to our main goal: we show that in the latter situation of lost self-organization, one can nevertheless steer the system (2.9) to return to stable energy levels by means of sparse feedback controls similar to those of Definition 2.4. This is another remarkable example, beside the one already presented in Section 2.1, of how certain social systems are naturally prone to sparse stabilization, explaining the effectiveness of parsimonious interventions of governments in societies.

Let us stress that, although here we follow a conceptually similar path as in the previous section, the structural differences between the models (1.8) and (2.9) are such that, not only the analysis of this system requires several nontrivial a priori estimates for stability, collected below in Section 2.3, which were not necessary for (1.8), but also the final result turns out to differ substantially. In particular, while the stabilization of Cucker-Smale systems by means of sparse feedback controls is unconditional with respect to the initial conditions (see Theorem 2.9), for the Cucker-Dong model our analysis guarantees stabilization only within certain total energy levels, which is suggesting that also stabilization can be conditional. However, the numerical experiments reported in Section 2.6 suggest that it is possible to exceed such an upper energy barrier in many cases, even if there are pathological situations for which there is no hope to steer the agents towards a cohesive configuration.

2.2.1 The model

The Cucker-Dong model [80] is given by the following system of differential equations

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = -b_i(t)v_i(t) + \displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|^2\right)(x_j(t) - x_i(t)) + \\ \qquad + \displaystyle\sum_{j \neq i}^{N} f\left(\|x_i(t) - x_j(t)\|^2\right)(x_i(t) - x_j(t)), \end{cases} \quad i = 1, \dots, N, \tag{2.10}$$

The evolution is governed by an *attraction* force, modeled by a function $a : \mathbb{R}_+ \to \mathbb{R}_+$, which is, for some fixed constant $H > 0$ and $\beta \geq 0$, of the form

$$a(r) = \frac{H}{(1+r)^{\beta}},$$

(notice that here we have r in place of r^2, since we write $a(\|x_i - x_j\|^2)$ in place of $a(\|x_i - x_j\|)$, hence a has the same form as (1.7)), though in general any Lipschitz-continuous, nonincreasing function with maximum in $a(0)$ suffices. This force is counteracted by a *repulsion* given by a locally Lipschitz continuous or $\mathcal{C}^1$, nonincreasing function $f : (0, +\infty) \to \mathbb{R}_+$. We request that

$$\int_\delta^{+\infty} f(r)\, dr < +\infty, \qquad \text{for every } \delta > 0, \tag{2.11}$$

A typical example of such a function is $f(r) = r^{-p}$ for every $p > 1$. The uniformly continuous, bounded functions $b_i : \mathbb{R}_+ \to [0, \Lambda]$, $i = 1, \ldots, N$, for a given $\Lambda \geq 0$, are interpreted as a friction which helps the system to stay confined.

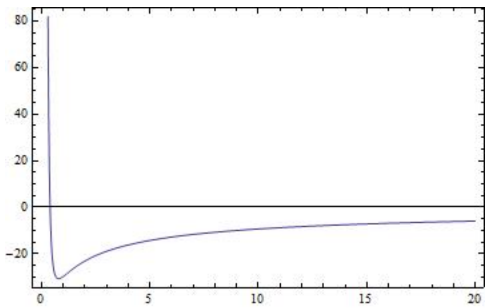


Figure 2.2: Sum of the attraction and repulsion forces $h(r) = f(r) - a(r)$ as a function of the distance $r > 0$. The parameters here are $H = 50$, $\beta = 0.7$, and $p = 4$.

At first glance it may seem perhaps a bit cumbersome to consider a rather arbitrary splitting of the force into two terms governed by the functions a and f instead of considering more naturally a unique function $h(r) \overset{\Delta}{=} f(r) - a(r)$ of the distance $r > 0$, as depicted in Figure 2.2. However, as we shall clarify later, the interplay of the polynomial decay of the function h to infinity and its singularity at 0 is fundamental in order to be able to characterize the confinement and collision avoidance of the dynamics, and such a splitting, emphasizing the individual role of these two properties, will turn out to be useful in our statements. As a matter of fact, several forces in nature do have similar behavior, for instance the van der Waals forces are governed by Lennard-Jones potentials for which $h(r) = \sigma_f / r^{13} - \sigma_a / r^7$, for suitable positive constants σ_f and σ_a. For such forces the polynomial decay corresponds to the parameter $\beta = 7$, hence a rather fast decay that makes the confinement of the system conditional to the initial total energy level (see Theorem 2.12 below).
It is easily seen how the above model can be rewritten as

$$\begin{cases} \dot{x}(t) = v(t), \\ \dot{v}(t) = -L(x(t))x(t) - v(t)b(t), \end{cases}$$

where for any $x \in \mathbb{R}^{dN}$

$$L(x) \overset{\Delta}{=} L^a(x) - L^f(x) \tag{2.12}$$

is the difference between the Laplacians of the two matrices $(a(\|x_i - x_j\|^2))_{i,j=1}^N$ and $(f(\|x_i - x_j\|^2))_{i,j=1}^N$, respectively, and we have set $vb \triangleq (v_i b_i)_{i=1}^N$ for any $b = (b_1, \ldots, b_N)$. To quantify the behavior of the system we introduce a quantity called the *total energy* which includes the kinetic and potential energies; for all $(x, v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ we define

$$E(x, v) \triangleq \sum_{i=1}^N \|v_i\|^2 + \frac{1}{2} \sum_{\substack{i,j=1 \\ i\neq j}}^N \int_0^{\|x_i - x_j\|^2} a(r)dr + \frac{1}{2} \sum_{\substack{i,j=1 \\ i\neq j}}^N \int_{\|x_i-x_j\|^2}^{+\infty} f(r)dr. \tag{2.13}$$

If $(x(t), v(t))$ describes a point of a trajectory of system (2.10), $E(t)$ will stand for $E(x(t), v(t))$.

Remark 2.10. Every term appearing in the definition of $E(x, v)$ is nonnegative (and well-defined by (2.11)), and hence we are allowed to bound from above every term appearing in the expression of $E(x, v)$ (as, for instance, $\|v_i\|^2$ or $\frac{1}{2}\int_{\|x_i-x_j\|^2}^{+\infty} f(r)dr$) by $E(x, v)$ itself.

The total energy $E(t) = E(x(t), v(t))$ is a Lyapunov functional for system (2.10) and, provided we are in presence of no friction at all (i.e., $\Lambda = 0$), it is a conserved quantity. Since the proof of this result, already derived in [80], will be helpful later on, we will report it in full details.

Proposition 2.11. *For every $t \geq 0$, we have*

$$\frac{d}{dt}E(t) = -2\sum_{i=1}^N b_i(t)\|v_i(t)\|^2.$$

Hence, if $\Lambda = 0$ then $\frac{d}{dt}E \equiv 0$.

Proof. Let us compute

$$\begin{aligned}\frac{d}{dt}E(t) = \frac{d}{dt}\sum_{i=1}^N \|v_i(t)\|^2 &+ \sum_{i,j=1}^N a(\|x_i(t) - x_j(t)\|^2)(x_i(t) - x_j(t)) \cdot (v_i(t) - v_j(t)) - \\ &- \sum_{i,j=1}^N f(\|x_i(t) - x_j(t)\|^2)(x_i(t) - x_j(t)) \cdot (v_i(t) - v_j(t)).\end{aligned} \tag{2.14}$$

The first term of the sum above is

$$\begin{aligned}\frac{d}{dt}\sum_{i=1}^N \|v_i(t)\|^2 &= 2\sum_{i=1}^N \dot{v}_i(t) \cdot v_i(t) \\ &= -2\sum_{i=1}^N b_i(t)\|v_i(t)\|^2 - \\ &\quad - \sum_{i=1}^N\sum_{j=1}^N a(\|x_i(t) - x_j(t)\|^2)(x_i(t) - x_j(t)) \cdot (v_i(t) - v_j(t)) + \\ &\quad + \sum_{i=1}^N\sum_{j=1}^N f(\|x_i(t) - x_j(t)\|^2)(x_i(t) - x_j(t)) \cdot (v_i(t) - v_j(t)),\end{aligned} \tag{2.15}$$

which, plugged into (2.14), yields

$$\frac{d}{dt}E(t) = -2\sum_{i=1}^{N} b_i(t)\|v_i(t)\|^2.$$

So, if $\Lambda = 0$ then $b_i \equiv 0$ for every $i = 1, \dots N$, and thus $\frac{d}{dt}E \equiv 0$, as stated. □

If the attraction force at far distance is very strong (for $\beta \leq 1$), despite an initial high level of kinetic energy and of repulsion potential energy, perhaps due to a space compression of the group of particles, the dynamics is guaranteed to keep confined and collision avoiding in space at all times. If the attraction force is instead weak at far distance, i.e., $\beta > 1$, then confinement and collision avoidance turn out to be properties of the dynamics only conditionally to *initial* low levels of kinetic energy and repulsion potential energy, meaning that the particles should not be initially too fast and too close to each other. This latter condition is formulated in terms of a total energy critical threshold

$$\vartheta \triangleq (N-1)\int_0^{+\infty} a(r)dr. \tag{2.16}$$

This fundamental dichotomy of the dynamics has been characterized in the following result.

Theorem 2.12 ([80, Theorem 2.1]). *Consider an initial datum $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ satisfying $\|x_i^0 - x_j^0\|^2 > 0$ for all $i \neq j$ and*

$$E(0) \triangleq E(x^0, v^0) < \frac{1}{2}\int_0^{+\infty} f(r)dr. \tag{2.17}$$

Then there exists a unique solution $(x(\cdot), v(\cdot))$ of system (2.10) with initial condition (x^0, v^0). Moreover, if one of the two following hypotheses holds:

(*i*) $\beta \leq 1$,

(*ii*) $\beta > 1$ *and* $E(0) < \vartheta$,

then the population is cohesive and collision-avoiding, i.e., there exist two constants $B_0, b_0 > 0$ such that, for all $t \geq 0$

$$b_0 \leq \|x_i(t) - x_j(t)\| \leq B_0, \qquad \textit{for all } 1 \leq i \neq j \leq N. \tag{2.18}$$

Motivated by Theorem 2.12, we will call the *consensus energy sublevel* or *consensus region* the set

$$C \triangleq \{w \in \mathbb{R} : w \leq \vartheta\}$$

We will say that the system (2.10) is *in the consensus region at time t* if $E(t) \in C$. It is an obvious corollary of Theorem 2.12 the fact that if system (2.10) is in the consensus region at time T, for some $T \geq 0$, then condition (2.18) is fulfilled for every $t \geq T$.

Remark 2.13. Theorem 2.12 is the Cucker-Dong counterpart of Theorem 1.11. Indeed, for the choice of a as in (1.7), Theorem 1.11 implies that

(i) if $\beta \leq 1/2$ then $a \notin L^1(\mathbb{R}_+)$, therefore consensus is achieved regardless of the initial conditions;

(ii) if $\beta > 1/2$ then $a \in L^1(\mathbb{R}_+)$, and consensus is guaranteed only if (1.14) is satisfied.

Remark 2.14. Let us stress again the fact that the word *consensus* must be intended here as a stable *cohesion and collision-avoiding* dynamics, in the spirit of the conclusion of Theorem 2.12. This is in contrast with the meaning of the word *consensus* in Chapter 1, which describes a situation where all the agents move according to the same velocity vector. We point out that this definition of *consensus* does not imply this particular feature, but it is rather intended to make a parallel between Theorem 2.12 and Theorem 1.11, as already done by the authors in [80, Remark 1].

2.2.2 Introducing the control

Notice the similarity of the present situation and that of Section 2.1: in both cases we have a system whose desired pattern can be enforced by decreasing a certain Lyapunov functional under the action of a sparse intervention. Given a positive constant M modeling the limited resources given to the external policy maker to influence instantaneously the dynamics, it is very natural to define the set of *admissible controls* precisely as in Definition 2.1: a control $u : \mathbb{R}_+ \to \mathbb{R}^{dN}$ is admissibile if it is a measurable functions which satisfies the $\ell_1^N - \ell_2^d$-norm constraint for every $t \geq 0$

$$\sum_{i=1}^{N} \|u_i(t)\| \leq M. \tag{2.19}$$

Hence, the *controlled* Cucker-Dong model is given by

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = -b_i(t)v_i(t) + \sum_{j=1}^{N} a\left(\|x_i(t)-x_j(t)\|^2\right)(x_j(t)-x_i(t)) + \\ \qquad + \sum_{\substack{j=1 \\ i\neq j}}^{N} f\left(\|x_i(t)-x_j(t)\|^2\right)(x_i(t)-x_j(t)) + u_i(t), \end{cases} \quad i = 1, \ldots, N, \tag{2.20}$$

where u is admissible. For this system we define again the total energy function E and the threshold ϑ as in (2.13) and (2.16), respectively.

The control should be exerted until $E(T) < \vartheta$ at some finite time T, and then it should be turned off, similarly to the sparse selection of Definition 2.4. Since we start from $E(0) > \vartheta$, then it is necessary that our control forces the total energy to decrease, for instance by ensuring $\frac{d}{dt}E < 0$. The following technical result helps us to identify the form of admissible controls satisfying this property. For the time being, we assume that a Filippov solution of (2.20) exists, later on we shall prove it.

Lemma 2.15. *Suppose there exists a solution of the system* (2.20) *and let* E *be the total energy function associated to it. Then, for every* $t \geq 0$ *we have*

$$\frac{d}{dt}E(t) = -2\sum_{i=1}^{N} b_i(t)\|v_i(t)\|^2 + 2\sum_{i=1}^{N} u_i(t) \cdot v_i(t). \tag{2.21}$$

Proof. The identity (2.14) in the proof of Proposition 2.11 is still valid for a solution to (2.20), while (2.15) changes as follows

$$\begin{aligned}\frac{d}{dt}\sum_{i=1}^{N}\|v_i(t)\|^2 = &-2\sum_{i=1}^{N} b_i(t)\|v_i(t)\|^2 - \\ &-\sum_{i=1}^{N}\sum_{j=1}^{N} a(\|x_i(t)-x_j(t)\|^2)(x_i(t)-x_j(t))\cdot(v_i(t)-v_j(t)) + \\ &+\sum_{i=1}^{N}\sum_{j=1}^{N} f(\|x_i(t)-x_j(t)\|^2)(x_i(t)-x_j(t))\cdot(v_i(t)-v_j(t)) + \\ &+2\sum_{i=1}^{N} u_i(t)\cdot v_i(t),\end{aligned}$$

because of the control term. Inserting this expression into (2.14) we get (2.21). □

From expression (2.21), it is clear that the best way our control can act on E in order to push it below the threshold is not acting on the mutual distances between agents, but according to the velocities v. Hence, we focus on the following family of controls, which closely resembles (2.4) .

Definition 2.16. Let $(x, v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ and $0 \leq \varepsilon \leq M/E(0)$. We define the *sparse feedback control* $u(x,v) = (u_1(x,v), \dots, u_N(x,v))^T \in \mathbb{R}^{dN}$ associated to (x, v) as

$$u_i(x,v) \triangleq \begin{cases} -\varepsilon E(x,v)\dfrac{v_{\hat{\imath}(x,v)}}{\|v_{\hat{\imath}(x,v)}\|} & \text{if } i = \hat{\imath}(x,v), \\ 0 & \text{otherwise.}\end{cases}$$

where $\hat{\imath}(x,v)$ is the minimum index such that

$$\|v_{\hat{\imath}(x,v)}\| = \max_{1\leq j\leq N}\|v_j\|.$$

Whenever the point (x,v) is a point of a curve $(x,v) : \mathbb{R}_+ \to \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, i.e. $(x,v) = (x(t), v(t))$ for some $t \geq 0$, we will replace everywhere $u(x,v) = u(x(t), v(t))$ and $\hat{\imath}(x,v) = \hat{\imath}(x(t), v(t))$ with $u(t)$ and $\hat{\imath}(t)$, respectively.
The parameter ε will help us to tune the control in order to ensure the convergence to the consensus region. Moreover, whenever it is clear from the context, we will omit the time (or the space) dependency of $\hat{\imath}$.

Remark 2.17. Definition 2.16 makes sense if $\|v_{\hat{\imath}(t)}(t)\| \neq 0$ for at least almost every $t \geq 0$. Notice that, if the latter condition were not holding, then $v_i(t) = 0$ for all $i = 1, \dots, N$ and for all $t \geq 0$, hence $\dot{v}_i(t) = 0$ for all $i = 1, \ldots, N$ and for all $t \geq 0$, hence the configuration of the system would be in a steady state and no control would be needed.

Remark 2.18. Besides this latter observation on the well-posedness of Definition 2.16, later we show that actually $\|v_{\hat{\imath}(t)}(t)\|$ keeps bounded away from 0 in the interval of time where we intend the control to act. In particular in the proof of Theorem 2.25 we prove that if the average speed at time 0 is nonzero, i.e., $\|\overline{v}(0)\| > 0$, and $E(0)$ is not "too far" from the threshold (in a precise sense), then we have $\|v_{\hat{\imath}(t)}(t)\| > 0$ and $E(t) \leq E(0)$ for all t such that the control is needed. Hence, for the choice of u as in Definition 2.16 with parameter $\varepsilon \leq M/E(0)$ we have

$$\sum_{i=1}^{N} \|u_i(t)\| = \varepsilon E(t) \leq \frac{M}{E(0)} E(t) \leq M,$$

whence the validity of the constraint (2.19). Therefore, our control is admissible.

2.3 Properties of sample-controlled Cucker-Dong systems

The main goal of this section is to investigate some properties of sampling solutions (see Definition 2.7) of Cucker-Dong systems associated with the sparse feedback control of Definition 2.16. We shall show that such solutions possess several monotonicity properties, which will let us conclude that each of these systems satisfying certain *regularity conditions* can be driven in the consensus region in finite time. This will pave the way to an existence result for systems modeled by (2.20) but associated to a class of controls containing properly the one introduced in Definition 2.16, due to necessary continuity properties.

Therefore, given u as in Definition 2.16 and a sampling time $\tau > 0$, consider the *sampling control*

$$\widetilde{u}_i(t) \triangleq u(n\tau) \quad \text{for every } t \in [n\tau, (n+1)\tau]. \tag{2.22}$$

Lemma 2.15, together with (2.22), gives us the following crucial fact concerning the growth of the energy function. Indeed, we can estimate from above the energy function inside every interval $[n\tau, (n+1)\tau]$ by a quadratic function on τ.

Lemma 2.19. *For all $\tau > 0$, for all $n \in \mathbb{N}$ and for all s, t such that $n\tau \leq s \leq t \leq (n+1)\tau$, it holds*

$$\sqrt{E(t)} \leq \sqrt{E(s)} + \varepsilon E(n\tau)(t - s). \tag{2.23}$$

Proof. From (2.21), for all $n\tau \leq t < (n+1)\tau$ we get

$$\begin{aligned}
\frac{d}{dt} E(t) &\leq 2 \sum_{i=1}^{N} \widetilde{u}_i(t) \cdot v_{\hat{\imath}}(t) \\
&= -2\varepsilon E(n\tau) \frac{v_{\hat{\imath}}(n\tau) \cdot v_{\hat{\imath}}(t)}{\|v_{\hat{\imath}}(n\tau)\|} \\
&\leq 2\varepsilon E(n\tau) \|v_{\hat{\imath}}(t)\|
\end{aligned}$$

$$\leq 2\varepsilon E(n\tau)\sqrt{E(t)}$$

from Cauchy-Schwarz inequality and the fact that $\|v_i(t)\| \leq \sqrt{E(t)}$ for all $i = 1, \dots, N$. Hence, integrating between s and t, where $n\tau \leq s \leq t \leq (n+1)\tau$, we get the desired inequality. □

A useful corollary of the lemma above tells us that, if the energy does not increase in the interval $[0, n\tau]$, then the agents do not collide until $t = (n+1)\tau$, i.e., we are able to bound from below the mutual distances of the agents of system (2.20) on the whole interval $[0, (n+1)\tau]$.

Corollary 2.20. *There are sufficiently small $\tau_0 > 0$ and $\omega > 0$ such that, for all $\tau < \tau_0$ there exists $d_0 > \omega$ satisfying*

$$\|x_i(t) - x_j(t)\| \geq d_0, \tag{2.24}$$

for all $1 \leq i \neq j \leq N$ and for all $t \in [0, \tau]$. Moreover, if E is nonincreasing in $[0, n\tau]$ for some $n \in \mathbb{N}$, then $d_0 \geq \omega$ can be chosen such that (2.24) is true for all $t \in [0, (n+1)\tau]$.

Proof. Hypothesis (2.17) ensures that there are sufficiently small $\tau_0 > 0$ and $\omega > 0$ such that

$$E(0) < -\left[\left(\sqrt{E(0)} + \varepsilon E(0)\tau_0\right)^2 - E(0)\right] + \frac{1}{2}\int_{\omega}^{+\infty} f(r)dr. \tag{2.25}$$

Take $\tau < \tau_0$. Then, from estimate (2.23) we have that, for all $t \in [0, \tau]$,

$$\sqrt{E(t)} \leq \sqrt{E(0)} + \varepsilon E(0)\tau, \tag{2.26}$$

so, plugging (2.26) into (2.25), we get

$$E(t) < \frac{1}{2}\int_{\omega}^{+\infty} f(r)dr.$$

From the definition of E follows that, for all $i, j = 1, \dots, N$ such that $i \neq j$, it holds

$$\int_{\|x_i(t)-x_j(t)\|^2}^{+\infty} f(r)dr < \int_{\omega}^{+\infty} f(r)dr;$$

the thesis then follows from the fact that f is nonincreasing.
Now suppose that E is nonincreasing in $[0, n\tau]$. Using this fact and estimate (2.23), we get that (2.26) holds for all $t \in [0, (n+1)\tau]$, and from the same argument as above, we obtain $d_0 \geq \omega$ satisfying (2.24) for all $t \in [0, (n+1)\tau]$. □

Corollary 2.21. *For any $n \in \mathbb{N}$, if E is nonincreasing in $[0, n\tau]$, then the function $L(x(t))$ defined in (2.12) is bounded from above for every $t \in [0, (n+1)\tau]$.*

Proof. Since the function a is bounded from above by $a(0)$ and, by Corollary 2.20, the function $f(\|x_i(t) - x_j(t)\|^2)$ is bounded from above by $f(\omega^2)$ for every $t \in [0, (n+1)\tau]$, then $L(x(t))$ is bounded from above in the whole interval $[0, (n+1)\tau]$. □

The following lemma, again a corollary of Lemma 2.19, says that the mutual distances of the agents grow at most quadratically in time.

Lemma 2.22. *Fix $\tau > 0$ and suppose that for all $i, j = 1, \dots, N$ it holds*

$$\|x_i(0) - x_j(0)\| \leq C_0\sqrt{E(0)},$$

for some $C_0 > 0$. Then we get

$$\|x_i(t) - x_j(t)\| \leq (\tau D(\tau) + C_0)\sqrt{E(0)}, \tag{2.27}$$

for all $i, j = 1, \dots, N$ and for all $t \in [0, \tau]$, where

$$D(\tau) \triangleq \frac{M}{\sqrt{E(0)}}\tau + 2.$$

Moreover, for all $k \in \mathbb{N}$, if E is nonincreasing in $[0, k\tau]$ then for all $t \in [k\tau, (k+1)\tau]$ and all $i, j = 1, \dots, N$ it holds

$$\|x_i(t) - x_j(t)\| \leq ((k+1)\tau D(\tau) + C_0)\sqrt{E(0)}. \tag{2.28}$$

Proof. The following inequalities follow easily from the equations in (2.20), the definition of the control, and the estimate (2.23): for all $k\tau \leq s \leq t \leq (k+1)\tau$ it holds

$$\begin{aligned}
\|x_i(t) - x_j(t)\| &\leq \int_s^t \|v_i(\theta) - v_j(\theta)\| d\theta + \|x_i(s) - x_j(s)\| \\
&\leq \int_s^t (\|v_i(\theta)\| + \|v_j(\theta)\|) d\theta + \|x_i(s) - x_j(s)\| \\
&\leq 2\int_s^t \sqrt{E(\theta)} d\theta + \|x_i(s) - x_j(s)\| \\
&\leq 2\int_s^t (\sqrt{E(s)} + \varepsilon E(n\tau)(\theta - s)) d\theta + \|x_i(s) - x_j(s)\| \\
&= 2\sqrt{E(s)}(t-s) + \varepsilon E(n\tau)(t-s)^2 + \|x_i(s) - x_j(s)\|,
\end{aligned}$$

which, taking $s = k\tau$, yields

$$\begin{aligned}
\|x_i(t) - x_j(t)\| &\leq 2\sqrt{E(k\tau)}(t - k\tau) + \varepsilon E(k\tau)(t-k\tau)^2 + \|x_i(k\tau) - x_j(k\tau)\| \\
&\leq 2\sqrt{E(k\tau)}\tau + \varepsilon E(k\tau)\tau^2 + \|x_i(k\tau) - x_j(k\tau)\|.
\end{aligned} \tag{2.29}$$

Thus, if $k = 0$ it holds

$$\begin{aligned}
\|x_i(t) - x_j(t)\| &\leq 2\sqrt{E(0)}\tau + \varepsilon E(0)\tau^2 + \|x_i(0) - x_j(0)\| \\
&\leq (\tau(\varepsilon\sqrt{E(0)}\tau + 2) + C_0)\sqrt{E(0)}\tau \\
&\leq (\tau D(\tau) + C_0)\sqrt{E(0)}\tau.
\end{aligned}$$

Now let us prove by induction on k that, if $t \in [k\tau, (k+1)\tau]$ then (2.28) holds assuming that E is nonincreasing in $[0, k\tau]$. If $k = 0$, then it follows from (2.27); so suppose (2.28) holds for all $j \leq k$. Then from (2.29) we get,

$$\begin{aligned}\|x_i(t) - x_j(t)\| &\leq 2\sqrt{E(k\tau)}\tau + \varepsilon E(k\tau)\tau^2 + \|x_i(k\tau) - x_j(k\tau)\| \\ &\leq 2\sqrt{E(0)}\tau + \varepsilon E(0)\tau^2 + (k\tau D(\tau) + C_0)\sqrt{E(0)} \\ &\leq ((k+1)\tau D(\tau) + C_0)\sqrt{E(0)}.\end{aligned}$$

This concludes the proof. □

The last property of the systems under investigation is the fact that if, at time $t = 0$ the norm of the average velocity of the system $\overline{v}$ is larger in norm than a positive scalar η, then the norm of $\overline{v}(t)$ is bounded from below by η until the time horizon $t = T^\star(\eta)$, defined below.

Lemma 2.23. *Suppose that $\|\overline{v}(0)\| > 0$ and let η be a positive estimate from below of $\|\overline{v}(0)\|$, i.e., $\|\overline{v}(0)\| \geq \eta > 0$. Define the following auxiliary time horizon:*

$$T^\star(\eta) \triangleq \frac{\|\overline{v}(0)\|^2 - \eta^2}{2\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)}.$$

Furthermore, let $\varepsilon \leq M/E(0)$, fix $\tau > 0$ and $n \leq [T^\star(\eta)/\tau]$, where $[x]$ stands for the integer part of x. Then, if E is nonincreasing in $[0, n\tau]$ we have

$$\|\overline{v}(t)\| \geq \eta \quad \textit{for every } t \in [0, n\tau]. \tag{2.30}$$

Proof. Let us compute, for all $t \in [k\tau, (k+1)\tau)$ with $k \leq n$, the derivative of $\|\overline{v}(t)\|^2$.

$$\begin{aligned}\frac{d}{dt}\|\overline{v}(t)\|^2 &= 2\frac{d}{dt}\overline{v}(t) \cdot \overline{v}(t) \\ &= -\frac{2}{N}\sum_{j=1}^{N} b_j(t)v_j(t) \cdot \overline{v}(t) - \frac{2\varepsilon E(k\tau)}{N}\frac{v_{\hat{\imath}(k\tau)}(k\tau) \cdot \overline{v}(t)}{\|v_{\hat{\imath}(k\tau)}(k\tau)\|} \\ &\geq -2\Lambda\|v_{\hat{\imath}(t)}(t)\|^2 - \frac{2\varepsilon E(k\tau)}{N}\|v_{\hat{\imath}(t)}(t)\|,\end{aligned}$$

using the fact that $\|v_{\hat{\imath}(t)}(t)\| \geq \max_{1\leq j\leq N}\{\|v_j(t)\|, \|\overline{v}(t)\|\}$ for all $t \geq 0$. Now we use again the inequality $\|v_i(t)\| \leq \sqrt{E(t)}$, valid for all $i = 1, \ldots, N$, to obtain

$$\begin{aligned}\frac{d}{dt}\|\overline{v}(t)\|^2 &\geq -2\|v_{\hat{\imath}(t)}(t)\|\left[\Lambda\|v_{\hat{\imath}(t)}(t)\| + \frac{\varepsilon E(k\tau)}{N}\right] \\ &\geq -2\sqrt{E(t)}\left[\Lambda\sqrt{E(t)} + \frac{\varepsilon E(k\tau)}{N}\right] \\ &\geq -2\sqrt{E(0)}\left[\Lambda\sqrt{E(0)} + \frac{\varepsilon E(0)}{N}\right] \\ &\geq -2\sqrt{E(0)}\left[\Lambda\sqrt{E(0)} + \frac{M}{N}\right],\end{aligned}$$

since E is nonincreasing in $[0, k\tau]$ for all $k \leq n$ and $\varepsilon \leq M/E(0)$. Furthermore, since the last quantity is constant, by integrating between $k\tau$ and t we get

$$\|\overline{v}(t)\|^2 \geq -2\sqrt{E(0)}\left[\Lambda\sqrt{E(0)} + \frac{M}{N}\right](t - k\tau) + \|\overline{v}(k\tau)\|^2, \tag{2.31}$$

for every $t \in [k\tau, (k+1)\tau]$. We will now prove by induction on $k \leq n$ that

$$\|\overline{v}(t)\|^2 \geq -2\sqrt{E(0)}\left[\Lambda\sqrt{E(0)} + \frac{M}{N}\right]t + \|\overline{v}(0)\|^2, \tag{2.32}$$

for every $t \in [k\tau, (k+1)\tau]$. This is straightforward for $k = 0$, so suppose (2.32) holds for $k \leq n$ and let us prove it for $k+1$. From (2.31) and from the inductive hypothesis we get

$$\begin{aligned}
\|\overline{v}(t)\|^2 &\geq -2\sqrt{E(0)}\left[\Lambda\sqrt{E(0)} + \frac{M}{N}\right](t - k\tau) + \|\overline{v}(k\tau)\|^2 \\
&\geq -2\sqrt{E(0)}\left[\Lambda\sqrt{E(0)} + \frac{M}{N}\right](t - k\tau) \\
&\quad - 2\sqrt{E(0)}\left[\Lambda\sqrt{E(0)} + \frac{M}{N}\right]k\tau + \|\overline{v}(0)\|^2 \\
&= -2\sqrt{E(0)}\left[\Lambda\sqrt{E(0)} + \frac{M}{N}\right]t + \|\overline{v}(0)\|^2.
\end{aligned}$$

Hence, if $t \leq n\tau \leq [T^\star(\eta)/\tau]\tau \leq T^\star(\eta)$, we get

$$\|\overline{v}(t)\|^2 \geq \eta^2,$$

which concludes the proof. □

Remark 2.24. We can get a larger interval of validity of (2.30) by using in the proof above the upper bound $\|v_i(t)\|^2 \leq E(t) - E_p(t)$ instead of $\|v_i(t)\|^2 \leq E(t)$, where

$$E_p(t) \triangleq \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{N}\int_0^{\|x_i(t)-x_j(t)\|^2} a(r)dr + \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{N}\int_{\|x_i(t)-x_j(t)\|^2}^{+\infty} f(r)dr$$

is the *potential part of* $E(t)$. In this case, given $b, B \geq 0$ such that $b \leq \|x_i(t) - x_j(t)\|^2 \leq B$ for every $i \neq j$ and $t \geq 0$, we can estimate $E_p(t)$ from below by

$$S(b,B) \triangleq \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{N}\int_0^{b} a(r)dr + \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{N}\int_{B}^{+\infty} f(r)dr.$$

Thus, given b and B as above we get the finer estimate $\|v_i(t)\|^2 \leq E(t) - S(b,B)$, and if we replace $\|v_i(t)\|^2 \leq E(t)$ with it we get

$$T^\star(\eta,b,B) \triangleq \frac{\|\overline{v}(0)\|^2 - \eta^2}{2\sqrt{E(0) - S(b,B)}\left(\Lambda\sqrt{E(0) - S(b,B)} + \frac{M}{N}\right)}.$$

Clearly, with the choice $b = 0$ and $B = +\infty$ we get back $T^\star(\eta,b,B) = T^\star(\eta)$, since

$S(b, B) = 0$.

For the convenience of the reader, we now list all the highlighted properties of controlled Cucker-Dong systems which will be useful in the proof of the main result of the following section:

(i) linear growth estimate of $\sqrt{E(t)}$ (Lemma 2.19);

($ii^\dagger$) the distance $\|x_i(t) - x_j(t)\|$ can be estimated from below by a constant $d_0 \geq \omega$ (Corollary 2.20);

($iii^\dagger$) quadratic growth estimate from above of $\|x_i(t) - x_j(t)\|$ (Lemma 2.22);

(iv) if $\|\overline{v}(0)\| \geq \eta > 0$ then there exists $T > 0$ such that $\|\overline{v}(t)\| \geq \eta$ for every $t \leq T$ (Lemma 2.23), provided that E is not increasing in $[0, T]$.

Each condition marked by a † holds in each starting interval $[0, \tau]$ and in every interval $[0, (n+1)\tau]$ provided that E is nonincreasing in $[0, n\tau]$.

2.4 Piecewise-constant sparse controls are effective

We pass to prove that the sampled sparse control strategy is indeed able to steer Cucker-Dong systems into the consensus region, provided they are not too *unruly* (in a sense specified below), from the properties listed in the section above.

Since in what follows, η will always be a fraction of $\|\overline{v}(0)\|$, i.e., $\eta = \|\overline{v}(0)\|/q$ for some $q > 1$, we will set

$$T^\star(q) \triangleq T^\star\left(\frac{1}{q}\|\overline{v}(0)\|\right) = \frac{\|\overline{v}(0)\|^2}{2\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)}\left(1 - \frac{1}{q^2}\right).$$

We can now present the main result of this section, which is the equivalent version of Theorem 2.8 for Cucker-Dong systems.

Theorem 2.25. *Fix $M > 0$. Let $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ be such that the following hold:*

(a) $0 < \|x_i^0 - x_j^0\| \leq C_0\sqrt{E(0)}$ *for all* $1 \leq i \neq j \leq N$, *for some* $C_0 > 0$,

(b) $\|\overline{v}(0)\| > 0$,

(c) *suppose there exists a $q > 1$ satisfying*

$$E(0)\ln\left(\frac{E(0)}{\vartheta}\right) < \frac{2M}{q}\|\overline{v}(0)\| T^\star(q).$$

Then there exist $\tau_0 > 0$, $\omega > 0$ and $\varepsilon > 0$ such that for every $\tau \leq \tau_0$

$$\frac{1}{2T^\star(q)\left(\frac{1}{q}\|\overline{v}(0)\| - \mu(\tau, q)\right)}\ln\left(\frac{E(0)}{\vartheta}\right) \leq \varepsilon \leq \frac{M}{E(0)} \tag{2.33}$$

holds, where for the sake of brevity we have set

$$\mu(\tau, q) \triangleq M\Lambda\tau^2 + \left(M + \Lambda\sqrt{E(0)} + (a(0) + f(\omega^2))\,((T^\star(q) + \tau)D(\tau) + C_0)\,N\sqrt{E(0)}\right)\tau. \tag{2.34}$$

Moreover, the sampling solution of system (2.20) *associated with the control* u *of Definition 2.16 with control parameter* ε *satisfying* (2.33)*, the sampling time* $\tau \leq \tau_0$ *and initial datum* (x^0, v^0) *reaches the consensus region before time* $T^\star(q)$*.*

Proof. Let $q > 1$ such that it satisfies condition (c) of the statement, and take τ_0 and ω sufficiently small to allow us to apply Corollary 2.20 and to satisfy the inequality

$$\frac{1}{q}\,\|\overline{v}(0)\| - \mu(\tau_0, q) \geq \frac{E(0)}{2M\; T^\star(q)} \ln\left(\frac{E(0)}{\vartheta}\right). \tag{2.35}$$

Notice that the existence of such τ_0 and ω follows from condition (c). Furthermore, (2.35) grants the existence of an $\varepsilon > 0$ for which (2.33) is satisfied and implies that the following inequalities are true:

$$\frac{1}{q}\,\|\overline{v}(0)\| - \mu(\tau_0, q) + T^\star(q)D(\tau_0) > 0, \tag{2.36}$$

$$\frac{1}{q}\,\|\overline{v}(0)\| - \mu(\tau_0, q) > 0. \tag{2.37}$$

Moreover, all the above inequalities are still true if we substitute τ_0 by any $\tau \leq \tau_0$.
Now, denote with $(x(\cdot), v(\cdot))$ the solution of (2.20) associated with the control u with ε as in (2.33), the sampling time $\tau \leq \tau_0$ and initial datum (x^0, v^0). As already pointed out in the proof of Lemma 2.19, for every $t \in [n\tau, (n+1)\tau)$ it holds

$$\frac{d}{dt}E(t) \leq -2\varepsilon E(n\tau)\frac{v_{\hat{\imath}}(n\tau)\cdot v_{\hat{\imath}}(t)}{\|v_{\hat{\imath}}(n\tau)\|},$$

where $\hat{\imath} \triangleq \hat{\imath}(n\tau)$. We define

$$\phi(t) \triangleq E(n\tau)\frac{v_{\hat{\imath}}(n\tau)\cdot v_{\hat{\imath}}(t)}{\|v_{\hat{\imath}}(n\tau)\|},$$

and we study its derivative with respect to $t \in [n\tau, (n+1)\tau)$. Notice that if E is nonincreasing in $[0, n\tau]$ and if $n\tau \leq T^\star(q)$, then

$$\phi(n\tau) = E(n\tau)\|v_{\hat{\imath}}(n\tau)\| \geq \frac{1}{q}\,\|\overline{v}(0)\|\,E(n\tau)$$

holds because we are under the hypothesis of Lemma 2.23 and we can use the estimate (2.30). We have the following estimate from below

$$\begin{aligned}\frac{d}{dt}\phi(t) &= E(n\tau)\frac{v_{\hat{\imath}}(n\tau)\cdot \frac{d}{dt}v_{\hat{\imath}}(t)}{\|v_{\hat{\imath}}(n\tau)\|}\\ &= \frac{E(n\tau)}{\|v_{\hat{\imath}}(n\tau)\|}\Bigg[-b_{\hat{\imath}}(t)v_{\hat{\imath}}(n\tau)\cdot v_{\hat{\imath}}(t) + \sum_{j=1}^{N} a\left(\|x_{\hat{\imath}}(t) - x_j(t)\|^2\right) v_{\hat{\imath}}(n\tau)\cdot(x_j(t) - x_{\hat{\imath}}(t)) +\end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=1}^{N} f\left(\|x_{\hat{\imath}}(t)-x_j(t)\|^2\right) v_{\hat{\imath}}(n\tau)\cdot(x_{\hat{\imath}}(t)-x_j(t)) - \frac{\varepsilon E(n\tau)\|v_{\hat{\imath}}(n\tau)\|^2}{\|v_{\hat{\imath}}(n\tau)\|}\Bigg] \\
&\geq E(n\tau)\Bigg[-b_{\hat{\imath}}(t)\|v_{\hat{\imath}}(t)\| - \sum_{j=1}^{N} a\left(\|x_{\hat{\imath}}(t)-x_j(t)\|^2\right)\|x_j(t)-x_{\hat{\imath}}(t)\| - \\
&-\sum_{j=1}^{N} f\left(\|x_{\hat{\imath}}(t)-x_j(t)\|^2\right)\|x_{\hat{\imath}}(t)-x_j(t)\| - \varepsilon E(n\tau)\Bigg] \\
&\geq -E(n\tau)\left[\Lambda\|v_{\hat{\imath}}(t)\| + \sum_{j=1}^{N}\left(a(0)+f\left(\|x_{\hat{\imath}}(t)-x_j(t)\|^2\right)\right)\|x_j(t)-x_{\hat{\imath}}(t)\| + \varepsilon E(n\tau)\right] \\
&\geq -E(n\tau)\left[\Lambda\sqrt{E(t)} + \sum_{j=1}^{N}\left(a(0)+f\left(\|x_{\hat{\imath}}(t)-x_j(t)\|^2\right)\right)\|x_j(t)-x_{\hat{\imath}}(t)\| + \varepsilon E(n\tau)\right],
\end{aligned}
$$

having used the Cauchy-Schwarz inequality and the fact that the damping functions b_i and the attraction force function a are bounded from above.
Now, for the sake of compact notation, set $r_{ij}(t) \triangleq \|x_i(t)-x_j(t)\|$. By the mean value theorem, there exists $\xi_n \in [n\tau, t]$ such that, using the estimate (2.23) in Lemma 2.19, we get

$$
\begin{aligned}
\phi(t) &\geq \phi(n\tau) - \tau E(n\tau)\left[\Lambda\sqrt{E(\xi_n)} + \sum_{j=1}^{N}(a(0)+f(r_{\hat{\imath}j}(\xi_n)^2))r_{\hat{\imath}j}(\xi_n) + \varepsilon E(n\tau)\right] \\
&\geq E(n\tau)\Bigg[\|v_{\hat{\imath}}(n\tau)\| - (\Lambda\varepsilon E(n\tau))\tau^2 - \\
&\quad -\left(\left(\varepsilon E(n\tau)+\Lambda\sqrt{E(n\tau)}\right) + \sum_{j=1}^{N}(a(0)+f(r_{\hat{\imath}j}(\xi_n)^2))r_{\hat{\imath}j}(\xi_n)\right)\tau\Bigg] \\
&\geq E(n\tau)\left[\|v_{\hat{\imath}}(n\tau)\| - \beta(n)\tau^2 - \left(\gamma(n) + \sum_{j=1}^{N}\left(a(0)+f(r_{\hat{\imath}j}(\xi_n)^2)\right)r_{\hat{\imath}j}(\xi_n)\right)\tau\right],
\end{aligned}
$$

where we have estimated ε from above by $M/E(0)$ and, again for the sake of brevity, we have set

$$
\begin{aligned}
\beta(k) &\triangleq \frac{\Lambda M E(k\tau)}{E(0)}, \\
\gamma(k) &\triangleq \frac{M E(k\tau)}{E(0)} + \Lambda\sqrt{E(k\tau)},
\end{aligned}
$$

for every $k \in \mathbb{N}$. So in the end we obtained the estimate

$$
\frac{d}{dt}E(t) \leq -2\varepsilon E(n\tau)\left[\|v_{\hat{\imath}}(n\tau)\| - \beta(n)\tau^2 - \left(\gamma(n) + \sum_{j=1}^{N}(a(0)+f(r_{\hat{\imath}j}(\xi_n)^2))r_{\hat{\imath}j}(\xi_n)\right)\tau\right],
$$

which is valid for all $t \in [n\tau, (n+1)\tau]$ where $n\tau \leq T^\star(q)$.

We will now prove by induction on n that E is nonincreasing in $[0,(n+1)\tau]$, where $n\tau \leq T^\star(q)$. First of all, if $n=0$ then $\xi_0 \in [0,\tau]$, and we can use Corollary 2.20 and Lemma 2.22 to bound $r_{\hat{\imath}j}(\xi_0)$ from below by ω and from above by $(\tau D(\tau)+C_0)\sqrt{E(0)}$. Moreover, since $\tau \leq T^\star(q)$, we can bound $\|v_{\hat{\imath}}(n\tau)\|$ from below by $\|\overline{v}(0)\|/q$ by Lemma 2.23. So by (2.36) we get

$$\begin{aligned}\frac{d}{dt}E(t) &\leq -2\varepsilon E(0)\left[\frac{1}{q}\|\overline{v}(0)\|-\beta(0)\tau^2-\left(\gamma(0)+(a(0)+f(\omega^2))(\tau D(\tau)+C_0)N\sqrt{E(0)}\right)\tau\right]\\ &= -2\varepsilon E(0)\left[\frac{1}{q}\|\overline{v}(0)\|-\mu(\tau,q)+T^\star(q)D(\tau)\right]\\ &< 0.\end{aligned}$$

Hence suppose that E is nonincreasing in $[0,n\tau]$ where $(n-1)\tau \leq T^\star(q)$, and let us prove that it is also true in $[n\tau,(n+1)\tau]$, if $n\tau \leq T^\star(q)$. By the fact that $\xi_n \geq n\tau$ and by the inductive hypothesis, we can use the bounds from Corollary 2.20, Lemma 2.22, and Lemma 2.23 as before to get

$$\begin{aligned}\frac{d}{dt}E(t) \leq -2\varepsilon E(0)&\left[\frac{1}{q}\|\overline{v}(0)\|-\beta(0)\tau^2-\right.\\ &\left.-\left(\gamma(0)+(a(0)+f(\omega^2))((n+1)\tau D(\tau)+C_0)N\sqrt{E(0)}\right)\tau\right]\\ \leq -2\varepsilon E(0)&\left[\frac{1}{q}\|\overline{v}(0)\|-\beta(0)\tau^2-\right.\\ &\left.-\left(\gamma(0)+(a(0)+f(\omega^2))((T^\star(q)+\tau)D(\tau)+C_0)N\sqrt{E(0)}\right)\tau\right]\\ \leq -2\varepsilon E(0)&\left[\frac{1}{q}\|\overline{v}(0)\|-\mu(\tau,q)\right]\\ < 0,&\end{aligned}$$

by condition (2.37). It is worth noticing that since E is nonincreasing in $[0,n\tau]$ then

$$\beta(n)\leq\beta(0)\quad\text{and}\quad\gamma(n)\leq\gamma(0).$$

Hence, E is nonincreasing in $[0,[T^\star(q)/\tau]\tau+1]$. From this fact follows that for all n such that $n\tau \leq T^\star(q)$ and for all $t\in[n\tau,(n+1)\tau]$ we have $E(t)\leq E(n\tau)$, which gives us, together with Corollary 2.20, Lemma 2.22, and Lemma 2.23, the following:

$$\begin{aligned}\frac{d}{dt}E(t) &\leq -2\varepsilon E(n\tau)\left[\frac{1}{q}\|\overline{v}(0)\|-\beta(n)\tau^2-\left(\gamma(n)+(a(0)+f(\omega^2))\sum_{j=1}^{N}d_{\hat{\imath},j}(\xi_n)\right)\tau\right]\\ &\leq -2\varepsilon E(t)\left[\frac{1}{q}\|\overline{v}(0)\|-\mu(\tau,q)\right].\end{aligned}\tag{2.38}$$

By integrating between 0 and t we obtain

$$E(t)\leq E(0)e^{-2\varepsilon\left[\frac{1}{q}\|\overline{v}(0)\|-\mu(\tau,q)\right]t},$$

so if

$$t \geq \frac{1}{2\varepsilon\left[\frac{1}{q}\|\overline{v}(0)\| - \mu(\tau,q)\right]} \ln\left(\frac{E(0)}{\vartheta}\right),$$

then $E(t) \leq \vartheta$ and thus the control steers the system into the consensus region in finite time.
Now we have to ensure that the energy level goes below the threshold before inequality (2.38) does not hold anymore; clearly, this happens if we impose that

$$T^\star(q) \geq \frac{1}{2\varepsilon\left[\frac{1}{q}\|\overline{v}(0)\| - \mu(\tau,q)\right]} \ln\left(\frac{E(0)}{\vartheta}\right),$$

and this is true if and only if

$$\varepsilon \geq \frac{1}{2T^\star(q)\left[\frac{1}{q}\|\overline{v}(0)\| - \mu(\tau,q)\right]} \ln\left(\frac{E(0)}{\vartheta}\right),$$

which holds because ε was chosen to satisfy (2.33).
Moreover, condition (2.19) is also satisfied since, for all $t \in [n\tau, (n+1)\tau]$

$$\begin{aligned}\sum_{i=1}^{N} \|u_i(t)\| &\leq \frac{\varepsilon E(n\tau)}{\|v_{\hat{\imath}}(n\tau)\|}\|v_{\hat{\imath}}(n\tau)\| \\ &\leq \varepsilon E(0) \\ &\leq M\end{aligned}$$

which follows from (2.33) and the fact that E is decreasing in the whole time interval of activity of the control. □

We now give a more operative reformulation of condition (c), which gives a clearer idea of when the above result holds.

Proposition 2.26. *Condition (c) of Theorem 2.25 is equivalent to the following request:*

(c') the threshold ϑ satisfies

$$\vartheta > E(0)\exp\left(-\frac{2\sqrt{3}}{9}\frac{M\|\overline{v}(0)\|^3}{E(0)\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)}\right).$$

Moreover, if (c') is satisfied, one can take

$$q \triangleq \frac{2M\|\overline{v}(0)\|^3}{3E(0)\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)\ln\left(\frac{E(0)}{\vartheta}\right)}$$

for condition (c) to hold true.

Proof. Remembering that

$$T^\star(q) = \frac{\|\overline{v}(0)\|^2}{2\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)}\left(1 - \frac{1}{q^2}\right),$$

we may rewrite condition (c) of Theorem 2.25 as

$$E(0)\ln\left(\frac{E(0)}{\vartheta}\right) < \frac{M\|\overline{v}(0)\|^3}{\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)}\left(\frac{q^2-1}{q^3}\right). \tag{2.39}$$

If we set

$$K \triangleq \frac{E(0)\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)}{M\|\overline{v}(0)\|^3}\ln\left(\frac{E(0)}{\vartheta}\right),$$

we obtain from (2.39) the equivalent polynomial inequality in q

$$Kq^3 - q^2 < -1.$$

We now show that the request $K < 2\sqrt{3}/9$ is equivalent to the existence of a $q > 0$ such that $Kq^3 - q^2 < -1$ holds, and additionally one can choose $q > 1$, as in the statement. If we study the derivative of the polynomial $\pi(q) = Kq^3 - q^2$ we find that the stationary points of $\pi(q)$ are $q = 0$ (which is a local maximum) and $\overline{q} = 2/3K$ (which is a local minimum). Moreover, since the roots of $\pi(q)$ are $q = 0$ and $q = 1/K$ then $\pi(q) \leq 0$ if and only if $q \leq 1/K$. Hence, there exists a $q > 0$ such that $\pi(q) < -1$ if and only if $\pi(\overline{q}) < -1$. But

$$\begin{aligned}\pi(\overline{q}) &= (K\overline{q} - 1)\overline{q}^2 \\ &= -\frac{4}{27K^2},\end{aligned}$$

so we have $\pi(\overline{q}) < -1$ if and only if

$$K < \frac{2\sqrt{3}}{9}. \tag{2.40}$$

The bound (2.40) trivially implies that $\overline{q} > 1$, and furthermore is equivalent to

$$\ln\left(\frac{E(0)}{\vartheta}\right) < \frac{2\sqrt{3}}{9}\frac{M\|\overline{v}(0)\|^3}{E(0)\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)}.$$

From this follows that hypothesis (c) of Theorem 2.25 holds if and only if (c') holds. □

Remark 2.27. Since we were under the hypothesis that the initial energy was *above* the threshold, from condition (c') we get that our control strategy is effective whenever

$$E(0) > \vartheta > E(0)\exp\left(-\frac{2\sqrt{3}}{9}\frac{M\|\overline{v}(0)\|^3}{E(0)\sqrt{E(0)}\left(\Lambda\sqrt{E(0)} + \frac{M}{N}\right)}\right),$$

which means that the threshold should not be too far away from the initial energy.

2.5 Existence of solutions of controlled Cucker-Dong systems

In order to prove an existence result for (2.20), as already claimed at the beginning of Section 2.3 we need to enlarge the set of sparse controls we want to work with. Indeed, as in the case of the controlled Cucker-Smale system, the family of sparse controls introduced in Definition 2.16 is too tight because, as the sampling time goes to 0, we may lose the sparsity we have enforced on our controls, and we have to weaken it to gain convergence for the sequence of controls.

2.5.1 Enlarging the set of controls

In what follows, we fix a Cucker-Dong system (2.20) satisfying conditions $(a)-(c)$ of Theorem 2.25, we indicate with $E(0)$ its initial energy and with

$$\gamma_q \triangleq \frac{\|\overline{v}(0)\|}{q},$$

where q is as in Proposition 2.26.
Each value of γ_q yields a partition of $\mathbb{R}^{dN} \times \mathbb{R}^{dN}$ into four disjoint sets:

$\mathcal{P}_1 \triangleq \{(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN} : \max_{1 \leq i \leq N} \|v_i\| < \gamma_q\}$,

$\mathcal{P}_2 \triangleq \{(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN} : \max_{1 \leq i \leq N} \|v_i\| = \gamma_q$ and $\exists k \geq 1$ and $i_1, \ldots, i_k \in \{1, \ldots N\}$ such that $\|v_{i_1}\| = \ldots = \|v_{i_k}\|$ and $\|v_{i_1}\| > \|v_j\|$ for every $j \notin \{i_1, \ldots, i_k\}\}$,

$\mathcal{P}_3 \triangleq \{(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN} : \max_{1 \leq i \leq N} \|v_i\| > \gamma_q$ and $\exists! i \in \{1, \ldots N\}$ such that $\|v_i\| > \|v_j\|$ for every $j \neq i\}$,

$\mathcal{P}_4 \triangleq \{(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN} : \max_{1 \leq i \leq N} \|v_i\| > \gamma_q$ and $\exists k > 1$ and $i_1, \ldots, i_k \in \{1, \ldots N\}$ such that $\|v_{i_1}\| = \ldots = \|v_{i_k}\|$ and $\|v_{i_1}\| > \|v_j\|$ for every $j \notin \{i_1, \ldots, i_k\}\}$,

Definition 2.28. For every $(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ we denote with $U(x,v) \subseteq \mathbb{R}^{dN}$ the set of all vectors $u(x,v) = (u_1(x,v), \ldots, u_N(x,v))^T \in \mathbb{R}^{dN}$, whose vector entries are of the form

$$u_i(x,v) = \begin{cases} -\varepsilon_i E(x,v) \dfrac{v_i}{\|v_i\|} & \text{if } \|v_i\| \neq 0, \\ 0 & \text{if } \|v_i\| = 0, \end{cases}$$

where the coefficients $\varepsilon_i \geq 0$ satisfy

$$\sum_{i=1}^{N} \varepsilon_i \leq \frac{M}{E(0)},$$

and

- if $(x,v) \in \mathcal{P}_1$ then $\varepsilon_i = 0$ for every $i = 1, \ldots, N$;

- if $(x, v) \in \mathcal{P}_2$ then indicating with $i_1, \ldots, i_k$ the indexes such that $\|v_{i_1}\| = \ldots = \|v_{i_k}\| = \gamma_q$ and $\|v_{i_1}\| > \|v_j\|$ for every $j \notin \{i_1, \ldots, i_k\}$, we have $\varepsilon_j = 0$ for every $j \notin \{i_1, \ldots, i_k\}$;
- if $(x, v) \in \mathcal{P}_3$ then, indicating with i the only index such that $\|v_i\| > \|v_j\|$ for every $j \neq i$, we have $\varepsilon_i = M/E(0)$ and $\varepsilon_j = 0$ for every $j \neq i$;
- if $(x, v) \in \mathcal{P}_4$ then, indicating with $i_1, \ldots, i_k$ the indexes such that $\|v_{i_1}\| = \ldots = \|v_{i_k}\|$ and $\|v_{i_1}\| > \|v_j\|$ for every $j \notin \{i_1, \ldots, i_k\}$, we have $\varepsilon_j = 0$ for every $j \notin \{i_1, \ldots, i_k\}$ and $\sum_{\ell=1}^{k} \varepsilon_{i_\ell} = M/E(0)$.

Remark 2.29. For every $0 \leq \varepsilon \leq M/E(0)$ and for every $t \geq 0$, the control $u(t)$ introduced in Definition 2.16 associated to ε belongs to $U(x(t), v(t))$ whenever $t \leq T^\star(q)$, since Lemma 2.23 guarantees $\max_{i \leq i \leq N} \|v_i(t)\| \geq \gamma_q$ for every $t \in [0, T^\star(q)]$.

The set $U(x, v)$ is closed and convex, and, moreover, has the following very elegant alternative variational interpretation, reminiscent of Definition 2.3.

Proposition 2.30. *For every $(x, v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ and for every $M \geq 0$, set*

$$M(x, v) \triangleq M \frac{E(x, v)}{E(0)} \quad \text{and} \quad K(x, v) \triangleq \left\{ u \in \mathbb{R}^{dN} : \sum_{i=1}^{N} \|u_i\| \leq M(x, v) \right\}.$$

Let $\mathcal{J} : \mathbb{R}^{dN} \to \mathbb{R}$ be the functional defined by

$$\mathcal{J}(u, v) \triangleq v \cdot u + \gamma_q \sum_{i=1}^{N} \|u_i\| . \tag{2.41}$$

Then

$$U(x, v) = \underset{u \in K(x,v)}{\arg\min}\, \mathcal{J}(u, v).$$

Proof. Without loss of generality one can minimize (2.41) componentwise, i.e., considering the minimization of the function

$$u_i \mapsto v_i \cdot u_i + \gamma_q \|u_i\| ,$$

for all $i = 1, \ldots, N$ individually. If $\|v_i\| < \gamma_q$ then, by Cauchy-Schwarz, we infer that

$$v_i \cdot u_i + \gamma_q \|u_i\| \geq 0, \qquad \text{for all } u \in K(x, v),$$

and in particular the choice $u_i = 0$ minimizes this component. Otherwise, if $\|v_i\| \geq \gamma_q$ then, due to the isotropy of the norm $\|u_i\|$, the minimal direction is the opposite to v_i, i.e., for $u_i = -\alpha_i v_i/\|v_i\|$, for some $\alpha_i \geq 0$, or equivalently there exists $\varepsilon_i \geq 0$ such that $\alpha_i = \varepsilon_i E(x, v)$. Hence, now the problem is how to allocate the weights ε_i in order to keep u within $K(x, v)$ and simultaneously to minimize $\mathcal{J}(\cdot, v)$. Clearly, for u to be in $K(x, v)$, the choice of ε_i has to be such that

$$\sum_{\{i \in \{1,\ldots,N\} : \|v_i\| \geq \gamma_q\}} \varepsilon_i \leq \frac{M}{E(0)}.$$

To conclude, from a direct check of the individual cases depending on the partition $\mathcal{P}_\ell$, $\ell = 1,\dots,4$, we can easily infer that $U(x,v)$ is the set of minimizers of $\mathcal{J}(\cdot,v)$ over the set $K(x,v)$. □

Another key property of the set U is that, as a set-valued function of (x,v), it is upper hemicontinuous.

Definition 2.31. A set-valued function F from a topological space X to $\mathscr{P}(Y)$, the power set of a real vector space Y, is said to be *upper hemicontinuous at* $x_0 \in X$ if and only if for every continuous linear functional $p : Y \to \mathbb{R}$ the function

$$x \mapsto \sup_{u \in F(x)} p(u)$$

is upper semicontinuous at $x_0 \in X$. F is called *upper hemicontinuous* if it is upper hemicontinuous at every $x_0 \in X$.

Proposition 2.32. *The function* $U : \mathbb{R}^{dN} \times \mathbb{R}^{dN} \to \mathscr{P}(\mathbb{R}^{dN})$ *which associates to every point* (x,v) *the set* $U(x,v)$ *is upper hemicontinuous.*

Proof. We have to prove that for every continuous linear functional p from $\mathbb{R}^{dN}$ to $\mathbb{R}$, the function $\sigma : \mathbb{R}^{dN} \times \mathbb{R}^{dN} \to \mathbb{R}$ defined as follows

$$\sigma : \ (x,v) \ \mapsto \sup_{u \in U(x,v)} p(u)$$

is upper semicontinuous, i.e., for every $(x,v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ and for every $\varepsilon > 0$ there exists a neighborhood $\mathcal{N}$ of (x,v) such that $\sigma(\widetilde{x},\widetilde{v}) \leq \sigma(x,v) + \varepsilon$ for all $(\widetilde{x},\widetilde{v}) \in \mathcal{N}$ whenever $\sigma(x,v) > -\infty$, and $\sigma(\widetilde{x},\widetilde{v})$ tends to $-\infty$ as $(\widetilde{x},\widetilde{v})$ tends to (x,v) whenever $\sigma(x,v) = -\infty$.
First of all, notice that since $U(x,v)$ is closed and bounded, then it is compact and hence $\sigma(x,v) > -\infty$.
Hence fix a continuous linear functional p from $\mathbb{R}^{dN}$ to $\mathbb{R}$, $(x,v) \in \mathbb{R}^{dN}$ and $\varepsilon > 0$. From Definition 2.28, an entry u_j of $u \in U(x,v)$ is zero if and only if $\|v_j\| < \max\{\gamma_q, \max_{1\leq i\leq N}\|v_i\|\}$, so if the latter condition holds, by the continuity of $\|\cdot\|$ we can find a neighborhood $\mathcal{N}$ of (x,v) such that for every $(\widetilde{x},\widetilde{v}) \in \mathcal{N}$ we have

$$\|\widetilde{v}_j\| < \max\left\{\gamma_q, \max_{1\leq i\leq N}\|\widetilde{v}_i\|\right\}.$$

Since given any two vectors $u, w \in U(x,v)$, u_i is zero if and only if w_i is, this implies that for every $(\widetilde{x},\widetilde{v}) \in \mathcal{N}$ and for every $\widetilde{u} \in U(\widetilde{x},\widetilde{v})$, the component $\widetilde{u}_i$ is zero only if u_i is zero for some $u \in U(x,v)$.
Moreover, if $\|v_{i_1}\| = \ldots = \|v_{i_k}\| = \max_{1\leq i\leq N}\|v_i\|$ and for some $(\widetilde{x},\widetilde{v}) \in \mathcal{N}$ there exists $1 \leq h \leq k$ such that

$$\max\left\{\left\|\widetilde{v}_{i_{j_1}}\right\|,\ldots,\left\|\widetilde{v}_{i_{j_h}}\right\|\right\} < \left\|\widetilde{v}_{i_{\ell_1}}\right\| = \ldots = \left\|\widetilde{v}_{i_{\ell_{k-h}}}\right\| = \max_{1\leq i\leq N}\|\widetilde{v}_i\|,$$

for $i_{\ell_s} \neq i_{j_r}$, for all $s = 1,\dots,k-h$ and $r = 1,\dots,h$, then $\widetilde{u}_{i_{j_1}} = \ldots = \widetilde{u}_{i_{j_h}} = 0$, while $\widetilde{u}_{i_{\ell_s}} \neq 0$ for every $1 \leq s \leq k-h$, for all $\widetilde{u} \in U(\widetilde{x},\widetilde{v})$.

These two facts together bring us to the conclusion that $U(\mathcal{N}) = U(x,v)$, so for every $(\widetilde{x},\widetilde{v}) \in \mathcal{N}$, $\sigma(\widetilde{x},\widetilde{v}) = \sigma(x,v) < \sigma(x,v)+\varepsilon$, proving the upper semicontinuity of σ. □

Remark 2.33. The problem with the family of controls introduced in Definition 2.16 is that if we were to take $U(x,v) = \{u(x,v)\}$ instead of the set in Definition 2.28, where u is as in Definition 2.16, then U would not have been upper hemicontinuous.

2.5.2 Existence of solutions for the differential inclusion

We are going to use the following convergence result concerning hemicontinuous set-valued functions.

Theorem 2.34 ([17, Theorem 1]). *Let F be a proper hemicontinuous map from a Hausdorff locally convex space X to the closed convex subsets of a Banach space Y. Let I be an interval of $\mathbb{R}$ and $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be measurable functions from I to X and Y respectively, satisfying*

- *(i) for almost every $t \in I$ and for every neighborhood $\mathcal{N}$ of 0 in $X \times Y$ there exists n_0 such that for every $n \geq n_0$, $y_n(t) \in F(x_n(t)) + \mathcal{N}$;*
- *(ii) $(x_n)_{n\in\mathbb{N}}$ converges almost everywhere to a function $x : I \to X$;*
- *(iii) $(y_n)_{n\in\mathbb{N}} \subseteq L^1(I;Y)$ and converges weakly to y in $L^1(I;Y)$.*

Then for almost every $t \in I$ it holds

$$y(t) \in F(x(t)).$$

Theorem 2.35. *For every $M > 0$ and for every initial condition $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, if system (2.20) satisfies conditions $(a)-(c)$ of Theorem 2.25 then there exist $q > 1$, $\tau_0 > 0$, $\omega > 0$, and $\varepsilon > 0$ such that*

$$\frac{1}{2T^\star(q)\left(\frac{1}{q}\|\overline{v}(0)\| - \mu(\tau_0,q)\right)} \ln\left(\frac{E(0)}{\vartheta}\right) \leq \varepsilon \leq \frac{M}{E(0)} \tag{2.42}$$

holds, where $\mu(\tau,q)$ is as in (2.34).
Moreover, for every $t \leq T^\star(q)$ there exists a solution of

$$\boldsymbol{z}(t) = \boldsymbol{z}^0 + \int_0^t \big(g(\boldsymbol{z}(s)) + u(\boldsymbol{z}(s))\big)ds, \quad \boldsymbol{z} = (x,v), \quad g(\boldsymbol{z}) = (v, -L(x)x - vb)$$

associated with a control u such that $u(t) \in U(\boldsymbol{z}(t))$ for every $t \leq T^\star(q)$. Here we use the compact notation introduced in (2.12).

Proof. Apply Theorem 2.25 to get $q > 1$, $\tau_0 > 0$, $\omega > 0$, and $\varepsilon > 0$ satisfying (2.42). Notice that, if $\tau \leq \tau_0$, then

$$\frac{1}{\frac{1}{q}\|\overline{v}(0)\| - \mu(\tau,q)} < \frac{1}{\frac{1}{q}\|\overline{v}(0)\| - \mu(\tau_0,q)}$$

which implies that ε satisfies also (2.33) for every $\tau \leq \tau_0$.
Hence, for every $n > 1/\tau_0$, Theorem 2.25 guarantees the existence of a sampling solution $\boldsymbol{z}_n(\cdot) \triangleq (x_n(\cdot), v_n(\cdot))$ of system (2.20) associated with the feedback control u satisfying Definition 2.16, the sampling time $\tau = 1/n$ and initial datum $\boldsymbol{z}^0 = (x^0, v^0)$. Setting $\widetilde{u}_n(t)$ to be the $\mathbb{R}^{dN} \times \mathbb{R}^{dN}$ vector $(0, \ldots, 0, u(\boldsymbol{z}([nt]/n)))^T$, we have that

$$\boldsymbol{z}_n(t) = \boldsymbol{z}^0 + \int_0^t \big(g(\boldsymbol{z}_n(s)) + \widetilde{u}_n(s)\big) ds. \tag{2.43}$$

By Corollary 2.21 the function $L(x(t))$ defined in (2.12) is bounded in the interval $[0, T^\star(q)]$, thus there is a constant $\alpha > 0$ such that the following linear growth estimate for g holds for every $n \in \mathbb{N}$ and $s \in [0, T^\star(q)]$:

$$\|g(\boldsymbol{z}_n(s))\| \leq \alpha(\|\boldsymbol{z}_n(s)\| + 1).$$

Plugging this estimate into (2.43) and using Gronwall's Lemma A.1, we get that the inequality

$$\|\boldsymbol{z}_n(t)\| \leq \big(\|\boldsymbol{z}^0\| + (\alpha + M)t\big)\, e^{\alpha t}$$

is true in $[0, T^\star(q)]$, hence the sequence of piecewise $\mathcal{C}^1$ functions $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ is pointwise equibounded by the function

$$C(t) \triangleq \big(\|\boldsymbol{z}^0\| + (\alpha + M)t\big)\, e^{\alpha t}, \qquad \text{for every } t \in [0, T^\star(q)].$$

Moreover, the sequence $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ is also equicontinuous in $[0, T^\star(q)]$ since for every $t, s \in [0, T^\star(q)]$ such that $t \leq s$, it holds

$$\|\boldsymbol{z}_n(t) - \boldsymbol{z}_n(s)\| \leq \int_s^t (\|g(\boldsymbol{z}_n(\theta))\| + M) d\theta \leq (t - s)(\alpha(C(T^\star(q)) + 1) + M).$$

Therefore, an application of the Ascoli-Arzelá Theorem yields a subsequence which we rename again as $(\boldsymbol{z}_n)_{n\in\mathbb{N}}$ converging uniformly to a continuous function $\boldsymbol{z} : [0, T^\star(q)] \to \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ and, by the continuity of g, $(g(\boldsymbol{z}_n))_{n\in\mathbb{N}}$ converges to $g(\boldsymbol{z})$. Concerning the sequence of controls $(u_n)_{n\in\mathbb{N}}$, since each u_n has norm bounded by M, it is a uniformly integrable sequence; hence it satisfies the hypothesis of the Dunford-Pettis Theorem (see [11, Theorem 1.38]), which means that, up to subsequences, it converges weakly in L^1 to a function $u \in L^1([0, T^\star(q)])$.
From Proposition 2.32, by taking $F = U$, $I = [0, T^\star(q)]$, $(x_n)_{n\in\mathbb{N}} = (\boldsymbol{z}_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}} = (u_n)_{n\in\mathbb{N}}$, $x = \boldsymbol{z}$ and $y = u$, we are under the hypothesis of Theorem 2.34 and so we can conclude that $u(t) \in U(\boldsymbol{z}(t))$ for every $t \leq T^\star(q)$ (since by Remark 2.29 we can guarantee $u_n(t) \in U(\boldsymbol{z}_n(t))$ only if $t \leq T^\star(q)$). Since $(u_n)_{n\in\mathbb{N}}$ converges weakly to u in $L^1([0, T^\star(q)])$, in particular we have

$$\lim_{n\to\infty} \int_0^t u_n(s) ds = \int_0^t u(s) ds$$

for every $t \leq T^\star(q)$. Hence passing to the limit in (2.43) we conclude the proof. □

We are thus able to prove existence for Filippov solutions to the differential inclusion associated to system (2.9) with controls belonging to U, in the spirit of Theorem 2.9.

Corollary 2.36. *For every $M > 0$ and every $(x, v) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ define*

$$F(x, v) \triangleq \{(v, -L(x)x - vb + u) : u \in U(x, v)\}.$$

Then, for every $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, if conditions $(a) - (c)$ of Theorem 2.25 are satisfied, the differential inclusion

$$(\dot{x}(t), \dot{v}(t)) \in F(x(t), v(t))$$

with initial condition $(x(0), v(0)) = (x^0, v^0)$ is well-posed as long as $t \in [0, T^\star(q)]$.

2.5.3 Decay rate estimate for sparse control strategy and instantaneous optimality

Let us stress again that in general the existence of a solution associated to a sparse control as in Definition 2.16 does not follow from Corollary 2.36, as we can say only that solutions exists within the class of feedback controls $U(x, v)$, which is much larger and contains nonsparse controls as well. Nevertheless, in what follows, we do suppose there exists a solution to system (2.20) associated to the control introduced in Definition 2.16 with $\varepsilon = M/E(0)$, and we analyze what could be the corresponding rate of convergence to the consensus region, compared with any other solution within the class of feedback controls $U(x, v)$.

Proposition 2.37. *If $\|\overline{v}(0)\| \geq \eta > 0$ then, for every $t \in [0, T^\star(q)]$ it holds*

$$E(t) \leq E(0) e^{-\frac{2\eta}{E(0)} M t}.$$

Proof. Before we start, notice that we can prove a continuous-time analogous of Lemma 2.23 in the same way, thus obtaining the fact that if $\|\overline{v}(0)\| \geq \eta > 0$ then $\|\overline{v}(t)\| \geq \eta$ for every $t \leq T^\star(\eta)$. Using Lemma 2.15 together with this fact we get

$$\begin{aligned}\frac{d}{dt} E(t) &\leq 2 \left(-\varepsilon E(t) \frac{v_{\hat{\imath}(t)}(t)}{\left\|v_{\hat{\imath}(t)}(t)\right\|} \right) \cdot v_{\hat{\imath}(t)}(t) \\ &= -2\varepsilon E(t) \left\|v_{\hat{\imath}(t)}(t)\right\| \\ &\leq -\frac{2\eta M}{E(0)} E(t),\end{aligned}$$

hence, integrating between 0 and t we get the desired estimate. □

The next result shows that this decay estimate is the best we can get among the controls introduced in Definition 2.28. It is the Cucker-Dong counterpart of Proposition 2.5.

Theorem 2.38. *The feedback control of Definition 2.16 associated to the solution of Theorem 2.25 is an instantaneous minimizer of*

$$\mathcal{D}(t, u) \triangleq \frac{d}{dt} E(t)$$

over all possible feedback controls in $U(x(t), v(t))$.

Proof. We have already proven in Lemma 2.15 that

$$\frac{d}{dt}E(t) = -2\sum_{i=1}^{N} b_i(t)\|v_i(t)\|^2 + 2\sum_{i=1}^{N} u_i(t)\cdot v_i(t),$$

so in order to minimize $\mathcal{D}(t,u)$ with respect to $u \in U(x(t), v(t))$ we have to work on the second term. Now, since

$$\sum_{i=1}^{N} u_i(t)\cdot v_i(t) = -E(t)\sum_{i=1}^{N} \varepsilon_i \,\|v_i(t)\|,$$

minimizing $\mathcal{D}(t,u)$ is equivalent to solve

$$\max \sum_{i=1}^{N} \varepsilon_i \,\|v_i(t)\| \quad \text{subject to} \quad \sum_{i=1}^{N} \varepsilon_i \leq \frac{M}{E(0)}. \tag{2.44}$$

By definition

$$\begin{aligned}\sum_{i=1}^{N} \varepsilon_i \,\|v_i(t)\| &\leq \left\|v_{\hat{\imath}(t)}(t)\right\| \sum_{i=1}^{N} \varepsilon_i \\ &\leq \frac{M}{E(0)} \left\|v_{\hat{\imath}(t)}(t)\right\|,\end{aligned}$$

which means that the choice of the control made in Definition 2.16 is a maximizer of (2.44). Moreover, the solution of the problem is unique whenever $\left\|v_{\hat{\imath}(t)}(t)\right\| > \|v_j(t)\|$ for all $j \neq \hat{\imath}(t)$. □

Similarly to what we have seen in the case of the Cucker-Smale system, the previous result shows that the most effective control strategy that the external policy maker can enact is to allocate all the resources at its disposal only on very few key agents in the system, in order to keep the dynamics bounded and collision avoiding. One of the most relevant differences with respect to Theorem 2.9 though is that for the Cucker-Smale model the stabilization can be achieved unconditionally, i.e., independently of the initial conditions (x^0, v^0). For the Cucker-Dong model, instead, a similar sparse control strategy yields only a conditional results, i.e., we obtain stabilization conditionally to an initial energy level

$$\vartheta < E(0) < c\vartheta,$$

for a constant $c > 1$, as stated in condition (c') of Proposition 2.26. Our numerical experiments, which follow below, suggest that it is possible to exceed such an upper energy barrier, but it is unclear whether this is just a matter of fortunate choices of good initial conditions or we can actually have a broader stabilization range than the one analytically derived in the previous sections.

2.6 Numerical validation of the sparse control strategy

In this section we will report the results of significant numerical simulations on Cucker-Dong systems in dimension $d = 2$ with and without the use of the sparse control strategy outlined in Definition 2.16. Throughout the section, we will keep fixed the number of agents ($N = 8$), the friction applied ($\Lambda = 0$, i.e., frictionless) and the form of the repulsive function ($f(r) = r^{-p}$). We restrict only to $N = 8$ simply for an easier visualization of the results. This means that we will vary the shape of the function a (i.e., we will act on β), the slope of the repulsion function (changing the value of p) and the maximum amount of strength of the sparse control (the parameter M). The parameter ε is always set equal to $M/E(0)$.

2.6.1 First case study: $\beta = 1.1$ and $p = 2$

Figure 2.3 displays the spatial evolution and speeds of the agents of a Cucker-Dong system with $\beta = 1.1$ and $p = 2$:

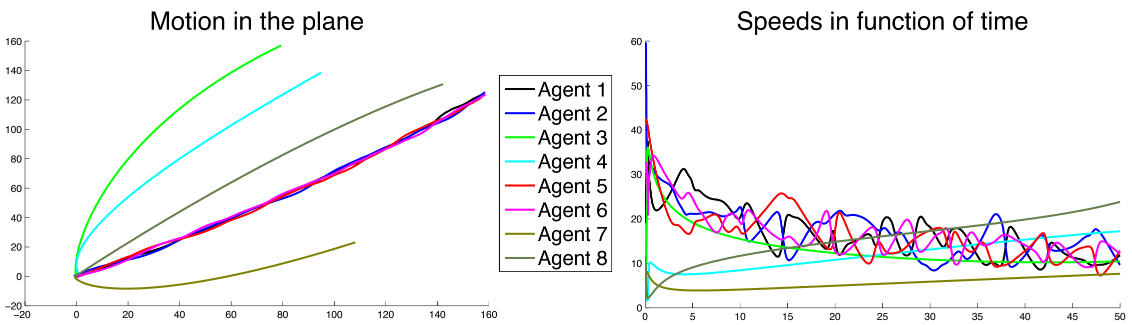


Figure 2.3: Space evolution and speeds of the uncontrolled system.

Though we can not infer the divergence of the system from this finite-time simulation, the portrayed situation seems far from going towards a flocking behavior. The only agents which seem to flock are Agent 1, Agent 2, Agent 5 and Agent 6 (resp. black, blue, red and magenta trajectories), as it is also visible by the corresponding speed graph, in which the speed of each agent is adjusted to the one of the other agents.

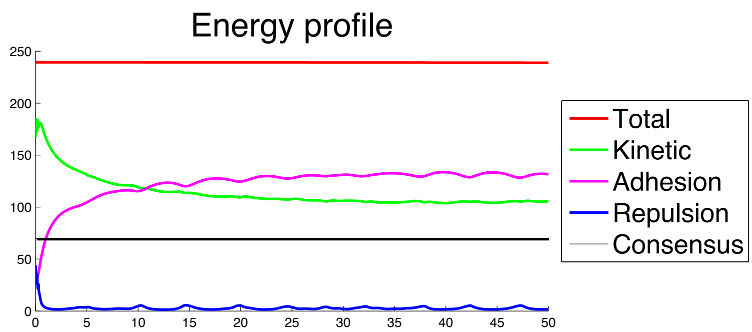


Figure 2.4: Energy profile of the uncontrolled system.

Figure 2.4 shows that the total energy E (the red line) is constant and far away from the consensus threshold ϑ (black line). The increase in the distances between particles is

reflected in an increase in the adhesion potential energy (the one due to a, see (2.13)) and in a decrease in the repulsive one (due to f).
If instead we apply our sparse control strategy with $M = 35$ on the same system with the same initial conditions, the situation gets immediately far better from a consensus point of view, as we can witness from Figure 2.5.

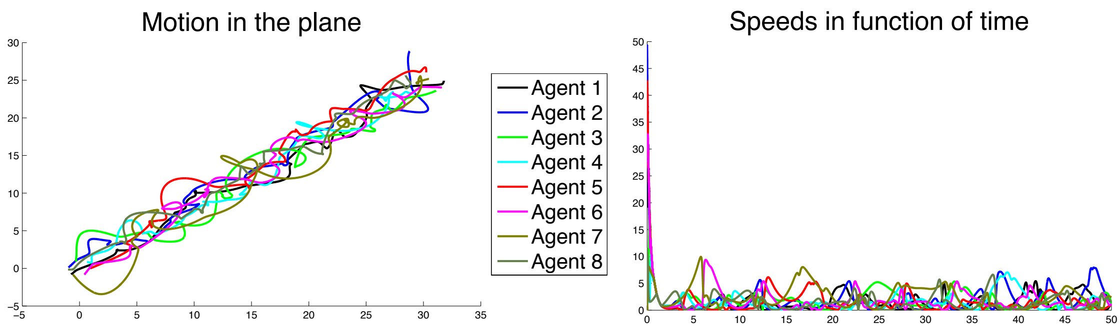


Figure 2.5: Space evolution and speeds of the controlled system.

The spatial evolution graph shows a braid movement which resembles a pattern near to flocking as it is commonly interpreted. The action of our control is evident from the energy profile of the system, portrayed in Figure 2.6, where the total energy is driven below the threshold in a very short time. The fall of the total energy is mainly due to its kinetic part (the green line), which is the only one directly affected by our control strategy. The sharp decrease of the kinetic energy is also witnessed in the graph showing the modulus of the speeds, where, after a quick, strong brake at the beginning, they stabilize at a very low level.

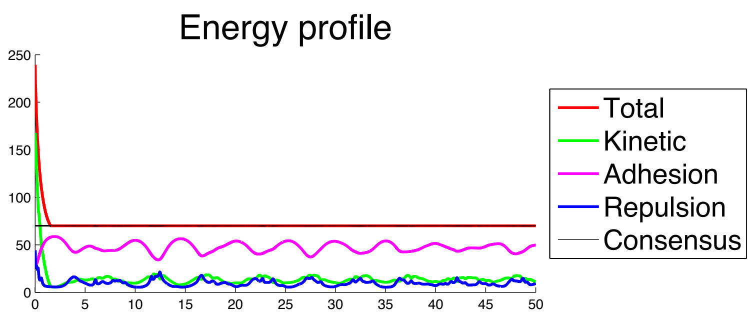


Figure 2.6: Energy profile of the controlled system.

2.6.2 Second case study: $\beta = 1.02$ and $p = 1.1$

The second case study takes into account a system with a weaker communication rate than before ($\beta = 1.02$) and with a different form of the repulsive function ($p = 1.1$), and we apply on it our control strategy with several values for M.
The top-left corner of Figure 2.7 is the uncontrolled system: it seems legitimate to suppose that it is very unlikely that the system will converge to consensus, especially looking at its energy profile graph (top-right corner of Figure 2.7), which shows an increase in the adhesion potential energy, phenomenon associated to an increase in the distance between

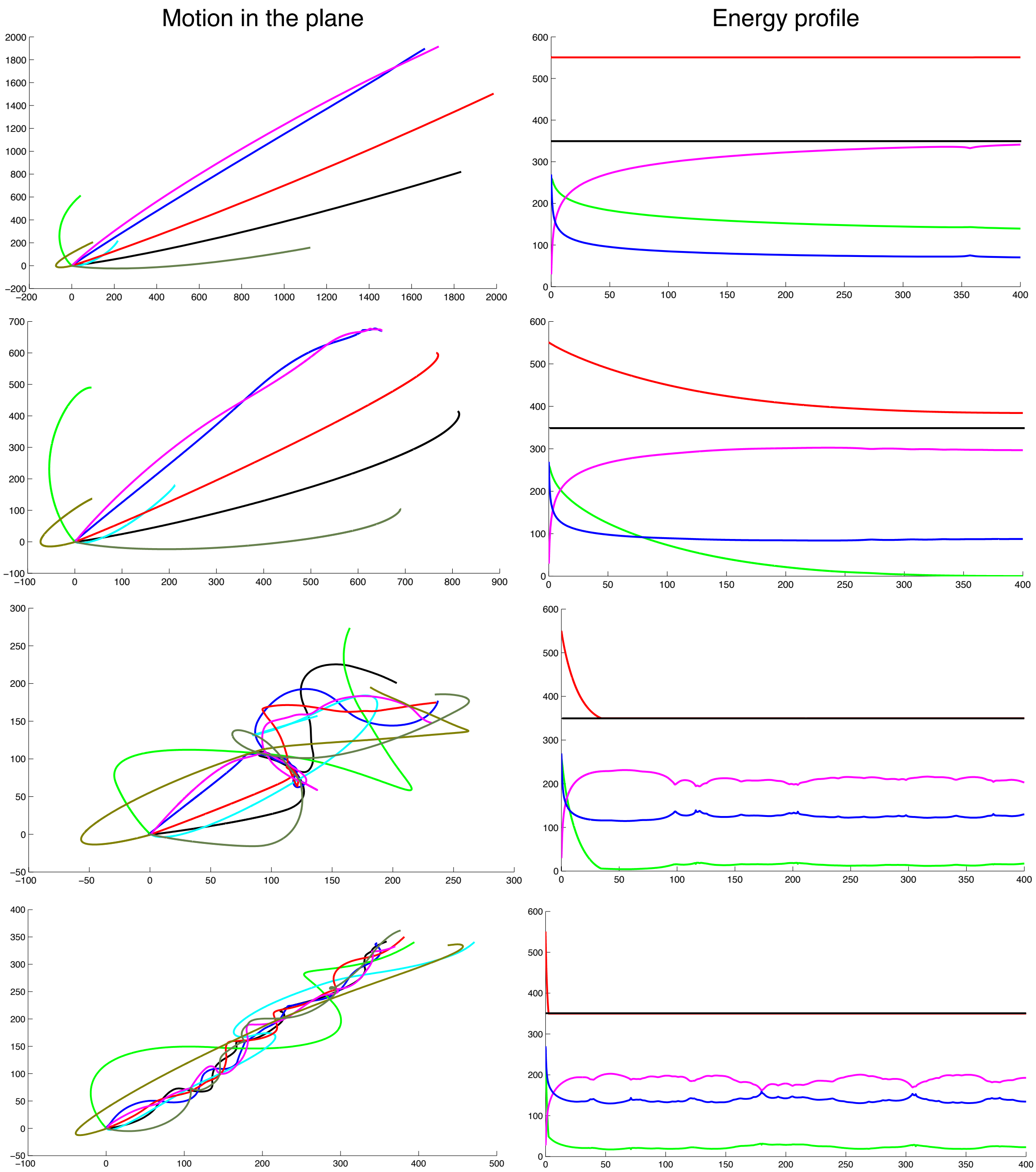


Figure 2.7: Spatial evolutions (left) and relative energy profiles (right). From top to bottom: $M = 0, M = 0.1, M = 1, M = 10$. The colors in the left column stand for: total energy (red), consensus region (black), adhesion energy (magenta), repulsion energy (blue), kinetic energy (green).

particles, as already pointed out. In the second line we see the spatial evolution graph of the same system but with the sparse control strategy acting with parameter $M = 0.1$, where the agents are starting to converge to consensus, as is also evident in their energy profile. The two bottom lines of Figure 2.7 display the action of controls with $M = 1$ and $M = 10$, respectively. It is clear how the situation goes better as M increases, which is due to the fact that the threshold is reached in shorter time (see the relative energy profile).

The right column of Figure 2.7 also clearly confirms the behavior of the decay rate of the energy as a function of M, as predicted by our analytical result of Proposition 2.37: $E(t)$ decreases as e^{-kMt}, for a certain constant $k > 0$.

It is interesting to notice that convergence to the consensus region occurs even if the hypothesis of Proposition 2.26 is not met, i.e., ϑ is very far away from $E(0)$, as it is likely to be a sub-optimal sufficient condition. Indeed, in all the case studies above

$$E(0)\exp\left(-\frac{2\sqrt{3}}{9}\frac{M\,\|\overline{v}(0)\|^3}{E(0)\sqrt{E(0)}\left(\Lambda\sqrt{E(0)}+\frac{M}{N}\right)}\right)\approx E(0),$$

but, nonetheless, we were able to steer each system to consensus in finite time.

2.6.3 A counterexample to unconditional sparse controllability

The last numerical experiment we report shows that in certain pathological situations the sparse control strategy can fail to steer a Cucker-Dong systems to consensus.

We consider $N = 2$ agents in dimension $d = 2$ and choose the interaction parameters as $H = 1$, $\beta = 2$, $p = 1.1$ and $M = 1$. In this situation, the force balance $f - a$ is completely in favor of the repulsive force, as Figure 2.8 shows: this means that, regardless of the mutual positions of the agents, they shall always be repelled from each other.

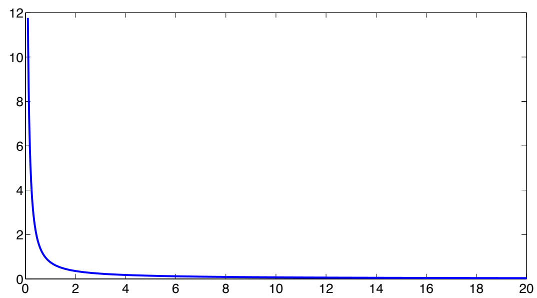


Figure 2.8: Sum of the attraction and repulsion forces $h(r) = f(r) - a(r)$ as a function of the distance $r > 0$ in the case study of Section 2.6.3

If we exert the sparse control strategy, the only result that we obtain is to freeze the agents where they are. Indeed, Figure 2.9 shows that the agents' speeds are rapidly reduced to values close to 0 as an effect of the control (also visible in the energy profile from the trajectory of the kinetic energy), but the total energy stays far away from the consensus region (the black line). The picture makes very clear that our control does not affect the potential energy of the system, as the sum of the adhesion and repulsion energies stays constant in time (the adhesion energy line is almost totally juxtaposed to the consensus line, as the theory prescribes in the case of $N = 2$ agents).

Furthermore, notice that, as soon as we shut down the control, the two agents will start to move again in opposite directions (very slowly, since the energy of a system without control stays constant), hence not only the total energy remains above the threshold, but also the system is not in consensus.

However, it must be observed that the control strategy fails in this situation due to the

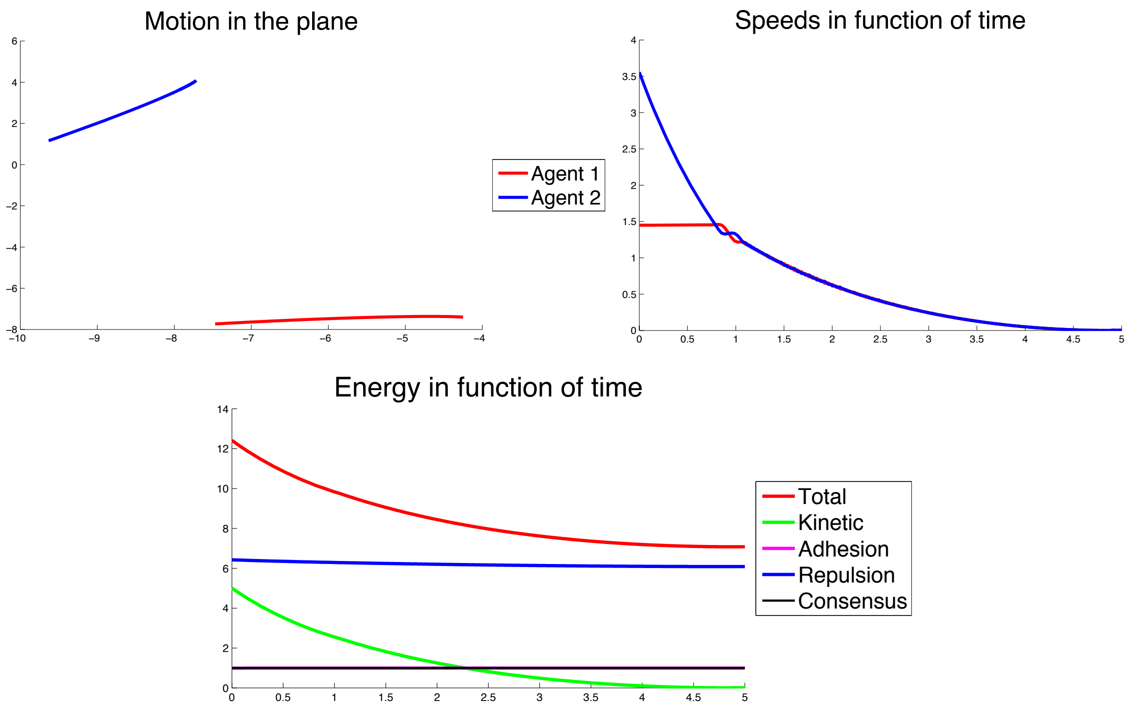


Figure 2.9: Space evolution, speeds and energy profile of the system considered in Section 2.6.3.

peculiar nature of the system. As a matter of fact, being the force balance strictly repulsive, the trajectories of any solution will never remain cohesive. This leaves open the question whether there exist “non-pathological” instances of the Cucker-Dong model (in the sense that their solutions are not doomed to diverge regardless of the initial condition) for which the sparse control strategy does not work.

CHAPTER 3

Sparse control in high dimension

In the previous chapter we have seen several features of sparse feedback controls: effectiveness, parsimony and efficiency. However, since feedback controls need at every instant the current state of the entire system to perform their task, processing this information to compute the sparse control strategy may quickly become unfeasible as d increases, an instance of the phenomenon that Richard Bellman in [24] named *curse of dimensionality*. In this chapter we shall see how we can circumvent such inconvenience by means of a dimensionality reduction via *Johnson-Lindenstrauss embeddings*.

3.1 The curse of dimensionality

To visualize the idea behind the curse of dimensionality consider, as an example, the problem of numerically simulating the familiar Cucker-Smale system

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right), \end{cases} \quad i = 1, \ldots, N, \tag{3.1}$$

where $(x_i(t), v_i(t)) \in \mathbb{R}^{2d}$ for every $t \geq 0$. When the number of agents N and their dimensionality d are relatively modest – say, around 100 each – this is far from being a challenge: any Runge-Kutta scheme will perform the task quickly and with a reasonable degree of accuracy. But when N or d are of the order of thousands, or worse of millions, such strategy could need hours or days to return meaningful results. And this is not caused by the particular numerical scheme one might choose: this is a severe limitation of the formulation of the problem. And this is what, in practice, the curse of dimensionality is: the cost of the numerical computations needed to find solutions of a problem blows up as the size of certain parameters increases.

The curse of dimensionality hits also the sparse feedback strategy of Definition 2.4: to know which is the agent $\hat{\imath}(t)$ that must be controlled at time t, we must compute the discrepancies from the mean $v_i^{\perp}(t)$ (which is a d-dimensional vector) for each agent $i = 1, \ldots, N$, again a procedure that scales horribly with N and d.

The issue of understanding how sparse controls can be implemented as $N \to \infty$ and d is modest has been investigated in [103, 106], and we shall have a closer look to these (and related) results in Chapter 4 and 6. In this chapter we instead focus on the other case, i.e., $d \to \infty$ and N remains relatively low.

One might argue that systems such as (3.1) have no application outside the three dimensional Euclidean space, since they were introduced to study the coordination of animals. However, there are several instances of "abstract alignment" which may occur in high-dimension: for instance, in [3] the authors consider an application of alignment models to predict the collective phenomenon of asset pricing and volatilities in financial markets, while in [23] similar models are used to describe qualitatively the evolution of systemic risk in the banking sector.

Furthermore, notice that solving such an issue is of paramount importance, since in view of the increasing technical ability of collecting large amounts of time-evolving data and of potentially modeling them into high-dimensional dynamical systems, the controllability of complex multi-agent interactions has an enormous social and economical impact. Therefore, in those circumstances where the dimensionality of the dynamics is very high, it becomes a relevant question whether it is possible to define control strategies of the dynamics overcoming the curse of dimensionality.

3.1.1 Dimensionality reduction via Johnson-Lindenstrauss embeddings

In recent years, several techniques have been developed in order to reduce the dimensionality of time-evolving point clouds, such as *diffusion maps* applied to networks changing in time [68] and geometric multiscale dimensionality reductions [40], just to mention few of them.

Besides these perhaps involved methods based on computationally demanding nonlinear embeddings of the high-dimensional clouds in lower dimension, Johnson-Lindenstrauss embeddings, introduced in the seminal work [131], have the remarkable property of being simple *linear operators* $M \in \mathbb{R}^{k \times d}$ preserving the distances between points in the cloud $\mathcal{P} \subset \mathbb{R}^d$ up to an ε-distortion, i.e.,

$$(1-\varepsilon)\|x - x'\|_{\ell_2^d} \leq \|Mx - Mx'\|_{\ell_2^k} \leq (1+\varepsilon)\|x - x'\|_{\ell_2^d}, \quad \text{for all } x, x' \in \mathcal{P},$$

where the projected dimension k satisfies

$$k \sim \varepsilon^{-2} \log(|\mathcal{P}|). \tag{3.2}$$

The above property is called *quasi-isometry property*, and as Johnson-Lindenstrauss embeddings with such scaling of the projected dimension are constructed by generating random projections, the quasi-isometry property on the point cloud is usually stated with a certain (high) probability. The random linear projection of high-dimensional systems governed by smooth nonlinearities depending on mutual distances has been investigated in [101]: roughly speaking, given a dynamical system in high dimension $d \gg 1$ governed by locally Lipschitz functions $f_i : \mathbb{R}_+^{N\times N} \to \mathbb{R}^d$

$$\dot{z}_i(t) = f_i((\|z_j(t) - z_\ell(t)\|)_{j,\ell=1}^N) \in \mathbb{R}^d, \quad i = 1, \ldots, N,$$

and its lower-dimensional counterpart

$$\dot{\zeta}_i(t) = M f_i((\|\zeta_j(t) - \zeta_\ell(t)\|)_{j,\ell=1}^N) \in \mathbb{R}^k, \quad i = 1, \ldots, N,$$

where $M : \mathbb{R}^d \to \mathbb{R}^k$ is a Johnson-Lindenstrauss linear embedding for $k \sim \varepsilon^{-2}\log(N)$, the following finite-time approximation holds with high probability

$$\|\zeta_i(t) - M z_i(t)\| \le C_T \varepsilon, \quad \text{for all } t \in [0, T]. \tag{3.3}$$

Notice here the key property of projected dynamical systems: intuitively, if the number of agents N is very small in comparison to their dimensionality d, we expect the projected dimensionality k given by (3.2) to be very small with respect to d, and hence we save a huge amount of time by simulating the projected system $\zeta_i(t)$ instead of the original one $z_i(t)$. Since, by (3.3), the distance $\|\zeta_i(t) - M z_i(t)\|$ is "small", we can then reconstruct the original dynamics $z_i(t)$ using compressed sensing techniques, see [101].

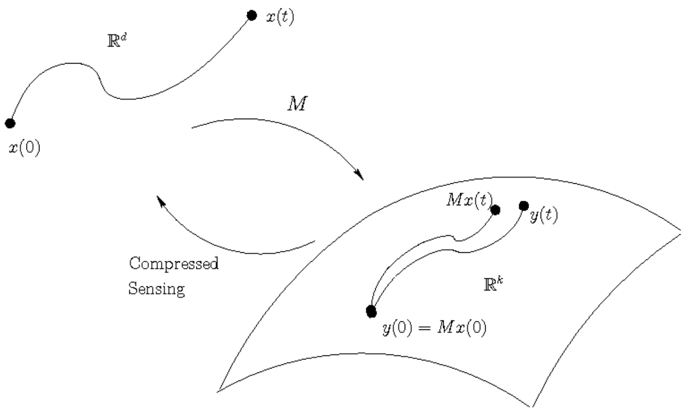


Figure 3.1: Dimensionality reduction of dynamical systems.

In this chapter we shall see how this method of projecting high-dimensional systems in low dimension can also be applied in the context of sparse controls, leading to the following

Meta-theorem. *For high dimensional multiagent systems, governed by smooth nonlinearities depending on mutual distances between the agents, one can construct low-dimensional representations of the dynamical system, which allow the computation of nearly optimal control strategies in high dimension with overwhelming confidence.*

As control is usually goal-oriented, hence highly dependent on the specific dynamical system, investigating the qualitative applicability of this statement in its full generality may risk to dilute its quantitative understanding. Thus, we shall prove a specific instance of it, which conveys nonetheless all the relevant aspects and technical issues potentially encountered in other situations: we shall focus on the sparse control of the Cucker-Smale model, for which we have already at our disposal all the knowledge accumulated in Chapters 1 and 2. We mention that similar results for the Cucker-Dong model are obtained in [27].

Intuitively, if we applied a Johnson-Lindenstrauss embedding $M \in \mathbb{R}^{k\times d}$ verbatim to each equation of a Cucker-Smale system (3.1), we would obtain the following approximation

$$\frac{d}{dt}Mv_i(t) = \frac{1}{N}\sum_{j=1}^{N} a\left(\|x_j(t)-x_i(t)\|\right)\left(Mv_j(t)-Mv_i(t)\right)$$
$$\overset{\text{wish}}{\sim} \frac{1}{N}\sum_{j=1}^{N} a\left(\|Mx_j(t)-Mx_i(t)\|\right)\left(Mv_j(t)-Mv_i(t)\right),$$

leading to the formulation of the low-dimensional system in $\mathbb{R}^k$

$$\begin{cases} \dot{y}_i(t) = w_i(t), \\ \dot{w}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|y_i(t)-y_j(t)\|\right)\left(w_j(t)-w_i(t)\right), \end{cases} \quad i = 1,\dots,N,$$

with initial conditions $(y^0, w^0) = (Mx^0, Mv^0) \in \mathbb{R}^{kN}\times\mathbb{R}^{kN}$, where $(x^0, v^0) \in \mathbb{R}^{dN}\times\mathbb{R}^{dN}$ is the initial datum of the d-dimensional system (3.1) and for the sake of brevity we have set

$$(Mx^0, Mv^0) \triangleq (Mx_1^0,\dots,Mx_N^0, Mv_1^0,\dots,Mv_N^0). \tag{3.4}$$

The first result of this chapter, refining and generalizing those in [101], is roughly summarized as follows.

Theorem 3.1. *Let $(x(\cdot), v(\cdot))$ be a solution of the d-dimensional Cucker-Smale system for the given initial datum $(x^0, v^0) \in \mathbb{R}^{dN}\times\mathbb{R}^{dN}$, and let $M \in \mathbb{R}^{k\times d}$ be a Johnson-Lindenstrauss matrix for a suitable $\varepsilon > 0$ distortion parameter and low dimension k depending on the logarithm of the number of agents N. Then the k-dimensional solution $(y(\cdot), w(\cdot))$ with initial values $(y^0, w^0) = (Mx^0, Mv^0)$ stays close to the projected d-dimensional trajectory $(Mx(\cdot), Mv(\cdot))$, i.e.,*

$$\|y(t)-Mx(t)\| + \|w(t)-Mv(t)\| \lesssim \varepsilon e^{Ct}, \quad \text{for every } t \le T. \tag{3.5}$$

As we highlight in details in Section 3.3, not only the approximation (3.5) holds for finite time, but, remarkably, the lower dimensional representation also shows a rather impressive faithfulness in terms of the asymptotic (long time) detection of collective behavior emergence, i.e., global alignment occurs in lower dimension k if and only if it occurs in high dimension d with high probability. The key technical tool for proving this result and the ones following is a weak form of the Johnson-Lindenstrauss Lemma, formulated below in Lemma 3.6, valid for continuous trajectories and not only for clouds of points. Similar results appear, to some extent in greater generality in [22, 94], but not in the weak form considered in this chapter.

3.1.2 Dimensionality reduction and sparse controls

As already mentioned, we combine the above analysis with the sparse controllability of Cucker-Smale systems outlined in Section 2.1, and show that a high-dimensional dynamical system of Cucker-Smale type can be *nearly optimally* stabilized towards consensus by means of a control strategy completely identified by the *optimal* control strategy in low dimension with high probability. More formally, we consider for a given initial datum (x^0, v^0) the high-dimensional controlled system

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \frac{1}{N}\sum_{j=1}^{N} a\left(\|x_j(t) - x_i(t)\|\right)\left(v_j(t) - v_i(t)\right) + u_i^h(t), \end{cases} \qquad i = 1, \ldots, N,$$

and its low-dimensional counterpart with initial datum $(y^0, w^0) = (Mx^0, Mv^0)$ in the sense of (3.4),

$$\begin{cases} \dot{y}_i(t) = y_i(t), \\ \dot{w}_i(t) = \frac{1}{N}\sum_{j=1}^{N} a\left(\|y_j(t) - y_i(t)\|\right)\left(w_j(t) - w_i(t)\right) + u_i^\ell(t). \end{cases} \qquad i = 1, \ldots, N,$$

The family of *sparse* control strategies applied to the systems are defined as follows: fix $\theta > 0$ and for every $t \geq 0$ set

$$\begin{aligned} u_i^\ell(t) &\triangleq \begin{cases} -\theta \dfrac{w_{\hat{\imath}}^\perp(t)}{\|w_{\hat{\imath}}^\perp(t)\|} & \text{if } i = \hat{\imath}(t), \\ 0 & \text{otherwise}, \end{cases} \\ u_i^h(t) &\triangleq \begin{cases} -\theta \dfrac{v_{\hat{\imath}}^\perp(t)}{\|v_{\hat{\imath}}^\perp(t)\|} & \text{if } i = \hat{\imath}(t), \\ 0 & \text{otherwise}, \end{cases} \end{aligned} \tag{3.6}$$

where $\hat{\imath}(t)$ is the smallest index such that $\|w_{\hat{\imath}}^\perp(t)\| = \max_{1\leq j\leq N} \|w_j^\perp(t)\|$.

Notice that:

- the control $u^\ell(t)$ is precisely the sparse feedback control of Definition 2.4;
- the control $u^h(t)$ is *sparse* (all the components are zero except one);
- $u^h(t)$ is defined exclusively through the following information: the index $\hat{\imath}(t)$ which is computed from the low-dimensional control problem, the consensus parameter $v_{\hat{\imath}}(t)$, which is actually the only information to be observed in high dimension, and the mean consensus parameter

$$\overline{v}(t) = \overline{v}(0) + \frac{1}{N}\sum_{i=1}^{N}\int_0^t u_i^h(s)ds,$$

which one does compute by integration and sums of previous controls.

The main result of this chapter reads as follows.

Theorem 3.2. *Let $M \in \mathbb{R}^{k\times d}$ and $\theta > 0$. Assume that $(x(\cdot), v(\cdot))$ and $(y(\cdot), w(\cdot))$ are solutions of the d-dimensional and k-dimensional controlled Cucker-Smale systems with* *the control strategy* (3.6) *and with initial values* (x^0, v^0) *and* (Mx^0, Mv^0)*, respectively. Further assume that* M *is a Johnson-Lindenstrauss matrix for a certain distortion* $\varepsilon > 0$ *and low dimension* k *(which depends exponentially on the number of agents, but not on the dimension* d*). Then both solutions*

(i) *stay close to each other, after the projection of the high-dimensional trajectories;*

(ii) *reach the consensus region of Theorem 1.11 in finite time, and*

(iii) *reach the consensus region when a certain parameter of the low-dimensional system falls below a known threshold.*

We consciously do not wish to be more detailed at this point than this rather general and perhaps rough explanation because the precise statements appear in the rest of the chapter in a rather technical form and we want here, in the introduction, mainly to convey their fundamental message: sparse controls can be effectively and efficiently implemented on a large class of dynamical systems regardless of the dimensionality.

3.2 A continuous Johnson-Lindenstrauss Lemma

As it will be made clear below, we intend to reduce the computational effort of extracting fundamental features of the dynamical system (3.1), for instance its asymptotic behavior, by projecting it to a k-dimensional space for $k \ll d$ by a linear mapping $M \in \mathbb{R}^{k\times d}$. In particular, as already outlined in the previous section, we apply such a matrix M to each equation of (3.1) and by setting $y_i = Mx_i$ as well as $w_i = Mv_i$ for $i = 1, \dots, N$, we obtain the system

$$\begin{cases} \dot{y}_i(t) = w_i(t), \\ \dot{w}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|y_i(t) - y_j(t)\|\right)\left(w_j(t) - w_i(t)\right), \end{cases} \quad i = 1, \dots, N,$$

where we formally applied the identity

$$\|x_i(t) - x_j(t)\|_{\ell_2^d} \overset{!}{=} \|Mx_i(t) - Mx_j(t)\|_{\ell_2^k} \left(= \|y_i(t) - y_j(t)\|_{\ell_2^k}\right). \tag{3.7}$$

For (3.7) to hold, at least approximately, we need that M is "nearly" an isometry (here we further refine and extend results from [101, Section 3]).

Definition 3.3. Let $M \in \mathbb{R}^{k\times d}$, $\delta > 0$, and $\varepsilon \in (0,1)$. Then we say, that M is fulfilling the *weak Johnson-Lindenstrauss property* of parameters ε and δ at $x \in \mathbb{R}^d$ if either

$$(1-\varepsilon)\|x\|_{\ell_2^d} \le \|Mx\|_{\ell_2^k} \le (1+\varepsilon)\|x\|_{\ell_2^d} \tag{3.8}$$

or

$$\|x\|_{\ell_2^d} \le \delta \quad \text{and} \quad \|Mx\|_{\ell_2^k} \le \delta \tag{3.9}$$

holds. We say that M is fulfilling the *(strong) Johnson-Lindenstrauss property* of parameter ε at $x \in \mathbb{R}^d$ if exclusively (3.8) holds at $x \in \mathbb{R}^d$.

Remark 3.4. The earliest result providing the existence of matrices M for which (3.8) holds for every $x \in \mathcal{P}$, where $\mathcal{P} \subseteq \mathbb{R}^d$ is such that $N = |\mathcal{P}|$ and for the dimensionality k scaling as

$$k \sim \varepsilon^{-2} \log N, \tag{3.10}$$

is the celebrated Johnson-Lindenstrauss Lemma from the seminal paper [131]. We refer to [94] for a rather general version of this result and to the references therein for an extended literature.

The only constructions of a matrix M fulfilling the (strong) Johnson-Lindenstrauss property with scaling (3.10) known up to now are stochastic, i.e., the matrix is randomly generated and satisfies (3.8) with high probability. One of the remarkable features of these embeddings, which we exploit extensively in this paper, is that for their construction there is no need to know the specific points in advance: given a fixed cloud of points (not necessarily explicitly given!) a random matrix drawn according to certain distributions will fulfill the (strong) Johnson-Lindenstrauss property with high probability. Let us recall briefly some well-known instances of such distributions:

(S1) $k \times d$ matrices M whose entries m_{ij} are independent realizations of Gaussian random variables, i.e.,

$$m_{ij} \sim \mathcal{N}\left(0, \frac{1}{k}\right);$$

(S2) $k \times d$ matrices M whose entries m_{ij} are independent realizations of scaled Bernoulli random variables, i.e.,

$$m_{ij} = \begin{cases} +\frac{1}{\sqrt{k}} & \text{with probability } \frac{1}{2}, \\ -\frac{1}{\sqrt{k}} & \text{with probability } \frac{1}{2}. \end{cases}$$

Notice that it holds $1 \leq \|M\|_{\ell_2^d \to \ell_2^k} \leq \sqrt{d}$.

(S3) $k \times d$ matrices M which are random projections and are scaled by a factor $\sqrt{d/k}$, see [88]. In particular, it holds $\|M\|_{\ell_2^d \to \ell_2^k} = \sqrt{d/k}$.

Remark 3.5. While the Johnson-Lindenstrauss Lemma is a result for a finite number of points, we need an analogous continuous result for projecting trajectories of dynamical systems. A result in this direction was given in [101, Theorem 3.3]: given $\varepsilon > 0$ and a $\mathcal{C}^1$-curve $\varphi : [0,1] \to \mathbb{R}^d$, if φ has finite curvature, i.e.,

$$\rho \triangleq \max_{t \in [0,1]} \frac{\|\varphi'(t)\|}{\|\varphi(t)\|} < +\infty, \tag{3.11}$$

then there exists a matrix $M \in \mathbb{R}^{k \times d}$ for $k \sim \varepsilon^{-2} \log(d \cdot \rho \cdot \varepsilon^{-1})$ such that

$$(1-\varepsilon)\|\varphi(t)\| \leq \|M\varphi(t)\| \leq (1+\varepsilon)\|\varphi(t)\| \quad \text{for all } t \in [0,1]. \tag{3.12}$$

As already announced at the beginning of this section, we want to use (3.12) for

$$\varphi(t) \triangleq x_i(t) - x_j(t) \quad \text{or} \quad \varphi(t) \triangleq v_i(t) - v_j(t)$$

being $(x_i(\cdot), v_i(\cdot))$ the trajectory of the i-th agent in (3.1).
Unfortunately, (3.11) does not hold in this case even if we assume that $\|x_i(0) - x_j(0)\| \geq c > 0$ for all $i \neq j$. Let us consider, for instance, Example 1.13 of a Cucker-Smale system of the type (3.1) of two agents moving on the real line with positions and velocities at time t given by $(x_1(t), v_1(t))$ and $(x_2(t), v_2(t))$. We already noticed that, by indicating with $x(t) = x_1(t) - x_2(t)$ the relative main state and by $v(t) = v_1(t) - v_2(t)$ the relative consensus parameter at time t, then it holds

$$\dot{x}(t) = v(t) = -\arctan(x(t)) + \arctan(x(0)) + v(0).$$

Now, if $x(0) < 0$ and $v(0) + \arctan(x(0)) \triangleq c_0 > 0$, then $v(t) > c_0$ as long as $x(t) < 0$. Hence, there has to be a $T > 0$ with $x(T) = 0$ and $v(T) = c_0$. Therefore (3.11) is violated for the curve $\varphi(t) = x_1(t) - x_2(t)$.
Let us stress that (3.11) is a necessary condition for (3.12) to hold (see [101, Remark 1]). This motivates the relaxation of the strong Johnson-Lindenstrauss property to its weak version in Definition 3.3. We now prove a result based on the more general weak Johnson-Lindenstrauss property which will be sufficient for us in the following.

From now on, given two normed spaces $\mathcal{X}$ and $\mathcal{Y}$ with norms $\|\cdot\|_{\mathcal{X}}$ and $\|\cdot\|_{\mathcal{Y}}$, respectively, by $\mathrm{Lip}(\mathcal{X};\mathcal{Y})$ we denote the set of *Lipschitz functions* $\varphi : \mathcal{X} \to \mathcal{Y}$, i.e., the functions for which the quantity

$$\mathrm{Lip}_{\mathcal{X}}(\varphi) \triangleq \sup_{\substack{x,y\in\mathcal{X} \\ x\neq y}} \frac{\|\varphi(x) - \varphi(y)\|_{\mathcal{Y}}}{\|x - y\|_{\mathcal{X}}}$$

exists and is finite.

Lemma 3.6. *Let $\varphi : [0,1] \to \mathbb{R}^d$ be a Lipschitz function with Lipschitz constant $L_\varphi \triangleq \mathrm{Lip}_{[0,1]}(\varphi)$, and fix $\delta > 0$ and $\varepsilon \in (0,1)$. Let k be such that a matrix $M \in \mathbb{R}^{k\times d}$ – stochastically generated as in (S2) or (S3) of Remark 3.4 – satisfies the (strong) Johnson-Lindenstrauss property of parameter $\tilde{\varepsilon} = \varepsilon/2$ at $\mathcal{N}$ arbitrary points with some (high) probability, where*

$$\mathcal{N} \geq \frac{4L_\varphi(\sqrt{d} + 2)}{\delta\varepsilon}. \tag{3.13}$$

Then the matrix M fulfills the weak Johnson-Lindenstrauss property of parameters ε and δ at $\varphi(t)$ for every $t \in [0,1]$ with the same high probability, i.e., either

$$(1 - \varepsilon)\|\varphi(t)\| \leq \|M\varphi(t)\| \leq (1 + \varepsilon)\|\varphi(t)\|$$

or

$$\|\varphi(t)\| \leq \delta \text{ and } \|M\varphi(t)\| \leq \delta$$

holds for all $t \in [0, 1]$.

Proof. We shall adapt the arguments from the proof of [101, Theorem 3.3]: set

$$t_i \triangleq \frac{i}{\mathcal{N}}, \qquad \text{for } i = 0, \dots, \mathcal{N} - 1,$$

and assume that $M : \mathbb{R}^d \to \mathbb{R}^k$ fulfills the (strong) Johnson-Lindenstrauss property with parameter $\tilde{\varepsilon} = \varepsilon/2$ at the points $(\varphi(t_i))_{i=0}^{\mathcal{N}-1}$, i.e., we have

$$(1 - \tilde{\varepsilon})\|\varphi(t_i)\| \le \|M\varphi(t_i)\| \le (1 + \tilde{\varepsilon})\|\varphi(t_i)\|$$

for all $i = 0, \dots, \mathcal{N} - 1$. Furthermore, we may assume $1 \le \|M\| \le \sqrt{d}$, see (S2) and (S3) of Remark 3.4.
Let $t \in [0, 1]$ and choose $j \in \{0, \dots, \mathcal{N} - 1\}$ such that $t \in [t_j, t_{j+1}]$. Let us for the moment assume that

$$\|\varphi(t_j)\| \le \frac{\delta}{2}. \tag{3.14}$$

Since $\varepsilon \in (0, 1)$, by (3.13) we have that

$$\mathcal{N} \ge \frac{4L_\varphi\sqrt{d}}{\delta}.$$

Using this latter inequality and the Lipschitz continuity of φ we obtain

$$\|\varphi(t)\| \le \|\varphi(t) - \varphi(t_j)\| + \|\varphi(t_j)\| \le \frac{L_\varphi}{\mathcal{N}} + \frac{\delta}{2} \le \delta,$$

and also

$$\|M\varphi(t)\| \le \|M\|\|\varphi(t) - \varphi(t_j)\| + \|M\varphi(t_j)\| \le \sqrt{d}\frac{L_\varphi}{\mathcal{N}} + (1 + \varepsilon')\|\varphi(t_j)\| \le \frac{\delta}{4} + \frac{3}{2} \cdot \frac{\delta}{2} \le \delta.$$

Now, suppose instead that (3.14) does not hold, i.e., we have

$$\|\varphi(t_j)\| > \frac{\delta}{2}.$$

Using again the Lipschitz continuity of φ we obtain the estimate

$$\|\varphi(t) - \varphi(t_j)\| \le \frac{L_\varphi}{\mathcal{N}} \le \frac{L_\varphi}{\mathcal{N}} \cdot \frac{2\|\varphi(t_j)\|}{\delta} \le \frac{\delta\varepsilon}{4(\sqrt{d} + 2)} \cdot \frac{2\|\varphi(t_j)\|}{\delta} \le \frac{\|\varphi(t_j)\|(\varepsilon - \tilde{\varepsilon})}{\|M\| + 1 + \varepsilon},$$

where in the last inequality we used that

$$\|M\| + 1 + \varepsilon \le \sqrt{d} + 2.$$

This estimate of the distance $\|\varphi(t) - \varphi(t_j)\|$ and the (strong) Johnson-Lindenstrauss property at $\varphi(t_j)$ enable us to extend the (strong) Johnson-Lindenstrauss property at $\varphi(t)$ as well, as a direct application of [101, Lemma 3.2], i.e.,

$$(1-\varepsilon)\|\varphi(t)\| \le \|M\varphi(t)\| \le (1+\varepsilon)\|\varphi(t)\|.$$

Both cases together show the (weak) Johnson-Lindenstrauss property at $\varphi(t)$ for every $t \in [0,1]$. □

In the following lemma, we show that the mean-square norm and the relative order of the magnitudes of points in a cloud in high dimension are nearly preserved when projected in lower dimension by a weak Johnson-Lindenstrauss embedding.

Lemma 3.7. *Let $a_1, \dots, a_N \in \mathbb{R}^d, b_1, \dots, b_N \in \mathbb{R}^k$ and $M \in \mathbb{R}^{k\times d}$ such that there is $\Delta > 0$ with the following properties:*

(i) The matrix M fulfills the weak Johnson-Lindenstrauss property with $\varepsilon = 1/2$ and $\delta = \Delta$ for the points a_i, i.e., either

$$\frac{1}{2}\|a_i\| \le \|Ma_i\| \le \frac{3}{2}\|a_i\| \tag{3.15}$$

or

$$\|a_i\| \le \Delta \quad \text{and} \quad \|Ma_i\| \le \Delta \tag{3.16}$$

holds for all $i = 1, \dots, N$.

(ii) We have the following approximation bound

$$\|Ma_i - b_i\| \le \Delta$$

for all $i = 1, \dots, N$.

Let $\hat\imath$ be the smallest index such that $\|b_{\hat\imath}\| \ge \|b_j\|$ for all $j = 1, \dots, N$ and let

$$A \triangleq \frac{1}{N}\sum_{j=1}^N \|a_j\|^2 \quad \text{and} \quad B \triangleq \frac{1}{N}\sum_{j=1}^N \|b_j\|^2.$$

If $\sqrt{B} \ge 2\Delta$, then, for $c \triangleq 1/\sqrt{289}$, it holds

$$\|a_{\hat\imath}\| \ge \frac{1}{4}\|b_{\hat\imath}\|, \quad \|a_{\hat\imath}\| \ge c\sqrt{A}, \quad \text{and} \quad B \le 16NA.$$

If $\sqrt{B} \le 2\Delta$, then, for $C \triangleq \sqrt{72}$, it holds

$$\sqrt{A} \le C\Delta.$$

Proof. First suppose that $\sqrt{B} \ge 2\Delta$: since $\|b_{\hat\imath}\|$ is maximal, we have $\|b_{\hat\imath}\| \ge \sqrt{B} \ge 2\Delta$. By (ii) it holds

$$\|Ma_{\hat\imath}\| \ge \|b_{\hat\imath}\| - \Delta \ge 2\Delta - \Delta \ge \Delta$$

and hence using (3.15) we get

$$\|a_{\hat\imath}\| \geq \frac{1}{2}\|Ma_{\hat\imath}\| \geq \frac{1}{2}(\|b_{\hat\imath}\| - \Delta) \geq \frac{1}{4}\|b_{\hat\imath}\|. \tag{3.17}$$

This shows the first estimate of the statement. Let us address the second estimate. Let $j \in \{1,\ldots,N\}$ for $j \neq \hat\imath$. If $\|b_j\| \geq 2\Delta$, then, using the same argument as above, we have $\|Ma_j\| \geq \Delta$ and thus by (3.15) we get

$$\|a_j\| \leq 2\|Ma_j\| \leq 2(\|b_j\| + \Delta) \leq 2 \cdot \frac{3}{2} \cdot \|b_j\| = 3\|b_j\|. \tag{3.18}$$

On the other hand, if $\|b_j\| < 2\Delta$, then $\|Ma_j\| \leq 3\Delta$. Then either (3.15) holds and we have

$$\|a_j\| \leq 2\|Ma_j\| \leq 6\Delta, \tag{3.19}$$

or (3.16) holds and automatically $\|a_j\| \leq \Delta$. Now we can estimate the mean-square norm A. We obtain

$$NA = \sum_{j=1}^{N} \|a_j\|^2 = \|a_{\hat\imath}\|^2 + \sum_{j\in A_1} \|a_j\|^2 + \sum_{j\in A_2} \|a_j\|^2,$$

where A_1 is the index set of all $j \in \{1,\ldots,N\} \setminus \{\hat\imath\}$ such that $\|b_j\| \geq 2\Delta$ and A_2 is the index set of all $j \in \{1,\ldots,N\}$ for which $\|b_j\| < 2\Delta$. Using (3.18) and (3.19) we obtain

$$\begin{aligned} NA &\leq \|a_{\hat\imath}\|^2 + 9\sum_{j\in A_1} \|b_j\|^2 + |A_2| \cdot 36\Delta^2 \\ &\leq \|a_{\hat\imath}\|^2 + 9NB + 9N\|b_{\hat\imath}\|^2 \\ &\leq \|a_{\hat\imath}\|^2 + N\|a_{\hat\imath}\|^2(9\cdot 16 + 9\cdot 16) \\ &\leq 289\|a_{\hat\imath}\|^2 N, \end{aligned}$$

using the maximality of $\|b_{\hat\imath}\|$ and the inequality (3.17). Furthermore, we have

$$NB = \sum_{j=1}^{N} \|b_j\|^2 \leq N\|b_{\hat\imath}\|^2 \leq 16N\|a_{\hat\imath}\|^2 \leq 16N^2 A,$$

whence the inequality $B \leq 16NA$.
Let now $\sqrt{B} \leq 2\Delta$. We can argue similarly to the second estimate of the first part:

- if $\|b_j\| \geq 2\Delta$, then as in (3.18) we have $\|a_j\| \leq 3\|b_j\|$;
- if $\|b_j\| \leq 2\Delta$, then by (3.19) and the arguments thereafter we get $\|a_j\| \leq 6\Delta$.

Putting both estimates together and using the notation $\tilde{A}_1$ for the index set of all $j \in \{1,\ldots,N\}$ such that $\|b_j\| \geq 2\Delta$ as well as $\tilde{A}_2$ for the index set of all $j \in \{1,\ldots,N\}$ such that $\|b_j\| < 2\Delta$ yield

$$NA = \sum_{j=1}^{N} \|a_j\|^2$$

$$
\begin{aligned}
&= \sum_{j \in \tilde{A}_1} \|a_j\|^2 + \sum_{j \in \tilde{A}_2} \|a_j\|^2 \\
&\leq 9 \sum_{j \in \tilde{A}_1} \|b_j\|^2 + |\tilde{A}_2| \cdot 36\Delta^2 \\
&\leq 9NB + 36N\Delta^2 \\
&\leq N(9B + 36\Delta^2) \\
&\leq N(36\Delta^2 + 36\Delta^2) \\
&\leq 72N\Delta^2.
\end{aligned}
$$

Taking the square root on both sides finishes the proof. □

3.3 Dimension reduction of the Cucker-Smale system without control

In this section we consider the projection of the Cucker-Smale system without control. We compare two quantities: first, we calculate the trajectory $(x(\cdot), v(\cdot))$ of the high-dimensional Cucker-Smale system and we project them by $M \in \mathbb{R}^{k\times d}$, to obtain $(Mx(\cdot), Mv(\cdot))$ in the sense of (3.4). Second, we project the initial configurations to dimension k by M, and then we compute from these initial values the trajectories of the corresponding low-dimensional Cucker-Smale system $(y(\cdot), w(\cdot))$. What we shall do in the upcoming Theorem 3.9 is to give a precise bound from above to the distance between the the two k-dimensional trajectories $(Mx(\cdot), Mv(\cdot))$ and $(y(\cdot), w(\cdot))$.

Formally, given $M \in \mathbb{R}^{k\times d}$ (where $k \leq d$) and initial conditions (x^0, v^0) for (3.1), we indicate with $(y(\cdot), w(\cdot))$ the solution of the $\mathbb{R}^k$-projected Cucker-Smale system

$$
\begin{cases}
\dot{y}_i(t) = w_i(t), \\
\dot{w}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|y_i(t) - y_j(t)\|\right)\left(w_j(t) - w_i(t)\right),
\end{cases}
\qquad i = 1, \ldots, N,
$$

with initial data $y^0 = \left(Mx_1^0, \ldots, Mx_N^0\right) \in \mathbb{R}^{kN}$ and $w^0 = \left(Mv_1^0, \ldots, Mv_N^0\right) \in \mathbb{R}^{kN}$.

We define the low-dimensional analogues of X and V (introduced in (1.11) and (1.12), respectively) by

$$
Y(t) \triangleq B(y(t), y(t)), \qquad W(t) \triangleq B(w(t), w(t)). \tag{3.20}
$$

Here the bilinear form B is intended to act on $\mathbb{R}^{kN}$ instead of $\mathbb{R}^{dN}$, but with the same meaning of the symbol as before.

Remark 3.8. By Corollary 1.19 we know that V and W are decreasing. Hence for all $i, j = 1, \ldots, N$, it holds

$$
\begin{aligned}
\|w_i(t) - w_j(t)\|^2 &\leq 2\left(\|w_i(t) - \overline{w}(t)\|^2 + \|w_j(t) - \overline{w}(t)\|^2\right) \\
&\leq 2\sum_{\ell=1}^{N} \|w_\ell(t) - \overline{w}(t)\|^2
\end{aligned}
$$

$$\leq 2NW(t)$$
$$\leq 2NW(0),$$

thus

$$\|w_i(t) - w_j(t)\| \leq \sqrt{2NW(0)}.$$

An analogous estimate holds for V and v. Furthermore, we have

$$\|x_i(0) - x_j(0)\| \leq \sqrt{2NX(0)}.$$

Theorem 3.9. *Let $\delta > 0$, let $\varepsilon \in (0,1)$, and let $M \in \mathbb{R}^{k\times d}$ be a matrix satisfying the weak Johnson-Lindenstrauss property of parameters ε and δ at the vectors $x_i(t) - x_j(t)$ for all $t \in [0,T]$ and all $i,j = 1,\ldots,N$.*
Define the following errors:

$$\begin{aligned} e_i^x(t) &\triangleq \|y_i(t) - Mx_i(t)\|, & e_i^v(t) &\triangleq \|w_i(t) - Mv_i(t)\|, \\ \mathcal{E}^x(t) &\triangleq \max_{1\leq i\leq N} e_i^x(t), & \mathcal{E}^v(t) &\triangleq \max_{1\leq i\leq N} e_i^v(t), \\ \mathcal{E}_2^x(t) &\triangleq \left(\frac{1}{N}\sum_{i=1}^N (e_i^x(t))^2\right)^{1/2}, & \mathcal{E}_2^v(t) &\triangleq \left(\frac{1}{N}\sum_{i=1}^N (e_i^v(t))^2\right)^{1/2}. \end{aligned} \tag{3.21}$$

Furthermore, let $L_a \triangleq \mathrm{Lip}_{\mathbb{R}_+}(a)$ and set

$$K_1 \triangleq L_a\sqrt{NW(0)}\sqrt{2X(0)}, \quad K_2 \triangleq 2L_a\sqrt{NW(0)}, \quad K_3 \triangleq \frac{L_a}{2}\sqrt{NW(0)}\sqrt{2V(0)}.$$

Then for all $t \in [0,T]$ the estimates

$$\mathcal{E}^x(t) + \mathcal{E}^v(t) \leq \sqrt{N}((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{t\|\mathcal{K}\|_{\ell_1\to\ell_1}}$$

and

$$\mathcal{E}_2^x(t) + \mathcal{E}_2^v(t) \leq ((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{t\|\mathcal{K}\|_{\ell_1\to\ell_1}}$$

hold, where

$$\mathcal{K} \triangleq \begin{pmatrix} 2a(0) & 2L_a\sqrt{NW(0)} \\ 1 & 0 \end{pmatrix}.$$

Moreover, we have

$$\mathcal{E}^v(t) \leq \sqrt{N}\min\left\{((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{t\|\mathcal{K}\|_{\ell_1\to\ell_1}}, \|M\|\sqrt{V(t)} + \sqrt{W(t)}\right\}.$$

Proof. We estimate the decay of $\mathcal{E}_2^x$ and $\mathcal{E}_2^v$ in order to use Gronwall's Lemma A.1. For the following estimates we may assume that – without loss of generality – $e_i^v(t) \neq 0$ for $t \in [0,T]$ and for every $i = 1,\ldots,N$: if this is not the case, either $e_i^v \equiv 0$ in a neighborhood of t or, by continuity, the estimates will also hold true at t. Hence we may assume that e_i^v

is differentiable at $t \in [0,T]$. By the Cauchy-Schwarz inequality it holds

$$\begin{aligned}\frac{d}{dt}e_i^v(t) &= \frac{(w_i(t)-Mv_i(t))\cdot\frac{d}{dt}(w_i(t)-Mv_i(t))}{\|w_i(t)-Mv_i(t)\|}\\ &\le \|\dot w_i(t)-M\dot v_i(t)\|\\ &\le \frac{1}{N}\sum_{j=1}^N \Big\| a(\|y_i(t)-y_j(t)\|)(w_j(t)-w_i(t))-\\ &\qquad - a(\|x_i(t)-x_j(t)\|)(Mv_j(t)-Mv_i(t))\Big\|.\end{aligned}$$

Using the triangle inequality, the Lipschitz continuity of a and its monotonicity, we obtain

$$\begin{aligned}\frac{d}{dt}e_i^v(t) \le \frac{1}{N}\sum_{j=1}^N \Big[&|a(\|y_i(t)-y_j(t)\|)-a(\|x_i(t)-x_j(t)\|)|\,\|w_j(t)-w_i(t)\|+\\ &+a(\|x_i(t)-x_j(t)\|)\,\|(w_j(t)-w_i(t))-(Mv_j(t)-Mv_i(t))\|\Big]\\ \le \frac{1}{N}\sum_{j=1}^N \Big[&L_a\,|\|y_i(t)-y_j(t)\|-\|x_i(t)-x_j(t)\||\,\|w_j(t)-w_i(t)\|+\\ &+a(0)\,\|(w_j(t)-w_i(t))-(Mv_j(t)-Mv_i(t))\|\Big].\end{aligned} \tag{3.22}$$

We can now estimate the derivative of $\mathcal{E}_2^v$. First of all, again by the Cauchy-Schwarz inequality it follows

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_2^v(t) &= \frac{d}{dt}\left(\frac{1}{N}\sum_{i=1}^N \|(w_i(t)-Mv_i(t))\|^2\right)^{1/2}\\ &= \frac{\left(\frac{1}{N}\sum_{i=1}^N \|(w_i(t)-Mv_i(t))\|\cdot\frac{d}{dt}\|(w_i(t)-Mv_i(t))\|\right)^{1/2}}{\left(\frac{1}{N}\sum_{i=1}^N \|(w_i(t)-Mv_i(t))\|^2\right)^{1/2}}\\ &\le \left(\frac{1}{N}\sum_{i=1}^N\left(\frac{d}{dt}\|(w_i(t)-Mv_i(t))\|\right)^2\right)^{1/2}\\ &= \left(\frac{1}{N}\sum_{i=1}^N\left(\frac{d}{dt}e_i^v(t)\right)^2\right)^{1/2}.\end{aligned} \tag{3.23}$$

If we insert (3.22) into the last inequality and we use the triangle as well as the Cauchy-Schwarz inequality in sequence, we get

$$\begin{aligned}\frac{d}{dt}\mathcal{E}_2^v(t) \le &\left(\frac{1}{N}\sum_{i=1}^N\left(\frac{1}{N}\sum_{j=1}^N L_a\,|\|y_i(t)-y_j(t)\|-\|x_i(t)-x_j(t)\|\,\|w_j(t)-w_i(t)\|\right)^2\right)^{1/2}+\\ &+\left(\frac{1}{N}\sum_{i=1}^N\left(\frac{1}{N}\sum_{j=1}^N a(0)\,\|(w_j(t)-w_i(t))-(Mv_j(t)-Mv_i(t))\|\right)^2\right)^{1/2}\end{aligned}$$

$$\leq L_a \left(\frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{N} \sum_{j=1}^{N} \left| \|y_i(t) - y_j(t)\| - \|x_i(t) - x_j(t)\| \right|^2 \right) \times \right.$$
$$\left. \times \left(\frac{1}{N} \sum_{j=1}^{N} \|w_j(t) - w_i(t)\|^2 \right) \right)^{1/2} +$$
$$+ a(0) \left(\frac{1}{N^2} \sum_{i,j=1}^{N} \|(w_j(t) - w_i(t)) - (Mv_j(t) - Mv_i(t))\|^2 \right)^{1/2}$$
$$\leq L_a \left(\frac{1}{N^2} \sum_{i,j=1}^{N} \left| \|y_i(t) - y_j(t)\| - \|x_i(t) - x_j(t)\| \right|^2 \right)^{1/2} \times$$
$$\times \max_{1 \leq i \leq N} \left\{ \left(\frac{1}{N} \sum_{j=1}^{N} \|w_j(t) - w_i(t)\|^2 \right)^{1/2} \right\} +$$
$$+ a(0) \left(\frac{1}{N^2} \sum_{i,j=1}^{N} \|(w_j(t) - w_i(t)) - (Mv_j(t) - Mv_i(t))\|^2 \right)^{1/2}.$$

We now estimate the first term of the sum. It holds

$$\begin{aligned} \left| \|y_i - y_j\| - \|x_i - x_j\| \right| &= \left| \|y_i - y_j\| - \|Mx_i - Mx_j\| + \|Mx_i - Mx_j\| - \|x_i - x_j\| \right| \\ &\leq \left| \|y_i - y_j\| - \|Mx_i - Mx_j\| \right| + \left| \|Mx_i - Mx_j\| - \|x_i - x_j\| \right| \\ &\leq \|y_i - Mx_i\| + \|y_j - Mx_j\| + \left| \|Mx_i - Mx_j\| - \|x_i - x_j\| \right| \\ &\leq e_i^x + e_j^x + \left| \|Mx_i - Mx_j\| - \|x_i - x_j\| \right|, \end{aligned} \tag{3.24}$$

and for all $i = 1, \ldots, N$, we have

$$\begin{aligned} \frac{1}{N} \sum_{j=1}^{N} \|w_j(t) - w_i(t)\|^2 &\leq \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{i-1} \|w_j(t) - w_i(t)\|^2 \\ &= \frac{1}{2N} \sum_{i,j=1}^{N} \|w_j(t) - w_i(t)\|^2 \\ &= NW(t). \end{aligned}$$

Furthermore, for the second summand we have

$$\left(\frac{1}{N^2} \sum_{i,j=1}^{N} \|(w_j(t) - w_i(t)) - (Mv_j(t) - Mv_i(t))\|^2 \right)^{1/2} \leq$$
$$\leq \left(\frac{1}{N} \sum_{i=1}^{N} \|(w_i(t) - Mv_i(t))\|^2 \right)^{1/2} + \left(\frac{1}{N} \sum_{j=1}^{N} \|(w_j(t) - Mv_j(t))\|^2 \right)^{1/2}$$
$$\leq 2\mathcal{E}_2^v(t).$$

Hence our computation yields

$$\frac{d}{dt}\mathcal{E}_2^v(t) \leq L_a\sqrt{NW(t)}\left(\frac{1}{N^2}\sum_{i,j=1}^N |\|Mx_i - Mx_j\| - \|x_i - x_j\||^2\right)^{1/2} + \\ + 2L_a\sqrt{NW(t)}\mathcal{E}_2^x(t) + 2a(0)\mathcal{E}_2^v(t).$$

Now we use the properties we assumed the matrix M has: for every $i, j = 1, \ldots, N$ and $t \in [0, T]$ either (3.8) holds, and then

$$|\|Mx_i(t) - Mx_j(t)\| - \|x_i(t) - x_j(t)\|| \leq \varepsilon\|x_i(t) - x_j(t)\| \\ \leq \varepsilon\left(\|x_i(0) - x_j(0)\| + \int_0^t \|v_i(s) - v_j(s)\|ds\right), \tag{3.25}$$

or (3.9) holds, and then

$$|\|Mx_i(t) - Mx_j(t)\| - \|x_i(t) - x_j(t)\|| \leq 2\delta,$$

so we always have

$$|\|Mx_i(t) - Mx_j(t)\| - \|x_i(t) - x_j(t)\|| \leq \varepsilon\Bigg(\|x_i(0) - x_j(0)\| + \\ + \int_0^t \|v_i(s) - v_j(s)\|\, ds\Bigg) + 2\delta.$$

Using the (vector-valued) Minkowski inequality and observing again that, by Corollary 1.19, V and W are decreasing, we derive

$$\begin{aligned} \frac{d}{dt}\mathcal{E}_2^v(t) &\leq \varepsilon L_a\sqrt{NW(t)}\left(\frac{1}{N^2}\sum_{i,j=1}^N \|x_i(0) - x_j(0)\|^2\right)^{1/2} + \\ &\quad + \varepsilon L_a\sqrt{NW(t)}\left(\frac{1}{N^2}\sum_{i,j=1}^N \left(\int_0^t \|v_i(s) - v_j(s)\|\, ds\right)^2\right)^{1/2} + \\ &\quad + 2L_a\sqrt{NW(t)}(\delta + \mathcal{E}_2^x(t)) + 2a(0)\mathcal{E}_2^v(t) \\ &\leq \varepsilon L_a\sqrt{NW(0)}\left(\sqrt{2X(0)} + \int_0^t \left(\frac{1}{N^2}\sum_{i,j=1}^N \|v_i(s) - v_j(s)\|^2\right)^{1/2} ds\right) + \\ &\quad + 2L_a\sqrt{NW(0)}(\delta + \mathcal{E}_2^x(t)) + 2a(0)\mathcal{E}_2^v(t) \\ &\leq \varepsilon L_a\sqrt{NW(0)}\left(\sqrt{2X(0)} + \int_0^t \sqrt{2V(s)}\, ds\right) + \\ &\quad + 2L_a\sqrt{NW(0)}(\delta + \mathcal{E}_2^x(t)) + 2a(0)\mathcal{E}_2^v(t) \qquad (3.26) \\ &\leq \varepsilon L_a\sqrt{NW(0)}\left(\sqrt{2X(0)} + t\sqrt{2V(0)}\right) + 2\delta L_a\sqrt{NW(0)} + \end{aligned}$$

$$+ 2L_a\sqrt{NW(0)}\mathcal{E}_2^x(t) + 2a(0)\mathcal{E}_2^v(t). \tag{3.27}$$

On the other hand, in the same way as in (3.23), we obtain

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}_2^x(t) &\leq \left(\frac{1}{N}\sum_{i=1}^N \left(\frac{d}{dt}\|(y_i(t) - Mx_i(t))\|\right)^2\right)^{1/2} \\
&\leq \left(\frac{1}{N}\sum_{i=1}^N \left\|\frac{d}{dt}(y_i(t) - Mx_i(t))\right\|^2\right)^{1/2} \\
&= \left(\frac{1}{N}\sum_{i=1}^N \|w_i(t) - Mv_i(t)\|^2\right)^{1/2} \\
&= \mathcal{E}_2^v(t).
\end{aligned}$$

Let $K_1 \triangleq L_a\sqrt{NW(0)}\sqrt{2X(0)}$, $K_2 \triangleq 2L_a\sqrt{NW(0)}$, $K_3 \triangleq 1/2 \cdot L_a\sqrt{NW(0)}\sqrt{2V(0)}$, and

$$\mathcal{K} \triangleq \begin{pmatrix} 2a(0) & 2L_a\sqrt{NW(0)} \\ 1 & 0 \end{pmatrix}$$

be as in the statement. If we integrate the two inequalities for $\frac{d}{dt}\mathcal{E}_2^x$ and $\frac{d}{dt}\mathcal{E}_2^v$ from 0 to t we obtain

$$\begin{aligned}
&\mathcal{E}_2^v(t) \leq \mathcal{E}_2^v(0) + (\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2 + \int_0^t \left(2a(0)\mathcal{E}^v(s) + 2L_a\sqrt{NW(0)}\mathcal{E}^x(s)\right) ds, \\
&\mathcal{E}_2^x(t) \leq \mathcal{E}_2^x(0) + \int_0^t \mathcal{E}^v(s) ds,
\end{aligned}$$

which can be rearranged in vector form in the following way:

$$\begin{pmatrix} \mathcal{E}_2^v(t) \\ \mathcal{E}_2^x(t) \end{pmatrix} \leq \begin{pmatrix} \mathcal{E}_2^v(0) + (\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2 \\ \mathcal{E}_2^x(0) \end{pmatrix} + \int_0^t \mathcal{K} \cdot \begin{pmatrix} \mathcal{E}^v(s) \\ \mathcal{E}^x(s) \end{pmatrix} ds.$$

Notice that

$$\|\mathcal{K}\|_{\ell_1 \to \ell_1} = \max\left\{2a(0) + 1, 2L_a\sqrt{NW(0)}\right\} \geq 1.$$

Now we apply the ℓ_1-norm to the inequality and we use Gronwall's Lemma A.1 to deduce

$$\left\|\begin{array}{c} \mathcal{E}_2^v(t) \\ \mathcal{E}_2^x(t) \end{array}\right\|_{\ell_1} \leq \left\|\begin{array}{c} \mathcal{E}_2^v(0) + (\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2 \\ \mathcal{E}_2^x(0) \end{array}\right\|_{\ell_1} \cdot e^{t\|\mathcal{K}\|_{\ell_1 \to \ell_1}},$$

thus

$$\begin{aligned}
\mathcal{E}_2^v(t) + \mathcal{E}_2^x(t) &\leq \left[\mathcal{E}_2^v(0) + \mathcal{E}_2^x(0) + (\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2\right] \cdot e^{t\|\mathcal{K}\|_{\ell_1 \to \ell_1}} \\
&= ((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{t\|\mathcal{K}\|_{\ell_1 \to \ell_1}},
\end{aligned} \tag{3.28}$$

since $\mathcal{E}^v(0) = \mathcal{E}^x(0) = 0$ by definition.

Moreover, since by Proposition 1.16 we have $\overline{v}(t) = \overline{v}(0)$ and $\overline{w}(t) = \overline{w}(0)$ for every $t \geq 0$, it holds

$$\begin{aligned}\|Mv_i(t) - w_i(t)\| &= \|Mv_i(t) - \overline{w}(0) + \overline{w}(0) - w_i(t)\| \\ &\leq \|Mv_i(t) - M\overline{v}(0)\| + \|\overline{w}(0) - w_i(t)\| \\ &\leq \|M\|\|v_i(t) - \overline{v}(t)\| + \|\overline{w}(t) - w_i(t)\|,\end{aligned}$$

and hence we have

$$\mathcal{E}_2^v(t) \leq \|M\| \left(\frac{1}{N}\sum_{i=1}^N \|v_i(t) - \overline{v}(t)\|^2\right)^{1/2} + \left(\frac{1}{N}\sum_{i=1}^N \|w_i(t) - \overline{w}(t)\|^2\right)^{1/2}.$$

Together with (3.28), we deduce the upper bound

$$\mathcal{E}_2^v(t) \leq \min\left\{((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{t\|\mathcal{K}\|_{\ell_1\to\ell_1}}, \|M\|\sqrt{V(t)} + \sqrt{W(t)}\right\}.$$

Using the trivial estimate of the ℓ_∞-norm by the ℓ_2-norm we conclude as well the estimate

$$\mathcal{E}^v(t) \leq \sqrt{N}\min\left\{((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{t\|\mathcal{K}\|_{\ell_1\to\ell_1}}, \|M\|\sqrt{V(t)} + \sqrt{W(t)}\right\}.$$

and

$$\mathcal{E}^v(t) + \mathcal{E}^x(t) \leq \sqrt{N}((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{t\|\mathcal{K}\|_{\ell_1\to\ell_1}}.$$

This finally concludes the proof. □

Remark 3.10. In the proof we used the fact that both V and W are decreasing. When in what follows we will consider controlled systems, however, we shall even have a better estimate on the integral of V. In particular we shall use the following: *assume, additionally to the hypotheses of Theorem 3.9 that*

$$\int_0^t \sqrt{2V(s)}\, ds \leq \alpha \text{ for all } t \leq T,$$

for a fixed $\alpha > 0$. Then for all $t \leq T$ we have

$$\mathcal{E}^x(t) + \mathcal{E}^v(t) \leq \sqrt{N}((\varepsilon(K_1 + K_4) + \delta K_2)t) \cdot e^{t\|\mathcal{K}\|_{\ell_1\to\ell_1}}$$

with $K_1, K_2, \mathcal{K}$ as in Theorem 3.9, and $K_4 \triangleq L_a\sqrt{NW(0)}\alpha$.
To verify the latter estimate, just consider the boundedness of $\int_0^t \sqrt{2V(s)}\, ds$ within the inequality (3.26) in the proof of Theorem 3.9, and then proceed further as before.

Remark 3.11. Among the hypotheses of Theorem 3.9, we assumed the existence of a matrix $M \in \mathbb{R}^{k\times d}$ fulfilling the weak Johnson-Lindenstrauss property for all curves of the form $x_i(t) - x_j(t)$, where $i, j = 1, \ldots, N$ and $t \in [0, T]$. We show now that M is such a matrix provided that the target dimension k is sufficiently large and that it fulfills the (strong) Johnson-Lindenstrauss property for all the (finite) vectors of the form $x_i(t_m) - x_j(t_m)$, where $i, j = 1, \ldots, N$, and

$$t_m = \frac{m}{\lceil T \cdot \mathcal{N}' \rceil}, \quad \text{for } m = 0, \ldots, \lceil T \cdot \mathcal{N}' \rceil - 1,$$

with

$$\mathcal{N}' \geq \frac{4\sqrt{2NV(0)}(\sqrt{d}+2)}{\delta\varepsilon}. \tag{3.29}$$

Indeed, we can adapt the proof of Lemma 3.6 in order to obtain a result valid *simultaneously* for all the curves $\varphi_{ij} : [0,T] \to \mathbb{R}^d$ given by $\varphi_{ij}(t) = x_i(t) - x_j(t)$. For each of these curves we have the Lipschitz estimate

$$\begin{aligned} \frac{\|(x_i(t_1) - x_j(t_1)) - (x_i(t_2) - x_j(t_2))\|}{|t_1 - t_2|} &\leq \sup_{t\in[0,T]} \left\| \frac{d}{dt}(x_i(t) - x_j(t)) \right\| \\ &= \sup_{t\in[0,T]} \|v_i(t) - v_j(t)\| \\ &\leq \sqrt{2NV(0)}, \end{aligned} \tag{3.30}$$

from which follows $\mathrm{Lip}_{[0,T]}(\varphi_{ij}) \leq \sqrt{2NV(0)}$. In order for the argument of the proof to work, for each curve φ_{ij} we need $\mathcal{N}' \cdot T$ points (where $\mathcal{N}'$ is as in (3.29), and the factor T is due to stretching the dynamics from a reference time domain $[0,1]$ to $[0,T]$) at which the (strong) Johnson-Lindenstrauss property must hold, bringing the total number of points $\mathcal{N}$ at which that property must be true to $\mathcal{N}' \cdot T \cdot N^2$. So it holds

$$\mathcal{N} \sim \sqrt{NV(0)} \cdot \frac{\sqrt{d}}{\delta\varepsilon} \cdot T \cdot N^2.$$

Thus, if M is a $k \times d$ matrix fulfilling the (strong) Johnson-Lindenstrauss property of parameter ε at these $\mathcal{N}$ points, where $k \geq k_0$ with

$$\begin{aligned} k_0 &\lesssim \varepsilon^{-2} \log(\mathcal{N}) \\ &\lesssim \varepsilon^{-2} \log\left(\sqrt{NV(0)} \cdot \frac{\sqrt{d}}{\delta\varepsilon} \cdot T \cdot N^2\right) \\ &\sim \varepsilon^{-2} \left(\log(T \cdot N \cdot d \cdot V(0)) + |\log(\delta\varepsilon)|\right), \end{aligned}$$

then M satisfies also the hypothesis of Theorem 3.9 for any $\delta > 0$.

Remark 3.12. In the remark above we calculated the necessary minimal dimension k_0 for a matrix M to satisfy the weak Johnson-Lindenstrauss property for all curves of the form $x_i(t) - x_j(t)$, where $i, j = 1, \ldots, N$ and $t \in [0,T]$. The dependency of k_0 on N and ε is quite natural, but the dependency on the dimension d, even only logarithmically, is perhaps not desirable, since we are under the hypothesis that d is very large. But one can circumvent this dependence on the dimension d using certain direct estimates within the proof of Theorem 3.9. In analogy to what we did before, take

$$t_m = \frac{m}{\mathcal{N}'}, \qquad \text{for } m = 0, \ldots, \lceil T \cdot \mathcal{N}' \rceil - 1,$$

and $\mathcal{N}'$ – the number of sampling points – is to be chosen large enough later on. Furthermore, we assume that the matrix M fulfills the (strong) Johnson-Lindenstrauss property at t_m, i.e., we require that M satisfies

$$(1-\varepsilon)\|x_i(t_m) - x_j(t_m)\| \leq \|M(x_i(t_m) - x_j(t_m))\| \leq (1+\varepsilon)\|x_i(t_m) - x_j(t_m)\|,$$

and

$$(1-\varepsilon)\|v_i(t_m)-v_j(t_m)\| \le \|M(v_i(t_m)-v_j(t_m))\| \le (1+\varepsilon)\|v_i(t_m)-v_j(t_m)\|,$$

for all $i,j=1,\ldots,N$ and $m=0,\ldots,\lceil T\cdot\mathcal{N}'\rceil-1$. Now, for any $t\in[0,T]$ choose $m\in\{0,\ldots,\lceil T\cdot\mathcal{N}'\rceil-1\}$ such that $t\in[t_m,t_{m+1}]$. We start at the estimate (3.24):

$$\begin{aligned}
|\|x_i(t)-x_j(t)\|-\|y_i(t)-y_j(t)\|| &\le |\|x_i(t_m)-x_j(t_m)\|-\|y_i(t_m)-y_j(t_m)\||+\\
&\quad+|\|x_i(t)-x_j(t)\|-\|x_i(t_m)-x_j(t_m)\||+\\
&\quad+|\|y_i(t)-y_j(t)\|-\|y_i(t_m)-y_j(t_m)\||\\
&\le |\|x_i(t_m)-x_j(t_m)\|-\|Mx_i(t_m)-Mx_j(t_m)\||+\\
&\quad+e_i^x(t_m)+e_j^x(t_m)+\frac{L_{x_i-x_j}}{\mathcal{N}'}+\frac{L_{y_i-y_j}}{\mathcal{N}'},
\end{aligned} \tag{3.31}$$

where $L_{x_i-x_j}\triangleq L_{[0,T]}(x_i(\cdot)-x_j(\cdot))$ and $L_{y_i-y_j}\triangleq L_{[0,T]}(y_i(\cdot)-y_j(\cdot))$ are the Lipschitz constants of the functions $x_i(\cdot)-x_j(\cdot)$ and $y_i(\cdot)-y_j(\cdot)$ on $[0,T]$, respectively. Furthermore, using the (strong) Johnson-Lindenstrauss property of M at t_m we get the same estimates as in (3.25), only with t_m instead of t:

$$\begin{aligned}
|\|Mx_i(t_m)-Mx_j(t_m)\|-\|x_i(t_m)-x_j(t_m)\|| &\le \varepsilon\|x_i(t_m)-x_j(t_m)\|\\
&\le \varepsilon\Bigg(\|x_i(0)-x_j(0)\|+\\
&\qquad+\int_0^{t_m}\|v_i(s)-v_j(s)\|\,ds\Bigg).
\end{aligned}$$

For the estimate of the last two terms in (3.31) we choose $\mathcal{N}'$ large enough so that

$$\mathcal{N}'\sim\frac{\max_{1\le i,j\le N}\left\{L_{x_i-x_j},L_{y_i-y_j}\right\}}{\delta}.$$

Thus we arrive at

$$\begin{aligned}
|\|x_i(t)-x_j(t)\|-\|y_i(t)-y_j(t)\|| &\le e_i^x(t_m)+e_j^x(t_m)+\varepsilon\Bigg(\|x_i(0)-x_j(0)\|+\\
&\qquad+\int_0^{t_m}\|v_i(s)-v_j(s)\|\,ds\Bigg)+2\delta.
\end{aligned}$$

Following the steps of the proof of Theorem 3.9, we can get an analogue of (3.27):

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}_2^v(t) &\le \varepsilon L_a\sqrt{NW(0)}\left(\sqrt{2X(0)}+t\sqrt{2V(0)}\right)+2\delta L_a\sqrt{NW(0)}+\\
&\quad+2L_a\sqrt{NW(0)}\mathcal{E}_2^x(t_m)+2a(0)\mathcal{E}_2^v(t).
\end{aligned}$$

So, the main difference is the replacement of $\mathcal{E}_2^x(t)$ with $\mathcal{E}_2^x(t_m)$ on the right-hand sides, with $t_m=m/\mathcal{N}'$. At this point in the proof of Theorem 3.9 we applied Gronwall's Lemma, see the estimates before (3.28). Now here we intend to use its discrete version, Lemma

A.2: if K_1, K_2, and K_3 are as in the statement of Theorem 3.9, then, integrating between t_m and t, we get

$$\begin{pmatrix} \mathcal{E}_2^v(t) \\ \mathcal{E}_2^x(t) \end{pmatrix} \leq \begin{pmatrix} \mathcal{E}_2^v(t_m) + (\varepsilon K_1 + \delta K_2)(t - t_m) + \varepsilon K_3(t^2 - t_m^2) + K_2 \mathcal{E}_2^x(t_m)(t - t_m) \\ \mathcal{E}_2^x(t_m) \end{pmatrix} + \\ + \int_{t_m}^t \mathcal{K}' \cdot \begin{pmatrix} \mathcal{E}^v(s) \\ \mathcal{E}^x(s) \end{pmatrix} ds,$$

where

$$\mathcal{K}' \triangleq \begin{pmatrix} 2a(0) & 0 \\ 1 & 0 \end{pmatrix}.$$

Now applying the ℓ_1-norm and Lemma A.2 we get

$$\left\| \begin{matrix} \mathcal{E}_2^v(t) \\ \mathcal{E}_2^x(t) \end{matrix} \right\|_{\ell_1} \leq \left\| \begin{matrix} \mathcal{E}_2^v(0) + (\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2 \\ \mathcal{E}_2^x(0) \end{matrix} \right\|_{\ell_1} \cdot e^{t\left(\|\mathcal{K}'\|_{\ell_1 \to \ell_1} + K_2\right)}.$$

This is a slightly worse estimate than the original one of Theorem 3.9 by a factor 2 in the exponential, since

$$\begin{aligned} \|\mathcal{K}\|_{\ell_1 \to \ell_1} &= \max\left\{2a(0) + 1, 2L_a\sqrt{NW(0)}\right\} \\ &\leq 2L_a\sqrt{NW(0)} + \left\| \begin{pmatrix} 2a(0) & 0 \\ 1 & 0 \end{pmatrix} \right\|_{\ell_1 \to \ell_1} \\ &= K_2 + \|\mathcal{K}'\|_{\ell_1 \to \ell_1} \\ &\leq 2\max\left\{2a(0) + 1, 2L_a\sqrt{NW(0)}\right\} \\ &= 2\|\mathcal{K}\|_{\ell_1 \to \ell_1}. \end{aligned}$$

So eventually we obtain

$$\begin{aligned} \mathcal{E}_2^v(t) + \mathcal{E}_2^x(t) &\leq \left[\mathcal{E}_2^v(0) + \mathcal{E}_2^x(0) + (\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2\right] \cdot e^{2t\|\mathcal{K}\|_{\ell_1 \to \ell_1}} \\ &= ((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{2t\|\mathcal{K}\|_{\ell_1 \to \ell_1}}. \end{aligned}$$

However, at the cost of a slightly worse estimate, we gain that the admissible lower dimensionality k of the matrix M does not depend anymore on the higher dimension d: indeed we used the (strong) Johnson-Lindenstrauss property on $\mathcal{N} = 2\lceil T \cdot \mathcal{N}' \rceil N^2$ points. Hence, it suffices to take the minimal target dimension k_0 such that $M \in \mathbb{R}^{k \times d}$ with $k \geq k_0$ for

$$k_0 \lesssim \varepsilon^{-2} \log\left(T \cdot \mathcal{N}' \cdot N^2\right).$$

Actually, in order to verify the independence of the dimension d, we have to estimate the number of sampling points $\mathcal{N}'$ independently of it. By (3.30) in Remark 3.11, we know that $L_{x_i - x_j} \leq \sqrt{2NV(0)}$. Analogously, for $L_{y_i - y_j}$ we have

$$\frac{\|(y_i(t_1) - y_j(t_1)) - (y_i(t_2) - y_j(t_2))\|}{|t_1 - t_2|} \leq \sqrt{2NW(0)}$$

$$\leq \sqrt{2NV(0)(1+\varepsilon)^2}$$
$$\leq \sqrt{8NV(0)},$$

since

$$\|w_i(0) - w_j(0)\| = \|Mv_i(0) - Mv_j(0)\| \leq (1+\varepsilon)\|v_i(0) - v_j(0)\|.$$

Hence we obtain

$$\begin{aligned} k &\lesssim \varepsilon^{-2} \log\left(T \cdot \mathcal{N}' \cdot N\right) \\ &\lesssim \varepsilon^{-2} \log\left(T \cdot \sqrt{NV(0)} \cdot \delta^{-1} \cdot N\right) \\ &\sim \varepsilon^{-2} \left(\log\left(T \cdot N \cdot V(0)\right) + |\log \delta|\right), \end{aligned}$$

and this confirms that there is no asymptotic dependence on d.

Remark 3.13. The estimate of Theorem 3.9

$$\mathcal{E}^v(t) \leq \sqrt{N} \min\left\{((\varepsilon K_1 + \delta K_2)t + \varepsilon K_3 t^2) \cdot e^{t\|\mathcal{K}\|_{\ell_1 \to \ell_1}}, \|M\|\sqrt{V(t)} + \sqrt{W(t)}\right\}.$$

explains the plot presented in [101, Figure 3.5], where surprisingly the error for large time was shown to decrease instead of exploding according to classical Gronwall's estimates. Indeed, since V and W are decreasing functions, there is a time when the bound swaps from the exponential Gronwall-type bound to the decreasing curve given by

$$\sqrt{N}\left(\|M\|\sqrt{V(t)} + \sqrt{W(t)}\right).$$

Moreover, if both the high-dimensional and the low-dimensional trajectories entered the consensus region already, then $V(t)$ and $W(t)$ approach 0 as t tends to $+\infty$, forcing $\mathcal{E}^v(t)$ to tend to 0. The vanishing of the discrepancy between the low-dimensional trajectory $(w_i(\cdot))_{i=1}^N$ of the consensus parameters and the projected trajectory $(Mv_i(\cdot))_{i=1}^N$ is a remarkable property of the Cucker-Smale system (3.1) as the initial mean-consensus parameter $\overline{w}(0) = M\overline{v}(0)$ is actually a conserved quantity by Proposition 1.16.

3.4 Dimension reduction of the Cucker-Smale system with control

In this section we consider a k-dimensional Cucker-Smale system (where $k \ll d$) having as initial conditions the projected initial configuration of the original d-dimensional system. The projection will be done by a matrix $M \in \mathbb{R}^{k\times d}$ fulfilling the (strong) Johnson-Lindenstrauss property for a certain amount of points. We then apply to both systems the sparse control strategy defined in (3.6). We shall prove the controlled analog of Theorem 3.9: the solution of the controlled k-dimensional system obtained in this way stays close to the projected dynamics of the controlled original d-dimensional system via the matrix M. This, in turn, shall allow us to prove our main result: if we gather the information of which is the furthest agent away from consensus in the k-dimensional system and we control this agent in the original high-dimensional system by the sparse strategy presented in Definition 2.4, then we will still be able to drive the high-dimensional system to the consensus region in finite time with a near-optimal rate.

One of the main consequences of this fact is that simulations following this strategy will save a relevant amount of computational time with respect to approaching directly the problem in high dimension: indeed, we present in Section 3.6 numerical tests, which show that we can take k even conspicuously smaller than d and still be able to implement a successful sparse control strategy steering the dynamics to the consensus region.

Formally, let us now consider a controlled version of the high-dimensional system

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)(v_j(t) - v_i(t)) + u_i^h(t), \end{cases} \quad i = 1, \dots, N, \tag{3.32}$$

with initial datum $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, and of the associated low-dimensional system

$$\begin{cases} \dot{y}_i(t) = w_i(t), \\ \dot{w}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|y_i(t) - y_j(t)\|\right)(w_j(t) - w_i(t)) + u_i^\ell(t), \end{cases} \quad i = 1, \dots, N, \tag{3.33}$$

with initial condition $(y^0, w^0) \in \mathbb{R}^{kN} \times \mathbb{R}^{kN}$, where $y_i^0 = Mx_i^0$ and $w_i^0 = Mv_i^0$ for every $i = 1, \dots, N$, and $M \in \mathbb{R}^{k \times d}$ is a matrix fulfilling the (strong) Johnson-Lindenstrauss property at certain points of the high-dimensional trajectories.

Remark 3.14. Notice this apparent logical loop: we have already stated that the control u^h in high dimension shall depend on u^ℓ, the low-dimensional one. Since the latter control is a function of the low-dimensional dynamics determined by the initial datum (y^0, w^0), which in turn depends on M, the trajectories of the high-dimensional dynamics depend on M as well.

As already stated before, given a set of N points, not necessarily explicitly, a random matrix generated by one of the constructions reported in Remark 3.4 fulfills the Johnson-Lindenstrauss property at these N points with a certain high probability. Unfortunately, in the current situation and differently from the one encountered in Section 3.3, the points on the trajectories at which the Johnson-Lindenstrauss property has to hold, seem to depend on the matrix M that we have generated!

As we shall see in detail in Section 3.5, we can resolve this dependency of the high-dimensional trajectories on the generated matrix M, by observing that the realization of the trajectories depends actually on a finite number of control switchings. Hence, for the moment, we just *assume* that the Johnson-Lindenstrauss property holds at certain points of the trajectory and we postpone to Section 3.5 the explanation of how this assumption can in fact hold true.

In what follows, we shall always indicate with $\theta > 0$ the maximal amount of resources that the external policy maker is allowed to spend at every instant to keep the system confined. This means that our controls u^h and u^ℓ will satisfy – respectively – the $\ell_1^N - \ell_2^d$

and the $\ell_1^N - \ell_2^k$ constraints

$$\sum_{i=1}^{N} \|u_i^h\|_{\ell_2^d} \leq \theta \quad \text{and} \quad \sum_{i=1}^{N} \|u_i^\ell\|_{\ell_2^k} \leq \theta.$$

Definition 3.15. Fix $T > 0$ and let $(x, v) : [0, T] \to \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ and $(y, w) : [0, T] \to \mathbb{R}^{kN} \times \mathbb{R}^{kN}$ be continuous functions. Let V and W be as in (1.12) and (3.20), respectively. Let us fix a $\Gamma \geq 0$ and define

$$T_0^c \triangleq \begin{cases} \inf\{t \in [0, T] : W(t) \leq \Gamma\} & \text{if the set is non-empty,} \\ T & \text{otherwise.} \end{cases}$$

We define the componentwise feedback controls u^h and u^ℓ as follows:

- if $t \leq T_0^c$, let $\hat{\imath}(t) \in \{1, \ldots, N\}$ be the smallest index $\hat{\imath}$ such that

$$\|w_{\hat{\imath}}^\perp(t)\| = \max_{1 \leq i \leq N} \|w_i^\perp(t)\|, \tag{3.34}$$

  then define the controls as

$$u_i^\ell(t) \triangleq \begin{cases} -\theta \dfrac{w_{\hat{\imath}(t)}^\perp(t)}{\|w_{\hat{\imath}(t)}^\perp(t)\|} & \text{if } i = \hat{\imath}(t), \\ 0 & \text{otherwise,} \end{cases}$$
$$u_i^h(t) \triangleq \begin{cases} -\theta \dfrac{v_{\hat{\imath}(t)}^\perp(t)}{\|v_{\hat{\imath}(t)}^\perp(t)\|} & \text{if } i = \hat{\imath}(t), \\ 0 & \text{otherwise.} \end{cases} \tag{3.35}$$

- if $t > T_0^c$, then $u^h(t) \triangleq 0$ and $u^\ell(t) \triangleq 0$. In this case, $\hat{\imath}(t)$ is undefined.

We say that the trajectory in low dimension *has entered the consensus region given by the threshold* Γ if $t \in [T_0^c, T)$.

Let us stress again the following observation.

Remark 3.16. Notice that the control u^h is *sparse* (all the components are zero at most except one) and defined exclusively through the following information: the index $\hat{\imath}$ which is computed from the low-dimensional control problem according to (3.34), the consensus parameter $v_{\hat{\imath}}$, which is actually the only information to be observed in high dimension and enters the definition (3.35), and the mean consensus parameter

$$\overline{v}(t) = \overline{v}(0) + \frac{1}{N} \sum_{i=1}^{N} \int_0^t u_i^h(s) ds,$$

which one easily computes by integration and sum of previous controls, and it is also used in (3.35).

As already noticed in Section 2.1, sparse controls can be in general very irregular in time. To overcome this drawback, we introduced the notion of *sampling solution* in Definition 2.7, which we use again now in this context. Let us fix a sampling time $\tau > 0$. In the following we shall consider d-dimensional and k-dimensional Cucker-Smale systems for $k \ll d$ and feedback controls u^h and u^ℓ, respectively, as introduced in Definition 3.15. We shall focus on their sampling solutions $(x(\cdot), v(\cdot))$ and $(y(\cdot), w(\cdot))$ associated with τ as defined in Definition 2.7, hence

$$\widetilde{u}^\ell(t) \triangleq u^\ell(n\tau) \ \text{ and } \ \widetilde{u}^h(t) \triangleq u^h(n\tau) \quad \text{ for } t \in [n\tau, (n+1)\tau).$$

Since we are only able to change the control at times which are multiples of τ, we define the *switch-off time of the sampled control associated with the threshold* Γ as

$$T_0 \triangleq \begin{cases} \inf_{n\in\mathbb{N}} \{n\tau : W(n\tau) \leq \Gamma\} & \text{if the set is non-empty,} \\ T & \text{otherwise.} \end{cases} \tag{3.36}$$

In the following, we shall show an estimate of the error between the projection of the sampled controlled high-dimensional system and the sampled controlled low-dimensional system, under the crucial assumption of the validity of the weak Johnson-Lindenstrauss property for M for the differences of trajectories of the system. This result is the controlled counterpart of Theorem 3.9.

Proposition 3.17. *Let $T > 0$, $\Delta > 0$ and $k \in \mathbb{N}$ with $k \leq d$. Let $\tau > 0$ be a sampling time, $\hat{T} > 0$ be such that $\hat{T} + \tau \leq T$ and let $M \in \mathbb{R}^{k\times d}$.*

Let $(x(\cdot), v(\cdot))$ be the sampling solution of the d-dimensional Cucker-Smale system (3.32) *with initial conditions (x^0, v^0) and $(y(\cdot), w(\cdot))$ be the sampling solution of the $\mathbb{R}^k$-projected Cucker-Smale system* (3.33) *with initial conditions (y^0, w^0) given by*

$$y^0 = \left(Mx_1^0, \ldots, Mx_N^0\right) \in \mathbb{R}^{kN} \quad \text{and} \quad w^0 = \left(Mv_1^0, \ldots, Mv_N^0\right) \in \mathbb{R}^{kN},$$

as defined in Definition 2.7, where u_i^h and u_i^ℓ are the controls from Definition 3.15 with threshold $\Gamma = (2\Delta)^2$. Moreover, let T_0 be as in (3.36)*.*

Suppose that W is non-increasing in time and that there exists a constant $\alpha \geq 0$ such that

$$\int_0^t \sqrt{2V(s)}\, ds \leq \alpha, \ \text{for all } t \in [0, \min\{\hat{T} + \tau, T_0\}].$$

Let $\varepsilon' \in (0, 1)$ be so small that

$$\frac{\Delta}{2} \geq \varepsilon' \sqrt{N} \left(4L_a\sqrt{NV(0)}\left(\sqrt{2X(0)} + \alpha\right) + \frac{\theta}{\sqrt{N}}\right) \times \\ \times (\hat{T} + \tau) e^{(\hat{T}+\tau)\left(\max\left\{2a(0)+1, 4L_a\sqrt{NV(0)}\right\}+8\theta/\Delta\right)},$$

and assume that the matrix M

(JL1) has the weak Johnson-Lindenstrauss property of parameter $\varepsilon = \varepsilon'$ and

$$\delta = \min\left\{\varepsilon' \frac{\sqrt{2X(0)} + \alpha}{2}, \frac{1}{2}\right\}$$

at all points $x_i(t) - x_j(t)$ for $i, j = 1, \dots, N$, and $t \in [0, \hat{T} + \tau]$;

(JL2) has the weak Johnson-Lindenstrauss property of parameter $\varepsilon = \varepsilon'$ and $\delta = \Delta$ at all points $v_i(n\tau) - \overline{v}(n\tau)$ for all $i = 1, \dots, N$ and $n = 0, \dots, \lfloor \hat{T}/\tau \rfloor + 1$;

(JL3) has the (strong) Johnson-Lindenstrauss property of parameter $\varepsilon = 1/2$ at the points $v_i(0) - \overline{v}(0)$ and $x_i(0) - \overline{x}(0)$ for all $i = 1, \dots, N$.

Finally, define the errors $e_i^x(t)$, $e_i^v(t)$, $\mathcal{E}^x(t)$, $\mathcal{E}^v(t)$, $\mathcal{E}_2^x(t)$, and $\mathcal{E}_2^v(t)$ as in (3.21).
Then for every $t \in [0, \min\{\hat{T} + \tau, T_0\}]$ it holds

$$\begin{aligned}\frac{\Delta}{2} &\geq \varepsilon'\sqrt{N}\left(4L_a\sqrt{NV(0)}\left(\sqrt{2X(0)} + \alpha\right) + \frac{\theta}{\sqrt{N}}\right) \cdot te^{t\left(\max\left\{2a(0)+1, 4L_a\sqrt{NV(0)}\right\}+8\theta/\Delta\right)} \\ &\geq \mathcal{E}^v(t) + \mathcal{E}^x(t).\end{aligned} \tag{3.37}$$

Proof. We argue by induction: we want to show that if (3.37) holds true for every $t \in \{0, \tau, \dots, n\tau\}$, then it is also true for $t \in [n, (n+1)\tau]$ (and in particular at $t = (n+1)\tau$) as long as $n\tau \leq \hat{T}$ and $n\tau < T_0$, i.e., the control is not switched off before $(n+1)\tau$. Obviously, (3.37) holds for $n = 0$, this means at $t = 0$, and actually arguing in the same way as in the following inductive step, the base step is verified.

So, let $t \in [n\tau, (n+1)\tau]$ for $n \in \mathbb{N}$. First, we consider the estimate on the agent on which the control is acting. We shall estimate the decay in order to use Gronwall Lemma as in Theorem 3.9. We have

$$\begin{aligned}\frac{d}{dt}e_{\hat{\imath}}^v(t) &\leq \left\|\frac{d}{dt}\left(w_{\hat{\imath}}(t) - Mv_{\hat{\imath}}(t)\right)\right\| \\ &\leq \frac{1}{N}\sum_{j=1}^N \Big\|a(\|y_{\hat{\imath}}(t) - y_j(t)\|)(w_j(t) - w_{\hat{\imath}}(t)) - \\ &\quad - a(\|x_{\hat{\imath}}(t) - x_j(t)\|)(Mv_j(t) - Mv_{\hat{\imath}}(t))\Big\| + \\ &\quad + \theta\left\|\frac{w_{\hat{\imath}}^\perp(n\tau)}{\|w_{\hat{\imath}}^\perp(n\tau)\|} - \frac{Mv_{\hat{\imath}}^\perp(n\tau)}{\|v_{\hat{\imath}}^\perp(n\tau)\|}\right\|.\end{aligned} \tag{3.38}$$

For all $i = 1 \dots, N$ such that $i \neq \hat{\imath}$ we have

$$\begin{aligned}\frac{d}{dt}e_i^v(t) &\leq \frac{1}{N}\sum_{j=1}^N \Big\|a(\|y_i(t) - y_j(t)\|)(w_j(t) - w_i(t)) - \\ &\quad - a(\|x_i(t) - x_j(t)\|)(Mv_j(t) - Mv_i(t))\Big\|.\end{aligned} \tag{3.39}$$

We now focus on the control term:

$$\begin{aligned}&\left\|\frac{w_{\hat{\imath}}^\perp(n\tau)}{\|w_{\hat{\imath}}^\perp(n\tau)\|} - \frac{Mv_{\hat{\imath}}^\perp(n\tau)}{\|v_{\hat{\imath}}^\perp(n\tau)\|}\right\| = \\ &= \frac{1}{\|w_{\hat{\imath}}^\perp(n\tau)\|\|v_{\hat{\imath}}^\perp(n\tau)\|}\left\|\|v_{\hat{\imath}}^\perp(n\tau)\|w_{\hat{\imath}}^\perp(n\tau) - \|w_{\hat{\imath}}^\perp(n\tau)\|Mv_{\hat{\imath}}^\perp(n\tau)\right\|\end{aligned}$$

$$= \frac{1}{\|w_{\hat\imath}^\perp(n\tau)\|\|v_{\hat\imath}^\perp(n\tau)\|}\Bigg\| \Big(\|v_{\hat\imath}^\perp(n\tau)\| - \|w_{\hat\imath}^\perp(n\tau)\|\Big) w_{\hat\imath}^\perp(n\tau) -$$

$$- \|w_{\hat\imath}^\perp(n\tau)\| \Big(Mv_{\hat\imath}^\perp(n\tau) - w_{\hat\imath}^\perp(n\tau)\Big)\Bigg\|$$

$$\leq \frac{1}{\|v_{\hat\imath}^\perp(n\tau)\|}\Big|\|v_{\hat\imath}^\perp(n\tau)\| - \|w_{\hat\imath}^\perp(n\tau)\|\Big| + \frac{1}{\|v_{\hat\imath}^\perp(n\tau)\|}\Big\|Mv_{\hat\imath}^\perp(n\tau) - w_{\hat\imath}^\perp(n\tau)\Big\|. \tag{3.40}$$

Since by assumption $n\tau < T_0$ and (3.37) holds at $n\tau$ by inductive hypothesis, we get

$$\|w_{\hat\imath}^\perp(n\tau)\| \geq \sqrt{W(n\tau)} > 2\Delta, \tag{3.41}$$

as well as

$$\begin{aligned}\|Mv_{\hat\imath}^\perp(n\tau) - w_{\hat\imath}^\perp(n\tau)\| &\leq \|Mv_{\hat\imath}(n\tau) - w_{\hat\imath}(n\tau)\| + \frac{1}{N}\sum_{j=1}^{N}\|Mv_j(n\tau) - w_j(n\tau)\| \\ &\leq 2\mathcal{E}^v(n\tau) \\ &\leq \Delta.\end{aligned} \tag{3.42}$$

Hence we have

$$\|Mv_{\hat\imath}^\perp(n\tau)\| > \Delta. \tag{3.43}$$

From assumption (JL2) and (3.43) it follows that the (strong) Johnson-Lindenstrauss property with parameter $\varepsilon = \varepsilon'$ holds at $v_{\hat\imath}^\perp$. Therefore

$$\begin{aligned}\Big|\|v_{\hat\imath}^\perp(n\tau)\| - \|w_{\hat\imath}^\perp(n\tau)\|\Big| &\leq \Big|\|v_{\hat\imath}^\perp(n\tau)\| - \|Mv_{\hat\imath}^\perp(n\tau)\|\Big| + \Big|\|Mv_{\hat\imath}^\perp(n\tau)\| - \|w_{\hat\imath}^\perp(n\tau)\|\Big| \\ &\leq \Big|\|v_{\hat\imath}^\perp(n\tau)\| - \|Mv_{\hat\imath}^\perp(n\tau)\|\Big| + \|Mv_{\hat\imath}^\perp(n\tau) - w_{\hat\imath}^\perp(n\tau)\| \\ &\leq \varepsilon'\|v_{\hat\imath}^\perp(n\tau)\| + \|Mv_{\hat\imath}^\perp(n\tau) - w_{\hat\imath}^\perp(n\tau)\|\end{aligned}$$

and

$$\|v_{\hat\imath}^\perp(n\tau)\| \geq \frac{1}{1+\varepsilon'}\|Mv_{\hat\imath}^\perp(n\tau)\| \geq \frac{\Delta}{2}.$$

Inserting these estimates into (3.40) and using (3.41) and (3.42) we get

$$\begin{aligned}\left\|\frac{w_{\hat\imath}^\perp(n\tau)}{\|w_{\hat\imath}^\perp(n\tau)\|} - \frac{Mv_{\hat\imath}^\perp(n\tau)}{\|v_{\hat\imath}^\perp(n\tau)\|}\right\| &\leq \varepsilon' + 2\frac{\|Mv_{\hat\imath}^\perp(n\tau) - w_{\hat\imath}^\perp(n\tau)\|}{\|v_{\hat\imath}^\perp(n\tau)\|} \\ &\leq \varepsilon' + \frac{8\mathcal{E}^v(n\tau)}{\Delta}.\end{aligned}$$

Now we add the estimates for the derivatives of $e_{\hat\imath}^v$ in (3.38) and e_i^v for $i \neq \hat\imath$ in (3.39). By (JL1) the weak Johnson-Lindenstrauss property of parameter $\varepsilon = \varepsilon'$ and

$$\delta = \min\left\{\varepsilon'\frac{\sqrt{2X(0)} + \alpha}{2}, \frac{1}{2}\right\}$$

holds at $x_i(t) - x_j(t)$ for $t \in [0, (n+1)\tau)$. Hence, for the first (uncontrolled) part of (3.38) and (3.39) we can use the same estimate as in (3.26). Therefore, by using the inequality $\mathcal{E}^v(n\tau) \leq \sqrt{N}\mathcal{E}_2^v(n\tau)$ and the definition of δ, we obtain the bound

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}_2^v(t) &\leq \left(\frac{1}{N}\sum_{i=1}^N \left(\frac{d}{dt}e_i^v(t)\right)^2\right)^{1/2} \\
&\leq \varepsilon' L_a\sqrt{NW(0)}\left(\sqrt{2X(0)} + \int_0^t \sqrt{2V(s)}\,ds\right) + \\
&\quad + 2L_a\sqrt{NW(0)}(\delta + \mathcal{E}_2^x(t)) + 2a(0)\mathcal{E}_2^v(t) + \\
&\quad + \frac{\theta}{\sqrt{N}}\left\|\frac{w_{\hat{\iota}}^\perp(n\tau)}{\|w_{\hat{\iota}}^\perp(n\tau)\|} - \frac{Mv_{\hat{\iota}}^\perp(n\tau)}{\|v_{\hat{\iota}}^\perp(n\tau)\|}\right\| \\
&\leq \varepsilon'\left(2L_a\sqrt{NW(0)}\left(\sqrt{2X(0)} + \alpha\right) + \frac{\theta}{\sqrt{N}}\right) + \\
&\quad + 2L_a\sqrt{NW(0)}\mathcal{E}_2^x(t) + 2a(0)\mathcal{E}_2^v(t) + \frac{8\theta}{\Delta}\mathcal{E}_2^v(n\tau).
\end{aligned}$$

For $\mathcal{E}_2^x$ we have

$$\begin{aligned}
\frac{d}{dt}\mathcal{E}_2^x(t) &\leq \left(\frac{1}{N}\sum_{i=1}^N \left(\frac{d}{dt}e_i^x(t)\right)^2\right)^{1/2} \\
&\leq \left(\frac{1}{N}\sum_{i=1}^N (e_i^v(t))^2\right)^{1/2} \\
&= \mathcal{E}_2^v(t).
\end{aligned}$$

By integrating the estimates for $\frac{d}{dt}\mathcal{E}_2^v$ and $\frac{d}{dt}\mathcal{E}_2^x$ between $n\tau$ and t we get the inequalities

$$\begin{aligned}
\mathcal{E}_2^v(t) &\leq \left(1 + \frac{8\theta}{\Delta}(t - n\tau)\right)\mathcal{E}_2^v(n\tau) + \varepsilon'\left(K_1 + \frac{\theta}{\sqrt{N}}\right)(t - n\tau) + \\
&\quad + \int_{n\tau}^t \left(2a(0)\mathcal{E}_2^v(s) + 2L_a\sqrt{NW(0)}\mathcal{E}_2^x(s)\right) ds, \\
\mathcal{E}_2^x(t) &\leq \mathcal{E}_2^x(n\tau) + \int_{n\tau}^t \mathcal{E}_2^v(s)ds,
\end{aligned}$$

which can be recast in vector form in the following way:

$$\begin{aligned}
\begin{pmatrix} \mathcal{E}_2^v(t) \\ \mathcal{E}_2^x(t) \end{pmatrix} &\leq \begin{pmatrix} \left(1 + \frac{8\theta}{\Delta}(t - n\tau)\right)\mathcal{E}_2^v(n\tau) + \varepsilon'\left(K_1 + \frac{\theta}{\sqrt{N}}\right)(t - n\tau) \\ \mathcal{E}_2^x(n\tau) \end{pmatrix} + \\
&\quad + \int_{n\tau}^t \mathcal{K} \cdot \begin{pmatrix} \mathcal{E}_2^v(s) \\ \mathcal{E}_2^x(s) \end{pmatrix} ds,
\end{aligned}$$

with

$$\mathcal{K} \triangleq \begin{pmatrix} 2a(0) & 2L_a\sqrt{NW(0)} \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad K_1 \triangleq 2L_a\sqrt{2NW(0)}\left(\sqrt{2X(0)} + \alpha\right).$$

Now we apply the ℓ_1-norm to the inequality and obtain

$$\mathcal{E}_2^v(t) + \mathcal{E}_2^x(t) \leq \varepsilon' \left(K_1 + \frac{\theta}{\sqrt{N}}\right)(t - n\tau) + \left(1 + \frac{8\theta}{\Delta}(t - n\tau)\right)\left(\mathcal{E}_2^v(n\tau) + \mathcal{E}_2^x(n\tau)\right) + \\ + \int_{n\tau}^t \|\mathcal{K}\|_{\ell_1 \to \ell_1} \cdot \left(\mathcal{E}_2^v(s) + \mathcal{E}_2^x(s)\right) ds.$$

The discrete Gronwall Lemma A.2 applied for

$$\beta_1(t) = \frac{8\theta}{\Delta} t, \qquad \beta_2(t) = \|\mathcal{K}\|_{\ell_1 \to \ell_1},$$
$$\rho(t) = \left(K_1 + \frac{\theta}{\sqrt{N}}\right)\varepsilon' t, \qquad u(t) = \mathcal{E}_2^v(t) + \mathcal{E}_2^x(t) \geq 0,$$

yields

$$\mathcal{E}_2^v(t) + \mathcal{E}_2^x(t) \leq [\mathcal{E}_2^v(0) + \mathcal{E}^x(0)] \cdot e^{t\left(\|\mathcal{K}\|_{\ell_1 \to \ell_1} + 8\theta/\Delta\right)} + \varepsilon' \left(K_1 + \frac{\theta}{\sqrt{N}}\right) t e^{t\left(\|\mathcal{K}\|_{\ell_1 \to \ell_1} + 8\theta/\Delta\right)} \\ = \varepsilon' \left(K_1 + \frac{\theta}{\sqrt{N}}\right) t e^{t\left(\|\mathcal{K}\|_{\ell_1 \to \ell_1} + 8\theta/\Delta\right)}$$

because the initial time errors are 0 by definition of the low-dimensional system. Hence using a trivial estimate of the ℓ_2-norm by the ℓ_∞-norm we conclude the induction and also the proof:

$$\mathcal{E}^v(t) + \mathcal{E}^x(t) \leq \varepsilon' \sqrt{N} \left(K_1 + \frac{\theta}{\sqrt{N}}\right) t e^{t\left(\|\mathcal{K}\|_{\ell_1 \to \ell_1} + 8\theta/\Delta\right)} \\ \leq \varepsilon' \sqrt{N} \left(4L_a \sqrt{NV(0)}\left(\sqrt{2X(0)} + \alpha\right) + \frac{\theta}{\sqrt{N}}\right) \times \\ \times t e^{t\left(\max\left\{2a(0)+1, 4L_a\sqrt{NV(0)}\right\} + 8\theta/\Delta\right)}$$

using that $\sqrt{W(0)} \leq (1 + 1/2)\sqrt{V(0)} \leq 2\sqrt{V(0)}$ by (JL3). □

Now we are in the position to show that we can steer both the low- and high-dimensional systems simultaneously to the consensus region using the sampled version of the control defined in Definition 3.15. We repeat that this means that we choose the index of the agent on which the sparse control acts from the low-dimensional system and use it to control the high-dimensional one.

The challenge here is ensuring that the control coming out of this procedure drives the high-dimensional system to consensus as well. For this we need the estimates from Proposition 3.17 to show that the error of the projection of the high-dimensional system and the low-dimensional system stay near to each other. Moreover, from Proposition 2.5 it is known that the low-dimensional system is steered *optimally* to the consensus region in finite time, if using the sampled version of the control introduced in Definition 3.15.

Theorem 3.18. *Let $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$, $k \leq d$ and $M \in \mathbb{R}^{k \times d}$ be given. Let c, C be the constants from Lemma 3.7 and define*

$$\overline{X} \triangleq 2X(0) + \frac{2N^2}{c^2\theta^2} V(0)^2, \tag{3.44}$$

where γ *is given by* (2.3)*, as well as*

$$\Delta \triangleq \min\left\{\frac{\gamma(\overline{X})}{C}, \frac{1}{2}\gamma\left(4\overline{X}\right)\right\}, \tag{3.45}$$

and

$$\hat{T} \triangleq \frac{2N}{\theta}\left(2\sqrt{V(0)} - 2\Delta\right). \tag{3.46}$$

Choose $\tau_0 > 0$ *to satisfy*

$$\tau_0\left(a(0)\sqrt{N}\sqrt{V(0)} + \theta\right) + \tau_0^2 a(0)\theta \leq \frac{\Delta}{4} \tag{3.47}$$

and fix $\tau \in (0, \tau_0]$*. Let* u^h *and* u^ℓ *be the controls as in Definition 3.15 with threshold* $\Gamma = (2\Delta)^2$*. Denote by* $(x(\cdot), v(\cdot))$ *the sampling solution of the d-dimensional system*

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^N a\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right) + u_i^h(t), \end{cases} \quad i = 1, \ldots, N,$$

with initial conditions (x^0, v^0) *associated with the sampling time* τ*, and by* $(y(\cdot), w(\cdot))$ *the sampling solution of the* $\mathbb{R}^k$*-projected system*

$$\begin{cases} \dot{y}_i(t) = w_i(t), \\ \dot{w}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^N a\left(\|y_i(t) - y_j(t)\|\right)\left(w_j(t) - w_i(t)\right) + u_i^\ell(t), \end{cases} \quad i = 1, \ldots, N,$$

with initial conditions given by $(y^0, w^0) = \left(Mx_1^0, \ldots, Mx_N^0, Mv_1^0, \ldots, Mv_N^0\right) \in \mathbb{R}^{kN} \times \mathbb{R}^{kN}$ *associated with the sampling time* τ*.*
Set $\alpha \triangleq \sqrt{2}N/c\theta$ *and choose* $\varepsilon' \in (0,1)$ *to satisfy*

$$\begin{aligned} \frac{\Delta}{2} \geq \varepsilon'\sqrt{N}\left(4L_a\sqrt{NV(0)}\left(\sqrt{2X(0)} + \alpha\right) + \frac{\theta}{\sqrt{N}}\right) \times \\ \times (\hat{T} + \tau)e^{(\hat{T}+\tau)\left(\max\left\{2a(0)+1, 4L_a\sqrt{NV(0)}\right\} + 8\theta/\Delta\right)}, \end{aligned} \tag{3.48}$$

and assume that the matrix M

(JL1) has the weak Johnson-Lindenstrauss property of parameter $\varepsilon = \varepsilon'$ *and*

$$\delta = \varepsilon'\frac{\sqrt{2X(0)} + \alpha}{2}$$

at all points $x_i(t) - x_j(t)$ *for all* $i, j = 1, \ldots, N$*,* $t \in [0, \hat{T} + \tau]$*;*

(JL2) has the weak Johnson-Lindenstrauss property of parameter $\varepsilon = \varepsilon'$ *and* $\delta = \Delta$ *at all points* $v_i(n\tau) - \overline{v}(n\tau)$ *for all* $i = 1, \ldots, N$ *and* $n = 0, \ldots, \lfloor \hat{T}/\tau \rfloor + 1$*;*

(JL3) *has the (strong) Johnson-Lindenstrauss property of parameter* $\varepsilon = 1/2$ *at the points* $v_i(0) - \overline{v}(0)$ *and* $x_i(0) - \overline{x}(0)$ *for all* $i = 1, \ldots, N$*.*

Then there exists an $n \in \mathbb{N}$ *such that* $\sqrt{W(n\tau)} \leq 2\Delta$*. Moreover, setting*

$$T_0 = n^* \tau, \qquad \text{where } n^* \triangleq \min \left\{ n \in \mathbb{N} : \sqrt{W(n\tau)} \leq 2\Delta \right\},$$

it holds that at T_0 *both the high-dimensional and the projected low-dimensional systems are in the consensus region defined by Theorem 1.11. Furthermore, we have the estimates*

$$T_0 \leq \frac{2N}{\theta} \left(\sqrt{W(0)} - 2\Delta \right) + \tau \leq \hat{T} + \tau$$

as well as

$$\begin{aligned} \max_{t \in [0,T_0]} \max_{1 \leq i,j \leq N} \|x_i(t) - x_j(t)\| &\leq 2\sqrt{N\overline{X}}, \\ \max_{t \in [0,T_0]} \max_{1 \leq i,j \leq N} \|v_i(t) - v_j(t)\| &\leq 2\sqrt{NV(0)}. \end{aligned} \tag{3.49}$$

Proof. The proof is divided into several steps.
First step: let

$$\overline{Y} \triangleq 2Y(0) + \frac{2N^2}{\theta^2} W(0)^2. \tag{3.50}$$

We shall prove that the following implication is true for every $n \in \mathbb{N}$ such that $n\tau \leq \hat{T}$: if $\sqrt{W(m\tau)} > 2\Delta$ for every $m = 0, \ldots, n$ and the subsequent assumptions $P_1(n)$, $P_2(n)$, and $P_3(n)$ depending on n hold

$P_1(n)$: for $t \in [0, n\tau)$ it holds

$$\frac{d}{dt} W(t) \leq -\frac{\theta}{N} \sqrt{W(t)} < 0 \quad \text{and} \quad \frac{d}{dt} V(t) \leq -\frac{c\theta}{N} \sqrt{V(t)} < 0;$$

$P_2(n)$: $Y(t) \leq \overline{Y}$ and $X(t) \leq \overline{X}$ hold in $[0, n\tau]$;

$P_3(n)$: it holds

$$\int_0^{n\tau} \sqrt{2V(s)}\, ds \leq \alpha,$$

then also $P_1(n+1)$, $P_2(n+1)$, and $P_3(n+1)$ hold true.
So let us assume $\sqrt{W(m\tau)} > 2\Delta$ for every $m = 0, \ldots, n$, which means that $T_0 \geq (n+1)\tau$ by definition of T_0, and assume $P_1(n)$, $P_2(n)$, and $P_3(n)$. We begin by computing the derivative of V and W for $t \in [n\tau, (n+1)\tau]$: by (2.5) follows

$$\frac{d}{dt} V(t) \leq \frac{2}{N} \sum_{i=1}^{N} u_i^h(n\tau) \cdot v_i^\perp(t) \quad \text{and} \quad \frac{d}{dt} W(t) \leq \frac{2}{N} \sum_{i=1}^{N} u_i^\ell(n\tau) \cdot w_i^\perp(t).$$

Remember that, by definition, $u^\ell_{\hat\imath}(n\tau) = -\theta w^\perp_{\hat\imath}(n\tau)/\|w^\perp_{\hat\imath}(n\tau)\|$ where $\hat\imath$ is the smallest index such that $\|w^\perp_{\hat\imath}(n\tau)\| \geq \|w^\perp_j(n\tau)\|$ for all $j = 1,\dots,N$, and $u^\ell_j(n\tau) = 0$ for every $j \neq \hat\imath$. Then $u^h_{\hat\imath}(n\tau) = -\theta v^\perp_{\hat\imath}(n\tau)/\|v^\perp_{\hat\imath}(n\tau)\|$ and $u^h_j(n\tau) = 0$ for every $j \neq \hat\imath$. Therefore, similarly to what we have done in the proof of Theorem 2.25 we may define

$$\phi^h(t) \triangleq \frac{v^\perp_{\hat\imath}(n\tau)\cdot v^\perp_{\hat\imath}(t)}{\|v^\perp_{\hat\imath}(n\tau)\|} \quad \text{and} \quad \phi^\ell(t) \triangleq \frac{w^\perp_{\hat\imath}(n\tau)\cdot w^\perp_{\hat\imath}(t)}{\|w^\perp_{\hat\imath}(n\tau)\|},$$

in order to study the decay estimate of the Lyapunov functionals V and W. Indeed, we immediately get

$$\frac{d}{dt}V(t) \leq -\frac{2\theta}{N}\phi^h(t) \quad \text{and} \quad \frac{d}{dt}W(t) \leq -\frac{2\theta}{N}\phi^\ell(t). \tag{3.51}$$

As we want to prove that $P_1(n+1)$ holds, we need to deduce suitable lower bounds on $\phi^h(t)$ and $\phi^\ell(t)$ to estimate the right-hand side of (3.51). To this purpose we need first to derive auxiliary bounds on the growth of $\sqrt{V(t)}$ and $\sqrt{W(t)}$, see formula (3.52) below: the general estimate

$$\frac{a\cdot b_i}{\|a\|} \leq \|b_i\| \leq \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^N \|b_i\|^2\right)^{1/2}$$

valid for arbitrary vectors $a, b_1, \dots, b_N$, yields

$$|\phi^\ell(s)| \leq \sqrt{N}\sqrt{W(s)}, \quad |\phi^h(s)| \leq \sqrt{N}\sqrt{V(s)}$$

for all $s \in [n\tau, (n+1)\tau]$. We use these bounds to estimate the right-hand side of (3.51)

$$\begin{aligned}\frac{d}{ds}\sqrt{V(s)} &= \frac{1}{2\sqrt{V(s)}}\frac{d}{ds}V(s) \leq \frac{\theta}{\sqrt{N}},\\ \frac{d}{ds}\sqrt{W(s)} &= \frac{1}{2\sqrt{W(s)}}\frac{d}{ds}W(s) \leq \frac{\theta}{\sqrt{N}}.\end{aligned}$$

An integration between $n\tau$ and $s \in [n\tau, (n+1)\tau]$ yields

$$\begin{aligned}\sqrt{W(s)} &\leq \sqrt{W(n\tau)} + (s-n\tau)\frac{\theta}{\sqrt{N}} \leq \sqrt{W(n\tau)} + \tau\frac{\theta}{\sqrt{N}},\\ \sqrt{V(s)} &\leq \sqrt{V(n\tau)} + (s-n\tau)\frac{\theta}{\sqrt{N}} \leq \sqrt{V(n\tau)} + \tau\frac{\theta}{\sqrt{N}}.\end{aligned} \tag{3.52}$$

With the help of (3.52) we now work out lower bounds for ϕ^ℓ and ϕ^h. It holds

$$\begin{aligned}\phi^\ell(t) &= \frac{w^\perp_{\hat\imath}(n\tau)\cdot w^\perp_{\hat\imath}(t)}{\|w^\perp_{\hat\imath}(n\tau)\|}\\ &= \frac{w^\perp_{\hat\imath}(n\tau)\cdot w^\perp_{\hat\imath}(n\tau)}{\|w^\perp_{\hat\imath}(n\tau)\|} - \frac{w^\perp_{\hat\imath}(n\tau)\cdot (w^\perp_{\hat\imath}(n\tau) - w^\perp_{\hat\imath}(t))}{\|w^\perp_{\hat\imath}(n\tau)\|}\\ &\geq \|w^\perp_{\hat\imath}(n\tau)\| - \|w^\perp_{\hat\imath}(n\tau) - w^\perp_{\hat\imath}(t)\|\\ &\geq \|w^\perp_{\hat\imath}(n\tau)\| - \int_{n\tau}^t \|\dot w^\perp_{\hat\imath}(s)\|\, ds.\end{aligned} \tag{3.53}$$

We now estimate the integrand. From

$$\begin{aligned}\dot{w}_{\hat{\imath}}^{\perp}(t) &= \frac{1}{N}\sum_{j=1}^{N} a(\|y_j(t)-y_{\hat{\imath}}(t)\|)(w_j^{\perp}(t)-w_{\hat{\imath}}^{\perp}(t)) + u_{\hat{\imath}}^{\ell}(n\tau) - \frac{1}{N}\sum_{j=1}^{N} u_j^{\ell}(n\tau)\\ &= \frac{1}{N}\sum_{j=1}^{N} a(\|y_j(t)-y_{\hat{\imath}}(t)\|)(w_j^{\perp}(t)-w_{\hat{\imath}}^{\perp}(t)) - \theta\frac{N-1}{N}\frac{w_{\hat{\imath}}^{\perp}(n\tau)}{\|w_{\hat{\imath}}^{\perp}(n\tau)\|},\end{aligned}$$

and the inequality

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}\|w_j^{\perp}-w_{\hat{\imath}}^{\perp}\| &\leq \left(\frac{1}{N}\sum_{j=1}^{N}\|w_k^{\perp}-w_{\hat{\imath}}^{\perp}\|^2\right)^{1/2}\\ &\leq \left(\frac{1}{N}\sum_{j=1}^{N}\sum_{j'=1}^{j-1}\|w_{j'}^{\perp}-w_j^{\perp}\|^2\right)^{1/2}\\ &= \left(\frac{1}{2N}\sum_{j=1}^{N}\sum_{j'=1}^{N}\|w_j^{\perp}-w_{j'}^{\perp}\|^2\right)^{1/2}\\ &= \sqrt{N}\sqrt{W},\end{aligned}$$

it follows

$$\|\dot{w}_{\hat{\imath}}^{\perp}(s)\| \leq a(0)\sqrt{N}\sqrt{W(s)} + \theta\frac{N-1}{N}, \quad \text{for all} \quad s\in[n\tau,(n+1)\tau).$$

Using (3.52) we get

$$\|\dot{w}_{\hat{\imath}}^{\perp}(s)\| \leq a(0)\sqrt{N}\left(\sqrt{W(n\tau)} + \tau\frac{\theta}{\sqrt{N}}\right) + \theta.$$

Plugging the last inequality into (3.53) we deduce

$$\begin{aligned}\phi^{\ell}(t) &\geq \|w_{\hat{\imath}}^{\perp}(n\tau)\| - \tau\left(a(0)\sqrt{N}\left(\sqrt{W(n\tau)}+\tau\frac{\theta}{\sqrt{N}}\right)+\theta\right)\\ &\geq \|w_{\hat{\imath}}^{\perp}(n\tau)\| - \tau\left(a(0)\sqrt{N}\sqrt{W(n\tau)}+\theta\right) - \tau^2 a(0)\theta.\end{aligned}$$

The same calculations give us

$$\phi^{h}(t) \geq \|v_{\hat{\imath}}^{\perp}(n\tau)\| - \tau\left(a(0)\sqrt{N}\sqrt{V(n\tau)}+\theta\right) - \tau^2 a(0)\theta.$$

Together with (3.51) this yields

$$\frac{d}{dt}W(t) \leq -\frac{2\theta}{N}\left(\|w_{\hat{\imath}}^{\perp}(n\tau)\| - \tau\left(a(0)\sqrt{N}\sqrt{W(n\tau)}+\theta\right) - \tau^2 a(0)\theta\right), \tag{3.54}$$

$$\frac{d}{dt}V(t) \leq -\frac{2\theta}{N}\left(\|v_{\hat{\imath}}^{\perp}(n\tau)\| - \tau\left(a(0)\sqrt{N}\sqrt{V(n\tau)}+\theta\right) - \tau^2 a(0)\theta\right). \tag{3.55}$$

By the assumption on $\tau \leq \tau_0$ in (3.47) and by assumption (JL3) we have

$$\tau\left(a(0)\sqrt{N}\sqrt{W(0)}+\theta\right)+\tau^2 a(0)\theta \leq \tau\left(2a(0)\sqrt{N}\sqrt{V(0)}+\theta\right)+\tau^2 a(0)\theta \leq \frac{\Delta}{2} \leq \Delta.$$

Applying this inequality and the fact that W is decreasing in $[0, n\tau]$ (which follows from $P_1(n)$) we use (3.54) to deduce the following upper bound

$$\begin{aligned}
\frac{d}{dt}W(t) &\leq -\frac{2\theta}{N}\left(\|w_{\hat{\imath}}^{\perp}(n\tau)\| - \tau\left(a(0)\sqrt{N}\sqrt{W(n\tau)}+\theta\right) - \tau^2 2a(0)\theta\right)\\
&\leq -\frac{2\theta}{N}\left(\sqrt{W(n\tau)} - \tau\left(a(0)\sqrt{N}\sqrt{W(0)}+\theta\right) - \tau^2 2a(0)\theta\right)\\
&\leq -\frac{2\theta}{N}\left(\sqrt{W(n\tau)} - \Delta\right).
\end{aligned}$$

Since we assumed that $\sqrt{W(n\tau)} > 2\Delta$, this shows that W is decreasing on $[n\tau, (n+1)\tau]$. Additionally, using this assumption we also can estimate

$$\frac{d}{dt}W(t) \leq -\frac{2\theta}{N}\left(\sqrt{W(n\tau)}-\Delta\right) \leq -\frac{\theta}{N}\sqrt{W(n\tau)} \leq -\frac{\theta}{N}\sqrt{W(t)}$$

for all $t \in [n\tau, (n+1)\tau]$. Together with $P_1(n)$ this shows the stated assertion for $\frac{d}{dt}W$ in $P_1(n+1)$.

In order to conclude the statement of $P_1(n)$ for $\frac{d}{dt}V$ we need to take advantage of the estimates of the lower dimensional dynamics, of Proposition 3.17, and Lemma 3.7. By assumption $P_3(n)$ it holds that

$$\int_0^{n\tau} \sqrt{2V(s)}\, ds \leq \alpha.$$

Thus, by the choice of ε' in (3.48) and the assumptions (JL1), (JL2), and (JL3), the hypotheses of Proposition 3.17 are fulfilled in the interval $[0, n\tau]$ (since $n\tau \leq T_0$ by definition of T_0 as the time where we switch the control to 0). Hence (3.37) holds, but in particular we have the validity of (3.42), which reads

$$\|Mv_i^{\perp}(n\tau) - w_i^{\perp}(n\tau)\| \leq \Delta.$$

This estimate and assumption (JL2) allow us to use Lemma 3.7 for the vectors $a_i = v_i^{\perp}(n\tau)$ and $b_i = w_i^{\perp}(n\tau)$. Together with $\sqrt{W(n\tau)} > 2\Delta$ this results in the following inequalities

$$\begin{aligned}
\|v_{\hat{\imath}}^{\perp}(n\tau)\| &\geq \frac{1}{4}\|w_{\hat{\imath}}^{\perp}(n\tau)\|,\\
\|v_{\hat{\imath}}^{\perp}(n\tau)\| &\geq \frac{1}{4}\sqrt{W(n\tau)} \geq \frac{\Delta}{2},\\
\|v_{\hat{\imath}}^{\perp}(n\tau)\| &\geq c\sqrt{V(n\tau)}.
\end{aligned}$$

By assumption $P_1(n)$ we know that V is decreasing in $[0, n\tau]$. Using the estimate (3.55) together with the choice of $\tau \leq \tau_0$ in (3.47) we obtain

$$\frac{d}{dt}V(t) \leq -\frac{2\theta}{N}\left(\|v_{\hat{\imath}}^{\perp}(n\tau)\| - \tau\left(a(0)\sqrt{N}\sqrt{V(n\tau)}+\theta\right) - \tau^2 a(0)\theta\right)$$

$$\leq -\frac{2\theta}{N}\left(\|v_{\hat{\iota}}^{\perp}(n\tau)\| - \tau\left(a(0)\sqrt{N})\sqrt{V(0)} + \theta\right) - \tau^2 a(0)\theta\right)$$
$$\leq -\frac{2\theta}{N}\left(\|v_{\hat{\iota}}^{\perp}(n\tau)\| - \frac{\Delta}{4}\right)$$
$$\leq -\frac{\theta}{N}\|v_{\hat{\iota}}^{\perp}(n\tau)\|$$
$$\leq -\frac{c\theta}{N}\sqrt{V(n\tau)},$$

for all $t \in [n\tau, (n+1)\tau]$. This shows that also V is decreasing in $[n\tau, (n+1)\tau]$ and hence

$$\frac{d}{dt}V(t) \leq -\frac{c\theta}{N}\sqrt{V(n\tau)} \leq -\frac{c\theta}{N}\sqrt{V(t)}$$

for all $t \in [n\tau, (n+1)\tau]$. Together with $P_1(n)$ this finishes the proof of $P_1(n+1)$.
We can now use Lemma 2.6 with $\eta = \theta/N$ and $\eta = c\theta/N$ to get the following estimates for Y and X, respectively:

$$Y(t) \leq \overline{Y} \quad \text{and} \quad X(t) \leq \overline{X}, \quad \text{for all } t \in [0, (n+1)\tau]$$

with $\overline{X}$ as defined in (3.44) and $\overline{Y}$ as defined in (3.50). This shows $P_2(n+1)$. Furthermore, $P_1(n+1)$ yields by integration

$$\int_0^{(n+1)\tau} \sqrt{2V(s)}ds \leq -\frac{\sqrt{2}N}{c\theta}\int_0^{(n+1)\tau}\frac{d}{ds}V(s)ds = \frac{\sqrt{2}N}{c\theta}\left(V(0) - V((n+1)\tau)\right) \leq \alpha V(0),$$

with $\alpha = \sqrt{2}N/c\theta$ as in the statement. Hence, under the assumptions $P_1(n)$, $P_2(n)$, and $P_3(n)$ we have shown $P_1(n+1)$, $P_2(n+1)$ as well as $P_3(n+1)$, provided that $\sqrt{W(m\tau)} > 2\Delta$ for every $m = 0, \ldots, n$.

Second step: we shall now prove that there exists an $n^* \in \mathbb{N}$ such that $n^*\tau \leq \hat{T} + \tau$ and $\sqrt{W(n^*\tau)} \leq 2\Delta$ holds, where $\hat{T}$ is defined as in (3.46). By definition of the threshold $\Gamma = (2\Delta)^2$, this implies the switching of the control to 0 at time $n^*\tau$. Assume on the contrary that

$$\sqrt{W((n+1)\tau)} > 2\Delta \tag{3.56}$$

for all $n \in \mathbb{N}$ with $n\tau \leq \hat{T}$. In the first step we showed that this yields in particular for $t \in [0, (n+1)\tau)$ the estimates

$$\frac{d}{dt}W(t) \leq -\frac{\theta}{N}\sqrt{W(t)} < 0 \quad \text{and} \quad \sqrt{W(t)} > 2\Delta.$$

Hence for all $t \in [0, (n+1)\tau)$ it holds

$$\sqrt{W(t)} \leq \sqrt{W(0)} + t \cdot \sup_{\xi\in(0,(n+1)\tau)} \frac{d}{d\xi}\sqrt{W(\xi)}$$
$$= \sqrt{W(0)} + t \cdot \sup_{\xi\in(0,(n+1)\tau)} \frac{d}{d\xi}W(\xi)\frac{1}{2\sqrt{W(\xi)}}$$
$$\leq \sqrt{W(0)} - t \cdot \frac{\theta}{2N}.$$

Taking $n_0 \in \mathbb{N}$ such that $n_0\tau \leq \hat{T} < (n_0+1)\tau$ and using (JL3) we have

$$\begin{aligned}
\sqrt{W((n_0+1)\tau)} &\leq \sqrt{W(\hat{T})} \\
&\leq \sqrt{W(0)} - \hat{T}\frac{\theta}{2N} \\
&= \sqrt{W(0)} - \frac{2N}{\theta}\left(2\sqrt{V(0)} - 2\Delta\right)\frac{\theta}{2N} \\
&\leq \sqrt{W(0)} - \frac{2N}{\theta}\left(\sqrt{W(0)} - 2\Delta\right)\frac{\theta}{2N} \\
&\leq 2\Delta.
\end{aligned} \tag{3.57}$$

This contradicts assumption (3.56). Therefore there exists an $n^* \in \mathbb{N}$ that satisfies $n^*\tau \leq \hat{T} + \tau$ and for which it holds

$$\sqrt{W(n^*\tau)} \leq 2\Delta. \tag{3.58}$$

Third step: We shall show that (3.58) implies that the trajectories of both the low- and high-dimensional systems are in the consensus region identified by Theorem 1.11 at time $n^*\tau$, i.e.,

$$\sqrt{W(n^*\tau)} \leq \gamma(Y(n^*\tau)) \quad \text{and} \quad \sqrt{V(n^*\tau)} \leq \gamma(X(n^*\tau)).$$

We shall start considering the low-dimensional system. Notice that by (JL3) it holds

$$Y(0) \leq \left(1 + \frac{1}{2}\right)^2 \cdot X(0) \leq 4X(0) \quad \text{and} \quad W(0) \leq 4V(0).$$

Hence, by the fact that the constant c from Lemma 3.7 is smaller than 1, we can estimate

$$\overline{Y} = 2Y(0) + \frac{2N^2}{\theta^2}W(0)^2 \leq 4\left(2X(0) + \frac{2N^2}{c^2\theta^2}V(0)^2\right) = 4\overline{X}.$$

This together with (3.58), the definition of Δ in (3.45), and $P_2(n^*)$ lead to

$$\sqrt{W(n^*\tau)} \leq 2\Delta \leq \gamma\left(4\overline{X}\right) \leq \gamma(\overline{Y}) \leq \gamma(Y(n^*\tau)).$$

It remains to prove that the high-dimensional system is in the consensus region identified by Theorem 1.11. Again, the conditions of Lemma 3.7 for the vectors $a_i = v_i^\perp(n^*\tau)$ and $b_i = w_i^\perp(n^*\tau)$ are fulfilled: as in the first step we have by Proposition 3.17 that (3.42) holds, i.e.,

$$\|Mv_i^\perp(n^*\tau) - w_i^\perp(n^*\tau)\| \leq \Delta,$$

and property (JL2) holds at $n^*\tau$. Thus, an application of Lemma 3.7 shows

$$\sqrt{V(n^*\tau)} \leq C\Delta.$$

The definition of Δ in (3.45) and $P_2(n^*)$ then yield

$$\sqrt{V(n^*\tau)} \leq C\Delta \leq \gamma(\overline{X}) \leq \gamma(X(n^*\tau)).$$

We conclude that both the trajectories of the systems are in the consensus region at time $n^*\tau$. By Theorem 1.11 we are allowed to switch the control to 0 and both systems tend to consensus autonomously.

Fourth step: In the second and third steps we have proven that both systems enters the consensus region at time $T_0 = n^*\tau$, where $n^*\tau \leq \hat{T} + \tau$. By the computations in (3.57), we have the following estimate

$$T_0 \leq \frac{2N}{\theta}\left(\sqrt{W(0)} - 2\Delta\right) + \tau \leq \hat{T} + \tau.$$

Moreover, by $P_2(n^*)$ we have for every $t \in [0, n^*\tau]$

$$\|x_i(t) - x_j(t)\|^2 \leq 2\|x_i(t) - \overline{x}(t)\|^2 + 2\|x_j(t) - \overline{x}(t)\|^2 \leq 4NX(t) \leq 4N\overline{X},$$

and from $P_1(n^*)$ it follows

$$\|v_i(t) - v_j(t)\|^2 \leq 2\|v_i(t) - \overline{v}(t)\|^2 + 2\|v_j(t) - \overline{v}(t)\|^2 \leq 4NV(t) \leq 4NV(0)$$

for every $t \in [0, n^*\tau]$. This shows (3.49) and the proof is concluded. □

3.5 How to find a Johnson-Lindenstrauss matrix

The main ingredient of Proposition 3.17 and Theorem 3.18 is the existence of a Johnson-Lindenstrauss matrix $M \in \mathbb{R}^{k\times d}$ for the trajectories. Let Δ and ε' be as in Theorem 3.18 and let us recall what we explicitly needed. Assume that $\hat{T}$ is an upper estimate for T_0, the switch off time of the control. Then we need to define a matrix $M \in \mathbb{R}^{k\times d}$ such that the following properties hold:

(JL1) Let $\varepsilon = \varepsilon'$ and

$$\delta = \frac{\sqrt{2}}{2}\varepsilon'\left(\sqrt{X(0)} + \frac{N}{c\theta}\right).$$

For all $t \in [0, \hat{T} + \tau]$ and $i, j = 1, \dots, N$, either we have

$$(1-\varepsilon)\|x_i(t) - x_j(t)\| \leq \|M(x_i(t) - x_j(t))\| \leq (1+\varepsilon)\|x_i(t) - x_j(t)\|$$

or

$$\|x_i(t) - x_j(t)\| \leq \delta \quad \text{and} \quad \|M(x_i(t) - x_j(t))\| \leq \delta.$$

(JL2) Let $\varepsilon = \varepsilon'$ and $\delta = \Delta$. For all $n = 0, \dots, \lfloor T/\tau \rfloor + 1$ and $i = 1, \dots, N$, either we have

$$(1-\varepsilon)\|v_i(n\tau) - \overline{v}(n\tau)\| \leq \|M(v_i(n\tau) - \overline{v}(n\tau))\| \leq (1+\varepsilon)\|v_i(n\tau) - \overline{v}(n\tau)\|$$

or

$$\|v_i(n\tau) - \overline{v}(n\tau)\| \leq \delta \quad \text{and} \quad \|M(v_i(n\tau) - \overline{v}(n\tau))\| \leq \delta.$$

(JL3) Let $\varepsilon = 1/2$. Then for all $i = 1, \dots, N$ we have

$$(1-\varepsilon)\|v_i(0) - \overline{v}(0)\| \leq \|M(v_i(0) - \overline{v}(0))\| \leq (1+\varepsilon)\|v_i(0) - \overline{v}(0)\|$$

and

$$(1-\varepsilon)\|x_i(0) - \overline{x}(0)\| \leq \|M(x_i(0) - \overline{x}(0))\| \leq (1+\varepsilon)\|x_i(0) - \overline{x}(0)\|.$$

In order to prove conditions (JL2) and (JL3) one can directly invoke the Johnson-Lindenstrauss Lemma as discussed in Remark 3.4, while for (JL1) one can use its continuous version, Lemma 3.6, which boils down again to the application of the Johnson-Lindenstrauss Lemma on points sampled from the trajectories.

However, the Johnson-Lindenstrauss Lemma applies on points which are fixed a priori before generating randomly the matrix $M \in \mathbb{R}^{k\times d}$. At a first look, due to the fact that the high-dimensional controls depend on the low-dimensional ones, which depend on the matrix M, the points on which we apply the Johnson-Lindenstrauss Lemma may be seen as directly depending on M as well.

In order to resolve this apparent paradox, we want to clarify that actually, due to the finite number of sampling times of the control and the finite number of agents, the number of possible realizable trajectories, and consequently the number of possible sampling points for the Johnson-Lindenstrauss Lemma, is finite and *independent* of the choice of the matrix M. Hence we are now left with the tasks of counting the number of such trajectories and of verifying that they fulfill the necessary Lipschitz continuity assumptions for applying Lemma 3.6.

Let us state again that the lower dimension k of $M \in \mathbb{R}^{k\times d}$ scales as

$$k \sim \varepsilon^{-2}\log(\mathcal{N}), \tag{3.59}$$

where $\varepsilon \in (0,1)$ is the allowed distortion and $\mathcal{N}$ is the number of sampling points on all possible trajectories.

We focus first in (3.59) on the dependence of $\varepsilon = \min\{\varepsilon', 1/2\}$ on N, the number of agents, and the dimension d. According to (3.48) in Theorem 3.18 the estimate on ε' scales exponentially with N, i.e., $\varepsilon' \lesssim e^{-N}$, since $\hat{T}$ scales (at least) linearly with N, see (3.46) (assuming that θ is independent of N and d). The positive aspect, however, is that the estimate for ε' does not involve the dimension d.

In order to compute $\mathcal{N}$ in (3.59) we need, first of all, to estimate the number of realizable trajectories. Since we are insisting on *sparse* controls acting at most on *one* agent at the time, at every switching time $n\tau$ with $n\tau \leq T_0$, i.e., as long as the control is not switched off, there are precisely only N possible controls and hence N possible branches of future developments of the trajectories. By Theorem 3.18 it holds $T_0 \leq \hat{T} + \tau$, and we can then estimate the number P of possible paths by

$$P \leq N^{\lfloor \frac{T}{\tau} \rfloor + 1}.$$

Surprisingly, accounting for all the possible future branches is sufficient to show that we can already deterministically fix points a priori on which later apply an independently randomly drawn matrix!

(*i*) In order to fulfill (JL1) for every possible trajectory, an application of Lemma 3.6 yields an estimate of the number of necessary sampling points

$$\mathcal{N}_1 = P(\hat{T}+\tau) \cdot \binom{N}{2} \cdot \frac{4L_x(\sqrt{d}+2)}{\delta\varepsilon}, \tag{3.60}$$

where the factor $P(\hat{T}+\tau)$ accounts for the number of trajectories and their time length, the factor $\binom{N}{2}$ accounts for the number of space trajectory differences $x_i(\cdot) - x_j(\cdot)$, and L_x is an upper estimate for the individual Lipschitz constants. L_x can be computed explicitly from an estimate similar to (3.30) and the result from Theorem 3.18 that V is decreasing until T_0 as follows

$$L_x = \max_{1\leq i,j\leq N} \left\{L_{x_i - x_j}(0,T_0) : i,j = 1,\dots,N\right\} \leq \sup_{t\in(0,T_0)} \sqrt{2NV(t)} \leq \sqrt{2NV(0)}.$$

(*ii*) To fulfill (JL2) we shall now count the necessary sampling points at every switching time $n\tau$. For $n = 0$ we have to consider N sampling points. For $n = 1$ there are already N possible paths to take into account and hence we need to take $N^2 = N \cdot N$ sampling points. Going on in this way, at time $n\tau$ we have N^n possible outcomes of the dynamical system and hence we have to take $N^{n+1} = N^n \cdot N$ sampling points, as long as $n\tau \leq \hat{T} + \tau$. Summing up the number of sampling points, we conclude

$$\mathcal{N}_2 = \sum_{n=0}^{\lfloor \frac{T}{\tau} \rfloor + 1} N^{n+1} \leq N^{\lfloor \frac{T}{\tau} \rfloor + 3}$$

(*iii*) To fulfill (JL3) we need only $\mathcal{N}_3 = 2N$ sampling points.

Hence we can eventually estimate $\mathcal{N}$ from above by

$$\begin{aligned}
\mathcal{N} &\leq \mathcal{N}_1 + \mathcal{N}_2 + \mathcal{N}_3 \\
&= N^{\lfloor \frac{T}{\tau} \rfloor + 1}(\hat{T}+\tau) \cdot \binom{N}{2} \cdot \frac{4\sqrt{2NV(0)}(\sqrt{d}+2)}{\delta\varepsilon} + N^{\lfloor \frac{T}{\tau} \rfloor + 3} + 2N.
\end{aligned}$$

Thus, we can choose the dimension k of a Johnson-Lindenstrauss matrix $M \in \mathbb{R}^{k\times d}$ as

$$\begin{aligned}
k &\sim \varepsilon^{-2} \cdot \log(\mathcal{N}) \\
&\sim \varepsilon^{-2} \left[\left(\frac{\hat{T}}{\tau} + 1 \right) \log N + \log(\hat{T}+\tau) + \log d + \log V(0) + |\log(\delta\varepsilon)| \right],
\end{aligned}$$

where

$$\varepsilon = \min\left\{\varepsilon', \frac{1}{2}\right\} \quad \text{and} \quad \delta = \min\left\{\Delta, \frac{\sqrt{2}}{2}\varepsilon' \left(\sqrt{X(0)} + \frac{N}{c\theta} \right) \right\}.$$

Since the estimate of ε' scales exponentially in N, i.e., $\varepsilon' \lesssim e^{-N}$, the dimension k grows exponentially in N. The positive aspect is that the estimate of k only scales logarithmically with the dimension d. Hence we have shown that at least for a large dimension $d \gg 1$ and a relatively small number of agents N our dimensionality reduction approach pays off in terms of the size of the projected dimension k.

However, as we shall show in Section 3.6, these theoretical bounds turn out to be by far over-pessimistic and, surprisingly, this method of dimensionality reduction works effectively with a lower dimensions k conspicuously smaller than d. Moreover, in the Remarks 3.20 and 3.21 below, we show two tricks to circumvent the exponential dependency of k with respect to N at the cost of using *sequences* of Johnson-Lindenstrauss matrices.

Remark 3.19. The $\log(d)$-dependency only comes into play when we derive (JL1) from Lemma 3.6. One can actually use a similar argument as in Remark 3.12 in order to get rid of this logarithmic dependency. We do not elaborate further on this issue which appears to us just a mere and perhaps unnecessary technicality at this point.

Remark 3.20. We observed that in the worst-case scenario the dimension k of the Johnson-Lindenstrauss matrix blows up exponentially with the number of agents N. A practical approach to circumvent this problem is to use not one, but a whole family of matrices $M_0, \dots, M_\ell$ to project the high-dimensional system. The matrix M_0 is used from time 0 up to a certain time $t_0 > 0$ and thus only needs to fulfill the Johnson-Lindenstrauss property in the short time interval $[0, t_0]$. At time t_0 a new matrix M_1 is chosen: we observe the positions and the consensus parameters in high dimension at t_0 and project the system to low dimension using the new matrix M_1. Then we use the new low-dimensional system to calculate the index of the control for the high-dimensional system from time t_0 up to time t_1, and eventually we repeat the procedure again with new matrices $M_2, M_3, \dots$

This approach has the advantage that it requires the Johnson-Lindenstrauss properties for M_i, $i = 1, \dots, \ell$, only for a short time interval. The disadvantage is that we have to observe the high-dimensional system and project it to low-dimension again at every time t_i, $i = 0, \dots, \ell - 1$.

Remark 3.21. We now show how to get rid of the mutual dependency between the matrix and the points of the trajectories using families of matrices as in Remark 3.20.

First, we take a matrix M_0 having the Johnson-Lindenstrauss properties (JL2) at $t = 0$ and (JL3). We compute the index $\hat{\imath}_0$ of the control (as defined in Definition 3.15) at $t = 0$ using the projection M_0.

Then we choose a matrix M_1 having the Johnson-Lindenstrauss properties (JL1) for all $t \in [0, \tau)$, (JL2) at $t = \tau$, and (JL3). We compute the low-dimensional system using the projection M_1 in $[0, \tau]$ and let the control act on the agent $\hat{\imath}_0$ calculated by M_0. This is the main trick of the procedure: the points of the high-dimensional system in $[0, \tau]$ are not influenced by the matrix M_1 and hence the mutual dependency is removed, meaning that there is no need to consider all trajectories P anymore, in contrast to (3.60).

Now, from the low-dimensional system, computed by M_1 with control acting on $\hat{\imath}_0$ in $[0, \tau]$, we choose the agent $\hat{\imath}_1$ at τ on which the control will act in the next interval $[\tau, 2\tau]$.

This procedure can be carried out using a family of matrices $\{M_p : p = 0, \dots, \ell\}$ such that M_p fulfils the Johnson-Lindenstrauss properties (JL1) for all $t \in [0, p\tau)$, (JL2) at $t = p\tau$, and (JL3). The agent $\hat{\imath}_p$ on which the control shall act in the interval $[p\tau, (p + 1)\tau)$ is computed at $p\tau$ using the low-dimensional system projected by M_p, while the control acts on $\hat{\imath}_q$ in $[q\tau,(q+1)\tau)$ for $q=0,\dots p-1$. Therefore, in $[0,p\tau]$ the index of the controlled agent and hence the trajectories of the high-dimensional system are independent of M_p.

3.6 Numerical implementation of sparse controls in high dimension

We shall now present some numerical experiments to confirm the theoretical observations of the interplay between the Cucker-Smale system, the dimension reduction by a Johnson-Lindenstrauss matrix and the quality of the control chosen from the low-dimensional (projected) system as defined in Definition 3.15.

For every $\ell=0,1,\dots$ we recursively solve the d-dimensional Cucker-Smale system

$$\begin{cases} \dot{x}_i(t)=v_i(t), \\ \dot{v}_i(t)=\dfrac{1}{N}\displaystyle\sum_{j=1}^{N}\frac{v_j(t)-v_i(t)}{\left(1+\|x_i(t)-x_j(t)\|^2\right)^{\beta}}+u_i(\ell\tau), \end{cases} \quad t\in[\ell\tau,(\ell+1)\tau], \quad i=1,\dots,N,$$

numerically, using as the initial value $(x(\ell\tau),v(\ell\tau))$ the solution of the preceding interval $[(\ell-1)\tau,\ell\tau]$ and as the starting value for $\ell=0$ the given values $x(0)=x^0$ and $v(0)=v^0$. The experiments are implemented by using Matlab applying a Runge-Kutta method of order 4 solving the systems of ODEs with step width τ. The following are the different control strategies u_i we compare in our experiments:

(SP) Sparse control implemented in the high-dimensional system: this is the sparse control strategy outlined in Definition 2.4. The control acts on the agent with consensus parameter furthest away from the mean consensus parameter as long as we are not in the consensus region given by Theorem 1.11: for every $\ell\in\mathbb{N}$ let $\hat{\imath}\in\{1,\dots,N\}$ be the smallest index such that

$$\|v_{\hat{\imath}}^{\perp}(\ell\tau)\|=\max_{1\le i\le N}\|v_i^{\perp}(\ell\tau)\|,$$

and define the control as

$$u_j(\ell\tau)\triangleq\begin{cases} -\theta\dfrac{v_{\hat{\imath}}^{\perp}(\ell\tau)}{\|v_{\hat{\imath}}^{\perp}(\ell\tau)\|} & \text{if } j=\hat{\imath}, \\ 0 & \text{otherwise,} \end{cases}$$

as long as $V(\ell\tau)>\gamma(X(\ell\tau))^2$. As soon as $V(n\tau)\le\gamma(X(n\tau))^2$ is satisfied for some $n\in\mathbb{N}$, we set $T_0\triangleq n\tau$ and the control to zero.

This control was shown to be *optimal* in Proposition 2.5 in terms of maximizing the rate of convergence to the consensus region, and shall be therefore employed as a benchmark to test the effectiveness of the other controls.

(U) Uniform control: this control strategy acts on every agent simultaneously using a control pointing towards the mean consensus parameter with norm equal to θ/N as long as $V(\ell\tau) > \gamma(X(\ell\tau))^2$. This means

$$u_j(\ell\tau) = -\frac{\theta}{N}\frac{v_j^\perp(\ell\tau)}{\|v_j^\perp(\ell\tau)\|} \quad \text{for all } j = 1, \dots, N.$$

Again, as soon as $V(n\tau) \leq \gamma(X(n\tau))^2$ is satisfied for some $n \in \mathbb{N}$, we set $T_0 \triangleq n\tau$ and the control to zero.

(R) Random sparse control: as long as $V(\ell\tau) > \gamma(X(\ell\tau))^2$, at every sampling time $\ell\tau$ we choose an index $i \in \{1, \dots, N\}$ at random following a uniform distribution. Then we define the control as

$$u_j(\ell\tau) \triangleq \begin{cases} -\theta\dfrac{v_i^\perp(\ell\tau)}{\|v_i^\perp(\ell\tau)\|} & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

As in the above controls, as soon as $V(n\tau) \leq \gamma(X(n\tau))^2$ is satisfied for some $n \in \mathbb{N}$, we set $T_0 \triangleq n\tau$ and the control to zero.

(DR) Dimension reduction sparse control chosen from the low-dimensional projected system: here $u_i(\ell\tau) = u_i^h(\ell\tau)$ is defined as in Definition 3.15. In order to test the performance of this control, and to avoid the stability complications arising from finite precision approximation, we calculate the trajectories of both the high- and the low-dimensional system: if the high-dimensional system enters the consensus region first (i.e., $V(n\tau) \leq \gamma(X(n\tau))^2$ for some $n \in \mathbb{N}$), then we set the control to zero and $T_0 \triangleq n\tau$. Instead, if the system in low dimension reaches the consensus region first (i.e., $W(\ell\tau) \leq \gamma(Y(\ell\tau))^2$ for some $\ell \in \mathbb{N}$), then we switch the control for the high-dimensional system to the random sparse control strategy (R) until $V(n\tau) \leq \gamma(X(n\tau))^2$ is eventually satisfied for some $n \in \mathbb{N}$.

Notice that all the controls above are time sparse, and only the uniform control strategy (U) is not componentwise sparse.

Remark 3.22. The reasons for using the random sparse control strategy in the end phase of (DR) (in the case that the low-dimensional system reaches the consensus region before the high-dimensional one) are of numerical and computational nature. In fact, the step width τ computed in Theorem 3.18 to ensure convergence to the consensus region in finite time is often way too small, and in our numerical experiments we need to exceed it. Moreover, as soon as the high-dimensional system enters the consensus region, the difference between consensus parameters becomes so small to render, for such a large time step, the choice of the sparse control highly inaccurate, leading to inefficient chattering phenomena, without steering the high-dimensional system to consensus.

As an alternative, we employ the random sparse control as soon as the low-dimensional system has reached the consensus region (if this happens before the high-dimensional system does). This procedure has the advantage of always steering the high-dimensional system to the consensus region, and to only slightly affect the overall time of the process, since it is usually necessary only for a very short time (provided that the dimension of the Johnson-Lindenstrauss matrix is sufficiently large).

3.6.1 Content of the Numerics

The following are the driving issues concerning the controls introduced above.

(i) Does the control steer the system to the consensus region given by Definition 1.4 in finite time?

(ii) If yes, how long does it take to steer the system to the consensus region?

In the following, for every experiment we fix the number of agents N, the dimension d, the control strength θ, the power β for the interaction kernel a as in (1.7) with $H = 1$ and $\sigma = 1$, the step width τ, and in particular the configuration $(x^0, v^0) \in \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ at the beginning. We report the maximal step width τ_0 (theoretically) allowed by the formula (3.47), and the estimate from above for the time to consensus $\hat{T}$ (taken from Theorem 3.18). We also report the quantity $V(0) - \gamma(X(0))^2$, accounting for the discrepancy of the original configuration from the consensus region.
For every configuration we shall present a table containing the performances of the different controls, measured by the time employed by the high-dimensional system to reach the consensus region T_0, and the time $T_{0.5}$ it takes to halve the "distance" to the consensus region: this means that $T_{0.5}$ is the minimal time satisfying

$$V(T_{0.5}) - \gamma(X(T_{0.5}))^2 \leq \frac{1}{2}\left(V(0) - \gamma(X(0))^2\right).$$

To test the performances of the control (DR) we shall use a variety of Bernoulli random matrices $M \in \mathbb{R}^{k\times d}$ for different dimensions k. For any of these dimensions, we also report the initial discrepancy $W(0) - \gamma(Y(0))^2$ from the consensus region of the projected system, and the switching time T_S at which the random sparse control replaces the original dimension reduction control strategy (if the high-dimensional system enters the consensus region before the low-dimensional one, we set $T_S \triangleq T_0$).
Figure 3.2 below shows the first two coordinates of the initial configurations used in each section.

3.6.2 Examples where (DR) performs second best after (SP)

3.6.2.1 Configuration with one outlier

We take into account $N = 9$ agents in dimension $d = 100$ for which the j-th spatial component of the i-th agent is given by the formula

$$(x_i^0)_j = \frac{1}{2}\cos(i + j\sqrt{2}) \quad \text{for} \quad j = 1, \ldots, d \quad \text{and} \quad i = 1, \ldots, N.$$

The result obtained is a set of points non-homogeneously distributed over an almost spherical configuration, which, projected in $\mathbb{R}^2$, resembles an ellipse. A similar configuration is used for the consensus parameter of each agent, for which we have

$$(v_i^0)_j = \sin(i\sqrt{3} - j) \quad \text{for} \quad j = 1, \ldots, d \quad \text{and} \quad i = 1, \ldots, N-1;$$

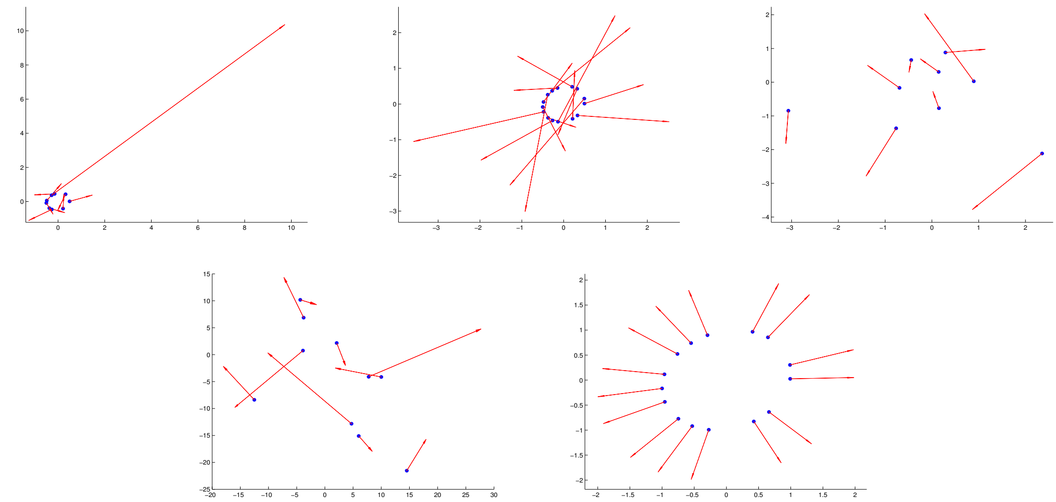


Figure 3.2: From top-left to bottom-right: first two coordinates of the initial configurations of Sections 3.6.2.1, 3.6.2.2, 3.6.2.3, 3.6.3.1, 3.6.3.2. The blue points are the positions of the agents, the red arrows their consensus parameters.

the initial consensus parameter of the N-th agent is instead the vector with all entries set equal to 10.

N	θ	β	d	τ_0	$\hat{T}$	τ	$V(0) - \gamma(X(0))^2$
9	5	0.6	100	$7.33 \cdot 10^{-4}$	115.17	10^{-2}	1031.3

The following table reports the performance of the different controls considered:

Control	(SP)	(U)	(R)	(R)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)
k	-	-	-	-	1	1	5	5	10	10	25	25	32	40	55	55
$W(0) - \gamma(Y(0))^2$	-	-	-	-	202.0	1014.2	509.9	870.2	1651.4	1072.3	1035.2	582.18	933.0	1273.1	1054.5	1046.5
T_0	27.78	87.21	87.90	88.79	69.75	30.98	44.27	30.47	35.55	29.65	27.8	44.75	30.65	32.49	28.20	28.19
$T_{0.5}$	5.44	22.96	22.64	23.21	5.92	5.44	5.44	5.44	5.44	5.44	5.44	5.44	5.44	5.44	5.44	5.44
T_S	-	-	-	-	13.47	22.25	17.81	22.39	32.62	29.03	27.8	23.14	26.2	28.94	28.20	28.15

We first observe that if the system is left alone, with no control acting on it, the quantity $V(t) - \gamma(X(t))^2$ decreases only from 1031.3 to 946.2 at time $t = 100$, from which we can infer that the system would not reach the consensus region without an external intervention. Notice that the Sparse Control (SP) is the fastest; this shall be a common feature of all our experiments, as expected by its optimality shown in Proposition 2.5. The uniform control (U) and the random control (R) perform similarly and both take more than three times longer to reach the consensus region as (SP). The control (DR) has comparable performances to (SP), and very surprisingly even when projecting to dimension $k = 1$ the system reaches the consensus region faster than with the controls (U) and (R).

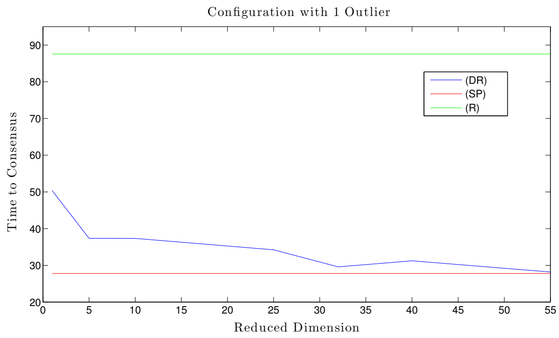


Figure 3.3: Time to consensus for (DR) in function of the projected dimension k, and comparison with (SP) and (R).

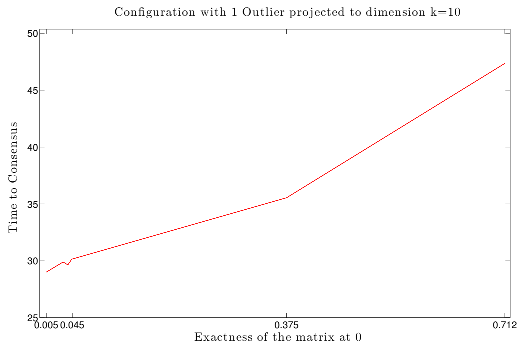


Figure 3.4: Time to consensus in function of the exactness of the matrix at 0 for the fixed dimension $k = 10$.

In Figure 3.3, we illustrate the time T_0 the system takes to reach the consensus region as a function of the projected dimension k for the control (DR). If multiple tests are made with the same dimension k, we consider an average of the results. We also report, in different colors, the values of T_0 we obtain with the control (SP) and the control (R) (blue and green line, respectively). It can be seen how the performance of (DR) is basically the same as (SP) even if we reduce the dimensionality by 80% .

Up to now, we don't have any procedure to test if the randomly generated matrix we use to implement the control (DR) satisfies the requested properties of Theorem 3.18. Moreover, to get a precise answer, we would need to gather information which belongs to the high-dimensional system beyond time $t = 0$, something which we are not allowed to know in advance. We claim, however, that the quantity, which we call *the exactness of the matrix M at 0*,

$$E_M \triangleq \left|1 - \frac{V(0)}{W(0)}\right| = \left|1 - \frac{\sum_{i=1}^N \|v_i^\perp(0)\|^2}{\sum_{i=1}^N \|Mv_i^\perp(0)\|^2}\right|.$$

is a measure of how "good" the matrix M is. To show that, we have considered six different $M \in \mathbb{R}^{k\times d}$ for $k = 10$ and their respective time to the consensus region: we report in Figure 3.4 the time to consensus for the system in function of the exactness of the matrices at 0. A correlation between how E_M is close to zero and how effective is the control, is clearly visible.

3.6.2.2 Configuration generated by a geometric distribution

In this section we consider a system where the locations are distributed as in the example before, while the consensus parameters are given by the formula

$$(v_i^0)_j = (1.2)^{(i-1)/2} \sin(i\sqrt{3} - j) \quad \text{for} \quad j = 1, \dots, d \quad \text{and} \quad i = 1, \dots, N - 1;$$

This results in a more heterogeneous situation at the beginning. We also increase the dimension d to 500, the strength of the force θ to 20 and β to 0.65.

N	θ	β	d	τ_0	$\hat{T}$	τ	$V(0)-\gamma(X(0))^2$
15	20	0.65	500	$1.26 \cdot 10^{-4}$	51.82	10^{-2}	1195.5

The following table summarizes the results of the experiments:

Control	(SP)	(U)	(R)	(R)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)
k	-	-	-	-	1	1	10	10	50	50	50	100	100
$W(0)-\gamma(Y(0))^2$	-	-	-	-	1194.2	1191.9	1194.3	1197.5	1007.2	1199.7	1178.2	1079.1	1204.7
T_0	23.45	38.02	38.10	39.82	40.96	45.41	26.66	29.81	27.45	24.33	26.48	26.88	24.02
$T_{0.5}$	5.49	7.60	7.68	7.66	7.455	9.04	5.64	5.86	5.55	5.5	5.59	5.59	5.50

If we let the system free to evolve, the quantity $V(t)-\gamma(X(t))^2$ decreases only from 1195.5 to 1122.3 at time $t = 30$. The slowness of the decay implies the necessity of a control. The uniform control (U) and the random control (R) perform similarly, as in the example before. However, the control (DR) overwhelms both when the projected dimension is large enough ($k \geq 10$). Figure 3.5 shows the performance of (DR) in function of k and compares it with (R) and (SP).

Time to consensus for (DR) in function of the projected dimension k, and comparison with (SP) and (R)

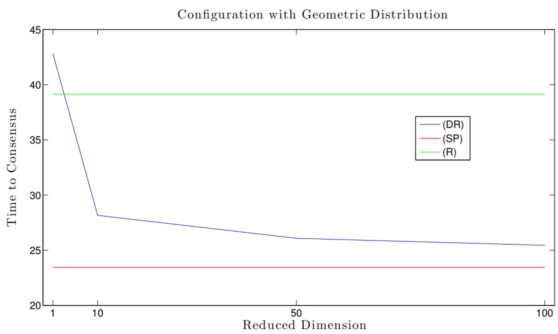


Figure 3.5

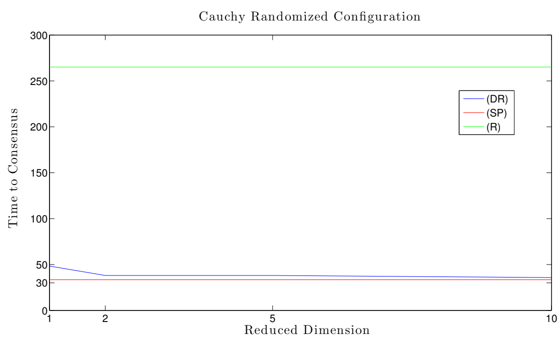


Figure 3.6

3.6.2.3 Configuration generated by a Cauchy distribution

For the system considered in this section, the initial configuration is calculated as follows: the j-th spatial component of the i-th agent is the value of a normal distribution with expected value 0 and standard deviation 1, independently selected for different i and j. The j-th component of the consensus parameter of the i-th agent is ruled by a Cauchy distribution, whose density is given by

$$f(x) = \frac{b}{\pi(b^2 + x^2)}.$$

We choose the height to be $b = 1/40$ (to get a reasonably large $V(0)$ in the computations). The initial configuration is generated once and then fixed for all the experiments with the

different controls (SP), (R), (U) and (DR).
Below we list the parameters we fix for this section:

N	θ	β	d	τ_0	$\hat{T}$	τ	$V(0) - \gamma(X(0))^2$
25	5	0.6	100	$3.77 \cdot 10^{-4}$	214.76	10^{-2}	464.03

The following table reports the performances of the various controls:

Control	(SP)	(U)	(R)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)
k	-	-	-	1	1	2	5	5	10	10
$W(0) - \gamma(Y(0))^2$	-	-	-	461.04	461.04	475.48	464.39	464.39	465.00	465.00
T_0	33.45	266.44	265.14	48.04	48.6	38.07	37.98	38.16	36.11	35.41
$T_{0.5}$	6.1	70.55	68.54	6.1	6.1	6.1	6.1	6.1	6.1	6.1

As in the examples before, the control (DR) clearly outperforms both (R) and (U), and in this case even for $k = 1$. Figure 3.6 compares the effectiveness of the controls (DR) (in function of k), (R) and (SP). We point out that, even in this situation, a control is necessary to steer the system to consensus since the quantity $V(t) - \gamma(X(t))^2$ decreases only from 464 to 436.5 in the time frame $[0, 50]$ if no control is applied.

3.6.3 Examples in which (R) and (U) are comparable to (DR)

3.6.3.1 Configuration generated by a normal distribution

In this example, the j-th spatial (resp., velocity) component of the i-th agent is independently generated by a normal distribution with expected value 0 and standard deviation 10 (resp., 8). As in Section 3.6.2.3, we generate the initial configuration once and we use it for all the experiments with the controls.

N	θ	β	d	τ_0	$\hat{T}$	τ	$V(0) - \gamma(X(0))^2$
10	20	0.65	500	$2.55 \cdot 10^{-6}$	165.68	$5 \cdot 10^{-3}$	27458

The parameters used for this configuration are listed in the table above, while on the one belowe we report the performances of the various controls.

Control	(SP)	(U)	(R)	(R)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)
dim. k	-	-	-	-	1	1	2	5	10	10	20	50	50	100	100
$W(0) - \gamma(Y(0))^2$	-	-	-	-	27496	27469	27421	27425	27458	27493	27464	27482	27481	27495	27498
T_0	82.65	84.56	85.82	85.25	129.28	153.02	115.91	99.79	95.31	100.18	96.7	89.79	91.02	91.67	89.60
$T_{0.5}$	24.09	24.13	24.15	24.13	36.195	42.76	29.28	26.47	26.75	25.25	25.32	24.7	24.74	24.44	24.32
T_S	-	-	-	-	68.06	62.94	76.91	80.51	80.55	77.30	81.57	82.25	82.25	82.49	82.57

This time the controls (R) and (U) are quasi-optimal, performing in almost the same way as the benchmark control (SP). Figure 3.7 shows that the control (DR) behaves similarly to (R) and (SP) up to a reduced dimension $k = 50$ (hence up to 10% of the original dimension): from that point on the efficiency rapidly deteriorates, making the control unfeasible.

Time to consensus for (DR) in function of the projected dimension k, and comparison with (SP) and (R)

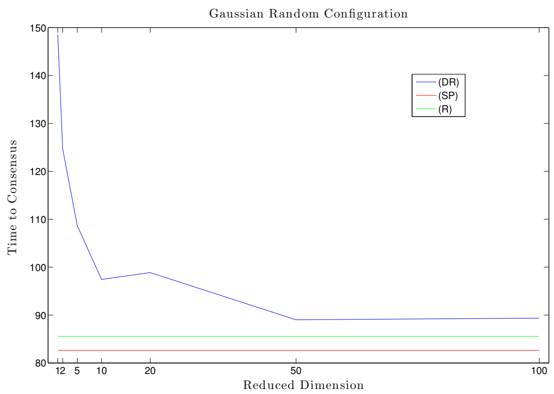


Figure 3.7

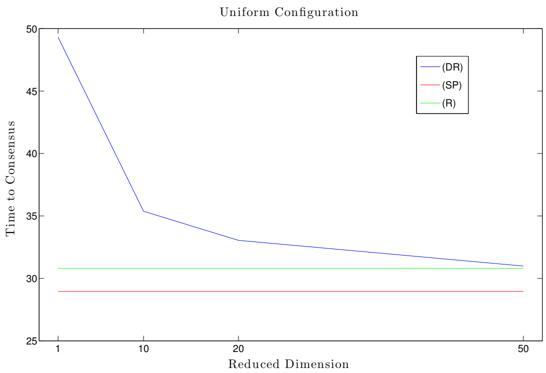


Figure 3.8

3.6.3.2 Uniform configuration

As last example we consider a configuration similar to the one of Section 3.6.2.1: the j-th spatial and consensus parameter components of the i-th agent are both given by

$$(x_i^0)_j = (v_i^0)_j = \cos(i + j\sqrt{2}) \quad \text{for} \quad j = 1, \ldots, d \quad \text{and} \quad i = 1, \ldots, N.$$

N	θ	β	d	τ_0	$\hat{T}$	τ	$V(0) - \gamma(X(0))^2$
15	5	0.8	200	$1.91 \cdot 10^{-5}$	59.48	10^{-3}	98.30

The above tables report the parameters of the configuration taken into account and the one below the outcomes of the experiments with the different controls.

Control	(SP)	(U)	(R)	(R)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)	(DR)
k	-	-	-	-	1	1	10	10	20	20	50	50
$W(0) - \gamma(Y(0))$	-	-	-	-	95.98	43.50	95.85	96.70	77.12	101.65	97.02	122.83
T_0	28.95	29.95	30.86	30.74	53.97	44.47	38.13	32.35	33.08	32.73	29.41	32.45
$T_{0.5}$	7.99	8.30	8.30	8.31	9.74	9.21	8.92	8.15	8.21	8.17	7.99	8.14

As before, (R) and (U) perform similarly to (SP); (DR) is able to compete up to a dimension reduction of 25% of the original dimension ($k = 50$). From there on, its efficiency steadily declines. This phenomenon can be witnessed in Figure 3.8.

3.6.4 Conclusions from the experiments

In this section we summarize the conclusions that can be drawn from the list of experiments reported in this numerical section.

(*i*) A common feature of all the experiments is that the control (DR) is highly competitive with respect to the benchmark control (SP) up to a reduced dimension which is 10% of the original one. Indeed, in this case (DR) takes between 5 to 22% more time than (SP) to steer the system to consensus. This suggests that the approach of dimension reduction works in general much better practically than theoretically, and that our analysis in Theorem 3.18 is quite conservative.

(*ii*) The dimension of the matrix is not the only necessary ingredient to obtain a competitive control: a matrix should also fulfill the Johnson-Lindenstrauss property for certain points of the high dimensional system. Since to check the latter condition we need information regarding the future development of the system, we need to design different criteria to distinguish "good" matrices from "bad" ones. In Section 3.6.2.1, we have seen that a possible sieve is the notion of *exactness of a matrix at* 0: the smaller this value is, the better the control shall perform, according to the empirical data we have gathered.

(*iii*) There is no proof yet that the random sparse control (R) forces the system to enter the consensus region almost surely for every configuration, but numerical experiments suggest this behavior. Furthermore, it is interesting to notice that the time to consensus obtained by the use of the uniform control (U) is always very close to the one we get by using the random sparse control strategy (R): this strongly hints that the expected value of the time to consensus of the random control (R) could be very near or even equal to the one of (U).

(*iv*) A common feature of the last two examples is the "relative homogeneity" of the consensus parameters with respect to the mean consensus parameter: by this we mean that the consensus parameters of all the agents compete to be the furthest away from it, and thus the sparse control will jump from one to another continuously, showing a chattering behavior. In contrast, all the first three experiments feature a relatively small subgroup of agents whose consensus parameters are the furthest away from the mean consensus parameter by a considerable margin. These are the cases where the controls (SP) and (DR) are substantially more efficient than (R) and (U): by firmly acting on the most "badly behaving" agents, we are able to steer the system to consensus faster than employing control strategies which are blind to the structure of the group. It is thus advisable to use sparse strategies only when the consensus parameters of the agents are sufficiently "asymmetric" at the starting point.

(*v*) In the next section we shall propose an *asymmetry measure* of a group of agents which tries to quantify how "heterogeneous" a configuration of agents is. This measure is designed to satisfy the following property: if the asymmetry of the initial configuration (x^0, v^0) is large, then (DR) markedly outperforms the control (R). If, instead, the asymmetry of (x^0, v^0) is small, the discrepancy between the performances of the controls (SP) and (R) is small, thus making (R) a viable substitute of it and of (DR). With such a tool, we shall be able to know in advance if it is reasonable to take on the computational effort to calculate (DR), or if a random choice of the agent to control is a feasible option.

3.6.5 The asymmetry measure of a configuration

We have seen in the numerical experiments of the last section that sometimes the Random control strategy (R) is not only sufficient to steer the system to consensus, but also performs almost as efficiently as the benchmark control (SP). Knowing in advance if a system reacts well to (R) or not can save a huge amount of computational time, both at the implementation level and for what concerns the speed of the convergence to consensus. Based on the empirical evidences we have collected in Section 3.6, in this section we propose a criterion to decide *a priori* if it is convenient to employ the Dimension Reduction control (DR) or the Random control (R).

We start by noticing that from (3.51) follows that whenever the control strategy we use acts on a single agent at a time, say the $\hat{\imath}$-th agent, then we can estimate the decay of the Lyapunov functional V as

$$\frac{d}{dt}V(t) \leq -\frac{2\theta}{N}\frac{v_{\hat{\imath}}^{\perp}(n\tau)\cdot v_{\hat{\imath}}^{\perp}(t)}{\|v_{\hat{\imath}}^{\perp}(n\tau)\|} \leq -\frac{2\theta}{N}\|v_{\hat{\imath}}^{\perp}(t)\|.$$

Hence, since we are trying to maximize this decay rate, the "right" agent to control is the farthest away from the mean consensus parameter. But how much does this choice pay off? This clearly depends on the overall *symmetry* of the configuration: if the distances of the agents from the mean $\overline{v}$ are more or less the same, i.e.,

$$\|v_i^{\perp}(t)\| \approx \|v_j^{\perp}(t)\|, \qquad \text{for all } i \neq j,$$

then it does not matter "too much" which agent should be controlled. However, whenever there is an agent whose distance from the mean consensus parameter is considerably larger than the others, it is crucial to pick precisely this agent in order to maximize the decay rate (which is also the idea behind Proposition 2.5).

To quantify this rough idea, define for every $p \in [1,\infty)$ the *p-normalized asymmetry measure at* 0 as

$$\overline{\sigma}_p \triangleq \left(\frac{1}{N}\sum_{i=1}^{N}\left|\|v_i^{\perp}(0)\| - \frac{1}{N}\sum_{j=1}^{N}\|v_j^{\perp}(0)\|\right|^p\right)^{1/p}.$$

One might be tempted to use $p = 2$ since then $\overline{\sigma}_2$ resembles the standard deviation of the moduli of the velocities. However, as we have already pointed out, a single outlier should "count" more than a large group of agents with similar distances from the mean velocity. And since it holds

$$\lim_{p\to\infty}\overline{\sigma}_p = \max_{1\leq i\leq N}\left|\|v_i^{\perp}(0)\| - \frac{1}{N}\sum_{j=1}^{N}\|v_j^{\perp}(0)\|\right|,$$

then a value $p \gg 2$ will be more susceptible to react to the presence of possible outliers. This is the reason why in what follows we choose $p = 100$.

We now show that there is a correlation between how large the asymmetry measure of an initial configuration (x^0, v^0) is and how well (R) performs to steer the system starting at (x^0, v^0) to the consensus region. To measure the difference between the performance of the benchmark control (SP) and the Random control (R), let $T_0^{(SP)}$ be the time to consensus when using (SP), and let $T_0^{(R)}$ be the time to consensus when using (R), averaged over several runs. We define the *relative time perfomance* T_1 as

$$T_1 \triangleq \frac{T_0^{(R)} - T_0^{(SP)}}{T_0^{(SP)}}.$$

Notice that, by the optimality of (SP), we expect T_1 to be a positive quantity.

We also measure the difference between the performances of (R) and (DR) (which, notice, might also be negative). If $T_0^{(DR)}$ is the average time to consensus using the control (DR) – averaged among all dimensions k in the examples – then the *relative time perfomance* T_2 is given by

$$T_2 \triangleq \frac{T_0^{(R)} - T_0^{(DR)}}{T_0^{(SP)}}.$$

Table 3.1 reports the values of $T_0^{(SP)}$, $T_0^{(DR)}$, $T_0^{(R)}$, T_1, T_2 and $\overline{\sigma}_{100}$ for the five examples of the last section. We can see that the first three examples have a high asymmetry and both T_1 and T_2 are positive. The last two examples have a relatively lower level of asymmetry, together low values of T_1 and T_2 (which is negative since, as we have seen in the last section, in these examples (R) is faster that (DR) in bringing the system to consensus).

Example Section	6.2.1	6.2.2	6.2.3	6.3.1	6.3.2
$T_0^{(SP)}$	27.78	23.45	33.45	82.65	28.95
$T_0^{(DR)}$	34.41	26.70	40.34	113.49	36.93
$T_0^{(R)}$	88.35	38.65	265.14	85.53	30.80
T_1	2.18	0.67	6.93	0.04	1.39
T_2	1.94	0.51	6.72	-0.34	-0.21
$\overline{\sigma}_{100}$	65.87	21.74	90.67	8.28	1.3943

Table 3.1: Values of $T_0^{(SP)}$, $T_0^{(DR)}$, $T_0^{(R)}$, T_1, T_2 and $\overline{\sigma}_{100}$ for the examples in Section 3.6.

We illustrate this further with two histograms using a richer data set of initial configurations, encompassing those of the previous sections. Figure 3.9 reports the correlation between T_1 and $\overline{\sigma}_{100}$, while Figure 3.10 shows the correlation between T_2 and $\overline{\sigma}_{100}$. The histograms clearly show that the asymmetry measure $\overline{\sigma}_{100}$ is a suitable detector of when the (DR) control strategy is needed or when the Random control (R) is sufficient in order to steer the system to consensus.

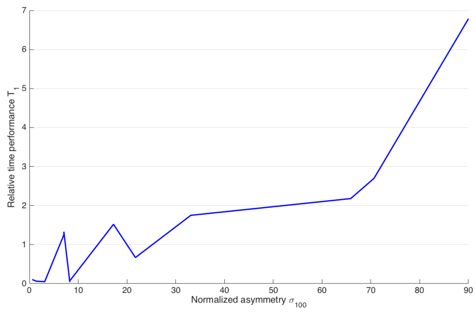


Figure 3.9: Correlation between T_1 and $\overline{\sigma}_{100}$.

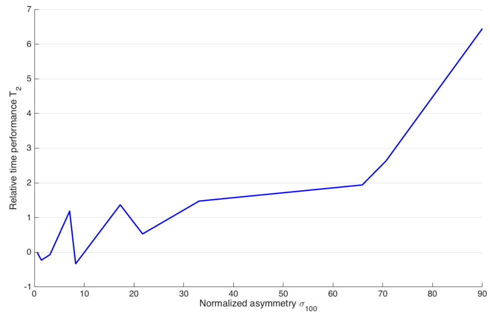


Figure 3.10: Correlation between T_2 and $\overline{\sigma}_{100}$.

3.7 Real-life approach

We conclude this chapter by listing several guidelines for the use of sparse controls and dimensionality reduction techniques. These points are summing up both the theoretical and numerical analysis we carried out in the last two chapters, and should be considered as a practical receipt to follow whenever sparse control on a dynamical system is needed but its dimensionality makes the implementation of the direct strategy problematic.

(i) Fix a threshold $A > 0$ for the maximal asymmetry level allowed, and another one $E > 0$ for the maximal exactness of the matrices at 0 (that is, E_M) allowed.

(ii) Given an initial configuration (x^0, v^0), compute $V(0)$ and $\gamma(X(0))$. If the system is already in the consensus region, then there is nothing to do. If not, compute the asymmetry $\overline{\sigma}_{100}$ of (x^0, v^0).

(iii) If $\overline{\sigma}_{100} < A$, then use the Random control (R). Notice that the only high-dimensional information needed at each step is the consensus parameter v_i of the agent to control, since i is randomly chosen.

(iv) If $\overline{\sigma}_{100} \geq A$, choose a feasible projected dimension $k \geq d$ (experiments suggest $k \sim d/10$), generate a random matrix $M \in \mathbb{R}^{k\times d}$ as in (S2) or (S3) of Remark 3.4 and compute E_M, its exactness at 0.

(v) If $E_M \geq E$ then repeat the step above. If $E_M < E$, enact the control strategy (DR) using the projected system obtained via M. Again, the only high-dimensional information needed at each step is the consensus parameter $v_{\hat{\imath}}$ of the agent to control, since $\hat{\imath}$ is chosen in low dimension.

CHAPTER 4

The mean-field approximation

We now make a brief pause from sparse controls and show how the curse of dimensionality can be circumvented when the number of agents N is very large while their dimension d remains modest. The strategy relies on a very simple idea: instead of considering the influence that the entire population has on a single agent, we substitute it with an averaged one (a process known in the literature as *mean-field approximation*), thus reducing a many-body problem to a one-body problem. This let us drop the original huge system of ODEs in favor of a single PDE for the probability density of agents in the state space: the price to pay is that, in this way, we only have an average knowledge of the behavior of the agents. We shall study in detail two approaches to derive rigorously this procedure: the mean-field limit and the grazing interaction limit. All the results presented here are slight variations of well-known results and are reported for the sake of completeness, since they will be employed at several times in the forthcoming chapters. We remand to the reviews [107, 130, 162] and the book [151] for further insights.

4.1 The Wasserstein space of probability measures

We start by introducing the basic definitions and properties regarding spaces of measures. For more details, we refer to the classical references [13, 176].

Let $\mathcal{M}_b(\mathbb{R}^{d_1};\mathbb{R}^{d_2})$ be the space of finite vector measures from $\mathbb{R}^{d_1}$ to $\mathbb{R}^{d_2}$ equipped with the total variation norm $\|\cdot\|_{\mathcal{M}_b(\mathbb{R}^{d_1};\mathbb{R}^{d_2})}$. We denote by $\mathcal{M}_b(\mathbb{R}^d) \triangleq \mathcal{M}_b(\mathbb{R}^d;\mathbb{R}_+)$ the set of finite positive measures defined on the Borel sets of $\mathbb{R}^d$, and we define $\mathcal{P}(\mathbb{R}^d)$ as the subset of $\mathcal{M}_b(\mathbb{R}^d)$ of all the probability measures on $\mathbb{R}^d$, i.e., all $\mu \in \mathcal{M}_b(\mathbb{R}^d)$ for which $\mu(\mathbb{R}^d) = 1$. The space $\mathcal{P}_p(\mathbb{R}^d)$ is the subset of $\mathcal{P}(\mathbb{R}^d)$ whose elements have finite p-th moment, i.e.,

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

Clearly, we can extend the notation to any subset $\Omega \subseteq \mathbb{R}^d$ and, whenever Ω is bounded, it holds $\mathcal{P}_p(\Omega) = \mathcal{P}(\Omega)$ for every $p \geq 1$. We denote by $\mathcal{P}_c(\mathbb{R}^d)$ the set of all $\mu \in \mathcal{P}(\mathbb{R}^d)$ with compact support, i.e., for which the set

$$\mathrm{supp}(\mu) \triangleq \overline{\{x \in \mathbb{R}^d : \text{for every open set } A \text{ such that } x \in A \text{ it holds } \mu(A) > 0\}}$$

is contained in a compact set of $\mathbb{R}^d$. Notice that $\mathcal{P}_c(\mathbb{R}^d) \subset \mathcal{P}_p(\mathbb{R}^d)$ for every $p \geq 1$.
For any $\mu \in \mathcal{P}(\mathbb{R}^{d_1})$ and any Borel function $f : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, we denote by $f_{\#}\mu \in \mathcal{P}(\mathbb{R}^{d_2})$ the *push-forward of* μ *through* f, defined as

$$f_{\#}\mu(B) \triangleq \mu(f^{-1}(B)) \quad \text{for every } B \subseteq \mathbb{R}^{d_2} \text{ Borel set.}$$

In particular, if one considers the projection operators π_1 and π_2 defined on the product space $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, for every $\rho \in \mathcal{P}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ we call *first* (resp., *second*) *marginal* of ρ the probability measure $\pi_{1\#}\rho$ (resp., $\pi_{2\#}\rho$). Given $\mu \in \mathcal{P}(\mathbb{R}^{d_1})$ and $\nu \in \mathcal{P}(\mathbb{R}^{d_2})$, we denote by $\Gamma(\mu, \nu)$ the subset of all probability measures in $\mathcal{P}(\mathbb{R}^{d_1} \times \mathbb{R}^{d_2})$ with first marginal μ and second marginal ν.
On the set $\mathcal{P}_p(\mathbb{R}^d)$ we shall consider the following distance, called the *Wasserstein or Monge-Kantorovich-Rubinstein* distance,

$$\mathcal{W}_p(\mu, \nu) \triangleq \inf_{\rho \in \Gamma(\mu,\nu)} \left\{ \int_{\mathbb{R}^{2d}} \|x - y\|^p d\rho(x, y) \right\}^{1/p}. \tag{4.1}$$

If $p = 1$, we have the following dual representation for the Wasserstein distance

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

We denote by $\Gamma_o(\mu, \nu)$ the subset of measures $\rho \in \Gamma(\mu, \nu)$ for which the minimum in (4.1) is attained, i.e.,

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

It is well-known that $\Gamma_o(\mu, \nu)$ is non-empty for every $(\mu, \nu) \in \mathcal{P}_p(\mathbb{R}^n) \times \mathcal{P}_p(\mathbb{R}^n)$, hence the infimum in (4.1) is actually a minimum.
For any $\mu \in \mathcal{M}_b(\mathbb{R}^d)$ and $f : \mathbb{R}^d \to \mathbb{R}^d$, the notations $f * \mu$ and $\langle f, \mu \rangle$ stand for the convolution and the duality pairing between f and μ, respectively, i.e.,

$$(f * \mu)(x) \triangleq \int_{\mathbb{R}^d} f(x - y) d\mu(y) \quad \text{and} \quad \langle f, \mu \rangle \triangleq \int_{\mathbb{R}^{2d}} f(x) d\mu(x),$$

whenever those quantities are well-defined (for example, whenever f is sublinear and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, or simply whenever f is bounded). Furthermore, if $\omega \in \mathcal{C}(\mathbb{R}^d; \mathbb{R}^d)$ is sublinear and $\mu \in \mathcal{P}_1(\mathbb{R}^d)$, we define the Radon vector measure $\omega\mu \in \mathcal{M}_b(\mathbb{R}^d; \mathbb{R}^d)$ as

$$\omega\mu(A) \triangleq \int_A \omega(x) d\mu(x) \quad \text{for every } A \subset \mathbb{R}^d \text{ bounded.} \tag{4.2}$$

We shall denote by $\int \omega\mu \triangleq \omega\mu(\mathbb{R}^d) = \langle \omega, \mu \rangle$.

For any $\beta \in \mathbb{N}^d$ we set $|\beta| \triangleq \sum_{i=1}^d \beta_i$, and for any function $h \in \mathcal{C}^p(\mathbb{R}^{2d};\mathbb{R})$, with $p \geq 0$ and $\beta \in \mathbb{N}^d$ such that $|\beta| \leq p$, we define for every $(x,v) \in \mathbb{R}^{2d}$

$$\partial_v^\beta h(x,v) \triangleq \frac{\partial^{|\beta|} h}{\partial^{\beta_1} v_1 \cdots \partial^{\beta_d} v_d}(x,v),$$

with the convention that if $\beta = (0,\ldots,0)$ then $\partial_v^\beta h(x,v) \triangleq h(x,v)$.

Later on in the chapter, we shall need the following class of test functions.

Definition 4.1 (Test functions)**.** For every $\delta > 0$, we denote by $\mathcal{T}_\delta$ the set of compactly supported functions φ from $\mathbb{R}^{2d}$ to $\mathbb{R}$ such that for any multi-index $\beta \in \mathbb{N}^d$ we have:

(i) if $|\beta| < 2$, then $\partial_v^\beta \varphi(x,\cdot)$ is continuous for every $x \in \mathbb{R}^d$;

(ii) if $|\beta| = 2$, then there exists an $M > 0$ such that

(a) $\partial_v^\beta \varphi(x,\cdot)$ is uniformly Hölder continuous of order δ for every $x \in \mathbb{R}^d$ with Hölder bound M, that is for every $x \in \mathbb{R}^d$ and for every $v, w \in \mathbb{R}^d$ it holds

$$\left\| \partial_v^\beta \varphi(x,v) - \partial_v^\beta \varphi(x,w) \right\| \leq M \left\| v - w \right\|^\delta;$$

(b) $\|\partial_v^\beta \varphi(x,v)\| \leq M$ for every $(x,v) \in \mathbb{R}^{2d}$.

Notice that $\mathcal{C}_c^\infty(\mathbb{R}^{2d};\mathbb{R}) \subseteq \mathcal{T}_\delta$ for every $\delta \in (0,1]$.

4.2 The mean-field equation and existence of solutions

Notice that by using the functional $G^{[a]} : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ defined as

$$G^{[a]}(x,v) \triangleq (v, -a(\|x\|)v), \qquad \text{for all } (x,v) \in \mathbb{R}^{2d},$$

we can rewrite the Cucker-Smale system (1.6) in the following compact form

$$(\dot{x}_i(t), \dot{v}_i(t)) = \frac{1}{N}\sum_{j=1}^N G^{[a]}(x_i(t) - x_j(t), v_i(t) - v_j(t)). \tag{4.3}$$

This shows that systems of this kind are all basically first-order, and it is thus not restrictive to focus our attention on systems of the form

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1}^N K(x_i(t) - x_j(t)) + g(t, x_i(t)), \tag{4.4}$$

at least for the sake of compactness. To shorten the notation further, we pass from dimension $2d$ as in (4.3) to simply d: hence in (4.4), the interaction kernel K is a map from $\mathbb{R}^d$ to $\mathbb{R}^d$. The extra term $g : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ stands for an (optional) external vector field applied to each agent and changing in time.

The situation of equation (4.4) in which the force exerted on each agent is of order 1 as $N \to \infty$ is known in the mathematical literature as the *mean-field* (or *weak coupling*) scaling.

In order to study the behavior of the solutions of system (4.4) as $N \to \infty$, we consider the family of ODE systems (indexed by $N \in \mathbb{N}$) with finite time horizon $T > 0$ given by

$$\begin{cases} \dot{x}_{i,N}(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} K(x_{i,N}(t) - x_{j,N}(t)) + g_N(t, x_{i,N}(t)) \quad \text{for } t \in (0,T], \\ x_{i,N}(0) = x^0_{i,N}, \end{cases} \quad i = 1, \dots, N. \tag{4.5}$$

Notice the label N on the term g_N, since its form may also depend on the number of agents.

As N increases, so does the complexity of system (4.5): this specifies, for instance, into a chaotic behavior – which means a strong sensitivity to small perturbations – and numerical intractability. Indeed, not even fast particles methods [111, 112] can handle the number of 10^{25} agents that are sometimes needed in several applications. However, a first milestone towards the complexity reduction of such systems comes from exploiting the mean-field scaling.

Definition 4.2 (Empirical measures)**.** The empirical distribution of system (4.5) is the curve $\mu_N : [0,T] \to \mathcal{P}_c(\mathbb{R}^d)$ defined as

$$\mu_N(t) \triangleq \frac{1}{N}\sum_{i=1}^{N} \delta_{x_{i,N}(t)} \quad \text{for every } t \in [0,T], \tag{4.6}$$

where $x_N : [0,T] \to \mathbb{R}^{dN}$ is the solution of system (4.5) and δ_x denotes the Dirac measure centered on $x \in \mathbb{R}^d$.

By means of empirical measures, system (4.5) can be rewritten as

$$\begin{cases} \dot{x}_{i,N}(t) = (K * \mu_N(t))(x_{i,N}(t)) + g_N(t, x_{i,N}(t)) \quad \text{for } t \in (0,T], \\ x_{i,N}(0) = x^0_{i,N}, \end{cases} \quad i = 1, \dots, N. \tag{4.7}$$

It is well-known that renormalized sums of Dirac measures like (4.6) are dense in $\mathcal{P}(\mathbb{R}^d)$, hence under further hypotheses on the atoms $(x_{i,N}(\cdot), \dots, x_{i,N}(\cdot))$ (which we overlook, for the moment) we may assume that there exists $\mu : [0,T] \to \mathcal{P}(\mathbb{R}^d)$ such that $\mu_N(t) \rightharpoonup \mu(t)$ for every $t \in [0,T]$. The approximation $\mu_N \approx \mu$ valid for N large implies that the solution of (4.7) is close to that of

$$\begin{cases} \dot{\xi}(t) = (K * \mu(t))(\xi(t)) + g(t, \xi(t)) \quad \text{for } t \in (0,T], \\ \xi(0) = x^0_{i,N}, \end{cases} \tag{4.8}$$

(provided that g_N converges uniformly to the function g). However, the above equation is nothing but a characteristic curve of the PDE

$$\frac{\partial \mu}{\partial t}(t) = -\nabla_x \cdot \Big((K * \mu(t) + g(t))\mu(t) \Big),$$

which is a particular instance of the general Vlasov-type transport equation

$$\frac{\partial \mu}{\partial t}(t) = -\nabla_x \cdot (V(t,\mu(t))\mu(t)) \tag{4.9}$$

for the choice of the vector field

$$V(t,\mu)(x) = (K * \mu)(x) + g(t,x).$$

Definition 4.3 (Weak solution of (4.9))**.** We say that a map $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ is a solution of (4.9) if the following holds:

(*i*) $\mu(\cdot)$ has uniformly compact support, i.e., there exists $R > 0$ such that $\operatorname{supp}(\mu(t)) \subseteq B(0,R)$ for every $t \in [0,T]$;

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

(*iii*) $\mu(\cdot)$ satisfies (4.9) in the weak sense, i.e. (see [13, Equation (8.1.4)]),

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

for every $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d;\mathbb{R})$ and every $t \in [0,T]$.

Remark 4.4 ([106, Section 4])**.** Since the linear span of functions of the type $\eta(t)\zeta(x,v)$ with $\eta \in \mathcal{C}_c^\infty([0,T];\mathbb{R})$ and $\zeta \in C_c^\infty(\mathbb{R}^d;\mathbb{R})$ is dense in $C_c^\infty([0,T]\times\mathbb{R}^d;\mathbb{R})$, it is not difficult to manipulate equation (4.9) to see it is equivalent to

$$\int_0^T \int_{\mathbb{R}^d} \left(\frac{\partial \varphi}{\partial t}(t,x) + \nabla_x\varphi(t,x) \cdot V(t,\mu(t))(x)\right) d\mu(t,x)dt = 0,$$

for every $\varphi \in C_c^\infty([0,T] \times \mathbb{R}^d;\mathbb{R})$. Another equivalent reformulation of (4.9) involving the duality pairing between continuous functions and measures is the following: for every $t \in [0,T]$ and $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^d;\mathbb{R})$ it holds

$$\langle \phi, \mu(t) - \mu(0)\rangle = \int_0^t \int_{\mathbb{R}^d} \nabla\phi(x) \cdot V(s,\mu(s))(x) d\mu(s,x)ds.$$

Motivated by the above insights, we consider the system

$$\begin{cases} \dfrac{\partial \mu}{\partial t}(t) = -\nabla_x \cdot \Big((K * \mu(t) + g(t))\mu(t)\Big) & \text{for } t \in (0,T], \\ \mu(0) = \mu^0, \end{cases} \tag{4.10}$$

and we pass to determine under which conditions the formal computations that led us to it actually hold true. As a preliminary result, we can show that solutions of system (4.5) are also solutions of system (4.10), whenever conveniently rewritten.

Proposition 4.5. *Let $N \in \mathbb{N}$ be given. Let $(x_{1,N},\dots,x_{N,N}) : [0,T] \to \mathbb{R}^{dN}$ be the solution of* (4.5) *with initial datum $(x^0_{1,N},\dots,x^0_{N,N}) \in \mathbb{R}^{dN}$. Then the empirical measure-valued* *curve* $\mu_N : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ *defined as in* (4.6) *is a solution of* (4.10) *with initial datum*

$$\mu^0 = \mu^0_N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x^0_{i,N}}.$$

Proof. It is straightforward to check that inserting $\mu_N(\cdot)$ into (4.10) and using the equations of (4.5), we get an identity. □

Remark 4.6. In view of the above result, there is no ambiguity whenever we say that the empirical measure-valued curve $\mu_N(\cdot)$ is a solution of the discrete system (4.5).

4.2.1 Existence results for convolution-type ODEs

To link solutions of (4.5) to those of (4.10), we first need two results concerning the regularity of the interaction kernel K.

Lemma 4.7. *If* $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$ *and* $\mu \in \mathcal{P}(\mathbb{R}^d)$ *then* $K * \mu \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$.

Proof. For any compact set $\Omega \subset \mathbb{R}^d$ and for every $x, y \in \Omega$ it holds

$$\begin{aligned}\|(K*\mu)(x) - (K*\mu)(y)\| &= \left\|\int_{\mathbb{R}^d} K(x-z)d\mu(z) - \int_{\mathbb{R}^d} K(y-z)d\mu(z)\right\| \\ &\le \mathrm{Lip}_{\widehat{\Omega}}(K)\|x-y\|,\end{aligned}$$

for any conveniently larger compact set $\widehat{\Omega} \subset \mathbb{R}^d$ depending only on Ω. This shows the local Lipschitz continuity of $K * \mu$. □

Lemma 4.8. *If* K *is sublinear and* $\mu \in \mathcal{P}_1(\mathbb{R}^d)$ *then* $K * \mu$ *is sublinear. Furthermore, it holds*

$$\|(K*\mu)(x)\| \le C_K\left(1 + \|x\| + \int_{\mathbb{R}^d}\|y\|d\mu(y)\right), \tag{4.11}$$

where C_K *stands for the sublinearity constant of* K.

Proof. If C_K is the sublinearity constant of K, an easy computation shows that

$$\begin{aligned}\|(K*\mu)(x)\| &= \left\|\int_{\mathbb{R}^d} K(x-y)d\mu(y)\right\| \\ &\le \int_{\mathbb{R}^d}\|K(x-y)\|d\mu(y) \\ &\le C_K\int_{\mathbb{R}^d}(1+\|x\|+\|y\|)d\mu(y) \\ &\le C_K\left(1+\|x\|+\int_{\mathbb{R}^d}\|y\|d\mu(y)\right) \\ &\le C(1+\|x\|),\end{aligned}$$

for a suitably large constant $C > 0$. In particular, (4.11) holds. □

By the two lemmas above, we can invoke the usual Cauchy-Lipschitz theory (in its Carathéodory version, see Theorems B.1 and B.3) to prove the existence and uniqueness of solutions of system (4.5). What is remarkable, and at the heart of the first rigorous results on the mean-field limit (see [41, 95, 148]), is that several peculiar properties of the solutions of system (4.5) are independent of N. This fact will turn out to be crucial when we shall extend these properties to solutions of system (4.10).

Proposition 4.9. *Assume that $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$ and is sublinear, and that the function $g_N \in \mathcal{C}([0,T] \times \mathbb{R}^d; \mathbb{R}^d)$ satisfies*

(i) there exists a constant $G > 0$ independent of N such that

$$g_N(t,x) \cdot x \leq G(1 + \|x\|^2) \quad \textit{for every } (t,x) \in [0,T] \times \mathbb{R}^d; \tag{4.12}$$

(ii) for every compact $\Omega \subset \mathbb{R}^d$ there exists $L_{N,\Omega} \in L^1([0,T])$ satisfying

$$\|g_N(t,x) - g_N(t,y)\| \leq L_{N,\Omega}(t)\|x - y\| \quad \textit{for every } t \in [0,T] \textit{ and } x, y \in \Omega; \tag{4.13}$$

(iii) for every compact $\Omega \subset \mathbb{R}^d$ there exists $G_\Omega > 0$ independent of N for which it holds

$$\|g_N\|_{L^\infty([0,T]\times\Omega)} \leq G_\Omega.$$

Then, for every $N \in \mathbb{N}$, system (4.5) admits a unique global solution defined on $[0,T]$ for every initial datum $(x^0_{1,N}, \ldots, x^0_{N,N}) \in \mathbb{R}^{dN}$.
Moreover, if there exists a bounded set $B \subset \mathbb{R}^d$ such that $x^0_{i,N} \in B$ for every $N \in \mathbb{N}$ and every $i = 1, \ldots, N$, then there exist two constants $L, R > 0$ independent of N such that

- *$\|x_{i,N}(t)\| \leq R$, for every $t \in [0,T]$ and $i = 1, \ldots, N$,*
- *$\|x_{i,N}(t) - x_{i,N}(s)\| \leq L|t - s|$, for every $t, s \in [0,T]$ and $i = 1, \ldots, N$.*

Proof. We begin by proving that, if $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$ then for any $N \in \mathbb{N}$ the Carathéodory function $F : \mathbb{R}^{dN} \to \mathbb{R}^{dN}$ defined for every $(x_1, \ldots, x_N) \in \mathbb{R}^{dN}$ as

$$F(x_1, \ldots, x_N) \triangleq \left(\frac{1}{N} \sum_{j=1}^N K(x_1 - x_j), \ldots, \frac{1}{N} \sum_{j=1}^N K(x_N - x_j) \right)^T,$$

satisfies $F \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^{dN})$. Indeed, for any $(x_1, \ldots, x_N), (y_1, \ldots, y_N) \in \Omega$ compact subset of $\mathbb{R}^{dN}$, it holds

$$\begin{aligned}
\|F(x_1, \ldots, x_N) - F(y_1, \ldots, y_N)\| &\leq \sum_{i=1}^N \left\| \frac{1}{N} \sum_{j=1}^N K(x_i - x_j) - \frac{1}{N} \sum_{j=1}^N K(y_i - y_j) \right\| \\
&\leq \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \|K(x_i - x_j) - K(y_i - y_j)\| \\
&\leq \frac{1}{N} \mathrm{Lip}_{\widehat{\Omega}}(K) \sum_{i=1}^N \sum_{j=1}^N \|x_i - x_j - y_i + y_j\|
\end{aligned}$$

$$\leq 2\,\mathrm{Lip}_{\widehat{\Omega}}(K)\sum_{i=1}^{N}\|x_i-y_i\|$$
$$\leq 2\sqrt{N}\,\mathrm{Lip}_{\widehat{\Omega}}(K)\,\|(x_1,\dots,x_N)-(y_1,\dots,y_N)\|$$

having used Lemma 4.7 for a sufficiently large compact subset $\widehat{\Omega}\subset\mathbb{R}^d$. The factor $\sqrt{N}$ comes from having estimated the ℓ_1-norm with the ℓ_2-norm.

We now invoke the sublinearity of K and Lemma 4.8 to show the sublinearity of F:

$$\begin{aligned}\|F(x_1,\dots,x_N)\| &\leq \sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{j=1}^{N}K(x_i-x_j)\right\|\\ &\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\|K(x_i-x_j)\|\\ &\leq \frac{C_K}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}(1+\|x_i\|+\|x_j\|)\\ &\leq C_K\left(N+2\sum_{i=1}^{N}\|x_i\|\right)\\ &\leq 2NC_K(1+\|(x_1,\dots,x_N)\|).\end{aligned}$$

By setting $H:\mathbb{R}^{dN}\to\mathbb{R}^{dN}$ to be

$$H(t,x_1,\dots,x_N)\triangleq(g_N(t,x_1),\dots,g_N(t,x_N))^T,$$

it follows by (i) and (ii) that $F+H$ satisfies the hypotheses of Theorems B.1 and B.3, which yield the existence of a unique Carathéodory solution $(x_{1,N},\dots,x_{N,N}):[0,T]\to\mathbb{R}^{dN}$ of system (4.5) for every $N\in\mathbb{N}$.

Now suppose there exists a bounded set $B\subset\mathbb{R}^d$ such that $x^0_{i,N}\in B$ for every $N\in\mathbb{N}$ and every $i=1,\dots,N$. Let us now fix $N\in\mathbb{N}$ and estimate the growth of $\|x_i^N(t)\|^2$ for $i=1,\dots,N$. From Lemma 4.8 and the hypotheses on K and g_N it follows

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\|x_{i,N}(t)\|^2 &= \dot{x}_{i,N}(t)\cdot x_{i,N}(t)\\ &= ((K*\mu_N(t))(x_i(t))+g_N(t,x_{i,N}(t)))\cdot x_{i,N}(t)\\ &\leq \|(K*\mu_N(t))(x_{i,N}(t))\|\|x_{i,N}(t)\|+g_N(t,x_{i,N}(t))\cdot x_{i,N}(t)\\ &\leq C_K\left(1+\|x_{i,N}(t)\|+\frac{1}{N}\sum_{j=1}^{N}\|x_{j,N}(t)\|\right)\|x_{i,N}(t)\|+G\|x_{i,N}(t)\|^2+G\\ &\leq C_1\max_{1\leq j\leq N}\|x_{j,N}(t)\|^2+C_2,\end{aligned}\tag{4.14}$$

with $C_1\triangleq C_K/2+G+2$ and $C_2\triangleq C_K/2+G$. We now denote by

$$\zeta(t)\triangleq\max_{1\leq j\leq N}\|x_{j,N}(t)\|^2\quad\text{for every } t\in[0,T].$$

Notice that, since $x_{i,N}(\cdot) \in W^{1,\infty}([0,T];\mathbb{R}^d)$, then $\zeta(\cdot) \in W^{1,\infty}([0,T];\mathbb{R})$. This implies that $\zeta(\cdot)$ is a.e. differentiable, hence by Stampacchia's Lemma (see for instance [134, Chapter 2, Lemma A.1]) for a.e. $t \in [0,T]$ there exists $j \in \{1,\dots,N\}$ such that

$$\dot{\zeta}(t) = \frac{d}{dt}\|x_{j,N}(t)\|^2,$$

which let us conclude the inequality

$$\dot{\zeta}(t) \leq 2C_1\zeta(t) + 2C_2.$$

Since $x^0_{i,N} \in B$, by setting

$$\delta(B) \triangleq \sup\left\{\|x\| : x \in B\right\},$$

from Gronwall's Lemma A.1 we obtain

$$\zeta(t) \leq (\zeta(0) + 2C_2 t)e^{2C_1 t} \leq (\delta(B) + 2C_2 t)e^{2C_1 t},$$

Therefore, $\|x_{j,N}(t)\| \leq R$ where

$$R \triangleq \sqrt{\delta(B) + 2C_2 T}\, e^{C_1 T}. \tag{4.15}$$

This implies that, for all $N \in \mathbb{N}$ and $i = 1,\dots,N$, from hypothesis (iii) we have

$$\begin{aligned}\|\dot{x}_{i,N}(t)\| &\leq \frac{1}{N}\sum_{j=1}^{N}\left\|K\big(t, x_{i,N}(t) - x_{j,N}(t)\big)\right\| + \|g_N(t, x_{i,N}(t))\| \\ &\leq C_K\left(1 + \|x_{i,N}(t)\| + \frac{1}{N}\sum_{j=1}^{N}\|x_{j,N}(t)\|\right) + \|g_N\|_{L^\infty([0,T]\times B(0,R))} \\ &\leq (1+2R)C_K + G_{B(0,R)},\end{aligned}$$

which, integrating between s and t implies

$$\|x_{i,N}(t) - x_{i,N}(s)\| \leq \Big((1+2R)C_K + G_{B(0,R)}\Big)|t-s|.$$

This gives us the uniform continuity of the curve $x_{i,N}(\cdot)$ with an upper bound for the Lipschitz constant given by $L \triangleq (1+2R)C_K + G_{B(0,R)}$, which is uniform in i and N. □

4.2.2 The mean-field limit

We now establish the existence of solutions of system (4.10). Informally, to do so we shall consider the solutions of the discrete convolution-type ODE systems (4.5), write them in the form of empirical measures μ_N, and then take the limit for $N \to \infty$ in the Wasserstein space of probability measures $\mathcal{P}_1(\mathbb{R}^d)$. This procedure, also known as *mean-field limit*, will let us extend the results obtained in Proposition 4.9 to solutions to (4.10).

We first need a preliminary estimate, variants of which are [106, Lemma 6.7] and [46, Lemma 4.7]

Lemma 4.10. *Let* $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$ *and let* $\mu : [0,T] \to \mathcal{P}_c(\mathbb{R}^d)$ *and* $\nu : [0,T] \to \mathcal{P}_c(\mathbb{R}^d)$ *be two continuous maps with respect to* $\mathcal{W}_1$ *satisfying*

$$\mathrm{supp}(\mu(t)) \cup \mathrm{supp}(\nu(t)) \subseteq B(0, \widehat{R}), \tag{4.16}$$

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

$$\|K * \mu(t) - K * \nu(t)\|_{L^\infty(B(0,r))} \leq L_{K,\widehat{R},r} \mathcal{W}_1(\mu(t), \nu(t)) \tag{4.17}$$

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

Proof. Fix $t \in [0,T]$ and take $\pi \in \Gamma_o(\mu(t), \nu(t))$. Since the marginals of π are by definition $\mu(t)$ and $\nu(t)$, it follows

$$\begin{aligned} K * \mu(t)(x) - K * \nu(t)(x) &= \int_{B(0,\widehat{R})} K(x-y) d\mu(t)(y) - \int_{B(0,\widehat{R})} K(x-z) d\nu(t)(z) \\ &= \int_{B(0,\widehat{R})^2} (K(x-y) - K(x-z)) \, d\pi(y,z) \end{aligned}$$

By hypothesis (4.16), $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$, and the definition of the Wasserstein distance $\mathcal{W}_1$ it follows

$$\begin{aligned} \|K * \mu(t) - K * \nu(t)\|_{L^\infty(B(0,r))} &\leq \operatorname*{ess\,sup}_{x \in B(0,r)} \int_{B(0,\widehat{R})^2} |K(x-y) - K(x-z)| \, d\pi(y,z) \\ &\leq \mathrm{Lip}_{B(0,\widehat{R}+r)}(K) \int_{B(0,\widehat{R})^2} |y-z| d\pi(y,z) \\ &= \mathrm{Lip}_{B(0,\widehat{R}+r)}(K) \mathcal{W}_1(\mu(t), \nu(t)), \end{aligned}$$

hence (4.17) holds by setting $L_{K,\widehat{R},r} \triangleq \mathrm{Lip}_{B(0,\widehat{R}+r)}(K)$. □

We are now ready to state under which conditions the approximation for N large of the discrete systems (4.5) by the mean-field dynamics (4.10) holds.

Theorem 4.11. *Let* $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$ *be given. Let* $(\mu_N^0)_{N \in \mathbb{N}} \subset \mathcal{P}_c(\mathbb{R}^d)$ *be a sequence of empirical measures of the form*

$$\mu_N^0 = \frac{1}{N} \sum_{i=1}^N \delta_{x_{i,N}^0}, \qquad \textit{for some } x_{i,N}^0 \in \mathrm{supp}(\mu^0)$$

satisfying the condition

$$\lim_{N \to \infty} \mathcal{W}_1(\mu_N^0, \mu^0) = 0,$$

and assume that the sequence $(g_N)_{N \in \mathbb{N}} \subset \mathcal{C}([0,T] \times \mathbb{R}^d; \mathbb{R}^d)$ *satisfying (i)–(iii) of Proposition 4.9 converges uniformly to* $g \in \mathcal{C}([0,T] \times \mathbb{R}^d; \mathbb{R}^d)$*. For every* $N \in \mathbb{N}$*, let* $\mu_N : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ *be the curve (4.6) where* $(x_{1,N}(\cdot), \ldots, x_{N,N}(\cdot))$ *is the unique solution of system (4.5) with initial datum* $(x_{1,N}^0, \ldots, x_{N,N}^0)$.

Then, there exists $R > 0$ depending only on T, K, G, and μ^0 such that the sequence $(\mu_N)_{N\in\mathbb{N}}$ converges uniformly in $\mathcal{P}_1(B(0,R))$ equipped with the Wasserstein metric $\mathcal{W}_1$ to a solution $\mu(\cdot)$ of (4.10) with initial datum μ^0 satisfying

$$\mathrm{supp}(\mu_N(t)) \cup \mathrm{supp}(\mu(t)) \subseteq B(0,R), \qquad \textit{for every } N \in \mathbb{N} \textit{ and } t \in [0,T].$$

Proof. First of all notice that, for every $N \in \mathbb{N}$, the unique solution $(x_{1,N}(\cdot), \dots, x_{N,N}(\cdot))$ of system (4.5) with initial datum $(x^0_{1,N}, \dots, x^0_{N,N})$ exists by Proposition 4.9.
Again by Proposition 4.9, the elements of the sequence $(\mu_N)_{N\in\mathbb{N}} \subset \mathcal{C}([0,T];\mathcal{P}_1(B(0,R)))$ have support uniformly contained in the ball $B(0,R)$, where R is given by (4.15), and they are Lipschitz continuous with uniform bound in N for a Lipschitz constant given by $L > 0$. We have thus found a sequence $(\mu_N)_{N\in\mathbb{N}} \subset \mathcal{C}([0,T],\mathcal{P}_1(B(0,R)))$ for which the following holds:

- $(\mu_N)_{N\in\mathbb{N}}$ is equicontinuous and is contained in a closed subset of $\mathcal{P}_1(B(0,R))$ equipped with the $\mathcal{W}_1$ metric, because of the uniform Lipschitz constant bound L;
- for every $t \in [0,T]$, the sequence $(\mu_N(t))_{N\in\mathbb{N}}$ is relatively compact in $\mathcal{P}_1(B(0,R))$ equipped with the $\mathcal{W}_1$ metric. This holds because $(\mu_N(t))_{N\in\mathbb{N}}$ is a tight sequence, since $B(0,R)$ is compact, and hence relatively compact with respect to weak convergence due to Prokhorov's Theorem. By [13, Proposition 7.1.5] and the uniform integrability of the first moments of the family $(\mu_N(t))_{N\in\mathbb{N}}$ follows the relative compactness also in the metric space $\mathcal{P}_1(B(0,R))$ equipped with $\mathcal{W}_1$.

Therefore, we can apply the Ascoli-Arzelá Theorem for functions with values in a metric space (see for instance, [133, Chapter 7, Theorem 18]) to infer the existence of a subsequence $(\mu_{N_k})_{k\in\mathbb{N}}$ of $(\mu_N)_{N\in\mathbb{N}}$ such that

$$\lim_{k\to\infty} \mathcal{W}_1(\mu_{N_k}(t), \mu(t)) = 0, \qquad \text{uniformly for a.e. } t \in [0,T], \tag{4.18}$$

for some $\mu \in \mathcal{C}([0,T],\mathcal{P}_1(B(0,R)))$ with Lipschitz constant bounded by L. The hypothesis $\lim_{N\to\infty} \mathcal{W}_1(\mu^0_N, \mu^0) = 0$ now obviously implies $\mu(0) = \mu^0$.
We are left with verifying that this curve $\mu(\cdot)$ is a solution of (4.10). For all $t \in [0,T]$, for all $N \in \mathbb{N}$ and for all $\phi \in \mathcal{C}^\infty_c(\mathbb{R}^d;\mathbb{R})$, it holds

$$\frac{d}{dt}\langle \phi, \mu_N(t)\rangle = \frac{1}{N}\frac{d}{dt}\sum_{i=1}^N \phi(x_{i,N}(t)) = \frac{1}{N}\sum_{i=1}^N \nabla\phi(x_{i,N}(t)) \cdot \dot{x}_{i,N}(t),$$

hence, by directly applying the substitution $\dot{x}_{i,N}(t) = (K * \mu_N(t))(x_{i,N}(t)) + g_N(t, x_{i,N}(t))$ from system (4.7), we have

$$\langle \phi, \mu_N(t) - \mu_N(0)\rangle = \int_0^t \left[\int_{\mathbb{R}^d} \nabla\phi(x) \cdot \Big((K * \mu_N(s))(x) + g_N(s,x)\Big) d\mu_N(s,x)\right] ds.$$

Since the uniform Wasserstein convergence (4.18) implies that for every $t \in [0,T]$ and $\phi \in \mathcal{C}^\infty_c(\mathbb{R}^d;\mathbb{R})$ it holds

$$\lim_{k\to\infty} \langle \phi, \mu_{N_k}(t) - \mu_{N_k}(0)\rangle = \langle \phi, \mu(t) - \mu(0)\rangle$$

then, by Remark 4.4, in order to show that $\mu(\cdot)$ is a solution to (4.10) we simply need to prove that

$$\lim_{k\to\infty}\int_0^t\left[\int_{\mathbb{R}^d}\nabla\phi(x)\cdot\Big((K*\mu_{N_k}(s))(x)+g_{N_k}(s,x)\Big)d\mu_{N_k}(s,x)\right]ds = \\ =\int_0^t\left[\int_{\mathbb{R}^d}\nabla\phi(x)\cdot\Big((K*\mu(s))(x)+g(s,x)\Big)d\mu(s,x)\right]ds.$$

To do so, notice that by Lemma 4.10 (with $r=\widehat{R}=R$) and the uniform $\mathcal{W}_1$ convergence of the μ_{N_k} to μ, it holds

$$\begin{aligned}\int_0^t\int_{\mathbb{R}^d}&\|((K*\mu_{N_k}(s))(x)-(K*\mu(s))(x))\cdot\nabla\phi(x)\|\,d\mu(s,x)ds\leq\\ &\leq L_{K,R,R}\int_0^t\mathcal{W}_1(\mu_{N_k}(s),\mu(s))\left[\int_{\mathbb{R}^d}\|\nabla\phi(x)\|d\mu(s,x)\right]ds\\ &\leq L_{K,R,R}\sup_{t\in[0,T]}\mathcal{W}_1(\mu_{N_k}(t),\mu(t))\int_0^t\left[\int_{\mathbb{R}^d}\|\nabla\phi(x)\|d\mu(s,x)\right]ds\\ &\to 0 \text{ as } k\to\infty,\end{aligned}$$

since $\nabla\phi$ is bounded and $\mu(\cdot)$ has uniform compact support in time. The same holds true for g_{N_k} and g in place of $K*\mu_{N_k}$ and $K*\mu$, respectively, due to the uniform convergence $g_N\to g$.
Therefore, from the dominated convergence theorem and the Lipschitz continuity of $K*\mu$ and g, we obtain

$$\langle\phi,\mu(t)-\mu(0)\rangle=\int_0^t\int_{\mathbb{R}^d}\nabla\phi(x)\cdot(K*\mu(s))(x)d\mu(s,x)ds=0,$$

which proves that $\mu(\cdot)$ is a solution of (4.10) with initial datum μ^0. The fact that the sequence $(\mu_N)_{N\in\mathbb{N}}$ converges to μ without passing to a subsequence shall follow from the uniqueness of μ, see Remark 4.18, which will be proven in the following section. □

Remark 4.12. A way to choose the initial data $x^0_{1,N},\ldots,x^0_{N,N}$ for every $N\in\mathbb{N}$ in order to guarantee the condition $\lim_{N\to\infty}\mathcal{W}_1(\mu^0_N,\mu^0)=0$, is the following: let $(x_i)_{i\in\mathbb{N}}$ be a sequence of independently and identically drawn vectors from $\mathrm{supp}(\mu^0)$ according to the probability law μ^0. If for every $N\in\mathbb{N}$ we define

$$x^0_{i,N}\overset{\Delta}{=}x_i,\qquad\text{for all } i=1,\ldots,N,$$

then the validity of $\lim_{N\to\infty}\mathcal{W}_1(\mu^0_N,\mu^0)=0$ follows from [102, Lemma 3.3].

4.3 The transport map and uniqueness of solutions

The *transport map* provides a direct way to describe the evolution in time of the initial measure μ^0 and is often employed in connection with stability estimates for the solutions of (4.10). It can be constructed by considering a characteristic curve of equation (4.10), which yields the system

$$\begin{cases} \dot{\xi}(t) = (K * \mu(t))(\xi(t)) + g(t, \xi(t)) & \text{for } t \in (0, T], \\ \xi(0) = \xi^0, \end{cases} \tag{4.19}$$

where ξ is a mapping from $[0, T]$ to $\mathbb{R}^d$. We assume here that $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ is sublinear and $g \in \mathcal{C}([0, T] \times \mathbb{R}^d; \mathbb{R}^d)$ satisfies (4.12)–(4.13). The map $\mu : [0, T] \to \mathcal{P}_1(\mathbb{R}^d)$ is continuous with respect to the Wasserstein distance $\mathcal{W}_1$ and satisfies $\mu(0) = \mu^0$ as well as $\mathrm{supp}(\mu(t)) \subseteq B(0, \overline{R})$ for every $t \in [0, T]$, for a given $\overline{R} > 0$.
The following preliminary result shows that system (4.19) has a unique solution $\xi(\cdot)$ defined on the whole time interval $[0, T]$.

Proposition 4.13. *Suppose that $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ is sublinear and that $g \in \mathcal{C}([0, T] \times \mathbb{R}^d; \mathbb{R}^d)$ satisfies (4.12)–(4.13). Fix $T > 0$, $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$, $\xi^0 \in \mathbb{R}^d$ and $\overline{R} > 0$. Then, for every map $\mu : [0, T] \to \mathcal{P}_1(\mathbb{R}^d)$ which is continuous with respect to $\mathcal{W}_1$ and satisfies both $\mu(0) = \mu^0$ and*

$$\mathrm{supp}(\mu(t)) \subseteq B(0, \overline{R}) \quad \text{for every } t \in [0, T],$$

there exists a unique solution of system (4.19) with initial value ξ^0 defined on the whole interval $[0, T]$.

Proof. By Lemma 4.8 follows that, for any compact set $\Omega \subset \mathbb{R}^d$ containing ξ^0, there exists a function $m_\Omega \in L^1([0, T])$ for which the function $G(t, y) = (K * \mu(t))(y) + g(t, y)$ satisfies (B.2) of Theorem B.1. Moreover, for fixed t this function is locally Lipschitz continuous in y, as follows from Lemma 4.7, thus G is a Carathéodory function.
From the hypothesis that $\mathrm{supp}(\mu(t))$ is contained in $B(0, \overline{R})$ uniformly in $t \in [0, T]$ and Lemma 4.8, follows the existence of a constant C depending on T, K and μ^0 such that

$$\|(K * \mu(t))(y)\| \leq C(1 + \|y\|)$$

holds for every $y \in \mathbb{R}^d$ and for every $t \in [0, T]$. Hence the function $G(t, y) = (K * \mu(t))(y) + g(t, y)$ satisfies (B.3) (see also Remark B.2). By considering a sufficiently large compact set Ω containing ξ^0, Theorem B.1 then guarantees the existence of a solution of system (4.19) defined on $[0, T]$.
To establish uniqueness, notice that by Lemma 4.7, the function $K * \mu(t)$ is locally Lipschitz. Adding the local Lipschitz continuity of g, uniqueness follows from Theorem B.3 by taking $g_1 = g_2 = G$, $y_1^0 = y_2^0 = \xi^0$ and $r = \|\xi^0\|$. □

Fix now $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ sublinear and $g \in \mathcal{C}([0, T] \times \mathbb{R}^d; \mathbb{R}^d)$ satisfying (4.12)–(4.13), and consider a solution $\mu(\cdot)$ of (4.10) with K and g (which exists by Theorem 4.11). Since $\mu(\cdot)$ satisfies the hypotheses of the proposition above, we can consider the family of flow maps $\mathcal{T}_t^\mu : \mathbb{R}^d \to \mathbb{R}^d$, indexed by $t \in [0, T]$, defined by

$$\mathcal{T}_t^\mu(\xi^0) \triangleq \xi(t),$$

where $\xi : [0, T] \to \mathbb{R}^d$ is the unique solution of (4.19) with initial datum ξ^0. It is well-known, see [46, Theorem 3.10], that a map $\mu : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ is a solution of (4.10) with initial datum μ^0 if and only if it is the unique fixed point of the *push-foward map* $\Gamma[\nu](t) \triangleq (\mathcal{T}_t^\nu)_\# \mu^0$, i.e., it holds

$$\mu(t) = (\mathcal{T}_t^\mu)_\# \mu^0 \quad \text{for every } t \in [0,T]. \tag{4.20}$$

A relevant basic property of the transport map is proved in the following

Proposition 4.14. *If $K \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ is sublinear and $g \in \mathcal{C}([0,T] \times \mathbb{R}^d;\mathbb{R}^d)$ satisfies* (4.12)–(4.13)*, then $\mathcal{T}_t^\mu$ is a locally Lipschitz map. Moreover, if $g \equiv 0$, then $\mathcal{T}_t^\mu$ is a locally bi-Lipschitz map, i.e., it is invertible, locally Lipschitz and with locally Lipschitz inverse.*

Proof. Let $\overline{R} > 0$ be sufficiently large such that $\mathrm{supp}(\mu^0) \subseteq B(0,\overline{R})$. Then, the choice $r = \overline{R}$ in Theorem B.3 trivially implies the following stability estimate

$$\left\|\mathcal{T}_t^\mu(x) - \mathcal{T}_t^\mu(y)\right\| \le e^{T\left(\mathrm{Lip}_{B(0,\overline{R})}(K)+\mathrm{Lip}_{[0,T]\times B(0,\overline{R})}(g)\right)}\|x-y\|, \quad \text{for every } x,y \in B(0,\overline{R}).$$

i.e., $\mathcal{T}_t^\mu$ is locally Lipschitz with

$$\mathrm{Lip}_{B(0,\overline{R})}(\mathcal{T}_t^\mu) \le e^{T\left(\mathrm{Lip}_{B(0,\overline{R})}(K)+\mathrm{Lip}_{[0,T]\times B(0,\overline{R})}(g)\right)}. \tag{4.21}$$

Suppose now that $g \equiv 0$. In view of the uniqueness of the solutions of the ODE (4.19), it is also clear that, for any $t_0 \in [0,T]$, the inverse of $\mathcal{T}_{t_0}^\mu$ is given by the transport map associated to the backward-in-time ODE

$$\begin{cases} \dot{\xi}(t) = (K * \mu(t))(\xi(t)) & \text{for } t \in [0,t_0), \\ \xi(t_0) = \xi^0. \end{cases}$$

However, this problem in turn can be cast into the form of an usual IVP simply by considering the reverse trajectory $\nu(t) = \mu(t_0 - t)$. Then $y(t) = \xi(t_0 - t)$ solves

$$\begin{cases} \dot{y}(t) = -\big(K * \nu(t)\big)(y(t)) & \text{for } t \in (0,t_0], \\ y(0) = \xi(t_0). \end{cases}$$

The corresponding stability estimate for this problem then yields that the inverse of $\mathcal{T}_t^\mu$ exists and is locally Lipschitz (with the same estimate for the local Lipschitz constant as (4.21)). □

Thanks to the transport map we can prove Theorem 4.16 below, stating the continuous dependance on initial data for system (4.10), which in turn yields the uniqueness of solutions. The following lemma and inequality (B.6) are the main ingredients of its proof.

Lemma 4.15. *Let $\mathcal{T}_1$ and $\mathcal{T}_2 \colon \mathbb{R}^n \to \mathbb{R}^n$ be two bounded Borel measurable functions. Then, for every $\mu \in \mathcal{P}_1(\mathbb{R}^n)$ one has*

$$\mathcal{W}_1((\mathcal{T}_1)_\#\mu, (\mathcal{T}_2)_\#\mu) \le \|\mathcal{T}_1 - \mathcal{T}_2\|_{L^\infty(\mathrm{supp}(\mu))}.$$

If in addition $\mathcal{T}_1$ is locally Lipschitz continuous, and μ, $\nu \in \mathcal{P}_1(\mathbb{R}^n)$ are both compactly supported on a ball $B(0,r)$ of $\mathbb{R}^n$, then

$$\mathcal{W}_1((\mathcal{T}_1)_{\#}\mu, (\mathcal{T}_1)_{\#}\nu) \le \mathrm{Lip}_{B(0,r)}(\mathcal{T}_1)\mathcal{W}_1(\mu,\nu).$$

Proof. See [46, Lemma 3.11] and [46, Lemma 3.13]. □

Theorem 4.16. *Fix $T > 0$, $K_1, K_2 \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ sublinear and $g_1, g_2 \in \mathcal{C}([0,T] \times \mathbb{R}^d;\mathbb{R}^d)$ satisfying (4.12)–(4.13). Let $\mu_1, \mu_2 : [0,T] \to \mathcal{P}_1(\mathbb{R}^d)$ be solutions of*

$$\begin{cases} \dfrac{\partial \mu_i}{\partial t}(t) = -\nabla_x \cdot \Big((K_i * \mu_i(t) + g_i(t))\mu_i(t)\Big) & \text{for } t \in (0,T], \\ \mu_i(0) = \mu_i^0, \end{cases} \qquad i = 1,2,$$

equi-compactly supported in a ball $B(0,\overline{R})$, for some $\overline{R} > 0$. Then the following estimate holds for every $t \in [0,T]$

$$\begin{aligned}\mathcal{W}_1(\mu_1(t),\mu_2(t)) \le \alpha\mathcal{W}_1(\mu_1^0,\mu_2^0) + \beta\int_0^t \Big(& L_{K_1,\overline{R},\overline{R}}\mathcal{W}_1(\mu_1(s),\mu_2(s)) + \\ &+ \|K_1 - K_2\|_{L^\infty(B(0,2\overline{R}))} + \|g_1(s) - g_2(s)\|_{L^\infty(B(0,\overline{R}))}\Big)\,ds,\end{aligned}$$

where the positive constants α and β depend only on μ_1^0, μ_2^0, K_1, K_2, g_1, g_2, $\overline{R}$ and T. Hence, if $K_1 = K_2$ and $g_1 = g_2$ it holds

$$\mathcal{W}_1(\mu_1(t),\mu_2(t)) \le \overline{C}\mathcal{W}_1(\mu_1^0,\mu_2^0) \quad \text{for every } t \in [0,T]. \tag{4.22}$$

In particular, solutions of (4.10) are uniquely determined by the initial datum.

Remark 4.17. The estimate (4.22) is known as *Dobrushin's inequality* and was first presented in [95].

Proof. Let $\mathcal{T}_t^1$ and $\mathcal{T}_t^2$ be the flow maps associated to system (4.19) with measure μ_1 and μ_2, respectively. By (4.20), the triangle inequality, Lemma 4.15 and (4.21) we have for every $t \in [0,T]$

$$\begin{aligned}\mathcal{W}_1(\mu_1(t),\mu_2(t)) &= \mathcal{W}_1((\mathcal{T}_t^1)_{\#}\mu_1^0, (\mathcal{T}_t^2)_{\#}\mu_2^0) \\ &\le \mathcal{W}_1((\mathcal{T}_t^1)_{\#}\mu_1^0, (\mathcal{T}_t^1)_{\#}\mu_2^0) + \mathcal{W}_1((\mathcal{T}_t^1)_{\#}\mu_2^0, (\mathcal{T}_t^2)_{\#}\mu_2^0) \\ &\le \mathrm{Lip}_{B(0,\overline{R})}(\mathcal{T}_t^1)\mathcal{W}_1(\mu_1^0,\mu_2^0) + \|\mathcal{T}_t^1 - \mathcal{T}_t^2\|_{L^\infty(B(0,\overline{R}))} \\ &\le \alpha\mathcal{W}_1(\mu_1^0,\mu_2^0) + \|\mathcal{T}_t^1 - \mathcal{T}_t^2\|_{L^\infty(B(0,\overline{R}))},\end{aligned} \tag{4.23}$$

where $\alpha = e^{T\left(\mathrm{Lip}_{B(0,\overline{R})}(K_1) + \mathrm{Lip}_{[0,T]\times B(0,\overline{R})}(g_1)\right)}$.
Using inequality (B.6) with $y_1^0 = y_2^0 = \xi^0$ we get

$$\|\mathcal{T}_t^\mu - \mathcal{T}_t^\nu\|_{L^\infty(B(0,\overline{R}))} \le \beta\int_0^t \|K_1 * \mu_1(s) + g_1(s) - K_2 * \mu_2(s) - g_2(s)\|_{L^\infty(B(0,\overline{R}))}\,ds, \tag{4.24}$$

for $\beta = \max_{i=1,2} e^{T\left(\mathrm{Lip}_{B(0,\overline{R})}(K_i) + \mathrm{Lip}_{[0,T]\times B(0,\overline{R})}(g_i)\right)}$.

To estimate the integrand, notice that

$$\begin{aligned}\|K_1 * \mu_2(s) - K_2 * \mu_2(s)\|_{L^\infty(B(0,\overline{R}))} &\le \sup_{x\in B(0,\overline{R})} \int_{\mathbb{R}^d} \|K_1(x-y) - K_2(x-y)\| d\mu_2(s,y) \\ &\le \sup_{x\in B(0,\overline{R})} \sup_{y\in B(0,\overline{R})} \|K_1(x-y) - K_2(x-y)\| \\ &\le \|K_1 - K_2\|_{L^\infty(B(0,2\overline{R}))},\end{aligned}$$

therefore we have

$$\begin{aligned}\|K_1 * \mu_1(s) - K_2 * \mu_2(s)\|_{L^\infty(B(0,\overline{R}))} &\le \|K_1 * \mu_1(s) - K_1 * \mu_2(s)\|_{L^\infty(B(0,\overline{R}))} + \\ &\quad + \|K_1 * \mu_2(s) - K_2 * \mu_2(s)\|_{L^\infty(B(0,\overline{R}))} \\ &\le L_{K_1,\overline{R},\overline{R}} \mathcal{W}_1(\mu_1(s),\mu_2(s)) + \|K_1 - K_2\|_{L^\infty(B(0,2\overline{R}))},\end{aligned}$$

where we have used Lemma 4.10 with $K = K_1$ and $r = \widehat{R} = \overline{R}$. By plugging the above estimate in (4.24) we get

$$\begin{aligned}\|\mathcal{T}_t^\mu - \mathcal{T}_t^\nu\|_{L^\infty(B(0,\overline{R}))} \le \beta \int_0^t \Big(& L_{K_1,\overline{R},\overline{R}} \mathcal{W}_1(\mu_1(s),\mu_2(s)) + \|K_1 - K_2\|_{L^\infty(B(0,2\overline{R}))} + \\ & + \|g_1(s) - g_2(s)\|_{L^\infty(B(0,\overline{R}))} \Big) ds,\end{aligned}$$

Combining (4.23) and the above estimate we obtain

$$\begin{aligned}\mathcal{W}_1(\mu_1(t),\mu_2(t)) \le \alpha \mathcal{W}_1(\mu_1^0,\mu_2^0) + \beta \int_0^t \Big(& L_{K_1,\overline{R},\overline{R}} \mathcal{W}_1(\mu_1(s),\mu_2(s)) + \\ & + \|K_1 - K_2\|_{L^\infty(B(0,2\overline{R}))} + \|g_1(s) - g_2(s)\|_{L^\infty(B(0,\overline{R}))} \Big) ds,\end{aligned}$$

which is the desired inequality. In particular, if $K_1 = K_2$ and $g_1 = g_2$, from Gronwall's Lemma A.1 we get (4.22) with $\overline{C} = \alpha e^{T\beta L_{K_1,\overline{R},\overline{R}}}$.
Consider now two solutions $\mu_1(\cdot)$ and $\mu_2(\cdot)$ of (4.10) with the same initial datum μ^0. The above estimate tells us that they are the same curve in $\mathcal{P}_1(\mathbb{R}^d)$. This concludes the proof. □

Remark 4.18. Going back to the application of the Ascoli-Arzelá Theorem in Theorem 4.11, consider another converging subsequence of the sequence $(\mu_N)_{N\in\mathbb{N}}$. It straightforwardly follows that its limit is another solution of (4.10) with the same initial datum. However, since there is only one of such solutions for Theorem 4.16, we have that all converging subsequences of $(\mu_N)_{N\in\mathbb{N}}$ have the same limit, from which follows that the sequence $(\mu_N)_{N\in\mathbb{N}}$ has limit μ too.

4.4 The Boltzmann equation

This section is devoted to the Boltzmann equation, which is a partial integro-differential equation describing the evolution of the particle density of a rarefied monatomic gas derived from the mathematical model of elastic balls by means of mechanical and statistical considerations.

The starting point of the Boltzmann analysis is the same that motivates the mean-field limit procedure: to renounce to study the gas in terms of the detailed motion of the particles which constitutes it because of their huge number, and start to investigate a nonnegative function $f(t) \in \mathcal{M}_b(\mathbb{R}^{2d})$ which represents the distribution of particles in the phase space $\mathbb{R}^{2d}$ at time t.

Boltzmann considered the situation of a system of N identical particles, viewed as hard spheres with radius 0, in which collisions are purely elastic. Taken any two particles with state parameters (x, v) and $(\hat{x}, \hat{v})$, the *post-collisional coordinates* (x^*, v^*) and $(\hat{x}^*, \hat{v}^*)$ are given by the change of variables

$$\begin{cases} v^* = v - [(v - \hat{v}) \cdot n]n, \\ \hat{v}^* = \hat{v} + [(v - \hat{v}) \cdot n]n, \end{cases} \tag{4.25}$$

where n is a unitary vector in $S^2_+ = \{n : (v - v_1) \cdot n \geq 0\}$. The equations (4.25) are consequence of the conservation of energy, momentum, and angular momentum.

If particles would not collide, then the dynamics for the distribution $f(\cdot)$ of a single particle should follow the law

$$\frac{\partial f}{\partial t}(t) + v \cdot \nabla_x f(t) = 0,$$

where the term $v \cdot \nabla_x f(t, x, v)$ represents the effect of free particle transport. However, due to collisions, we have an equation like

$$\frac{\partial f}{\partial t}(t) + v \cdot \nabla_x f(t) = G(t) - L(t),$$

where the gain term $G(t)$ "counts" the particles that acquire the velocity v at time t due to the collisions of other particle, i.e., how much the quantity $f(t, x, v)\ dx\ dv$ increases. Similarly, the loss term $L(t)$ describes the loss of $f(t, x, v)\ dx\ dv$ due to all collisions in which particles with velocity v acquire a different velocity.

The Boltzmann equation is therefore given by

$$\frac{\partial f}{\partial t}(t) + v \cdot \nabla_x f(t) = Q(f, f)(t), \tag{4.26}$$

where the *collision operator* Q accounts for the bilateral loss and gain terms as follows

$$Q(f, f)(t, x, v) \triangleq \int_{\mathbb{R}^3} \int_{S^2_+} (v - \hat{v}) \cdot n \Big(f(t, x, v^*) f(t, x, \hat{v}^*) - f(t, x, v) f(t, x, \hat{v}) \Big)\ d\hat{v}\ dn.$$

Remark 4.19. In the mathematical literature concerning the Boltzmann equation (see, for instance, [151, Section 2.3]), when integrating against the distribution of agents $f(t)$ it is customary to write $f(t, x, v)\ dx\ dv$, instead of $df(t, x, v)$, even if $f(t)$ is not absolutely continuous with respect to the Lebesgue measure $\mathcal{L}^{2d}$ on $\mathbb{R}^{2d}$. This is because, in this case, one sees $f(t)$ as a *generalized function* rather than a measure. We shall also adhere to this tradition whenever dealing with Boltzmann-like systems.

One of the advantages of the Boltzmann approach is its flexibility: it is naturally prone to generalizations in contexts outside physics, like biology and social sciences, and has experienced great success in economics as a cornerstone of *econophysics* (see, for instance, the classic [141]).

Another advantage is provided by the availability of *binary interaction algorithms* to solve numerically the nonlocal PDE (4.26): these meshless Monte Carlo methods have a reduced computational cost with respect to mesh-based methods and guarantee high level of accuracy [7, 151]. We shall make use of them in Chapter 7.

4.4.1 The Boltzmann approach outside physics

Mathematically speaking, nothing prevents us to use a different change of variables from (4.25) and study the Boltzmann-like equation related to it. Actually, nobody even prescribes that interactions are only allowed within agents of the same population f. We can therefore consider two distributions $f, g \in \mathcal{M}_b(\mathbb{R}^{2d})$ (which can, of course, coincide), take an agent from f with state parameters $(x, v) \in \mathbb{R}^{2d}$ and another one from g with coordinates $(\hat{x}, \hat{v}) \in \mathbb{R}^{2d}$: we then prescribe their post-interaction states to satisfy the following general change of variables

$$\begin{cases} v^* = v + \eta\left[\alpha(x, v)\xi + \beta(x, v) + \gamma(x, v, \hat{x}, \hat{v})\right], \\ \hat{v}^* = \hat{v} + \eta\left[\alpha'(\hat{x}, \hat{v})\xi' + \beta'(x, v) + \gamma'(\hat{x}, \hat{v}, x, v)\right], \end{cases} \tag{4.27}$$

where η is the strength of interaction and ξ and ξ' are random variables whose entries are i.i.d. following a normal distribution with mean 0, variance ς^2 and taking values in a set $\mathcal{B} \subseteq \mathbb{R}^d$. The functions α and α' are stochastic self-propulsion terms (since they are acting on the random variables ξ and ξ'), while β and β' are deterministic drifts. The interaction kernels γ and γ' determine the reaction of the agents to the collision. Likely to (4.25), in what follows we assume that it holds

(Inv) system (4.27) constitutes an invertible change of variables from $(v, \hat{v})$ to $(v^*, \hat{v}^*)$.

Following Boltzmann's *ansatz* (see [58, 160]), we introduce the *Boltzmann-Povzner collisional operator* $Q(f, g)$ associated to (4.27) as follows

$$Q(f, g)(t, x, v) \triangleq \mathbb{E}\left[\int_{\mathbb{R}^{2d}} \left(\frac{1}{\mathfrak{J}} f(t, x_*, v_*) g(t, \hat{x}_*, \hat{v}_*) - f(t, x, v) g(t, \hat{x}, \hat{v})\right) d\hat{x}\, d\hat{v}\right]. \tag{4.28}$$

In the above expression

- the pair (x_*, v_*) is the pre-interaction states that generates (x, v) via (4.27), and $\mathfrak{J}$ is the Jacobian of the change of variables given by (4.27) (well-defined by (Inv));
- the expected value $\mathbb{E}$ is computed with respect to $\xi \in \mathcal{B}$.

The time evolution of $f(\cdot)$ is then given by the following Boltzmann-type equation

$$\frac{\partial f}{\partial t}(t) + v \cdot \nabla_x f(t) = \lambda Q(f, g)(t), \tag{4.29}$$

where λ stands for the interaction frequencies among the two populations f and g.

Notice that, since the interaction (4.27) do not correspond at this point to any physical phenomenon, the conservation of the mass of f and g, i.e., the quantities

$$\rho[f](t) \triangleq \int_{\mathbb{R}^{2d}} f(t,x,v)\; dx\; dv \quad \text{and} \quad \rho[g](t) \triangleq \int_{\mathbb{R}^{2d}} g(t,x,v)\; dx\; dv, \tag{4.30}$$

is not anymore guaranteed. Therefore, in specifying the definition of a solution of (4.29), we have to require that $\rho[f]$ remains constant.

Definition 4.20. Fix $T > 0$, $\delta > 0$, and let $g \in L^2([0,T]; \mathcal{M}_b(\mathbb{R}^{2d}))$ be such that $\rho[g](t) = \rho[g](0) < +\infty$ for every $t \in [0,T]$. By a *δ-weak solution* of the initial value problem for the equation (4.29) with initial datum $f^0 \in \mathcal{M}_b(\mathbb{R}^{2d})$ in the interval $[0,T]$, we mean any $f \in L^2([0,T]; \mathcal{M}_b(\mathbb{R}^{2d}))$ satisfying

(*i*) $f(0,x,v) = f^0(x,v)$ for every $(x,v) \in \mathbb{R}^{2d}$;

(*ii*) there exists $R > 0$ depending only on f^0, T, δ, g, α, β, and γ, such that $\operatorname{supp}(f(t)) \subseteq B(0,R)$ for every $t \in [0,T]$;

(*iii*) $\rho[f](t) = \rho[f](0)$ for every $t \in [0,T]$;

(*iv*) $f(\cdot)$ satisfies the weak form of the equation (4.29), i.e., for every $\varphi \in \mathcal{T}_\delta$ and $t \in (0,T]$ it holds

$$\frac{\partial}{\partial t} \langle f(t), \varphi \rangle + \langle f(t), v \cdot \nabla_x \varphi \rangle = \lambda \langle Q(f,g)(t), \varphi \rangle,$$

where

$$\langle Q(f,g)(t), \varphi \rangle = \mathbb{E}\left[\int_{\mathbb{R}^{4d}} (\varphi(x,v^*) - \varphi(x,v))\, f(t,x,v) g(t,\hat{x},\hat{v})\, dx\, dv\, d\hat{x}\, d\hat{v} \right], \tag{4.31}$$

and where v^* is given by (4.27).

Remark 4.21. In this thesis, we shall not be concerned about the existence or uniqueness of solutions in the sense of Definition 4.20 for system (4.29), but only about their mean-field approximation. General well-posedness results for Boltzmann-like equations can be found in [9, 26, 93].

4.4.2 The grazing interaction limit

The brand-new point of view adopted with the Boltzmann approach raises several issues concerning the phenomena it tries to model.

(*i*) The Boltzmann equation is not a mean-field equation, since it still retains the microscopic interaction rules between agents given by (4.27). Is it possible to pass to a mean-field approximation by means of some procedure?

(*ii*) The analysis of equation (4.29) is in general quite difficult due to the highly convoluted operator Q. Is there a way to obtain a more regular operator, from which it is possible to determine explicitly the evolution of $f(\cdot)$ and its asymptotic behavior?

(*iii*) How is the binary interaction model related to the usual multiagent system

$$\begin{cases} \dot{x}(t) = v(t), \\ \dot{v}(t) = \alpha(x(t), v(t))\zeta + \beta(x(t), v(t)) + \displaystyle\int_{\mathbb{R}^{2d}} \gamma(x(t), v(t), \hat{x}, \hat{v}) g(t, \hat{x}, \hat{v}) \, d\hat{x} \, d\hat{v}, \end{cases} \tag{4.32}$$

where ζ is a d-dimensional random variable with normal distribution $\mathcal{N}(0, \sigma^2)$? To see that (4.32) is a generalization of already encountered multiagent models, consider

$$f(t) = g(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{(x_i(t), v_i(t))}.$$

where, for every $i = 1, \ldots, N$, $(x_i(\cdot), v_i(\cdot))$ solves

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \displaystyle\frac{1}{N} \sum_{j=1}^{N} a(\|x_i(t) - x_j(t)\|)(v_j(t) - v_i(t)). \end{cases}$$

i.e., $(x(\cdot), v(\cdot))$ is a solution of the Cucker-Smale system (1.6). Then $(x_i(\cdot), v_i(\cdot))$ also solves (4.32) for the choices $\gamma(x, v, \hat{x}, \hat{v}) = a(\|x - \hat{x}\|)(\hat{v} - v)$ and $\alpha = \beta \equiv 0$.

To answer the above questions, let us have a closer look to (4.32). If we consider the measure

$$f(t) = \sum_{i=1}^{N} \delta_{(x_i(t), v_i(t))}$$

where $(x_i(\cdot), v_i(\cdot))$ is a solution of (4.32) for every $i = 1, \ldots, N$, then it can be shown, by direct substitution and using Itô's Lemma, that $f(\cdot)$ solves the following *Fokker-Planck*-type equation [165]

$$\frac{\partial f}{\partial t}(t) + v \cdot \nabla_x f(t) = -\nabla_v \cdot \Big(\mathcal{H}[g](t) f(t) \Big) + \frac{1}{2} \sigma^2 \Delta_v (\alpha^2 f(t)), \tag{4.33}$$

where we have set

$$\mathcal{H}[g](t, x, v) \triangleq \beta(x, v) + \int_{\mathbb{R}^{2d}} \gamma(x, v, \hat{x}, \hat{v}) g(t, \hat{x}, \hat{v}) \, d\hat{x} \, d\hat{v}. \tag{4.34}$$

The integro-differential operator

$$FP(f, g)(t, x, v) \triangleq -\nabla_v \cdot \Big(\mathcal{H}[g](t, x, v) f(t, x, v) \Big) + \frac{1}{2} \sigma^2 \Delta_v (\alpha^2(x, v) f(t, x, v))$$

is called the *Fokker-Planck operator* and consists of a Vlasov-type term describing the interaction between f and g plus a diffusion term which accounts for the random walk of each particle. Equation (4.33) is a mean-field type equation which, as shown, e.g., in [8, 169], can be easily manipulated in order to obtain long time stationary profiles of the dynamics. The three questions above can then be roughly reformulated into a single one:

is there a way to play with λ, η and ς of (4.27) and (4.29) in such a way that, denoting by $Q_{\lambda,\eta,\varsigma}(f^{\lambda,\eta,\varsigma}, g)$ the collisional operator associated to the choice of the parameters, the limit

$$\lambda\langle Q_{\lambda,\eta,\varsigma}(f^{\lambda,\eta,\varsigma}, g)(t), \varphi\rangle \to \langle FP(f, g)(t), \varphi\rangle$$

holds for all *nice enough* test functions φ? The regime under which this limit holds is called *grazing interaction limit.* This technique, analogous to the so-called "grazing collision limit" in plasma physics, has been thoroughly studied in [91, 175] and allows, as pointed out in [151], to obtain a mean-field approximation for the binary Boltzmann description introduced in the previous section.

The mean-field approximation via grazing interaction limit of the Boltzmann-like equation (4.29) holds under a proper scaling of the quantities λ, η and ς that resembles the mean-field scaling $1/N$. Indeed, we have to tune the frequency λ and the strength η of interactions in such a way that, as the number of interactions increases, their intensity diminish proportionally. Hence, in both cases the overall force acting on a single agent turns out to be of order 1.

Formally, we assume that the interaction strength η scales according to a parameter ε, the interaction frequency λ scales as $1/\varepsilon$, and we let $\varepsilon \to 0$. In order to avoid losing the diffusion term in the limit, we also scale the variance of the noise term ς^2 as $1/\varepsilon$. More precisely, we set

$$\eta \overset{\Delta}{=} \varepsilon, \qquad \lambda \overset{\Delta}{=} \frac{1}{\varepsilon}, \qquad \varsigma^2 \overset{\Delta}{=} \frac{\sigma^2}{\varepsilon}. \tag{4.35}$$

We now pass to study the Boltzmann equation (4.29) to see if simplifications may occur under the above scaling assumptions.

Let us fix $T > 0$, $\delta > 0$, $\varepsilon > 0$, an initial datum $f^0 \in \mathcal{M}_b(\mathbb{R}^{2d})$, and a measure-valued curve $g \in L^2([0,T]; \mathcal{M}_b(\mathbb{R}^{2d}))$ such that $\rho[g](t) = \rho[g](0) < +\infty$ for every $t \in [0,T]$. Consider a δ-weak solution $f(\cdot) = f^\varepsilon(\cdot)$ of equation (4.29) with initial datum f^0 and let the scaling (4.35) holds for the chosen ε.

Remark 4.22. Fix $t \in [0,T]$. Since, from (*ii*)–(*iii*) of Definition 4.20, the distributions $\{f^\varepsilon(t)\}_{\varepsilon\geq 0}$ have equal constant mass and uniformly bounded support, it follows that this class is tight. Therefore, by virtue of Prokhorov Theorem, any sequence $\{f^{\varepsilon_n}(t)\}_{n\in\mathbb{N}}$ admits a subsequence converging weakly to some limit distribution $f(t)$.

Following the ideas in [54, 70], if we denote by $Q(f,g) \overset{\Delta}{=} Q^\varepsilon(f^\varepsilon, g)$ the collisional operator (4.28) associated to the scaling (4.35) for the chosen ε, then we can expand $\varphi(x, v^*)$ inside (4.31) in Taylor' series of $v^* - v$ up to the second order, to get

$$\langle Q(f,g)(t), \varphi\rangle = \mathbb{E}\Bigg[\int_{\mathbb{R}^{4d}} \nabla_v\varphi(x,v)\cdot(v^*-v)\, f(t,x,v)g(t,\hat{x},\hat{v})\; dx\; dv\; d\hat{x}\; d\hat{v} +$$

$$+ \frac{1}{2}\int_{\mathbb{R}^{4d}} \left(\sum_{i,j=1}^{d} \partial_v^{(i,j)}\varphi(x,v)\,(v^*-v)_i\,(v^*-v)_j\right) f(t,x,v)g(t,\hat{x},\hat{v})\; dx\; dv\; d\hat{x}\; d\hat{v}\Bigg] + R_\varphi,$$

where the remainder R_φ of the Taylor expansion has the form

$$R_\varphi = \mathbb{E}\Bigg[\frac{1}{2}\int_{\mathbb{R}^{4d}}\left(\sum_{i,j=1}^{d}\Big(\partial_v^{(i,j)}\varphi(x,v) - \partial_v^{(i,j)}\varphi(x,\overline{v})\Big)\,(v^*-v)_i\,(v^*-v)_j\right)\times$$
$$\times\, f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v}\Bigg],$$

with $\overline{v} = \vartheta v^* + (1-\vartheta)v$ for some $\vartheta \in [0,1]$. By using the substitution given by the interaction rule (4.27), i.e.

$$v^* - v = \eta\left[\alpha(x,v)\xi + \beta(x,v) + \gamma(x,v,\hat{x},\hat{v})\right],$$

we obtain

$$\int_{\mathbb{R}^{4d}}\nabla_v\varphi(x,v)\cdot(v^*-v)\,f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v} =$$
$$= \eta\int_{\mathbb{R}^{4d}}\nabla_v\varphi(x,v)\cdot\left[\alpha(x,v)\xi + \beta(x,v) + \gamma(x,v,\hat{x},\hat{v})\right]f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v}$$
$$= \eta\rho[g](t)\left\langle f(t), \nabla_v\varphi\cdot\alpha\xi\right\rangle + \eta\left\langle f(t), \nabla_v\varphi\cdot\mathcal{H}[g](t)\right\rangle, \tag{4.36}$$

where $\mathcal{H}[g]$ is as in (4.34). By taking the expected value with respect to $\xi \in \mathcal{B}$ of (4.36), the term $\eta\rho[g](t)\left\langle f(t), \nabla_v\varphi\cdot\alpha\xi\right\rangle$ is canceled out, since $\mathbb{E}\left[\xi\right] = 0$.
The same substitution yields for the second order term

$$\frac{1}{2}\int_{\mathbb{R}^{4d}}\left(\sum_{i,j=1}^{d}\partial_v^{(i,j)}\varphi(x,v)\,(v^*-v)_i\,(v^*-v)_j\right)f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v} =$$
$$= \frac{1}{2}\eta^2\int_{\mathbb{R}^{4d}}\Bigg[\sum_{i,j=1}^{d}\partial_v^{(i,j)}\varphi(x,v)\Big(\alpha(x,v)\xi + \beta(x,v) + \gamma(x,v,\hat{x},\hat{v})\Big)_i\times$$
$$\times\Big(\alpha(x,v)\xi + \beta(x,v) + \gamma(x,v,\hat{x},\hat{v})\Big)_j\Bigg]f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v}.$$

Notice that if we also compute the expected value of the expression above, all the cross terms $\xi_i\xi_j$ vanish, since they are drawn independently from each other. Hence we obtain

$$\lambda\left\langle Q(f,g)(t),\varphi\right\rangle = \lambda\eta\left\langle f(t),\nabla_v\varphi\cdot\mathcal{H}[g](t)\right\rangle +$$
$$+\frac{1}{2}\lambda\eta^2\varsigma^2\left\langle f(t),\Delta_v(\alpha^2\varphi)\right\rangle + \frac{1}{2}\lambda\eta^2\Xi_\varphi[f,g] + \lambda R_\varphi,$$

where the notation $\Xi[f,g]$ stands for

$$\Xi_\varphi[f,g] \triangleq \int_{\mathbb{R}^{4d}}\left(\sum_{i,j=1}^{d}\partial_v^{(i,j)}\varphi(x,v)\Big(\beta(x,v)+\gamma(x,v,\hat{x},\hat{v})\Big)_i\Big(\beta(x,v)+\gamma(x,v,\hat{x},\hat{v})\Big)_j\right)\times$$
$$\times\, f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v}.$$

If we now apply the rescaling rules (4.35), the following simplifications occur

$$\lambda\eta = 1, \quad \lambda\eta^2\varsigma^2 = \sigma^2, \quad \lambda\eta^2 = \varepsilon.$$

Hence, if we let $\varepsilon \to 0$ and we assume that, for every $\varphi \in \mathcal{T}_\delta$,

$$\lim_{\varepsilon\to 0}\left(\frac{\varepsilon}{2}\left\|\Xi_\varphi[f,g]\right\| + \frac{1}{\varepsilon}\left\|R_\varphi\right\|\right) = 0 \tag{4.37}$$

holds true, we obtain, up to subsequences, the following weak formulation of a Fokker-Planck-type equation for the dynamics of the limit distribution $f(\cdot)$

$$\frac{\partial}{\partial t}\langle f(t),\varphi\rangle + \langle f(t), v\cdot\nabla_x\varphi\rangle = \left\langle f(t), \nabla_v\varphi\cdot\mathcal{H}[g](t) + \frac{1}{2}\sigma^2\Delta_v(\alpha^2\varphi)\right\rangle. \tag{4.38}$$

Since φ has compact support, equation (4.38) can be recast in strong form by means of integration by parts. The resulting PDE is the Fokker-Planck-type equation (4.33).

The details of the proof of the limit (4.37) are given in the following section.

4.4.3 Estimates for the remainder terms

Motivated by the results in [169], we shall estimate the quantity $\|R_\varphi\|$ as follows: since $\overline{v} = \vartheta v^* + (1-\vartheta)v$ implies $\|v-\overline{v}\| \le \vartheta\,\|v-v^*\|$, then for every $\varphi \in \mathcal{T}_\delta$ it follows that

$$\left\|\partial_v^{(i,j)}\varphi(x,v) - \partial_v^{(i,j)}\varphi(x,\overline{v})\right\| \le M\left\|v-\overline{v}\right\|^\delta \le M\vartheta^\delta\left\|v-v^*\right\|^\delta,$$

for some $M > 0$. Hence we get

$$\begin{aligned}
\frac{1}{\varepsilon}\left\|R_\varphi\right\| &\le \frac{1}{2\varepsilon}M\vartheta^\delta\mathbb{E}\left[\int_{\mathbb{R}^{4d}}\left\|v^*-v\right\|^{2+\delta} f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v}\right] \\
&= \frac{1}{2\varepsilon}M\vartheta^\delta\eta^{2+\delta}\mathbb{E}\left[\int_{\mathbb{R}^{4d}}\left\|\alpha(x,v)\xi + \beta(x,v) + \gamma(x,v,\hat{x},\hat{v})\right\|^{2+\delta}\times\right. \\
&\qquad\qquad\left.\times f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v}\right] \\
&= \frac{1}{2}M\vartheta^\delta\varepsilon^{1+\delta}\mathbb{E}\left[\int_{\mathbb{R}^{4d}}\left\|\alpha(x,v)\xi + \beta(x,v) + \gamma(x,v,\hat{x},\hat{v})\right\|^{2+\delta}\times\right. \\
&\qquad\qquad\left.\times f(t,x,v)g(t,\hat{x},\hat{v})\;dx\;dv\;d\hat{x}\;d\hat{v}\right].
\end{aligned}$$

From the inequality

$$\begin{aligned}
&\left\|\alpha(x,v)\xi + \beta(x,v) + \gamma(x,v,\hat{x},\hat{v})\right\|^{2+\delta} \le \\
&\qquad\le 2^{2+2\delta}\left(\left\|\alpha(x,v)\xi\right\|^{2+\delta} + \left\|\beta(x,v)\right\|^{2+\delta} + \left\|\gamma(x,v,\hat{x},\hat{v})\right\|^{2+\delta}\right),
\end{aligned}$$

we obtain

$$\|R_\varphi\| \leq 2^{1+2\delta} M \vartheta^\delta \varepsilon^{1+\delta} \Bigg(\rho[g](t)\ \mathbb{E}\left[\|\xi\|^{2+\delta}\right] \int_{\mathbb{R}^{2d}} \|\alpha(x,v)\|^{2+\delta} f(t,x,v)\ dx\ dv + \\ + \rho[g](t) \int_{\mathbb{R}^{2d}} \|\beta(x,v)\|^{2+\delta} f(t,x,v)\ dx\ dv + \\ + \int_{\mathbb{R}^{4d}} \|\gamma(x,v,\hat{x},\hat{v})\|^{2+\delta} f(t,x,v) g(t,\hat{x},\hat{v})\ dx\ dv\ d\hat{x}\ d\hat{v} \Bigg).$$

Similar computations performed on the term $\Xi_\varphi[f,g]$ yield the inequality

$$\frac{\varepsilon}{2} \|\Xi_\varphi[f,g]\| \leq \frac{\varepsilon}{2} M \vartheta \Bigg(\rho[g](t) \int_{\mathbb{R}^{2d}} \|\beta(x,v)\|^2 f(t,x,v)\ dx\ dv + \\ + \int_{\mathbb{R}^{4d}} \|\gamma(x,v,\hat{x},\hat{v})\|^2 f(t,x,v) g(t,\tilde{x},\tilde{v})\ dx\ dv\ d\hat{x}\ d\hat{v} \Bigg).$$

Remembering that solutions of (4.29) have support uniformly bounded in time and that we asked for a g with finite constant mass (i.e., $\rho[g](t) = \rho[g](0) < +\infty$), the following result follows from Remark 4.22.

Theorem 4.23. *Fix $T > 0$, $\delta > 0$, and $g \in L^2([0,T]; \mathcal{M}_b(\mathbb{R}^{2d}))$ such that $\rho[g](t) = \rho[g](0) < +\infty$ for every $t \in [0,T]$. For every $\varepsilon > 0$, let $f^\varepsilon(\cdot)$ be a δ-weak solution of* (4.29) *with initial datum $f^0 \in \mathcal{M}_b(\mathbb{R}^{2d})$, and where the quantities η, λ and ς^2 are rescaled with respect to ε according to* (4.35)*. Suppose that*

- *(i)* $\mathbb{E}\left[\|\xi\|^{2+\delta}\right]$ *is finite, and*
- *(ii) the functions α, β, and γ belong to L^p_{loc} for $p = 2, 2+\delta$.*

Then, as $\varepsilon \to 0$, the solutions $f^\varepsilon(\cdot)$ converge pointwise, up to a subsequence, to $f(\cdot)$ which satisfies the Fokker-Planck-type equation (4.33) *with initial datum f^0.*

Remark 4.24. We stress again that the above result is not concerned about the well-posedness of system (4.29) (see Remark 4.21): it only addresses the relationship between the solutions of (4.29) and those of (4.33) under the grazing interaction regime.

CHAPTER 5

Learning from observations of evolutive systems

Continuity equations with nonlocal interaction kernels like (4.10) are currently the subject of intensive research towards the modeling of the biological and social behavior of microorganisms, animals, humans, etc. (we refer to [51, 54] for recent overviews on this subject). Despite the tremendous theoretical success of such research directions in terms of mathematical results regarding the well-posedness and the asymptotic behavior of solutions, one of the issues which is so far scarcely addressed is their actual applicability. With this we mean that the interaction functions appearing in biological and social contexts are likely to differ significantly from those employed in well-known models in physics, and most of the results at disposal address a purely qualitative analysis given certain properties of the kernel without looking at the precise interaction function that is needed in applications. Besides, the correct knowledge of the interaction kernel is also of paramount importance for the optimal controllability of such systems, as will be stressed in Chapter 6. A solid mathematical framework which establishes the conditions of “learnability” of the interaction kernels from observations of the dynamics is currently not available and it will be the main subject of this chapter.

5.1 The learning problem

Many time-dependent phenomena in physics, biology and social sciences can be modeled by systems of differential equations of the type

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j=1}^{N} K(x_i(t) - x_j(t)), \quad i = 1, \dots, N,$$

for a proper choice of the interaction kernel $K : \mathbb{R}^d \to \mathbb{R}^d$. Very often, in the study of such phenomena, the kernel K is just chosen to reproduce, at least approximately or qualitatively, some of the macroscopical effects of the observed dynamics, such as the formation of certain emerging patterns. The goal of this chapter is to provide a learning framework to find the kernel K matching at best the data from real-life observations. We shall focus our attention on alignment models (see Section 1.1) of the form

$$\dot{x}_i^{[a]}(t) = \frac{1}{N}\sum_{j=1}^{N} a(\|x_i^{[a]}(t) - x_j^{[a]}(t)\|)(x_j^{[a]}(t) - x_i^{[a]}(t)), \quad i = 1, \dots, N. \tag{5.1}$$

Notice that $a : \mathbb{R}_+ \to \mathbb{R}$ can also be negative, meaning that it may give rise to repulsive forces between agents. Hence, the kernel to be learned is of the form $K = F^{[a]}$ where

$$F^{[a]}(x) \triangleq -a(\|x\|)x \quad \text{for every } x \in \mathbb{R}^d.$$

The function a characterizes the dynamics of the system and it is the crucial information we want to retrieve from the observation of the evolution of the solution $x^{[a]} = (x_1^{[a]}, \dots, x_N^{[a]}) : [0,T] \to \mathbb{R}^{dN}$ of system (5.1).

We restrict our attention to interaction functions a belonging to the following *set of admissible kernels*

$$X \triangleq \left\{ b : \mathbb{R}_+ \to \mathbb{R} : \; b \in L^\infty(\mathbb{R}_+) \cap W^{1,\infty}_{\mathrm{loc}}(\mathbb{R}_+) \right\}.$$

In particular, every $a \in X$ is weakly differentiable, and its local Lipschitz constant $\mathrm{Lip}_\Omega(a)$ is finite for every compact set $\Omega \subset \mathbb{R}_+$. Notice that the lemma below, together with Proposition 4.9, implies that (5.1) is well-posed for any given initial datum $x^0 \in \mathbb{R}^{dN}$ and for every $N \in \mathbb{N}$, motivating the restriction to the kernels in X.

Lemma 5.1. *If $a \in X$ then $F^{[a]} \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ and is sublinear.*

Proof. For any compact set $\Omega \subset \mathbb{R}^d$ and for every $x, y \in \Omega$ it holds

$$\begin{aligned}
\|F^{[a]}(x) - F^{[a]}(y)\| &= \|a(\|x\|)x - a(\|y\|)y\| \\
&\le |a(\|x\|)|\|x - y\| + |a(\|x\|) - a(\|y\|)|\|y\| \\
&\le (|a(\|x\|)| + \mathrm{Lip}_\Omega(a)\|y\|)\|x - y\|,
\end{aligned}$$

and since $a \in L^\infty(\mathbb{R}_+)$ and $y \in \Omega$, it follows that $F^{[a]}$ is locally Lipschitz with Lipschitz constant depending only on a and Ω. Also the sublinearity trivially follows from $a \in L^\infty(\mathbb{R}_+)$ since

$$\|F^{[a]}(x)\| \le \|a\|_{L^\infty(\mathbb{R}_+)}\|x\|.$$

Therefore the sublinearity constant is given by $\|a\|_{L^\infty(\mathbb{R}_+)}$. □

The fundamental question that we shall address in this chapter is: can we recover $a \in X$ with high accuracy given some observations of the realized evolutions $x^{[a]}(\cdot)$? This question is prone to several specifications: for instance, we may want to assume that the initial conditions x^0 are generated according to a certain probability distribution or they are chosen deterministically ad hoc to determine at best a; that the observations are complete or incomplete, etc. As one quickly realizes, this is a very broad field to explore with many possible developments. We remand to [20, 21, 57, 124, 127, 140] and references therein for recent studies on the inference of social rules in collective behavior.

5.1.1 The optimal control approach and its drawbacks

Let us introduce an approach which perhaps would be naturally considered at a first instance. Given a certain solution $x^{[a]}(\cdot)$ of system (5.1) depending on the unknown parameter function a, one might decide to design the recovery of a as an optimal control problem [42]: for instance, one may seek a parameter function $\widehat{a}$ which minimizes

$$\mathcal{E}^{[a]}(\widehat{a}) \triangleq \frac{1}{T}\int_0^T \left[\frac{1}{N}\|x^{[a]}(t) - x^{[\widehat{a}]}(t)\|^2 + \mathcal{R}(\widehat{a})\right] dt, \tag{5.2}$$

being $x^{[\widehat{a}]}(\cdot)$ the solution of (5.1) with interaction kernel $\widehat{a}$ in place of a, and $\mathcal{R}(\cdot)$ is a suitable regularization functional, which restricts the possible minimizers of (5.2) to a specific class of kernels. Notice that here $\|\cdot\|$ denotes the ℓ_2–norm in $\mathbb{R}^{dN}$.

The first fundamental problem one immediately encounters with this formulation is the strongly nonlinear dependency of $x^{[\widehat{a}]}(\cdot)$ on $\widehat{a}$, which results in a strong non-convexity of the functional (5.2). This also implies that a direct minimization of (5.2) would risk to lead to suboptimal solutions, and even the computation of first order optimality conditions in terms of the Pontryagin's Maximum Principle would not characterize uniquely the minimal solutions. Besides these fundamental hurdles, the numerical implementation of either strategy (direct optimization or solution of the first order optimality conditions) is expected to be computationally unfeasible to reasonable degree of accuracy as soon as the number of agents N is significantly large (an instance of the already encountered *curse of dimensionality*).

5.1.2 A variational approach towards learning interaction kernels

Instead of the nonconvex optimal control problem above, we propose an alternative, direct approach which is both computationally very efficient and guarantees accurate approximations under reasonable assumptions. In particular, we consider as an estimator of the kernel a a minimizer of the following *discrete error functional*

$$\mathcal{E}^{[a],N}(\widehat{a}) \triangleq \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left\| \frac{1}{N}\sum_{j=1}^N \Big(\widehat{a}(\|x_i^{[a]}(t) - x_j^{[a]}(t)\|)(x_i^{[a]}(t) - x_j^{[a]}(t)) - \dot{x}_i^{[a]}(t)\Big)\right\|^2 dt, \tag{5.3}$$

among all competitors $\widehat{a} \in X$. Actually, the minimization of $\mathcal{E}^{[a],N}$ has a close connection to the optimal control problem of the previous section, since it also promotes the minimization of the discrepancy $\|x^{[a]}(t) - x^{[\widehat{a}]}(t)\|^2$ in (5.2), as the following result shows.

Proposition 5.2. *If $a, \widehat{a} \in X$ then there exist a constant $C > 0$ depending on T, $\widehat{a}$, a, and x^0 such that*

$$\frac{1}{N}\|x^{[a]}(t) - x^{[\widehat{a}]}(t)\|^2 \leq C\mathcal{E}^{[a],N}(\widehat{a}), \tag{5.4}$$

for all $t \in [0,T]$, where $x^{[a]}$ and $x^{[\widehat{a}]}$ are the solutions of (5.1) with initial datum x^0 and for the interaction kernels a and $\widehat{a}$, respectively.

Proof. Before starting, notice that for every $i = 1, \dots, N$ and $t \in [0,T]$ we can write

$$\frac{1}{N}\sum_{j=1}^{N}\left(\widehat{a}(\|x_i^{[a]}(t) - x_j^{[a]}(t)\|)(x_i^{[a]}(t) - x_j^{[a]}(t)) - \dot{x}_i^{[a]}(t)\right) = \left((F^{[\widehat{a}]} - F^{[a]}) * \mu_N(t)\right)(x_i^{[a]}(t)), \tag{5.5}$$

having used the expression of the dynamics (5.1) and the empirical measure

$$\mu_N(t) = \frac{1}{N}\sum_{i=1}^{N}\delta_{x_i^{[a]}(t)} \qquad \text{for every } t \in [0,T]. \tag{5.6}$$

Let us denote by $x \triangleq x^{[a]}$ and $\widehat{x} \triangleq x^{[\widehat{a}]}$. Two applications of Proposition 4.9, one for the choice $K = F^{[a]}$ and $g_N \equiv 0$, the other for $K = F^{[\widehat{a}]}$ and $g_N \equiv 0$, yield the existence of an $R > 0$ depending only on T, $\widehat{a}$, a and x^0 for which $\|x_i(t)\|, \|\widehat{x}_i(t)\| \leq R$ holds for every $t \in [0,T]$. Therefore, we can estimate by Jensen (or Hölder) inequality

$$\begin{aligned}
\frac{1}{N}\|x(t) - \widehat{x}(t)\|^2 &= \frac{1}{N}\left\|\int_0^t (\dot{x}(s) - \dot{\widehat{x}}(s))ds\right\|^2 \\
&\leq t\int_0^t \frac{1}{N}\|\dot{x}(s) - \dot{\widehat{x}}(s)\|^2 ds \\
&\leq 2t\int_0^t \left[\frac{1}{N}\sum_{i=1}^{N}\left\|\left((F^{[a]} - F^{[\widehat{a}]}) * \mu_N(s)\right)(x_i(s)))\right\|^2 + \right. \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{N}\sum_{j=1}^{N}\widehat{a}(\|x_i(s) - x_j(s)\|)\Big((\widehat{x}_j(s) - x_j(s)) + (x_i(s) - \widehat{x}_i(s))\Big) + \right. \\
&\quad \left.\left. + \Big(\widehat{a}(\|\widehat{x}_i(s) - \widehat{x}_j(s)\|) - \widehat{a}(\|x_i(s) - x_j(s)\|)\Big)(\widehat{x}_j(s) - \widehat{x}_i(s))\right\|^2\right] ds \\
&\leq 2T^2\mathcal{E}^{[a],N}(\widehat{a}) + \int_0^t 8T(\|\widehat{a}\|^2_{L^\infty(\Omega)} + (R\,\mathrm{Lip}_\Omega(\widehat{a}))^2)\|x(s) - \widehat{x}(s)\|^2 ds,
\end{aligned}$$

where $\Omega \triangleq [0,2R]$. An application of Gronwall's Lemma A.1 yields the estimate

$$\frac{1}{N}\|x(t) - \widehat{x}(t)\|^2 \leq 2T^2 e^{8T^2(\|\widehat{a}\|^2_{L^\infty(\Omega)} + (R\,\mathrm{Lip}_\Omega(\widehat{a}))^2)}\mathcal{E}^{[a],N}(\widehat{a})$$

for any $t \in [0,T]$, which is the desired bound. □

Therefore, if $\widehat{a}$ makes $\mathcal{E}^{[a],N}(\widehat{a})$ small, the trajectories $x^{[\widehat{a}]}(\cdot)$ of system (5.1) with interaction kernel $\widehat{a}$ instead of a are a good approximation of the trajectories $x^{[a]}(\cdot)$ at finite time. For simplicity of notations, we may choose to ignore below the dependence on a of the trajectories, and write $x \triangleq x^{[a]}$ when such a dependence is clear from the context. Additionally, whenever we consider the limit $N \to \infty$, we may denote the dependency of the trajectory on the number of agents $N \in \mathbb{N}$ by setting $x_N \triangleq x = x^{[a]}$.

Contrary to the optimal control approach, the functional $\mathcal{E}^{[a],N}$ is convex and can be easily computed from witnessed trajectories $x(\cdot)$ and $\dot{x}(\cdot)$. We may even consider discrete-time approximations of the time derivatives $\dot{x}_i$ (e.g., by finite differences) and we shall assume that the data of the problem is the full set of observations $(x(t))_{t\in[0,T]}$, for a prescribed finite time horizon $T > 0$. Furthermore, being a simple quadratic functional, its minimizers can be efficiently numerically approximated on a finite element space: given a finite dimensional space $V \subset X$, we pick

$$\widehat{a}_{N,V} \in \underset{\widehat{a}\in V}{\arg\min}\, \mathcal{E}^{[a],N}(\widehat{a}). \tag{5.7}$$

The fundamental mathematical question addressed in this chapter can then again be restated as

(Q) *for which choice of the approximating spaces $V \in \Lambda$ (we assume here that Λ is a countable family of invading subspaces of X) does $\widehat{a}_{N,V} \to a$ for $N \to \infty$ and $V \to X$ and in which topology should this convergence hold?*

We show now how to address this issue by means of a variational approach, seeking a limit functional for which techniques of Γ-convergence [86] – whose general aim is establishing the convergence of minimizers for a sequence of equi-coercive functionals to minimizers of a target functional – may provide a clear charaterization of the limits for the sequence of minimizers $(\widehat{a}_{N,V})_{N\in\mathbb{N},V\in\Lambda}$. By means of (5.5), we rewrite the functional (5.3) as follows:

$$\begin{aligned}\mathcal{E}^{[a],N}(\widehat{a}) &= \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left\| \frac{1}{N}\sum_{j=1}^N \big(F^{[\widehat{a}]} - F^{[a]}\big)(x_i(t) - x_j(t)) \right\|^2 dt \\ &= \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left\| \Big((F^{[\widehat{a}]} - F^{[a]}) * \mu_N(t) \Big)(x) \right\|^2 d\mu_N(t,x)dt.\end{aligned} \tag{5.8}$$

This formulation of the functional makes it easy to recognize the candidate for a Γ-limit of $\mathcal{E}^{[a],N}(\widehat{a})$. Indeed, it is well-known by Chapter 4 that for $N \to \infty$ a mean-field approximation holds of system (5.1): by Theorem 4.11 and Remark 4.12, if the initial conditions $x^0_{i,N}$ are independently and identically distributed according to a compactly supported probability measure $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$ for every $i = 1, \dots, N$ and for every $N \in \mathbb{N}$, then the empirical measures curves $(\mu_N)_{N\in\mathbb{N}}$ given by (5.6) weakly converge for $N \to \infty$ to the probability measure-valued trajectory $\mu(\cdot)$ satisfying

$$\begin{cases} \dfrac{\partial \mu}{\partial t}(t) = -\nabla_x \cdot ((F^{[a]} * \mu(t))\mu(t)) & \text{for } t \in (0,T], \\ \mu(0) = \mu^0, \end{cases} \tag{5.9}$$

in the sense of Definition 4.9 with vector field $V(t,\mu)(x) = (F^{[a]} * \mu)(x)$.
The candidate Γ-limit is then

$$\mathcal{E}^{[a]}(\widehat{a}) \triangleq \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left\| \Big((F^{[\widehat{a}]} - F^{[a]}) * \mu(t) \Big)(x) \right\|^2 d\mu(t,x)dt, \tag{5.10}$$

where $\mu(\cdot)$ is the above weak solution with $\mu(0) = \mu^0$, as soon as the initial conditions

$x_{i,N}^0$ are identically and independently distributed according to μ^0.

Remark 5.3. For the rest of the section, we shall assume that the initial conditions $x_{i,N}^0$ are i.i.d. according to μ^0 for every $i = 1, \dots, N$ and $N \in \mathbb{N}$. Therefore we are under the hypotheses of Theorem 4.11 and, in particular, there exists $R > 0$ (given by (4.15)) for which it holds $\operatorname{supp}(\mu_N(t)) \cup \operatorname{supp}(\mu(t)) \subseteq B(0, R)$ for every $N \in \mathbb{N}$ and $t \in [0, T]$.

5.1.3 The coercivity condition

Although all of this is very natural, several issues need to be addressed at this point. The first one is to establish the space where a result of Γ-convergence may hold and the identification of a can take place. As the trajectories $x^{[a]}(\cdot)$ do not explore the whole space in finite time, we expect that such a space may *not* be independent of the initial probability measure μ^0, as we clarify immediately. By Jensen inequality we have

$$\begin{aligned}\mathcal{E}^{[a]}(\widehat{a}) &\leq \frac{1}{T}\int_0^T \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |\widehat{a}(\|x-y\|) - a(\|x-y\|)|^2 \|x-y\|^2 d\mu(t,x)d\mu(t,y)dt \\ &= \frac{1}{T}\int_0^T \int_{\mathbb{R}_+} \big|\widehat{a}(r) - a(r)\big|^2 r^2 d\varrho(t,r)dt\end{aligned} \tag{5.11}$$

where $\varrho(t)$ is the pushforward of $\mu(t) \otimes \mu(t)$ by the Euclidean distance map $d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ defined by $d : (x, y) \mapsto \|x - y\|$. In other words, $\varrho : [0, T] \to \mathcal{P}_1(\mathbb{R}_+)$ is defined for every Borel set $A \subset \mathbb{R}_+$ as

$$\varrho(t)(A) \triangleq (\mu(t) \otimes \mu(t))\big(d^{-1}(A)\big). \tag{5.12}$$

The mapping $t \in [0, T] \mapsto \varrho(t)(A)$ is lower semi-continuous for every open set $A \subseteq \mathbb{R}_+$, and it is upper semi-continuous for any compact set A (see Lemma 5.7). We may therefore define a time-averaged probability measure $\overline{\rho}$ on the Borel σ-algebra of $\mathbb{R}_+$ by averaging $\varrho(t)$ over $t \in [0, T]$: for any open set $A \subseteq \mathbb{R}_+$ we define

$$\overline{\rho}(A) \triangleq \frac{1}{T}\int_0^T \varrho(t)(A)dt, \tag{5.13}$$

and extend this set function to a probability measure on all Borel sets. Finally we define

$$\rho(A) \triangleq \int_A r^2 d\overline{\rho}(r), \tag{5.14}$$

for any Borel set $A \subseteq \mathbb{R}_+$, to take into account the polynomial weight r^2 as appearing in (5.11). Then one can reformulate (5.11) in a very compact form as follows

$$\mathcal{E}^{[a]}(\widehat{a}) \leq \int_{\mathbb{R}_+} \big|\widehat{a}(r) - a(r)\big|^2 d\rho(r) = \|\widehat{a} - a\|^2_{L^2(\mathbb{R}_+,\rho)}. \tag{5.15}$$

Notice that ρ is defined through μ which depends on the initial probability measure μ^0.

To establish the coercivity of the learning problem it is essential to assume that there exists $c_T > 0$ such that also the following additional bound holds

$$c_T\|\widehat{a} - a\|^2_{L^2(\mathbb{R}_+,\rho)} \leq \mathcal{E}^{[a]}(\widehat{a}), \tag{5.16}$$

for all relevant $\widehat{a} \in X \cap L^2(\mathbb{R}_+, \rho)$. This crucial assumption eventually determines also the natural space $X \cap L^2(\mathbb{R}_+, \rho)$ for the solutions, which therefore depends on the choice of the initial conditions μ^0 via the measure ρ. In particular the constant $c_T \geq 0$ might not be nondegenerate for all the choices of μ^0 and one has to pick the initial distribution so that (5.16) can hold for $c_T > 0$. In Section 5.3 we show that for some specific choices of a and rather general choices of $\widehat{a} \in X$ one can construct probability measure-valued trajectories $\mu(\cdot)$ which allow to validate (5.16).
In order to ensure the compactness of the sequence of minimizers of $\mathcal{E}^{[a],N}$, we shall need to restrict the sets of possible solutions to classes of the type

$$X_{M,\Omega} \triangleq \left\{ b \in W^{1,\infty}(\Omega) : \|b\|_{L^\infty(\Omega)} + \|b'\|_{L^\infty(\Omega)} \leq M \right\},$$

where $M > 0$ is some predetermined constant and $\Omega \subset \mathbb{R}_+$ is a suitable compact set.
We now introduce the key property that a family of approximation spaces V_N must possess in order to ensure that the minimizers of the functionals $\mathcal{E}^{[a],N}$ over V_N converge to minimizers of $\mathcal{E}^{[a]}$.

Definition 5.4. Let $M > 0$ and $\Omega = [0, 2R]$ interval in $\mathbb{R}_+$ be given. We say that a family of closed subsets $V_N \subset X_{M,\Omega}$, $N \in \mathbb{N}$ has the *uniform approximation property* in $L^\infty(\Omega)$ if for all $b \in X_{M,\Omega}$ there exists a sequence $(b_N)_{N\in\mathbb{N}}$ converging uniformly to b on Ω and such that $b_N \in V_N$ for every $N \in \mathbb{N}$.

We are ready to state the main result of the chapter, which in addition answers the question (Q) raised in Section 5.1.2.

Theorem 5.5. *Assume $a \in X$, fix $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$ and let $\Omega = [0, 2R] \subset \mathbb{R}_+$ with $R > 0$ as in Theorem 4.11. Set*

$$M \geq \|a\|_{L^\infty(\Omega)} + \|a'\|_{L^\infty(\Omega)}.$$

For every $N \in \mathbb{N}$, let $x^0_{1,N}, \dots, x^0_{N,N}$ be i.i. μ^0-distributed and define $\mathcal{E}^{[a],N}$ as in (5.8) *for the solution μ_N of the equation* (5.9) *with initial datum*

$$\mu^0_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{x^0_{i,N}}.$$

For $N \in \mathbb{N}$, let $V_N \subset X_{M,\Omega}$ be a sequence of subsets with the uniform approximation property as in Definition 5.4 and consider

$$\widehat{a}_N \in \operatorname*{arg\,min}_{\widehat{a} \in V_N} \mathcal{E}^{[a],N}(\widehat{a}).$$

Then the sequence $(\widehat{a}_N)_{N\in\mathbb{N}}$ has a subsequence converging uniformly on Ω to some continuous function $\widehat{a} \in X_{M,\Omega}$ such that $\mathcal{E}^{[a]}(\widehat{a}) = 0$.
If we additionally assume the coercivity condition (5.16)*, then $\widehat{a} = a$ in $L^2(\mathbb{R}_+, \rho)$. Moreover, in this latter case, if there exist rates $\alpha, \beta > 0$, constants $C_1, C_2 > 0$, and a sequence $(a_N)_{N\in\mathbb{N}}$ of elements $a_N \in V_N$ such that*

$$\|a - a_N\|_{L^\infty(\Omega)} \leq C_1 N^{-\alpha}, \tag{5.17}$$

and

$$\mathcal{W}_1(\mu_N^0, \mu^0) \leq C_2 N^{-\beta}, \tag{5.18}$$

then there exists a constant $C_3 > 0$ *such that*

$$\|a - \widehat{a}_N\|^2_{L^2(\mathbb{R}_+,\rho)} \leq C_3 N^{-\min\{\alpha,\beta\}}, \tag{5.19}$$

for all $N \in \mathbb{N}$*. In particular, in this case, it is the entire sequence* $(\widehat{a}_N)_{N\in\mathbb{N}}$ *(and not only subsequences) to converge to* a *in* $L^2(\mathbb{R}_+, \rho)$*.*

Remark 5.6. We remark that the $L^2(\mathbb{R}_+, \rho)$ used in the above result is useful when ρ has positive density on large intervals of $\mathbb{R}_+$. Notice that the main result, under the validity of the coercivity condition, not only ensures the identification of a on the support of ρ, but it also provides a prescribed rate of convergence. For functions a in $X_{M,\Omega}$ and for finite element spaces V_N of continuous piecewise linear functions constructed on regular meshes of size N^{-1} a simple sequence $(a_N)_{N\in\mathbb{N}}$ realizing (5.17) with $\alpha = 1$ and $C_1 = M$ is the piecewise linear approximation to a which interpolates a on the mesh nodes. Indeed, for the approximation estimate (5.18) there are plenty of results concerning such rates and we refer to [90] and references therein. Roughly speaking, for μ_N^0 the empirical measure obtained by sampling N times independently from μ^0, one expects β to be of order $1/d$ (more precisely see [90, Theorem 1]), which is a manifestation of the aforementioned curse of dimensionality. While it is in general relatively easy to increase α independently of d as the smoothness of a increases (since a is a function of one variable only), obtaining $\beta > 1/d$ is in general not possible unless μ^0 has very special properties, see [101, Sections 4.4 and 4.5].

5.1.4 Numerical implementation of the variational approach

The strength of the result obtained via the variational approach followed in Section 5.1.2 is the total arbitrariness of the sequence V_N except for the assumed *uniform approximation property*, and that the result holds – deterministically – with respect to the uniform convergence, which is quite strong. However, the condition that the spaces V_N are to be picked as subsets of $X_{M,\Omega}$ requires the prior knowledge of the bound $M \geq \|a\|_{L^\infty(\Omega)} + \|a'\|_{L^\infty(\Omega)}$. Therefore, the finite dimensional optimization (5.7) is not anymore a simple *unconstrained* least squares (as claimed in the paragraph before (5.7)), but a problem constrained by a uniform bound on both the solution and its derivative. Nevertheless, as we clarify in Section 5.5, for $M > 0$ fixed and choosing V_N made of piecewise linear continuous functions, imposing the uniform L^∞ bound in the least square problem does not constitute a severe difficulty. Also the tuning of the parameter $M > 0$ turns out to be rather simple. In fact, for N fixed, numerical evidences suggest that the minimizers $\widehat{a}_N = \widehat{a}_{N,M}$ obtained with different values of the constraint parameter M have the property that the map

$$M \mapsto \mathcal{E}^{[a],N}(\widehat{a}_{N,M})$$

is monotonically decreasing as a function of M and it becomes constant for $M \geq M^*$, where $M^* > 0$ does not depend on N. We claim that this special value M^* is indeed the "right" parameter for the L^∞ bound. For such a choice, we show also numerically that, as expected, if we let N grow, the minimizers $\widehat{a}_N$ approximates better and better the unknown potential a.

Despite the fact that both the tuning of $M > 0$ and the constrained minimization over $X_{M,\Omega}$ requiring L^∞ bounds are not severe issues, it would be way more efficient to perform an unconstrained least squares over X. In the upcoming paper [35] a strategy based on the approach developed by DeVore et al. in [29, 30] towards universal algorithms for learning regression functions from independent samples drawn according to an unknown probability distribution is used. This extension presents several challenges including the lack of independence of the samples collected in our framework and the nonlocality of the scalar products of the corresponding least squares. This has a price to pay: the spaces V_N need to be carefully chosen and the convergence result holds only with high probability.

5.1.5 Extensions to second order systems

The learning framework developed in the previous sections can be extended to second-order dynamical systems, like the Cucker-Smale one (1.6), by considering the following error functional

$$\mathcal{E}^{[a],N}(\widehat{a}) \triangleq \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left\| \frac{1}{N}\sum_{j=1}^N \left(\widehat{a}(\|x_i^{[a]}(t) - x_j^{[a]}(t)\|)(v_i^{[a]}(t) - v_j^{[a]}(t)) - \dot{v}_i^{[a]}(t)\right)\right\|^2 dt, \tag{5.20}$$

where $(x^{[a]}, v^{[a]}) : [0,T] \to \mathbb{R}^{dN} \times \mathbb{R}^{dN}$ is the solution of system (1.6) with the true a as interaction kernel. Figure 5.1 below shows the reconstruction of a Cucker-Smale kernel a by successive minimizations of the above error functional for increasing values of N. Also in this case, an improvement in the quality of the approximation as the number of agents increases is a clearly visible trend.

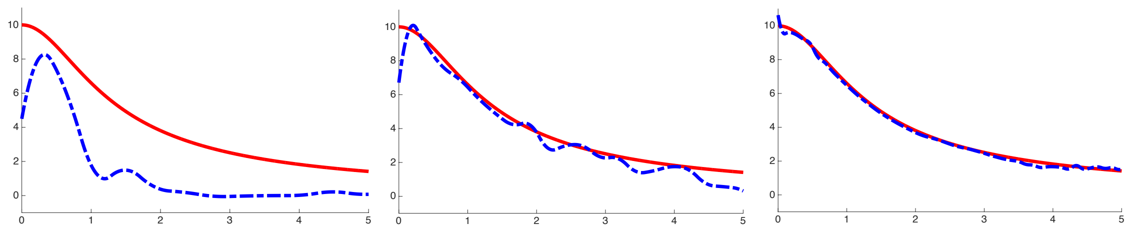


Figure 5.1: Recostruction of the Cucker-Smale kernel (1.7) with parameters $H = 10$, $\sigma = 1$ and $\beta = 0.6$ by minimization of the error functional (5.20) for $N = 10$, 20, 40.

However, despite the above promising preliminary results, the extension of the variational approach to the second-order case is not trivial and a careful analysis of its feasibility has not yet been performed.

5.2 The measure $\overline{\rho}$

As already explained in the above sections, our goal is to learn $a \in X$ from the observation of the dynamics $x^{[a]}(\cdot)$, which is the solution of system (5.1) with a as interaction kernel, $(x_1^0, \dots, x_N^0)$ as initial datum and T as finite time horizon. Remember that, by Remark 4.6, it is equivalent to speak about $x^{[a]}(\cdot)$ or the empirical measure-valued curve $\mu_N(\cdot)$ defined by (5.6).

Roughly speaking, we should pick $\widehat{a}$ among those functions in X which would give rise to a dynamics close to μ_N: more precisely, as discussed in Section 5.1.2, we choose $\widehat{a}_N \in X$ as a minimizer of the discrete error functional given by (5.3), which we rewrite for convenience:

$$\mathcal{E}^{[a],N}(\widehat{a}) = \frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left\| \frac{1}{N}\sum_{j=1}^N \left(\widehat{a}(\|x_i^{[a]}(t) - x_j^{[a]}(t)\|)(x_i^{[a]}(t) - x_j^{[a]}(t)) - \dot{x}_i^{[a]}(t)\right)\right\|^2 dt.$$

Let us remind that, by Proposition 5.2, this optimization guarantees also that any minimizer $\widehat{a}_N$ produces good trajectory approximations $x^{[\widehat{a}_N]}(t)$ to the "true" ones $x^{[a]}(t)$ at least at finite time $t \in [0, T]$.
In order to rigorously introduce the coercivity condition (5.16), we need to explore finer properties of the family of measures $(\varrho(t))_{t\in[0,T]}$ given by (5.14).

Lemma 5.7. *For every open set $A \subseteq \mathbb{R}_+$ the mapping $t \in [0, T] \mapsto \varrho(t)(A)$ is lower semi-continuous, whereas for any compact set A it is upper semi-continuous.*

Proof. As a first step we show that for every given sequence $(t_n)_{n\in\mathbb{N}}$ converging to $t \in [0, T]$ we have the weak convergence $\varrho(t_n) \rightharpoonup \varrho(t)$ for $n \to \infty$. We first note that $\mu(t_n) \otimes \mu(t_n) \rightharpoonup \mu(t) \otimes \mu(t)$, since $\mu(t_n) \rightharpoonup \mu(t)$ holds because of the continuity of $\mu(t)$ in the Wasserstein metric $\mathcal{W}_1$. This implies the claimed weak convergence $\varrho(t_n) \rightharpoonup \varrho(t)$, since for any function $f \in \mathcal{C}(\mathbb{R}_+)$, it holds $f \circ d \in \mathcal{C}(\mathbb{R}^d \times \mathbb{R}^d)$, and hence

$$\begin{aligned}\int_{\mathbb{R}_+} f\, d\varrho(t_n) &= \int_{\mathbb{R}^{2d}} (f \circ d)(x, y) d(\mu(t_n) \otimes \mu(t_n))(x, y)\\ &\overset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^{2d}} (f \circ d)(x, y) d(\mu(t) \otimes \mu(t))(x, y) = \int_{\mathbb{R}_+} f\, d\varrho(t).\end{aligned}$$

The claim now follows from general results for weakly* convergent sequences of Radon measures, see e.g. [11, Proposition 1.62]. □

Lemma 5.7 justifies the following

Definition 5.8. The probability measure $\overline{\rho}$ on the Borel σ-algebra on $\mathbb{R}_+$ is defined for any Borel set $A \subseteq \mathbb{R}_+$ as follows

$$\overline{\rho}(A) \triangleq \frac{1}{T}\int_0^T \varrho(t)(A) dt. \tag{5.21}$$

Notice that Lemma 5.7 shows that (5.21) is well-defined only for sets A that are open or compact in $\mathbb{R}_+$. This directly implies that $\overline{\rho}$ can be extended to any Borel set A, since both families of sets provide a basis for the Borel σ-algebra on $\mathbb{R}_+$. Moreover $\overline{\rho}$ is a regular measure on $\mathbb{R}_+$, since Lemma 5.7 also implies that for any Borel set A

$$\overline{\rho}(A) = \sup\left\{\overline{\rho}(F) : F \subseteq A,\ F \text{ compact}\right\} = \inf\left\{\overline{\rho}(G) : A \subseteq G,\ G \text{ open}\right\}.$$

The measure $\overline{\rho}$ "measures" which – and how much – regions of $\mathbb{R}_+$ (the set of inter-point distances) are explored during the dynamics of the system. Highly explored regions are

where our learning process ought to be successful, since these are the areas where we do have enough samples from the dynamics to reconstruct the function a.
We now show the absolute continuity of $\overline{\rho}$ with respect to $\mathcal{L}^1 \llcorner_{\mathbb{R}_+}$, the Lebesgue measure on $\mathbb{R}$ restricted to $\mathbb{R}_+$.

Lemma 5.9. *Let μ^0 be absolutely continuous with respect to the d-dimensional Lebesgue measure $\mathcal{L}^d$. Then $\mu(t)$ is absolutely continuous with respect to $\mathcal{L}^d$ for every $t \in [0,T]$.*

Proof. Both μ^0 and $\mu(t)$ are supported in $B(0,R)$, with R as in (4.15). By (4.20) and Proposition 4.14, the measure $\mu(t)$ is the pushforward of μ^0 under the locally bi-Lipschitz map $\mathcal{T}_t^{\mu}$. Since $\mathcal{T}_t^{\mu}$ has Lipschitz inverse on $B(0,R)$, this inverse maps $\mathcal{L}^d$-null sets to $\mathcal{L}^d$-null sets, so μ^0-null sets are not only $\mathcal{L}^d$-null sets by assumption, but are also $\mu(t)$-null sets. □

Lemma 5.10. *Let μ^0 be absolutely continuous with respect to $\mathcal{L}^d$. Then, for all $t \in [0,T]$, the measures $\varrho(t)$ and $\overline{\rho}$ are absolutely continuous with respect to $\mathcal{L}^1 \llcorner_{\mathbb{R}_+}$.*

Proof. Fix $t \in [0,T]$. By Lemma 5.9 we already know that $\mu(t)$ is absolutely continuous with respect to $\mathcal{L}^d$, and so $\mu(t) \otimes \mu(t)$ is absolutely continuous with respect to $\mathcal{L}^{2d}$. It hence remains to show that $\mathcal{L}^{2d}$ is absolutely continuous with respect to $\mathcal{L}^1 \llcorner_{\mathbb{R}_+}$, where d denotes the distance function, but this follows easily by observing that $d^{-1}(A) = 0$ for every $\mathcal{L}^1 \llcorner_{\mathbb{R}_+}$-null set A, and by an application of Fubini's Theorem. The absolute continuity of $\overline{\rho}$ now follows immediately from the one of $\varrho(t)$ for every $t \in [0,T]$ and its definition as an integral average (5.21). □

As an easy consequence of the fact that the dynamics of our system has support uniformly bounded in time, we get the following crucial properties of the measure $\overline{\rho}$.

Lemma 5.11. *Let $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$. Then the measure $\overline{\rho}$ is finite and has compact support.*

Proof. We have

$$\overline{\rho}(\mathbb{R}_+) = \frac{1}{T} \int_0^T \varrho(t)(\mathbb{R}_+) dt = \frac{1}{T} \int_0^T \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\| d\mu(t,x) d\mu(t,y) dt < +\infty,$$

since the distance function is continuous and the support of μ is uniformly bounded in time. This shows that $\overline{\rho}$ is bounded. Since the supports of the measures $\varrho(t)$ are the subsets of $\Omega = \{\|x-y\| : x, y \in B(0,R)\} = [0, 2R]$, where R is given by (4.15), by construction we also have $\operatorname{supp}(\overline{\rho}) \subseteq \Omega$. □

5.3 On the coercivity assumption

With the measure $\overline{\rho}$ at disposal, we may define ρ as in (5.14). An easy consequence of Lemma 5.11 is that if $a \in X$, then

$$\|a\|^2_{L^2(\mathbb{R}_+,\rho)} = \int_{\mathbb{R}_+} \left|a(r)\right|^2 d\rho(r) \leq \rho(\mathbb{R}_+) \|a\|^2_{L^\infty(\operatorname{supp}(\rho))},$$

and therefore $X \subseteq L^2(\mathbb{R}_+, \rho)$. As already mentioned in the previous section, for $N \to \infty$ a natural mean-field approximation to the learning functional is given by

$$\mathcal{E}^{[a]}(\widehat{a}) = \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left\| \left((F^{[\widehat{a}]} - F^{[a]}) * \mu(t) \right)(x) \right\|^2 d\mu(t,x)dt,$$

where $\mu(\cdot)$ is a weak solution to (4.10). Remember that using the definition of ρ we obtain

$$\mathcal{E}^{[a]}(\widehat{a}) \leq \frac{1}{T}\int_0^T \int_{\mathbb{R}_+} \left|\widehat{a}(r) - a(r)\right|^2 r^2 d\varrho(t,r)dt = \|\widehat{a} - a\|^2_{L^2(\mathbb{R}_+,\rho)}.$$

This inequality suggested in turn the coercivity condition (5.16):

$$\mathcal{E}^{[a]}(\widehat{a}) \geq c_T \|\widehat{a} - a\|^2_{L^2(\mathbb{R}_+,\rho)}.$$

The main reason this condition is of interest to us is given by the following

Proposition 5.12. *Assume $a \in X$ and that the coercivity condition* (5.16) *holds. Then any minimizer of $\mathcal{E}^{[a]}$ in X coincides ρ-a.e. with a.*

Proof. Notice that $\mathcal{E}^{[a]}(a) = 0$, and since $\mathcal{E}^{[a]}(\widehat{a}) \geq 0$ for all $\widehat{a} \in X$ this implies that a is a minimizer of $\mathcal{E}^{[a]}$. Now suppose that $\mathcal{E}^{[a]}(\widehat{a}) = 0$ for some $\widehat{a} \in X$. By (5.16) we obtain that $\widehat{a} = a$ in $L^2(\mathbb{R}_+, \rho)$, and therefore they coincide ρ-almost everywhere. □

5.3.1 Coercivity is "generically" satisfied

We make the case that while "degeneracies" would cause our coercivity condition to fail (i.e., $c_T = 0$), in a "generic" case the coercivity inequality holds. On the one hand, we show that if we could model the misfit $\mathcal{K}(r) \triangleq (a(r) - \widehat{a}(r))r$ to behave randomly, in a sufficiently independent manner, over a finite set of trajectory distances, then the coercivity condition holds with high probability. While the needed independence assumptions will typically be too strong to hold in practice, the arguments we provide are by far not the most general possible, and we view them as one possible notion of a "generic" case. On the other hand, in the next subsection we also present a more rigorous *deterministic* argument to verify the coercivity condition for particular choices of a.

With the notation of the misfit just introduced, the coercivity condition reads

$$\frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left\| \int_{\mathbb{R}^d} \mathcal{K}(\|x-y\|)\frac{x-y}{\|x-y\|} d\mu(t,x) \right\|^2 d\mu(t,y) \geq$$
$$\geq \frac{c_T}{T}\int_0^T \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \left|\mathcal{K}(\|x-y\|)\right|^2 d\mu(t,x)d\mu(t,y).$$

If the inequality holds without the time average for a fixed t_0,

$$\int_{\mathbb{R}^d} \left\| \int_{\mathbb{R}^d} \mathcal{K}(\|x-y\|)\frac{x-y}{\|x-y\|} d\mu(t_0,x) \right\|^2 d\mu(t_0,y) \geq$$
$$\geq c'_{t_0} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \left|\mathcal{K}(\|x-y\|)\right|^2 d\mu(t_0,x)d\mu(t_0,y),$$

then by a continuity argument it can be extended to a nontrivial time interval. We will therefore freeze time and investigate the inequality at this fixed time. Additionally, for the moment we restrict our attention to the case where $\mu(t_0)$ is a discrete measure $\mu_N = \sum_{i=1}^N \delta_{x_i}/N$ (we drop t_0 since it is now fixed), so that the inequality reads

$$\frac{1}{N}\sum_{i=1}^N \left\| \frac{1}{N}\sum_{j=1}^N \mathcal{K}(\|x_i - x_j\|)\frac{x_i - x_j}{\|x_i - x_j\|} \right\|^2 \geq \frac{c'_{t_0}}{N^2}\sum_{i=1}^N\sum_{j=1}^N |\mathcal{K}(\|x_i - x_j\|)|^2 . \tag{5.22}$$

We argue now that this ("instantaneous") inequality holds with high probability as soon as the matrix $\mathbf{K} \triangleq (\mathcal{K}(\|x_i - x_j\|))_{i,j=1}^N$ is modeled as a random matrix. Although it is not completely plausible to argue statistical independence of the entries of such a matrix because it comes from evaluating a smooth function over distances of non-random points, this model is not completely unreasonable: after all, $\mathbf{K}$ involves the difference of our estimator $\widehat{a}$ and the target influence function a. For least squares estimators this difference is random with the samples used to construct the estimator, often with nearly independent, perhaps even Gaussian, fluctuations.

Proposition 5.13. *Let $\mathbf{K} \triangleq (\mathcal{K}(\|x_i - x_j\|))_{i,j=1}^N$, where $\mathcal{K}(r) = (a(r) - \widehat{a}(r))r$, be a random matrix with independent Gaussian rows, each with variance $\sigma^2 I_N$. Then with high probability the coercivity bound (5.22) holds with a constant $c'_{t_0} = O(1/N)$.*
Moreover, if there is $\alpha > 0$ such that $\mathcal{K}(\|x_i - x_j\|)\|x_i - x_j\|^{\frac{\alpha}{2}}$ is distributed as η_{ij}, where η_{ij} are independent standard normal distributions, and the quantity $1/N\sum_{j=1}^N \|x_i - x_j\|^{-\alpha} \in [c_1, c_2]$ for all $i = 1, \ldots, N$, and for all N large enough, with $c_1, c_2 > 0$, then coercivity holds with a constant c'_{t_0} independent of N.

Proof. By assumption $\mathbf{K}$ has independent Gaussian rows, each with variance $\sigma^2 I_N$. Since the estimates below and the bounds we wish to obtain are scale invariant, we may assume $\sigma = 1$. Also, having fixed a given time t_0, we shall omit it from the notation. Let $\mathbf{X}_i \in \mathbb{R}^{N\times d}$ be the matrix whose j-th row is the vector $(x_i - x_j)\|x_i - x_j\| \in \mathbb{R}^d$, and let $\mathbf{K}(i,:) \in \mathbb{R}^N$ be the i-th row of $\mathbf{K}$. The coercivity inequality (5.22) can be re-written as

$$\frac{1}{N}\sum_{i=1}^N \left| \frac{1}{N}\mathbf{K}(i,:)\mathbf{X}_i \right|^2 \geq \frac{c'_t}{N^2}\|\mathbf{K}\|_{\mathbb{F}}^2, \tag{5.23}$$

where $\|\cdot\|_{\mathbb{F}}$ is the Frobenius matrix norm. We have

$$\begin{aligned}
\mathbb{E}\left[|\mathbf{K}(i,:)\mathbf{X}_i|^2\right] &= \sum_{l=1}^d \sum_{j,j'=1}^N \mathbb{E}\left[\mathcal{K}(\|x_i - x_j\|)\mathcal{K}(\|x_i - x_{j'}\|)|\right] \left(\frac{x_i - x_j}{\|x_i - x_j\|}\right)_l \left(\frac{x_i - x_{j'}}{\|x_i - x_{j'}\|}\right)_l \\
&= \sum_{j=1}^N \mathbb{E}\left[\mathcal{K}(\|x_i - x_j\|)^2\right] \sum_{l=1}^d \left(\frac{x_i - x_j}{\|x_i - x_j\|}\right)_l^2 \\
&= \sum_{j=1}^N \mathbb{E}\left[\mathcal{K}(\|x_i - x_j\|)^2\right] \\
&= N,
\end{aligned}$$

where we used independence, and in the last step we used the fact that every row of $\mathbf{X}_i$ is a unit vector. By concentration arguments (e.g., [172]) one readily obtains that with high probability

$$\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{N}\mathbf{K}(i,:)\mathbf{X}_i\right|^2 \geq \frac{C}{N}.$$

One the other hand, since $\mathbb{E}[||\mathbf{K}||_{\mathbb{F}}^2] \leq CN^2$ by standard random matrix theory results (e.g., [172]), and in fact not just in expectation but also with high probability, the right hand side of (5.23) is bounded by $c'_{t_0}C$ from above. Choosing c'_{t_0} small enough (and at least as small as $O(1/N)$, as a function of N), we obtain (5.23) with high-probability.

For the second case, we proceed exactly as above, and note that in this case it holds

$$\mathbb{E}\left[|\mathbf{K}(i,:)\mathbf{X}_i|^2\right] = \sum_{j=1}^{N}\|x_i - x_j\|^{-\alpha}.$$

Continuing as above, we obtain that with high probability

$$\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{N}\mathbf{K}(i,:)\mathbf{X}_i\right|^2 \geq \frac{1}{N}\sum_{i=1}^{N}\frac{C}{N}\sum_{j=1}^{N}\|x_i - x_j\|^{-2\alpha} \geq C,$$

which implies the claim. □

Let us comment on the second part of the statement. First of all there are two key additional assumptions. The first one captures the situation that the quantity $1/N\sum_{j=1}^{N}\|x_i - x_j\|^{-\alpha}$ ought to have, at least for some α, and under reasonable assumptions, a positive limit as N tends to infinity, and so in particular it is bounded above and below, and away from 0, for N large. The second one requires that the estimator has errors that decrease with the distance $\|x_i - x_j\|$ in the fashion prescribed: if the kernel and its gradient decrease with the distance r, then it is indeed natural to expect that the estimator would have this property. Under these assumptions, we obtain that the coercivity condition holds with a constant independent of N, which in turns suggests it also holds in the continuous limit.

The argument may be generalized to other models of random matrices, for example with sub-Gaussian rows (for $\mathbf{K}$) and uniformly lower-bounded smallest singular values. One may also consider $\mathbf{X}_i$ random, sufficiently uncorrelated with $\mathbf{K}$, and obtain similar results. Also, the continuous case is not substantially different from the discrete case, as it may be derived by smoothing discrete approximations. We do not pursue these generalizations, as our purpose here is to show that the coercivity assumption is "generically" satisfied.

5.3.2 The deterministic case

We construct now deterministic examples of trajectories $\mu(\cdot)$ for which the coercivity condition (5.16) holds. Again it is convenient to write $\mathcal{K}(r) \triangleq (a(r) - \widehat{a}(r))r$, so that the coercivity condition for a discrete system with N agents in this case can be reformulated in the following way

$$\frac{1}{T}\int_0^T \frac{1}{N}\sum_{i=1}^N \left\| \frac{1}{N}\sum_{j=1}^N \mathcal{K}(\|x_i(t)-x_j(t)\|)\frac{x_i(t)-x_j(t)}{\|x_i(t)-x_j(t)\|}\right\|^2 dt \geq$$
$$\geq \frac{c_T}{N^2 T}\int_0^T \sum_{i=1}^N\sum_{j=1}^N |\mathcal{K}(\|x_i(t)-x_j(t)\|)|^2 dt. \tag{5.24}$$

We start with the simple case of two agents, i.e., $N=2$, for which no specific assumptions on $a, \widehat{a}$ are required to verify (5.16) other than their boundedness in 0. Now, let us omit the time dependency and observe more closely the integrand on the left-hand-side. If for every $i=1,2$ we denote by $j(i)\in\{1,2\}$ the index such that $j(i)\neq i$, we obtain

$$\begin{aligned}
\frac{1}{2}\sum_{i=1}^2 \left\|\frac{1}{2}\sum_{j=1}^2 \mathcal{K}(\|x_i-x_j\|)\frac{x_i-x_j}{\|x_i-x_j\|}\right\|^2 &= \frac{1}{2}\sum_{i=1}^2 \left\|\frac{1}{2}\sum_{j\neq i}^2 \mathcal{K}(\|x_i-x_j\|)\frac{x_i-x_j}{\|x_i-x_j\|}\right\|^2 \\
&= \frac{1}{4}\sum_{i=1}^2 \left\|\mathcal{K}(\|x_i-x_{j(i)}\|)\frac{x_i-x_{j(i)}}{\|x_i-x_{j(i)}\|}\right\|^2 \\
&= \frac{1}{4}\sum_{i=1}^2 \left|\mathcal{K}(\|x_i-x_{j(i)}\|)\right|^2 \\
&= \frac{1}{4}\sum_{i=1}^2\sum_{j=1}^2 |\mathcal{K}(\|x_i-x_j\|)|^2.
\end{aligned}$$

Integrating over time the latter equality yields (5.24) for $N=2$ with an actual equal sign and $c_T=1$. Notice that here we have not made any specific assumptions on the trajectories $x_i(\cdot)$ for $i=1,2$.

Let us then consider the case of $N=3$ agents. Already in this simple case the angles between particles may be rather arbitrary and analyzing the many possible configurations becomes an involved exercise. (Notice that we circumvented this problem in the random model in Section 5.3.1 thanks to the assumed independence of the entries of the rows of $\mathbf{K}$.) To simplify the problem we assume that $d=2$ and that at a certain time t the particles are disposed precisely at the vertexes of a equilateral triangle of edge length $\overline{r}$. This makes the computation of the angles very simple. We also assume that $\mathcal{K}$ gets its maximal absolute value precisely at $\overline{r}$, hence

$$\frac{1}{9}\sum_{i=1}^3\sum_{j=1}^3 |\mathcal{K}(\|x_i(t)-x_j(t)\|)|^2 \leq \|\mathcal{K}\|^2_{L^\infty(\mathbb{R}_+)} = \mathcal{K}(\overline{r})^2 \quad \text{for every } t\in[0,T].$$

Notice that, independently of the behavior of the agents at any other time $t\in[0,T]$, it holds also

$$\frac{1}{9T}\int_0^T \sum_{i=1}^3\sum_{j=1}^3 |\mathcal{K}(\|x_i(t)-x_j(t)\|)|^2 dt \leq \|\mathcal{K}\|^2_{L^\infty(\mathbb{R}_+)} = \mathcal{K}(\overline{r})^2. \tag{5.25}$$

A direct computation in the case of agents disposed at the vertexes of a equilateral triangle shows that

$$\frac{1}{3}\sum_{i=1}^{3}\left\|\frac{1}{3}\sum_{j=1}^{3}\mathcal{K}(\|x_i(t)-x_j(t)\|)\frac{x_i(t)-x_j(t)}{\|x_i(t)-x_j(t)\|}\right\|^2 = \frac{1}{3}\mathcal{K}(\overline{r})^2, \qquad (5.26)$$

and therefore

$$\frac{1}{3}\sum_{i=1}^{3}\left\|\frac{1}{3}\sum_{j=1}^{3}\mathcal{K}(\|x_i(t)-x_j(t)\|)\frac{x_i(t)-x_j(t)}{\|x_i(t)-x_j(t)\|}\right\|^2 \geq \frac{1}{18}\sum_{i=1}^{3}\sum_{j=1}^{3}|\mathcal{K}(\|x_i(t)-x_j(t)\|)|^2.$$

Unfortunately the assumption that $\mathcal{K}$ achieves its maximum in absolute value at $\overline{r}$ does not allow us yet to conclude by a simple integration over time the coercivity condition as we did for the case of two particles. In order to extend the validity of the inequality to arbitrary functions taking maxima at other points, we need to integrate over time by assuming now that the particles are vertexes of equilateral triangles with time dependent edge length, say from $r = 0$ growing in time up to $r = 2R > 0$. This will allow the trajectories to explore any possible distance within a given interval and to capture the maximal absolute value of any kernel. More precisely, let us now assume that $\mathcal{K}$ is an arbitrary bounded continuous function, achieving its maximal absolute value over $[0, 2R]$, say at $\overline{r} \in (0, 2R)$ and we can assume that this is obtained corresponding to the time $t_0 \in [0, T]$ when the particles form precisely the equilateral triangle of side length $\overline{r}$. Now we need to make a stronger assumption on $\widehat{a}$, i.e., we require $\widehat{a}$ to belong to a class of equi-continuous functions, for instance functions which are Lipschitz continuous with uniform Lipschitz constant (such as the functions in $X_{M,\Omega}$). Under this equi-continuity assumption, there exist $\varepsilon > 0$ and a constant $c_{T,\varepsilon} > 0$ independent of $\mathcal{K}$ (but perhaps depending only on its modulus of continuity) such that

$$\begin{aligned}
&\frac{1}{T}\int_0^T \frac{1}{3}\sum_{i=1}^{3}\left\|\frac{1}{3}\sum_{j=1}^{3}\mathcal{K}(\|x_i(t)-x_j(t)\|)\frac{x_i(t)-x_j(t)}{\|x_i(t)-x_j(t)\|}\right\|^2 dt \geq \\
&\qquad \geq \frac{1}{T}\int_{t_0-\varepsilon}^{t_0+\varepsilon} \frac{1}{3}\sum_{i=1}^{3}\left\|\frac{1}{3}\sum_{j=1}^{3}\mathcal{K}(\|x_i(t)-x_j(t)\|)\frac{x_i(t)-x_j(t)}{\|x_i(t)-x_j(t)\|}\right\|^2 dt \\
&\qquad \geq \frac{c_{T,\varepsilon}}{3}\mathcal{K}(\overline{r})^2 \\
&\qquad \geq \frac{c_{T,\varepsilon}}{18T}\int_0^T \sum_{i=1}^{3}\sum_{j=1}^{3}|\mathcal{K}(\|x_i(t)-x_j(t)\|)|^2 dt.
\end{aligned}$$

In the latter sequence of inequalities we have used (5.26) and (5.25). Hence, also in this case, one can construct examples for which the coercivity assumption is verifiable. Actually this construction can be extended to any group of N particles disposed on the vertexes of regular polygons. As an example of how one should proceed, let us consider the case of $N = 4$ particles disposed instantaneously at the vertexes of a square of side length

$\sqrt{2}\overline{r} > 0$. In this case one directly verifies that

$$\frac{1}{4}\sum_{i=1}^{4}\left\|\frac{1}{4}\sum_{j=1}^{4}\mathcal{K}(\|x_i - x_j\|)\frac{x_i - x_j}{\|x_i - x_j\|}\right\|^2 = \frac{1}{16}(\mathcal{K}(2\overline{r}) + \sqrt{2}\mathcal{K}(\sqrt{2}\overline{r}))^2. \tag{5.27}$$

Let us assume that the maximal absolute value of $\mathcal{K}$ in absolute value is attained precisely at $\sqrt{2}\overline{r}$. Then the minimum of the expression on the right-hand side of (5.27) is attained for the case where $\mathcal{K}(2\overline{r}) = -\mathcal{K}(\sqrt{2}\overline{r})$ yielding the following estimate from below

$$\frac{1}{4}\sum_{i=1}^{4}\left\|\frac{1}{4}\sum_{j=1}^{4}\mathcal{K}(\|x_i - x_j\|)\frac{x_i - x_j}{\|x_i - x_j\|}\right\|^2 \geq \frac{3-2\sqrt{2}}{16}\mathcal{K}(\sqrt{2}\overline{r})^2.$$

Hence, also in this case, we can apply the continuity argument above to eventually show the coercivity condition. Similar procedures can be followed for any $N \geq 5$. However, as $N \to \infty$ one can show numerically that the lower bound vanishes quite rapidly, making it impossible, perhaps not surprisingly, to conclude the validity of the coercivity condition for the uniform distribution over the circle.

All the examples presented so far are based on discrete measures $\mu_N(t) = \sum_{i=1}^{N}\delta_{x_i(t)}/N$ supported on agents lying on the vertexes of polytopes. However, one can consider a g_ε for which $g_\varepsilon \to \delta_0$ as $\varepsilon \to 0$, and the corresponding regularized probability measure

$$\mu_\varepsilon(t,x) \triangleq g_\varepsilon * \mu_N(t,x) = \frac{1}{N}\sum_{i=1}^{N} g_\varepsilon(x - x_i(t)) \quad \text{for every } x \in \mathbb{R}^d.$$

This diffuse measure approximates μ_N in the sense that $\mathcal{W}_1(\mu_\varepsilon,\mu_N) \to 0$ for $\varepsilon \to 0$, hence, in particular, integrals against Lipschitz functions φ can be well-approximated

$$\left|\int_{\mathbb{R}^d}\varphi(x)d\mu_\varepsilon(t,x) - \int_{\mathbb{R}^d}\varphi(x)d\mu_N(t,x)\right| \leq \mathrm{Lip}_{\mathbb{R}^d}(\varphi)\mathcal{W}_1(\mu_\varepsilon(t),\mu_N(t)).$$

Under the additional assumption that $\mathrm{Lip}_\Omega(\widehat{a}) \sim \|\widehat{a}\|_{L^\infty(\Omega)}$ (and this is true whenever $\widehat{a}$ is a piecewise polynomial function over a finite partition of $\mathbb{R}_+$, with the constant of the equivalence depending on the particular partition) one can extend the validity of the coercivity condition for μ_N (5.24) to μ_ε as follows

$$\begin{aligned}\frac{1}{T}\int_0^T\int_{\mathbb{R}^d}\left\|\int_{\mathbb{R}^d}\mathcal{K}(\|x-y\|)\frac{y-x}{\|y-x\|}d\mu_\varepsilon(t,x)\right\|^2 d\mu_\varepsilon(t,y)dt \geq \\ \geq \frac{c_{T,\varepsilon}}{T}\int_0^T\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}|\mathcal{K}(\|x-y\|)|^2 d\mu_\varepsilon(t,x)d\mu_\varepsilon(t,y)dt,\end{aligned}$$

for a constant $c_{T,\varepsilon} > 0$ for $\varepsilon > 0$ small enough.

Summing up the present section, we showed that the coercivity condition holds for "generic" cases as well as for highly structured deterministic ones. In practice we can numerically verify that it holds in many situations, see Section 5.5.3, and from now on we assume it without further concerns.

5.4 Γ-convergence of $\mathcal{E}^{[a],N}$ to $\mathcal{E}^{[a]}$

This section is devoted to a proof of Theorem 5.5 using the methods of Γ-convergence. We start with its definition and several basic properties. We redirect the interested reader to the classical monograph [86] for further insights on the topic.

Definition 5.14 (Γ-convergence, [86, Definition 4.1, Proposition 8.1])**.** Let $\mathbb{X}$ be a metrizable separable space and consider for every $N \in \mathbb{N}$ a functional $\mathcal{J}_N : \mathbb{X} \to (-\infty, +\infty]$. We say that $(\mathcal{J}_N)_{N\in\mathbb{N}}$ Γ*-converges* to $\mathcal{J}$ for a given $\mathcal{J} : \mathbb{X} \to (-\infty, +\infty]$, whenever the following hold:

(*i*) $\Gamma - \liminf$ *condition:* for every $u \in \mathbb{X}$ and every sequence $u_N \to u$,

$$\mathcal{J}(u) \leq \liminf_{N\to\infty} \mathcal{J}_N(u_N).$$

(*ii*) $\Gamma - \limsup$ *condition:* for every $u \in \mathbb{X}$, there exists a sequence $u_N \to u$, called *recovery sequence*, such that

$$\mathcal{J}(u) \geq \limsup_{N\to\infty} \mathcal{J}_N(u_N).$$

We call the sequence $(\mathcal{J}_N)_{N\in\mathbb{N}}$ *equi-coercive* if for every $c \in \mathbb{R}$ there is a compact set $\Omega \subseteq \mathbb{X}$ such that $\{u \in \mathbb{X} : \mathcal{J}_N(u) \leq c\} \subseteq \Omega$ for all $N \in \mathbb{N}$.

As a direct consequence of Γ-convergence and equi-coercivity, assuming for all $N \in \mathbb{N}$ that $u^*_N \in \arg\min_{u\in\mathbb{X}} \mathcal{J}_N(u) \neq \emptyset$, we can infer that there exist a subsequence $(u^*_{N_k})_{k\in\mathbb{N}}$ and an element $u^* \in \mathbb{X}$ such that

$$u^*_{N_k} \to u^* \in \arg\min_{u\in\mathbb{X}} \mathcal{J}(u),$$

i.e., the minimizers of the functionals $\mathcal{J}_N$ converge to a minimizer of $\mathcal{J}$.
In order to establish the Γ-convergence of $\mathcal{E}^{[a],N}$ to $\mathcal{E}^{[a]}$, we start with the analysis of the discrete error functional $\mathcal{E}^{[a],N}$.

5.4.1 Existence of minimizers of $\mathcal{E}^{[a],N}$

The following result, which is a straightforward consequence of the Ascoli-Arzelá Theorem, indicates the right ambient space where to state an existence result for the minimizers of $\mathcal{E}^{[a],N}$.

Proposition 5.15. *Fix $M > 0$ and $\Omega = [0, 2R] \subset \mathbb{R}_+$ for any $R > 0$. Recall the set*

$$X_{M,\Omega} = \left\{ b \in W^{1,\infty}(\Omega) : \|b\|_{L^\infty(\Omega)} + \|b'\|_{L^\infty(\Omega)} \leq M \right\}.$$

The space $X_{M,\Omega}$ is relatively compact with respect to the uniform convergence on Ω.

Proof. Consider $(\widehat{a}_n)_{n\in\mathbb{N}} \subset X_{M,\Omega}$. The Fundamental Theorem of Calculus (applicable to functions in $W^{1,\infty}$, see [11, Theorem 2.8]) implies that the functions $\widehat{a}_n$ are all Lipschitz continuous with uniformly bounded Lipschitz constant, and are therefore equi-continuous.

Since they are also pointwise uniformly equi-bounded, by the Ascoli-Arzelá Theorem there exists a subsequence converging uniformly on Ω to some $\widehat{a} \in X_{M,\Omega}$. □

Proposition 5.16. *Assume $a \in X$. Fix $M > 0$ and $\Omega = [0, 2R] \subset \mathbb{R}_+$ for $R > 0$ as in Theorem 4.11. Let V be a closed subset of $X_{M,\Omega}$ with respect to the uniform convergence. Then, the optimization problem*

$$\min_{\widehat{a} \in V} \mathcal{E}^{[a],N}(\widehat{a})$$

admits a solution.

Proof. For proving the statement we apply the direct methods in the Calculus of Variations. Since $\inf \mathcal{E}^{[a],N} \geq 0$, we can consider a minimizing sequence $(\widehat{a}_n)_{n\in\mathbb{N}} \subset V$, i.e., such that

$$\lim_{n\to\infty} \mathcal{E}^{[a],N}(\widehat{a}_n) = \inf_{\widehat{a}\in V} \mathcal{E}^{[a],N}(\widehat{a}).$$

By Proposition 5.15 there exists a subsequence of $(\widehat{a}_n)_{n\in\mathbb{N}}$ (labelled again as $(\widehat{a}_n)_{n\in\mathbb{N}}$) converging uniformly on Ω to a function $\widehat{a} \in V$ (since V is closed). We now show that $\lim_{n\to\infty} \mathcal{E}^{[a],N}(\widehat{a}_n) = \mathcal{E}^{[a],N}(\widehat{a})$, from which it follows that the functional $\mathcal{E}^{[a],N}$ attains its minimum in V.

As a first step, notice that the uniform convergence of $(\widehat{a}_n)_{n\in\mathbb{N}}$ to $\widehat{a}$ on Ω and the compactness of Ω imply that the functionals $F^{[\widehat{a}_n]}(x-y)$ converge uniformly to $F^{[\widehat{a}]}(x-y)$ on $B(0,R) \times B(0,R)$ (where R is as in (4.15)) for $n \to \infty$. Moreover, we have the uniform bound

$$\begin{aligned}
\sup_{x,y\in B(0,R)} \|F^{[\widehat{a}_n]}(x-y) - F^{[a]}(x-y)\| &= \sup_{x,y\in B(0,R)} |\widehat{a}_n(\|x-y\|) - a(\|x-y\|)|\|x-y\| \\
&\leq 2R \sup_{r\in\Omega} |\widehat{a}_n(r) - a(r)| \\
&\leq 2R \left(M + \|a\|_{L^\infty(\Omega)}\right). \qquad (5.28)
\end{aligned}$$

As the measure-valued curves $\mu_N(\cdot)$ are compactly supported in $B(0,R)$ uniformly in time, the bound (5.28) allows us to apply three times the dominated convergence theorem to obtain

$$\begin{aligned}
\lim_{n\to\infty} \mathcal{E}^{[a],N}(\widehat{a}_n) &= \lim_{n\to\infty} \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left\| \int_{\mathbb{R}^d} \left(F^{[\widehat{a}_n]}(x-y) - F^{[a]}(x-y)\right) d\mu_N(t,y) \right\|^2 d\mu_N(t,x)dt \\
&= \frac{1}{T}\int_0^T \lim_{n\to\infty} \int_{\mathbb{R}^d} \left\| \int_{\mathbb{R}^d} \left(F^{[\widehat{a}_n]}(x-y) - F^{[a]}(x-y)\right) d\mu_N(t,y) \right\|^2 d\mu_N(t,x)dt \\
&= \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left\| \lim_{n\to\infty} \int_{\mathbb{R}^d} \left(F^{[\widehat{a}_n]}(x-y) - F^{[a]}(x-y)\right) d\mu_N(t,y) \right\|^2 d\mu_N(t,x)dt \\
&= \frac{1}{T}\int_0^T \int_{\mathbb{R}^d} \left\| \int_{\mathbb{R}^d} \left(F^{[\widehat{a}]}(x-y) - F^{[a]}(x-y)\right) d\mu_N(t,y) \right\|^2 d\mu_N(t,x)dt \\
&= \mathcal{E}^{[a],N}(\widehat{a}),
\end{aligned}$$

which proves the statement. □

5.4.2 Uniform convergence estimates

The following technical result is the main ingredient of the forthcoming Γ-convergence result in Corollary 5.18.

Lemma 5.17. *Under the assumptions of Theorem 5.5, let* $(b_N)_{N\in\mathbb{N}} \subset X_{M,\Omega}$ *be a sequence of continuous functions and* $b \in X_{M,\Omega}$*, for* $\Omega = [0, 2R]$ *with* $R > 0$ *as in* (4.15)*. Then there exists* $c_1, c_2 > 0$ *independent of* N *for which it holds*

$$\left|\mathcal{E}^{[a],N}(b_N) - \mathcal{E}^{[a]}(b)\right| \le c_1 \mathcal{W}_1(\mu_N^0, \mu^0) + c_2\|b_N - b\|_{L^\infty(\Omega)}. \tag{5.29}$$

Proof. By Dobrushin's inequality (4.22) it holds

$$\mathcal{W}_1(\mu(t), \mu_N(t)) \le \overline{C}\mathcal{W}_1(\mu^0, \mu_N^0)$$

uniformly in $t \in [0, T]$ for some constant $\overline{C} > 0$ independent of N. Moreover, for all $x, y, y' \in B(0, R)$, by the triangle inequality we have

$$\begin{aligned}\|(F^{[a]} - F^{[b]})(x - y') - &(F^{[a]} - F^{[b]})(x - y)\| \le \\ &\le \Big(2R\big(\operatorname{Lip}_\Omega(a) + \operatorname{Lip}_\Omega(b)\big) + \|a\|_{L^\infty(\Omega)} + \|b\|_{L^\infty(\Omega)}\Big)\|y - y'\|,\end{aligned}$$

which implies the Lipschitz continuity of $(F^{[a]} - F^{[b]})(x - \cdot)$ in $B(0, R)$ for fixed $x \in B(0, R)$. Since $a, b \in X_{M,K}$, it follows

$$\operatorname{Lip}_{B(0,R)}\Big(\|(F^{[a]} - F^{[b]})(x - \cdot)\|\Big) \le 2(2R + 1)M, \tag{5.30}$$

uniformly with respect to $x \in B(0, R)$. Consequently, we have

$$\begin{aligned}\left\|\int_{\mathbb{R}^d}\big(F^{[b]} - F^{[a]}\big)(x - y)d\mu_N(t, y) - \right.&\left.\int_{\mathbb{R}^d}\big(F^{[b]} - F^{[a]}\big)(x - y)d\mu(t, y)\right\| \le \\ &\le \operatorname{Lip}_{B(0,R)}\Big(\|(F^{[a]} - F^{[b]})(x - \cdot)\|\Big)\,\mathcal{W}_1(\mu_N(t), \mu(t)) \\ &\le 2\overline{C}(2R + 1)M\,\mathcal{W}_1(\mu_N^0, \mu^0),\end{aligned}$$

uniformly with respect to $t \in [0, T]$ and $x \in B(0, R)$. Furthermore, we also have the estimates

$$\sup_{x,y\in B(0,R)} \|F^{[b_N]}(x - y) - F^{[b]}(x - y)\| \le 2R\|b_N - b\|_{L^\infty(\Omega)}, \tag{5.31}$$

$$\sup_{x,y\in B(0,R)} \|F^{[a]}(x - y) - F^{[b]}(x - y)\| \le 2R\|a - b\|_{L^\infty(\Omega)}. \tag{5.32}$$

Hence we further obtain

$$\begin{aligned}&\left|\left\|\int_{\mathbb{R}^d}\big(F^{[b_N]} - F^{[a]}\big)(x - y)d\mu_N(t, y)\right\| - \left\|\int_{\mathbb{R}^d}\big(F^{[b]} - F^{[a]}\big)(x - y)d\mu(t, y)\right\|\right| \le \\ &\qquad \le \left\|\int_{\mathbb{R}^d}\big(F^{[b_N]} - F^{[a]}\big)(x - y)d\mu_N(t, y) - \int_{\mathbb{R}^d}\big(F^{[b]} - F^{[a]}\big)(x - y)d\mu(t, y)\right\|\end{aligned}$$

$$
\begin{aligned}
&\leq \left\| \int_{\mathbb{R}^d} \big(F^{[b_N]} - F^{[b]}\big)(x-y) d\mu_N(t,y) \right\| + \\
&\quad + \left\| \int_{\mathbb{R}^d} \big(F^{[b]} - F^{[a]}\big)(x-y) d\mu_N(t,y) - \int_{\mathbb{R}^d} \big(F^{[b]} - F^{[a]}\big)(x-y) d\mu(t,y) \right\| \\
&\leq 2R\|b_N - b\|_{L^\infty(\Omega)} \int_{\mathbb{R}^d} d\mu_N(t,y) + 2(2R+1)M\, \mathcal{W}_1(\mu_N(t), \mu(t)) \\
&\leq 2R\|b_N - b\|_{L^\infty(\Omega)} + 2\overline{C}(2R+1)M\, \mathcal{W}_1(\mu_N^0, \mu^0). \qquad (5.33)
\end{aligned}
$$

We define for every $(t,x) \in [0,T] \times B(0,R)$ the following quantities

$$
H_N(t,x) \triangleq \left\| \int_{\mathbb{R}^d} \big(F^{[b_N]} - F^{[a]}\big)(x-y) d\mu_N(t,y) \right\|^2, \qquad G_N(t) \triangleq \int_{\mathbb{R}^d} H_N(t,x) d\mu_N(t,x),
$$

$$
H(t,x) \triangleq \left\| \int_{\mathbb{R}^d} \big(F^{[b]} - F^{[a]}\big)(x-y) d\mu(t,y) \right\|^2, \qquad G(t) \triangleq \int_{\mathbb{R}^d} H(t,x) d\mu(t,x).
$$

Then immediately it follows

$$
\begin{aligned}
|G_N(t) - G(t)| \leq & \left| \int_{\mathbb{R}^d} H(t,x) d\mu_N(t,x) - \int_{\mathbb{R}^d} H(t,x) d\mu(t,x) \right| + \\
& + \int_{\mathbb{R}^d} |H_N(t,x) - H(t,x)|\, d\mu_N(t,x).
\end{aligned} \qquad (5.34)
$$

From (5.32) and (5.30) we obtain

$$
\begin{aligned}
\mathrm{Lip}_{B(0,R)}(H(t,\cdot)) &\leq 2 \left(\sup_{x,y \in B(0,R)} \|F^{[a]}(x-y) - F^{[b]}(x-y)\| \right) \times \\
&\quad \times \mathrm{Lip}_{B(0,R)} \Big((F^{[a]} - F^{[b]})(\cdot - y) \Big) \\
&\leq 4R\|a-b\|_{L^\infty(\Omega)} \cdot 2(2R+1)M,
\end{aligned}
$$

and therefore

$$
\begin{aligned}
\left| \int_{\mathbb{R}^d} H(t,x) d\mu_N(t,x) - \int_{\mathbb{R}^d} H(t,x) d\mu(t,x) \right| &\leq \mathrm{Lip}_{B(0,R)}(H(t,\cdot)) \mathcal{W}_1(\mu_N(t), \mu(t)) \\
&\leq 8R(2R+1)\overline{C}M\|a-b\|_{L^\infty(\Omega)} \mathcal{W}_1(\mu_N^0, \mu^0) \\
&\leq 16R(2R+1)\overline{C}M^2 \mathcal{W}_1(\mu_N^0, \mu^0)
\end{aligned} \qquad (5.35)
$$

uniformly in $t \in [0,T]$. Similarly, (5.31), (5.30), and (5.33) imply

$$
\begin{aligned}
\big|H_N(t,x) - H(t,x)\big| &\leq \Big(2R\|b_N - b\|_{L^\infty(\Omega)} + 2\overline{C}(2R+1)M\, \mathcal{W}_1(\mu_N^0, \mu^0) \Big) \times \\
&\quad \times 2R \Big(\|b_N - a\|_{L^\infty(\Omega)} + \|b - a\|_{L^\infty(\Omega)} \Big) \\
&\leq 8RM \Big(2R\|b_N - b\|_{L^\infty(\Omega)} + 2\overline{C}(2R+1)M\, \mathcal{W}_1(\mu_N^0, \mu^0) \Big)
\end{aligned} \qquad (5.36)
$$

uniformly in $t \in [0,T]$ and $x \in B(0,R)$. A combination of (5.34), (5.35) and (5.36) yields

$$|G_N(t) - G(t)| \leq 32\overline{C}R(2R+1)M^2\mathcal{W}_1(\mu_N^0, \mu^0) + 16R^2M\|b_N - b\|_{L^\infty(\Omega)}$$

uniformly in $t \in [0,T]$. Thus we finally arrive at

$$\begin{aligned}\left|\mathcal{E}^{[a],N}(b_N) - \mathcal{E}^{[a]}(b)\right| &= \left|\frac{1}{T}\int_0^T \big(G_N(t) - G(t)\big)dt\right| \\ &\leq 32\overline{C}R(2R+1)M^2\mathcal{W}_1(\mu_N^0, \mu^0) + 16R^2M\|b_N - b\|_{L^\infty(\Omega)}.\end{aligned}$$

This proves the claim for $c_1 = 32\overline{C}R(2R+1)M^2$ and $c_2 = 16R^2M$. □

As a corollary, we immediately obtain the following Γ-convergence result. Notice that we implicitly use the identity

$$\min_{\widehat{a}\in V_N} \mathcal{E}^{[a],N}(\widehat{a}) = \min_{\widehat{a}\in X_{M,\Omega}} \left(\mathcal{E}^{[a],N}(\widehat{a}) + I_{V_N}(\widehat{a})\right),$$

where, for any $V \subseteq X_{M,\Omega}$, the *indicator function* $I_V : X_{M,\Omega} \to \{0, +\infty\}$ is defined as

$$I_V(b) = \begin{cases} 0 & \text{if } b \in V, \\ +\infty & \text{if } b \notin V. \end{cases}$$

Corollary 5.18. *Under the assumptions of Theorem 5.5, the sequence of functionals* $(\mathcal{J}_N)_{N\in\mathbb{N}}$ *defined for every* $b \in X_{M,\Omega}$ *as*

$$\mathcal{J}_N(b) \triangleq \mathcal{E}^{[a],N}(b) + I_{V_N}(b)$$

Γ-converges to the functional

$$\mathcal{J}(b) \triangleq \mathcal{E}^{[a]}(b).$$

Proof. First notice that, under the assumptions of Theorem 5.5, if $(b_N)_{N\in\mathbb{N}} \subset X_{M,\Omega}$ is a sequence of continuous functions uniformly converging to $b \in X_{M,\Omega}$ on $\Omega = [0, 2R]$ with $R > 0$ as in (4.15), then it holds

$$\lim_{N\to\infty} \mathcal{E}^{[a],N}(b_N) = \mathcal{E}^{[a]}(b). \tag{5.37}$$

Indeed, this follows immediately from the estimate (5.29), upon noticing that we have $\mathcal{W}_1(\mu^0, \mu_N^0) \to 0$ for $N \to \infty$ as a consequence of Remark 4.12.

Therefore, we start by fixing the metrizable space of Definition 5.14 as $\mathbb{X} \triangleq X_{M,\Omega}$ endowed with the metric of uniform convergence, and we pass to prove the $\Gamma - \liminf$ condition. Let us fix a uniformly convergent sequence of functions $b_N \to b$ in $X_{M,\Omega}$. Then, since $b \in X_{M,\Omega}$, it trivially follows

$$0 = I_{X_{M,\Omega}}(b) \leq \liminf_{N\to\infty} I_{V_N}(b_N),$$

whence from (5.37)

$$\mathcal{E}^{[a]}(b) = \mathcal{E}^{[a]}(b) + I_{X_{M,\Omega}}(b) \leq \liminf_{N\to\infty}\left(\mathcal{E}^{[a],N}(b_N) + I_{V_N}(b_N)\right).$$

We now prove the $\Gamma - \limsup$ condition. We fix $b \in X_{M,\Omega}$ and consider the recovery sequence $(b_N)_{N\in\mathbb{N}}$ guaranteed by the uniform approximation property (see Definition 5.4). This implies that $b_N \in V_N$ for every $N \in \mathbb{N}$ and $b_N \to b$ uniformly on Ω. Therefore, since $I_{V_N}(b_N) = 0$ for every $N \in \mathbb{N}$, by (5.37) again follows

$$\mathcal{E}^{[a]}(b) = \mathcal{E}^{[a]}(b) + I_{X_{M,\Omega}}(b) \geq \limsup_{N\to\infty} \left(\mathcal{E}^{[a],N}(b_N) + I_{V_N}(b_N)\right),$$

and this proves the claimed Γ-convergence. □

5.4.3 Proof of Theorem 5.5

The sequence of minimizers $(\widehat{a}_N)_{N\in\mathbb{N}}$ is by definition a subset of $X_{M,\Omega}$, which is a compact set with respect to the uniform convergence by Proposition 5.15. Therefore, the sequence $(\mathcal{J}_N)_{N\in\mathbb{N}}$ is equicoercive, and $(\widehat{a}_N)_{N\in\mathbb{N}}$ admits a subsequence $(\widehat{a}_{N_k})_{k\in\mathbb{N}}$ uniformly converging to a function $\widehat{a} \in X_{M,\Omega}$.

To show the optimality of $\widehat{a}$ in $X_{M,\Omega}$ we need to prove that it is actually a minimizer of $\mathcal{E}^{[a]}$. To this end, we use the fact that $\mathcal{E}^{[a]}$ is the Γ-limit of the sequence $(\mathcal{E}^{[a],N} + I_{V_N})_{N\in\mathbb{N}}$ as proved in Corollary 5.18. Let $b \in X_{M,\Omega}$ be an arbitrary function and let $(b_N)_{N\in\mathbb{N}}$ be a recovery sequence given by the $\Gamma - \limsup$ condition, so that

$$\mathcal{E}^{[a]}(b) \geq \limsup_{N\to\infty} \left(\mathcal{E}^{[a],N}(b_N) + I_{V_N}(b_N)\right),$$

for every $N \in \mathbb{N}$. By the optimality of $\widehat{a}_{N_k}$ in V_{N_k} and the $\Gamma - \liminf$ condition, it follows

$$\begin{aligned}
\mathcal{E}^{[a]}(b) &\geq \limsup_{k\to\infty} \left(\mathcal{E}^{[a],N_k}(b_{N_k}) + I_{V_{N_k}}(b_{N_k})\right) \\
&\geq \liminf_{k\to\infty} \left(\mathcal{E}^{[a],N_k}(b_{N_k}) + I_{V_{N_k}}(b_{N_k})\right) \\
&\geq \liminf_{k\to\infty} \left(\mathcal{E}^{[a],N_k}(\widehat{a}_{N_k}) + I_{V_{N_k}}(\widehat{a}_{N_k})\right) \\
&\geq \mathcal{E}^{[a]}(\widehat{a}).
\end{aligned}$$

We can therefore conclude that for every $b \in X_{M,\Omega}$

$$\mathcal{E}^{[a]}(b) \geq \mathcal{E}^{[a]}(\widehat{a}). \tag{5.38}$$

In particular, (5.38) applies to $b = a \in X_{M,\Omega}$ (by the particular choice of M), which finally implies

$$0 = \mathcal{E}^{[a]}(a) \geq \mathcal{E}^{[a]}(\widehat{a}) \geq 0 \quad \Longrightarrow \quad \mathcal{E}^{[a]}(\widehat{a}) = 0,$$

showing that $\widehat{a}$ is also a minimizer of $\mathcal{E}^{[a]}$. When the coercivity condition (5.16) holds, it follows that $\widehat{a} = a$ in $L^2(\mathbb{R}_+, \rho)$, by Proposition 5.12.

Assume now that (5.17) and (5.18) hold together with (5.16). Then, by these latter conditions, two applications of (5.29), the minimality of $\widehat{a}_N$, and the optimality of a in the sense that $\mathcal{E}^{[a]}(a) = 0$, we obtain the following chain of estimates

$$\|\widehat{a}_N - a\|^2_{L^2(\mathbb{R}_+,\rho)} \leq \frac{1}{c_T}\mathcal{E}^{[a]}(\widehat{a}_N)$$

$$
\begin{aligned}
&\leq \frac{1}{c_T}\left(\mathcal{E}^{[a],N}(\widehat{a}_N) + (\mathcal{E}^{[a]}(\widehat{a}_N) - \mathcal{E}^{[a],N}(\widehat{a}_N))\right)\\
&\leq \frac{1}{c_T}\left(\mathcal{E}^{[a],N}(\widehat{a}_N) + c_1\mathcal{W}_1(\mu_N^0,\mu^0)\right)\\
&\leq \frac{1}{c_T}\left(\mathcal{E}^{[a],N}(a_N) + \underbrace{\mathcal{E}^{[a]}(a)}_{=0} + c_1\mathcal{W}_1(\mu_N^0,\mu^0)\right)\\
&\leq \frac{1}{c_T}\left(2c_1\mathcal{W}_1(\mu_N^0,\mu^0) + c_2\|a-a_N\|_{L^\infty(\Omega)}\right)\\
&\leq C_3 N^{-\min\{\alpha,\beta\}}.
\end{aligned}
$$

This concludes the proof.

5.5 Numerical experiments

In this section we report several numerical experiments to document the validity and applicability of Theorem 5.5. We will first show how the reconstruction of the unknown kernel a gets better as the number of agents N increases, in accordance with the Γ-convergence result reported in Section 5.4. This feature holds true also for at least some interaction kernels not lying in the function space X, as shown in Figure 5.3. We will then investigate empirically the validity of the coercivity condition (5.16) comparing the functional $\mathcal{E}^{[a],N}(\widehat{a}_N)$ with $\|a-\widehat{a}_N\|^2_{L^2(\mathbb{R}_+,\rho_N)}$ where ρ_N is constructed as ρ but referring to the empirical measures μ_N. Then we address the behavior of $\mathcal{E}^{[a],N}(\widehat{a}_{N,M})$ for N fixed, while we let the constraint constant M vary (here $\widehat{a}_{N,M} = \widehat{a}_N$). Finally, we show how we can get a very satisfactory reconstruction of the unknown interaction kernel by keeping N fixed and averaging the minimizers of the functional $\mathcal{E}^{[a],N}$ obtained from several samples of the initial data distribution μ^0.

5.5.1 Numerical framework

All experiments rely on a common numerical set-up, which we pass to clarify. All the initial data μ_N^0 are drawn from a common probability distribution μ^0 which is the uniform distribution on the d-dimensional cube $[-L,L]^d$. For every μ_N^0, we simulate the evolution of the system starting from μ_N^0 until time T, and we shall denote by R the maximal distance between particles reached during the time frame $[0,T]$. Notice that we have at our disposal only a finite sequence of snapshots of the dynamics: if we denote by $0 = t_0 < t_1 < \ldots < t_m = T$ the time instants at which these snapshots are taken, we can consider the *discrete-time error functional*

$$
\mathcal{E}_\Delta^{[a],N}(\widehat{a}) \triangleq \frac{1}{m}\sum_{k=1}^{m}\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{1}{N}\sum_{i=1}^{N}\widehat{a}(\|x_j(t_k)-x_i(t_k)\|)(x_j(t_k)-x_i(t_k)) - \dot{x}_i(t_k)\right\|^2,
$$

which is the time-discrete counterpart of the continuous-time error functional $\mathcal{E}^{[a],N}$. As already mentioned in the introduction, derivatives $\dot{x}_i(t_k)$ appearing in $\mathcal{E}_\Delta^{[a],N}$ are actually approximated by finite differences: in our experiments we use the simplest approximation

$$
\dot{x}_i(t_k) \triangleq \frac{x_i(t_k)-x_i(t_{k-1})}{t_k-t_{k-1}} \quad \text{for every } k \geq 1.
$$

Regarding the reconstruction procedure, we fix the constraint level $M > 0$ and consider the sequence of invading subspaces V_N of $X_{M,\Omega}$ ($\Omega = [0, 2R]$ here) generated by a B-spline basis with $D(N)$ elements supported on $[0, 2R]$: for every element $\widehat{a} \in V_N$ it holds

$$\widehat{a}(r) = \sum_{\lambda=1}^{D(N)} a_\lambda \varphi_\lambda(r) \qquad \text{for every } r \in [0, 2R].$$

In order for V_N to increase in N and invade $X_{M,\Omega}$, we let $D(N)$ be a strictly increasing function of N. For the sake of simplicity, we shall employ a *linear uniform* B-spline basis supported on the interval $[0, 2R]$ with 0-smoothness conditions at the boundary, see [89]. Whenever $\widehat{a} \in V_N$, we can rewrite the functional $\mathcal{E}_\Delta^{[a],N}$ as

$$\begin{aligned}
\mathcal{E}_\Delta^{[a],N}(\widehat{a}) &= \frac{1}{m}\sum_{k=1}^{m}\frac{1}{N}\sum_{j=1}^{N}\left\|\frac{1}{N}\sum_{i=1}^{N}\sum_{\lambda=1}^{D(N)} a_\lambda \varphi_\lambda(\|x_j(t_k) - x_i(t_k)\|)(x_j(t_k) - x_i(t_k)) - \dot{x}_i(t_k)\right\|^2 \\
&= \frac{1}{m}\sum_{k=1}^{m}\frac{1}{N}\sum_{j=1}^{N}\left\|\sum_{\lambda=1}^{D(N)} a_\lambda \frac{1}{N}\sum_{i=1}^{N}\varphi_\lambda(\|x_j(t_k) - x_i(t_k)\|)(x_j(t_k) - x_i(t_k)) - \dot{x}_i(t_k)\right\|^2 \\
&= \frac{1}{mN}\left\|\mathbf{C}\vec{a} - \vec{v}\right\|_{\ell_2}^2,
\end{aligned}$$

where $\vec{a} = (a_1, \ldots, a_{D(N)})$, $\vec{v} = (\dot{x}_1(t_1), \ldots, \dot{x}_N(t_1), \ldots, \dot{x}_1(t_m), \ldots, \dot{x}_N(t_m))$ and the tensor $\mathbf{C} \in \mathbb{R}^{d \times Nm \times D(N)}$ satisfies for every $j = 1, \ldots, N$, $k = 1, \ldots, m$, $\lambda = 1, \ldots, D(N)$

$$\mathbf{C}(jk, \lambda) = \frac{1}{N}\sum_{i=1}^{N}\varphi_\lambda(\|x_j(t_k) - x_i(t_k)\|)(x_j(t_k) - x_i(t_k)) \in \mathbb{R}^d.$$

We shall numerically implement the constrained minimization with the software CVX [109, 110], which allows constraints and objectives to be specified using standard MATLAB expression syntax. In order to use it, we need to rewrite the constraint of our minimization problem, which reads

$$\|a\|_{L^\infty([0,2R])} + \|a'\|_{L^\infty([0,2R])} \leq M,$$

using only the minimization variable of the problem, which is the vector of coefficients of the B-spline basis $\vec{a}$. Notice that the property of being a linear B-spline basis implies that, for every $\lambda = 1, \ldots, D(N) - 1$, the property $\operatorname{supp}(\varphi_\lambda) \cap \operatorname{supp}(\varphi_{\lambda+j}) \neq \emptyset$ holds if and only if $j = 1$. Hence, for every $a \in V_N$ we have

$$\begin{aligned}
\|a\|_{L^\infty([0,2R])} &= \max_{r \in [0,2R]}\left|\sum_{\lambda=1}^{D(N)} a_\lambda \varphi_\lambda(r)\right| \leq \max_{1 \leq \lambda \leq D(N)-1}(|a_\lambda| + |a_{\lambda+1}|) \leq 2\|\vec{a}\|_{\ell_\infty}, \\
\|a'\|_{L^\infty([0,2R])} &= \max_{r \in [0,2R]}\left|\sum_{\lambda=1}^{D(N)} a_\lambda \varphi'_\lambda(r)\right| \leq \max_{1 \leq \lambda \leq D(N)-1}|a_{\lambda+1} - a_\lambda| = \|\mathbf{B}\vec{a}\|_{\ell_\infty},
\end{aligned}$$

where, in the last line, $\mathbf{B}$ is the standard finite difference matrix

$$\mathbf{B} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \\ 0 & 0 & 0 & \dots & 1 & -1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

We therefore replace the constrained minimization problem

$$\min_{\widehat{a}\in V_N} \mathcal{E}^{[a],N}(\widehat{a}) \quad \text{subject to} \quad \|\widehat{a}\|_{L^\infty([0,2R])} + \|\widehat{a}'\|_{L^\infty([0,2R])} \leq M,$$

by

$$\min_{\vec{a}\in\mathbb{R}^{D(N)}} \frac{1}{mN} \|\mathbf{C}\vec{a} - \vec{v}\|_{\ell_2}^2 \quad \text{subject to} \quad 2\|\vec{a}\|_{\ell_\infty} + \|\mathbf{B}\vec{a}\|_{\ell_\infty} \leq M\,, \tag{5.39}$$

which has weaker constraints, but is amenable to numerical solution. The byproduct of the time discretization and the reformulation of the constraint is that minimizers of problem (5.39) may not be precisely the minimizers of the original one. This is the price to pay for this simple numerical implementation of the L^∞-constraints. Despite such a crude discrete model, we still observe all the approximation properties proved in the previous sections and the implementation results both simple and effective.

5.5.2 Varying N

In Figure 5.2 we show the reconstruction of a truncated Lennard-Jones type interaction kernel obtained with different values of N. Table 5.1 reports the values of the different parameters.

d	L	T	M	N	$D(N)$
2	3	0.5	100	$[10, 20, 40, 80]$	$2N$

Table 5.1: Parameter values for Figure 5.2 and Figure 5.3.

It is clearly visible how the piecewise linear approximant (displayed in blue) gets closer and closer to the potential to be recovered (in red), as predicted by the theoretical results of the previous sections. What is however surprising is that the same behavior is witnessed in Figure 5.3, where the algorithm is applied to an interaction kernel a not belonging to the function space X (due to its singularity at the origin) with the same specifications reported in Table 5.1. In particular, the algorithm performs an excellent approximation despite the highly oscillatory nature of the function a and produce a natural numerical homogenization when the discretization is not fine enough.

5.5.3 Numerical validation of the coercivity condition

We now turn our attention to the coercivity constant c_T appearing in (5.16) and thoroughly discussed in Section 5.3. In Figure 5.4 we see a comparison between the evolution of the

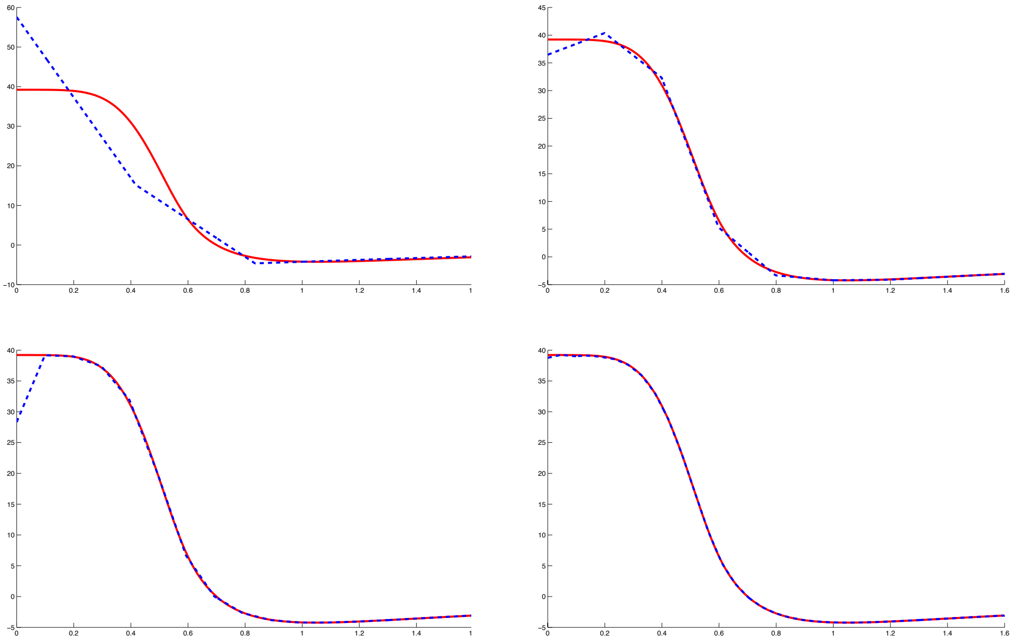


Figure 5.2: Iterative reconstruction of a potential with different values of N. In red: the unknown kernel. In blue: its reconstruction by minimization of $\mathcal{E}^{[a],N}$. From left-top to right-bottom: reconstruction with $N = 10, 20, 40, 80$ agents. We notice that the uniform convergence at 0 is slower in view of the quadratic polynomial weight r^2 as in (5.14) and because less information is actually gathered around 0.

value of the error functional $\mathcal{E}^{[a],N}_{\Delta}(\widehat{a}_N)$ and of the $L^2(\mathbb{R}_+, \rho_N)$-error $\|a - \widehat{a}_N\|^2_{L^2(\mathbb{R}_+,\rho_N)}$ for different values of N.

In this experiment, the potential a to be retrieved is the truncated Lennard-Jones type interaction kernel of Figure 5.2 and the parameters used in the algorithm are reported in Table 5.2.

d	L	T	M	N	$D(N)$
2	5	0.5	100	$[3, 4, \dots, 12]$	$3N - 5$

Table 5.2: Parameter values for Figure 5.4.

For every value of N, we have obtained the minimizer $\widehat{a}_N$ of problem (5.39) and we have computed the errors $\mathcal{E}^{[a],N}(\widehat{a}_N)$ and $\|a - \widehat{a}_N\|^2_{L^2(\mathbb{R}_+,\rho_N)}$. The $L^2(\mathbb{R}_+, \rho_N)$-error multiplied by a factor $1/10$ lies entirely below the curve of $\mathcal{E}^{[a],N}(\widehat{a}_N)$, which let us empirically estimate the value of c_T around that value (see Figures 5.4 and 5.5).

5.5.4 Tuning the constraint M

Figure 5.6 shows what happens when we modify the value of M in problem (5.39). More specifically, we generate μ^0_N as explained in Section 5.5.1 once, and we simulate the system starting from μ^0_N until time T. With the data of this single evolution, we solve problem

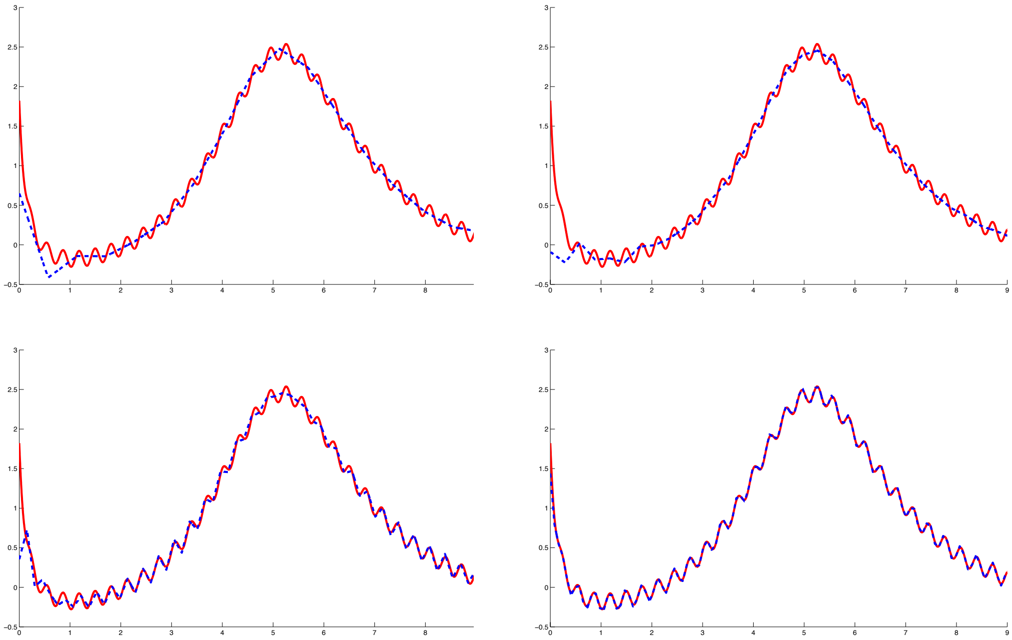


Figure 5.3: Iterative reconstruction of a potential with a singularity at the origin and highly oscillatory behavior. In red: the unknown kernel. In blue: its reconstruction by minimization of $\mathcal{E}^{[a],N}$. From left-top to right-bottom: reconstruction with $N = 10, 20, 40, 80$ agents.

(5.39) for several values of M and we denote with $\widehat{a}_M \triangleq \widehat{a}_{N,M} \triangleq \widehat{a}_N$ the minimizer obtained with a specific value of M. On the left side of Figure 5.6 we show how the reconstruction $\widehat{a}_M$ gets closer and closer to the true potential a (in white) as M increases, while on the right side we illustrate how the original trajectories (again, in white) used for the inverse problem are approximated better and better by those generated with the computed approximation $\widehat{a}_M$, if we let M grow. Table 5.3 reports the values of the parameters of these experiments.

	d	L	T	M	N	$D(N)$
First row	2	3	1	$2.7 \times [10, 15, \ldots, 40]$	20	60
Second row	2	3	1	$1.25 \times [10, 15, \ldots, 40]$	20	150

Table 5.3: Parameter values for Figure 5.6.

So far we have no *a priori* criteria to sieve those values of M, which enable a successful reconstruction of a potential $a \in X$. However, the tuning *a posteriori* of the parameter $M > 0$ turns out to be rather easy. In fact, for N fixed the minimizers $\widehat{a}_{N,M}$ have the property that the map

$$M \mapsto \mathcal{E}^{[a],N}(\widehat{a}_{N,M})$$

is monotonically decreasing as a function of the constraint parameter M and it becomes constant for $M \geq M^*$, for a certain $M^* > 0$ which, as shown empirically, does not depend on N. This special value M^* is indeed the "right" parameter for the L^∞ bound. For such a choice, we show that, if we let N grow, the minimizers $\widehat{a}_N$ approximates better and better the unknown potential a. Figure 5.7 documents precisely this expected behavior.

Figure 5.4: Plot in logarithmic scale of $\mathcal{E}^{[a],N}(\widehat{a}_N)$ and $1/10 \cdot \|a - \widehat{a}_N\|^2_{L^2(\mathbb{R}_+,\rho_N)}$ for different values of N. In this experiment, we can estimate the constant c_T with $1/10$.

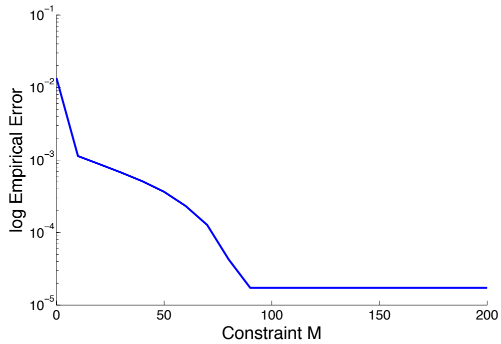


Figure 5.5: Values in logarithmic scale of $\mathcal{E}^{[a],N}_{\Delta}(\widehat{a}_{N,M})$ for fixed $N = 50$ for different values of $M \in [0, 200]$.

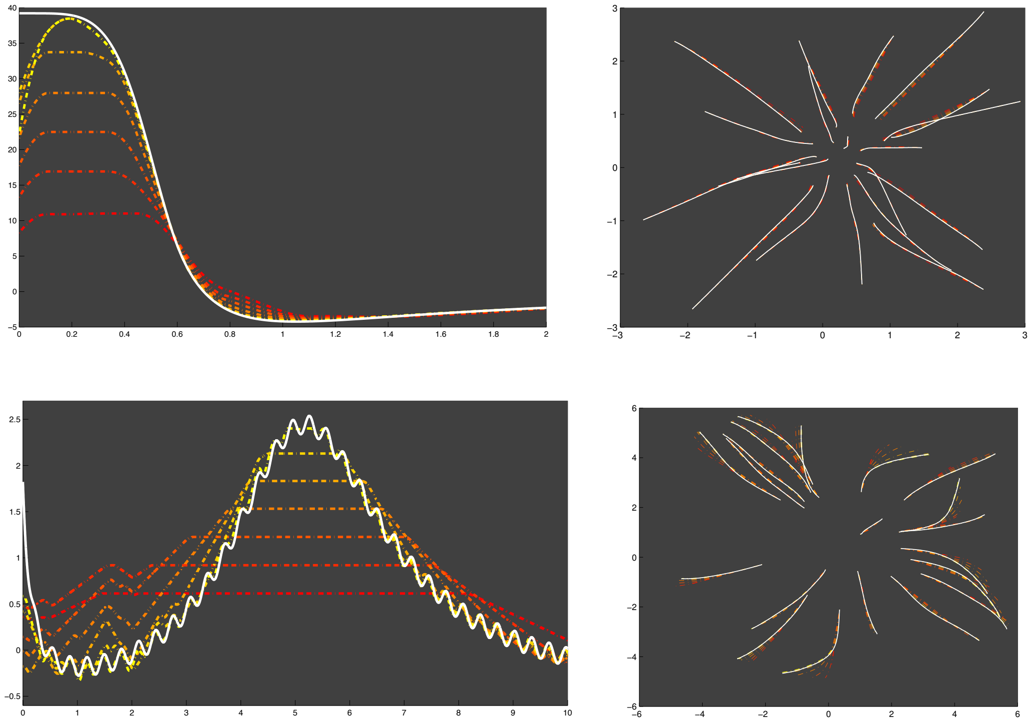


Figure 5.6: Different reconstructions of a potential for different values of M. On the left column: the true kernel in white and its reconstructions for different M; the brighter the curve, the larger the M. On the right column: the true trajectories of the agents in white, the trajectories associated to the reconstructed potentials with the same color.

5.5.5 Monte Carlo-like reconstructions for N fixed

We mimic now the mean-field reconstruction strategy, by multiple randomized draw of N particles as initial conditions i.i. distributed according to μ^0 for N fixed and relatively small. Indeed, problem (5.39) can swiftly become computationally unfeasible when N is moderately large, also because the dimension of the approximating subspaces V_N needs to increase with N too. We consider, for a fixed N, several discrete initial data $(\mu^0_{N,\theta})^{\Theta}_{\theta=1}$ all independently drawn from the same distribution μ^0 (in our case, the Lebsgue measure $\mathcal{L}^d$ restricted to the cube $[-L, L]^d$). For every $\theta = 1, \dots, \Theta$, we simulate the system until time T and, with the trajectories we obtained, we solve problem (5.39). At the end of this procedure, we have a family of reconstructed potentials $(\widehat{a}_{N,\theta})^{\Theta}_{\theta=1}$, all approximating the same true kernel a. Empirically averaging these potentials, we obtain an approximation

$$\widehat{a}_N(r) = \frac{1}{\Theta}\sum_{\theta=1}^{\Theta}\widehat{a}_{N,\theta}(r) \qquad \text{for every } r \in [0, 2R],$$

which we claim to be a better approximation to the true kernel a than any single snapshot. To support this claim, we report in Figure 5.8 the outcome of an experiment whose data can be found in Table 5.4.

d	L	T	M	N	$D(N)$	Θ
2	2	0.5	1000	50	150	5

Table 5.4: Parameter values for the experiment of Figure 5.8.

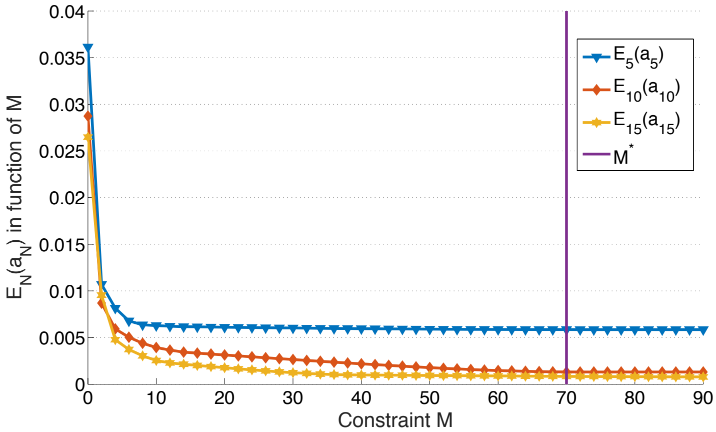


Figure 5.7: Behavior of $\mathcal{E}^{[a],N}(\widehat{a}_{N,M})$ as a function of the constraint M for different N.

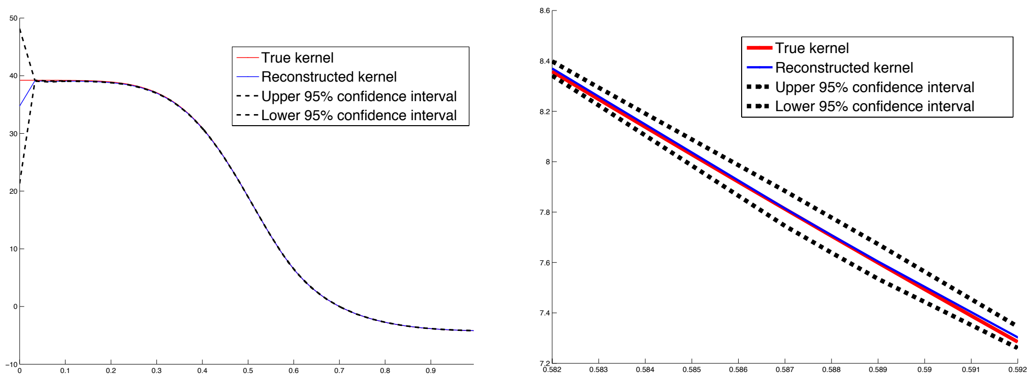


Figure 5.8: Reconstruction of a obtained by averaging 5 solutions of the minimization of $\mathcal{E}^{[a],N}_{\Delta}$ for $N = 50$. In red: the unknown kernel. In blue: the average of reconstructions. In black: 95% confidence interval for the parameter estimates returned by the Matlab function `normfit`. The figure on the right shows a zoom of the left figure.

CHAPTER 6

Mean-field sparse optimal control

The results obtained in Chapter 2 on the optimality of sparse controls were conditional to the restriction on our ability to forecast the evolution of the dynamical system. When this limitation is lifted, we gain the power to predict the reaction of such systems to external stimuli, and hence to address the problem of finding the *optimal control strategy* to steer the system to consensus. However, the optimal strategy may not be sparse at all, and moreover, due to the curse of dimensionality, as the number of agents increases the computational complexity of such strategies may blow up very fast.

To overcome these problems, in this chapter we shall introduce the notion of *mean-field sparse optimal control*, which are sparse optimal strategies that behave well in the regime $N \to \infty$. We shall first discuss under which conditions such controls exist and then provide optimality conditions for their computation.

6.1 Mean-field optimal controls

Consider again the case of the controlled Cucker-Smale system (2.1) (which we rewrite for convenience)

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} a\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right) + u_i(t), \end{cases} \qquad i = 1, \dots, N. \tag{6.1}$$

In Chapter 2, we identified the sparse feedback control strategy of Definition 2.4 to be the “best” (in terms of maximizing the rate of convergence to the consensus region) parsimonious control strategy *if the policy maker is not allowed to forecast the evolution of the system.* However, when this restriction is removed, we can easily show with the following example that sparser and more economical control strategies steering the system to consensus exist.

Example 6.1. Consider an instance of the Cucker-Smale system (6.1) without control in dimension $d = 1$ with $N = 2$ agents, where the interaction function $a : \mathbb{R}_+ \to \mathbb{R}_+$ is of the form

$$a(r) = \begin{cases} M & \text{if } r \leq R, \\ f(r) & \text{if } r \geq R, \end{cases}$$

for some given $R > 0$ and $f : \mathbb{R}_+ \to \mathbb{R}_+$ positive continuous function satisfying

$$f(R) = M \quad \text{and} \quad \int_R^{+\infty} f(r)\, dr = \varepsilon < +\infty.$$

The constant $M > 0$ is to be properly chosen later on. Assume that the initial state and consensus parameters of the two agents are $(x_1^0, v_1^0) = (-R/2, v^0)$ and $(x_2^0, v_2^0) = (R/2, -v^0)$ respectively, for some $v^0 > \varepsilon/2$.
Due to the nature of the situation, is fairly easy to check if condition (1.14) of Theorem 1.11 is satisfied or not. Indeed we have $X(0) = R^2/4$ and $V(0) = (v^0)^2$, and, by the particular form of a, after a change of variables the computation below follows

$$\int_{\frac{R}{2}}^{+\infty} a(2r)\, dr = \frac{1}{2}\int_R^{+\infty} a(r)\, dr = \frac{1}{2}\int_R^{+\infty} f(r)\, dr = \frac{\varepsilon}{2}.$$

Therefore at time $t = 0$ we are not in the consensus region given by (1.14), since

$$\int_{\sqrt{X(0)}}^{+\infty} a(\sqrt{4}r)\, dr = \frac{\varepsilon}{2} < v^0 = \sqrt{V(0)}.$$

We now show that there exists a time $T > 0$ such that

$$\int_{\sqrt{X(T)}}^{+\infty} a(\sqrt{4}r)\, dr \geq \sqrt{V(T)}, \tag{6.2}$$

i.e., the system enters the consensus region autonomously at time T.
To do so, we first compute a lower bound for the integral. Notice that, since we are considering a Cucker-Smale system with mean consensus parameter $\overline{v} = 0$, the speeds $|v_1(t)|$ and $|v_2(t)|$ are decreasing. Therefore, we can estimate from above the time until $|x_1(t) - x_2(t)| \leq R$ holds by $T^* \triangleq R/2v^0$ (since the agents are moving on the real line in opposite directions). Hence $X(t) \leq X(0) = R^2/4$ for every $t \in [0, T^*]$, which yields the following lower bound

$$\int_{\sqrt{X(t)}}^{+\infty} a(\sqrt{4}r)\, dr \geq \int_{\sqrt{X(0)}}^{+\infty} a(\sqrt{4}r)\, dr = \frac{\varepsilon}{2}$$

valid for any $t \leq T^*$.
We now compute an upper bound for the functional $\sqrt{V(t)}$ for $t \in [0, T^*]$. Notice that

$$a(\sqrt{4X(t)}) \geq a(\sqrt{4X(0)}) = a(R) = M,$$

hence by (1.23) we have

$$\frac{d}{dt}V(t) \leq -2MV(t)$$

which, by integration, implies that $\sqrt{V(t)} \leq v^0 e^{-Mt}$ for every $t \in [0, T^*]$.
We now plug together the two bounds. In order for (6.2) to hold at some time $T < T^*$, simply choose

$$M = M_T \triangleq \frac{1}{T} \log\left(\frac{2v^0}{\varepsilon}\right).$$

For this choice of M, it follows

$$\int_{\sqrt{X(T)}}^{+\infty} a(\sqrt{4}r)\, dr \geq \frac{\varepsilon}{2} = v^0 e^{-M_T T} \geq \sqrt{V(T)}.$$

From Theorem 1.11 we can then conclude that any solution of the above system tends autonomously to consensus.

The above example also shows that Cucker-Smale systems can converge to consensus even if the hypothesis of Theorem 1.11 is not satisfied.
Suppose now, that we are looking for a control that drives the system in Example 6.1 to consensus in a sufficiently large time frame $[0, T]$, but which also uses the least amount of resources possible, i.e., a control u for which the quantity

$$\int_0^T \sum_{i=1}^N \|u_i(t)\|\, dt \tag{6.3}$$

is minimal. If we were to use the sparse control strategy, the above integral would be equal to θT_0, where θ is the $\ell_1^N - \ell_2^d$ bound (2.2) and T_0 is the time of entrance into the consensus regions. Nonetheless, the control strategy that minimizes the cost (6.3) and steers the system to consensus is clearly given by the zero control $u_i \equiv 0$ for every $i = 1, \ldots, N$, since the above integral is then equal to 0 and we know that the system converges autonomously to consensus. Ironically, notice that in this case the optimal strategy is also the sparsest, since it is equal to 0 at every instant! However, it is clear that the knowledge that no control is necessary follows from the ability to predict that the two agents will interact with enough strength for a sufficiently long time T^*.
This example also stresses the paramount role that the knowledge of the interaction kernel a has in this analysis. The results collected in Chapter 5, therefore, justify an increased predictive power on first- and second-order systems (see Section 5.1.5)

Remark 6.2. Notice that the knowledge of a is not vital when using the sparse control of Definition 2.4: indeed, one can ignore the stopping criterion $\gamma(B(x,x)) \geq \max_{1 \leq i \leq N} \|v_i^{\perp}\|$ and exert ruthlessly the control until satisfied of the outcome.

As the above analysis shows, the ability to forecast the evolution of a system means that we can calculate the effect that different control strategies u have on its evolution, and consequently choose the optimal one. We therefore consider the *optimal control problem*

$$\min_{u \in \mathrm{Adm}} \mathcal{J}_N(x_1, \ldots, x_N, v_1, \ldots, v_N, u) \quad \text{subject to system (6.1).} \tag{6.4}$$

Notice that the optimization variable is only the control $u = (u_1, \dots, u_N) \in \mathrm{Adm}$, since the trajectories $(x_1, \dots, x_N, v_1, \dots, v_N) : [0,T] \to \mathbb{R}^{2dN}$ are fully determined by u and the dynamics (6.1). The set of *admissible controls* Adm is a crucial design choice: it should possess good compactness properties to let problem (6.4) be well-posed, and nonetheless it must be sufficiently large to include all interesting control strategies.

The functional $\mathcal{J}_N$ encodes the target that we want to reach with our control. For instance, in Section 2.1 we were addressing the problem of forcing the alignment of the agents, which can be obtained by minimizing the functional

$$\mathcal{J}_N(x_1, \dots, x_N, v_1, \dots, v_N, u) = \int_0^T \frac{1}{N} \sum_{i=1}^N \left(\left\| v_i(t) - \frac{1}{N} \sum_{j=1}^N v_j(t) \right\|^2 + \lambda \|u_i(t)\|_{\ell_p}^p \right) dt, \tag{6.5}$$

in a proper set of admissible controls Adm. A minimizer of the functional (6.5) is a control u for which the associated trajectory $(x_1(\cdot), \dots, x_N(\cdot), v_1(\cdot), \dots, v_N(\cdot))$ satisfies the property that, at every instant, the distance between $v_i(t)$ and the average $\overline{v}(t)$ is as small as possible, under the constraint of the dynamics (2.1) and the penalization of too high values of u (measured by an ℓ_p-cost, in this case). The case $p = 1$ has the effect of enforcing *sparsity* on the control u, see [49]. Notice the dependency of $\mathcal{J}_N$ on the number of agents N in the system and that the optimization is performed in the time frame $[0, T]$.

Remark 6.3. In this chapter we shall focus on the optimization in a finite time frame, i.e., $[0, T]$ with $T < +\infty$. This choice has only minimal impact on the applicability of the results that follow (since, rephrasing John Maynard Keynes' words, *for T sufficiently large we are all dead*!) while enabling the access to the results in Chapter 4 on mean-field approximation.

6.1.1 Mean-field limit equations with control

As seen in Chapter 4, whenever the number of agents N becomes too high, dealing with system (6.1) becomes computationally unfeasible due to the curse of dimensionality. We have already shown how, when no control is involved, one can simplify the multitude of interactions of the other agents on any given individual by a single averaged effect. This results in considering the evolution of the agent density distribution in the state space $\mu(t, x, v)$, leading to the mean-field PDE

$$\frac{\partial \mu}{\partial t}(t) + v \cdot \nabla_x \mu(t) = -\nabla_v \cdot \Big((G^{[a]} * \mu(t)) \mu(t) \Big) \quad \text{for } t \in [0, T],$$

where $G^{[a]}(x, v) \triangleq (v, -a(\|x\|)v)$ for any $(x, v) \in \mathbb{R}^{2d}$, to stress the dependance on a.

The proper definition of a limit dynamics when an external control is added to the system and it is supposed to have some sparsity features is currently still under debate. Indeed, the concept of sparse control has to be handled with care when trying to generalize it at the level of a mean-field dynamics, since 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.

The most immediate approach, very much in the spirit of Theorem 4.11, would be to assign to the finite dimensional control u an empirical vector valued measure curve

$$\nu_N(t) = \frac{1}{N}\sum_{i=1}^{N} u_i \delta_{(x_i(t),v_i(t))} \quad \text{for } t \in [0,T],$$

and consider a suitable limit ν for $N \to \infty$, leading to the PDE with control ν

$$\frac{\partial \mu}{\partial t}(t) + v \cdot \nabla_x \mu(t) = -\nabla_v \cdot \Big((G^{[a]} * \mu(t))\mu(t) + \nu(t)\Big) \quad \text{for } t \in [0,T]. \tag{6.6}$$

Unfortunately, despite the fact that ν is supposed to be the minimizer of a certain cost functional (which may grant the necessary compactness to derive the limit $\nu_N \to \nu$), the design of a general cost functional with a proper meaning in the finite dimensional model, like (6.5), and at the same time guaranteeing a good behavior of the measure ν remains a challenging task: for instance, in the optimal control problems considered in [49, Section 5] the above limit procedure does not prevent ν to be singular with respect to μ, which means that in the weak formulation of equation (6.6) the role of ν is essentially mute, losing completely its steering purpose. Therefore, we consider in (6.6) only the absolutely continuous part of ν with respect to μ, i.e., $\nu_a = u\mu$ (provided it has one), which leads us to an equation of the type

$$\frac{\partial \mu}{\partial t}(t) + v \cdot \nabla_x \mu(t) = -\nabla_v \cdot \Big((G^{[a]} * \mu(t) + u(t))\mu(t)\Big) \quad \text{for } t \in [0,T], \tag{6.7}$$

where now $u(t)$ is an $L^1(\mathbb{R}^{2d}, \mu(t))$ force field for a.e. $t \in [0,T]$. However, such regularity is in general not enough to establish existence and stability of solutions for equations of the type (6.7), see [10], an issue that has to be taken into account when choosing the class Adm. Notice that equation (6.7) is a particular instance of the Vlasov-type equation (4.9) for the choice of the vector field

$$V(t,\mu)(x,v) = (G^{[a]} * \mu)(x,v) + u(t,x,v).$$

On top of system (6.1), we now have the functional $\mathcal{J}_N$ too, whose form has to allow for the possibility to perform the limit $N \to \infty$. This may happen, for instance, when the state variables appearing in $\mathcal{J}_N$ can be manipulated to obtain a *moment* of the empirical measure μ_N. In the specific case of the functional (6.5), we can indeed rewrite it as

$$\begin{aligned}
&\mathcal{J}_N(x_1,\ldots,x_N,v_1,\ldots,v_N,u) = \\
&\qquad = \int_0^T \int_{\mathbb{R}^{2d}} \left(\left\| v - \int_{\mathbb{R}^{2d}} w \, d\mu_N(t,y,w) \right\|^2 + \lambda \|u(t,x,v)\|_{\ell_p}^p \right) d\mu_N(t,x,v)dt \\
&\qquad = \mathcal{J}_N(\mu_N, u),
\end{aligned}$$

thus making explicit its dependance on μ_N. In general, we shall assume that the functional $\mathcal{J}_N$ has the specific form

$$\mathcal{J}_N(\mu_N,u) \triangleq \int_0^T \left\{ \int_{\mathbb{R}^{2d}} \ell(x,v,\mu_N(t)) d\mu_N(t,x,v) + \int_{\mathbb{R}^{2d}} \psi(u(t,x,v)) d\mu_N(t,x,v) \right\} dt,$$

where ℓ is the so-called *Lagrangian* part (or *running cost*) of $\mathcal{J}_N$ and ψ is the *control cost*. The functional (6.5), for instance, falls into this category. This means that the corresponding limit functional $\mathcal{J} : \mathcal{P}_1(\mathbb{R}^{2d}) \times \mathrm{Adm} \to (-\infty, +\infty]$ of the sequence $(\mathcal{J}_N)_{N\in\mathbb{N}}$ for $N \to \infty$ is

$$\mathcal{J}(\mu, u) \triangleq \int_0^T \left\{ \int_{\mathbb{R}^{2d}} \ell(x, v, \mu(t)) d\mu(t, x, v) + \int_{\mathbb{R}^{2d}} \psi(u(t, x, v)) d\mu(t, x, v) \right\} dt.$$

6.1.2 Finite and infinite dimensional optimal control problems

So far, we have identified two different kinds of optimal control problems. The discrete one is the following, each instance of which is labeled by the number of agents $N \in \mathbb{N}$.

Problem 1 (Discrete/finite dimensional optimal control problem)**.** For $T > 0$ fixed, find $u_N^* \in \mathrm{Adm}$ minimizing the cost functional

$$\mathcal{J}_N(\mu_N, u) = \int_0^T \left\{ \int_{\mathbb{R}^{2d}} \ell(x, v, \mu_N(t)) d\mu_N(t, x, v) + \int_{\mathbb{R}^{2d}} \psi(u(t, x, v)) d\mu_N(t, x, v) \right\} dt,$$

where $\mu_N : [0, T] \to \mathcal{P}_1(\mathbb{R}^{2d})$ is given by

$$\mu_N(t) = \frac{1}{N} \sum_{i=1}^N \delta_{(x_i(t), v_i(t))} \qquad \text{for } t \in [0, T],$$

and $(x(\cdot), v(\cdot)) : [0, T] \to \mathbb{R}^{2dN}$ solves

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = \dfrac{1}{N} \displaystyle\sum_{j=1}^N a\left(\|x_i(t) - x_j(t)\|\right)\left(v_j(t) - v_i(t)\right) + u(t, x_i(t), v_i(t)), \end{cases} \qquad i = 1, \ldots, N,$$

for the given initial datum $(x(0), v(0)) = (x^0, v^0) \in \mathbb{R}^{2dN}$.

In Section 6.1.1 we discussed how, when $N \to \infty$, a reasonable candidate for a "limit" of the discrete problems above is the following continuous optimal control problem.

Problem 2 (Continuous/infinite dimensional optimal control problem)**.** For $T > 0$ fixed, find $u^* \in \mathrm{Adm}$ minimizing the cost functional

$$\mathcal{J}(\mu, u) = \int_0^T \left\{ \int_{\mathbb{R}^{2d}} \ell(x, v, \mu(t)) d\mu(t, x, v) + \int_{\mathbb{R}^{2d}} \psi(u(t, x, v)) d\mu(t, x, v) \right\} dt,$$

where $\mu : [0, T] \to \mathcal{P}_1(\mathbb{R}^{2d})$ solves

$$\frac{\partial \mu}{\partial t}(t) + v \cdot \nabla_x \mu(t) = -\nabla_v \cdot \Big((G^{[a]} * \mu(t) + u(t))\mu(t) \Big),$$

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

Our wish is to use the continuous optimal control Problem 2 as an approximation of the discrete instances of Problem 1, in order to circumvent the curse of dimensionality. As a consequence, we are not interested in all possible optima of Problem 2, but mainly on those which arise as limits of optimal strategies of the discrete problems. We call this subclass of the set of optima *mean-field optimal controls*.

Definition 6.4 (Mean-field optimal control for Problem 2)**.** Let $\mu^0 \in \mathcal{P}_c(\mathbb{R}^{2d})$ be given. An optimal control u^* for Problem 2 with initial datum μ^0 is a *mean-field optimal control* if there exists a sequence $(u^*_N)_{N\in\mathbb{N}} \subset \mathrm{Adm}$ and a sequence $(\mu^0_N)_{N\in\mathbb{N}} \subset \mathcal{P}_c(\mathbb{R}^d)$ such that

(*i*) the sequence of empirical measures $(\mu^0_N)_{N\in\mathbb{N}}$ defined for every $N \in \mathbb{N}$ as

$$\mu^0_N \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{(x^0_{i,N},v^0_{i,N})} \qquad \text{for some } (x^0_{i,N}, v^0_{i,N}) \in \mathrm{supp}(\mu^0),$$

satisfies $\mu^0_N \rightharpoonup \mu^0$ weakly* in the sense of measures;

(*ii*) for every $N \in \mathbb{N}$, u^*_N is a solution of Problem 1 with initial datum $(x^0_N, v^0_N) = (x^0_{1,N}, \dots, x^0_{N,N}, v^0_{1,N}, \dots, v^0_{N,N})$;

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

Notice that (*i*) and (*iii*) imply that, under certain regularity hypotheses on the interaction kernel a (like $a \in X$ as in Chapter 5), the discrete system (6.1) admits equation (6.7) as its mean-field limit by Theorem 4.11. This justifies the words *mean-field* to refer to this class of controls.

The main question is then, under which hypotheses on the interaction kernel a, the Lagrangian ℓ, the control cost ψ, and the set of admissible controls Adm, we can prove the existence of a mean-field optimal control for Problem 2. A first solution was given in [106], which identified the following class of functions as a valid candidate for Adm.

Definition 6.5. For a time horizon $T > 0$, and an exponent $q \in [1, +\infty)$ we fix a control bound function $\mathfrak{b} \in L^q([0,T])$. The class of *admissible control functions* $\mathcal{F}_\mathfrak{b}([0,T])$ is defined as the set of Carathéodory functions $u : [0,T] \times \mathbb{R}^{2d} \to \mathbb{R}^d$ satisfying

(*i*) $u(t,\cdot) \in W^{1,\infty}_{\mathrm{loc}}(\mathbb{R}^{2d};\mathbb{R}^d)$ for almost every $t \in [0,T]$,

(*ii*) $\|u(t,0)\| + \mathrm{Lip}_{\mathbb{R}^d}(u(t,\cdot)) \le \mathfrak{b}(t)$ for almost every $t \in [0,T]$.

Functions in $\mathcal{F}_\mathfrak{b}([0,T])$ are locally Lipschitz feedback controls with respect to the state variables with uniform Lipschitz bound given by the L^q function $\mathfrak{b}$. This mild regularity gifts the class $\mathcal{F}_\mathfrak{b}([0,T])$ with the following compactness result, which is a generalization of the Dunford-Pettis Theorem for equi-integrable families of functions with values in a reflexive and separable Banach spaces.

Proposition 6.6 ([106, Corollary 2.7])**.** *Let $p \in (1,+\infty)$. Assume that $(u_n)_{n\in\mathbb{N}}$ is a sequence of functions in $\mathcal{F}_\mathfrak{b}([0,T])$ for a given $\mathfrak{b} \in L^q(0,T)$ with $q \in [1,+\infty)$. Then there exist a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ and a function $u \in \mathcal{F}_\mathfrak{b}([0,T])$, such that*

$$\lim_{k\to\infty} \int_0^T \langle \phi(t), u_{n_k}(t,\cdot) - u(t,\cdot)\rangle dt = 0, \tag{6.8}$$

for all $\phi \in L^{q'}([0,T];H^{-1,p'}(\mathbb{R}^{2d},\mathbb{R}^d))$ *such that* $\operatorname{supp}(\phi(t)) \subseteq B(0,r)$ *for all* $t \in [0,T]$*, for some* $r > 0$*. The symbol* $\langle\cdot,\cdot\rangle$ *denotes the duality between* $W^{1,p}$ *and its dual* $W^{-1,p'}$*.*

Combining the above weak compactness of $\mathcal{F}_{\mathfrak{b}}([0,T])$ with some regularity hypotheses on a, ℓ and ψ, the following existence result for mean-field optimal controls is obtained. Remarkably, it is proven by means of a simultaneous Γ-limit and mean-field limit for $N \to \infty$ of the finite dimensional optimal control Problems 1 to the corresponding infinite dimensional Problem 2.

Theorem 6.7 ([106, Theorem 5.1])**.** *Suppose we are given maps* a*,* ℓ*, and* ψ *satisfying*

(*i*) $a \in X$ *in the sense of Chapter 5;*

(*ii*) *the Lagrangian* $\ell : \mathbb{R}^{2d} \times \mathcal{P}_1(\mathbb{R}^{2d}) \to \mathbb{R}_+$ *is continuous with respect to the product topology generated by the Euclidean distance on* $\mathbb{R}^{2d}$ *and the Wasserstein distance* $\mathcal{W}_1$ *on* $\mathcal{P}_1(\mathbb{R}^{2d})$*;*

(*iii*) *the control cost* $\psi : \mathbb{R}^d \to [0,+\infty)$ *is a nonnegative convex function satisfying the property: there exist* $C \geq 0$ *and* $q \in [1,+\infty)$ *such that for all* $R > 0$ *it holds*

$$\operatorname{Lip}_{B(0,R)}(\psi) \leq CR^{q-1}. \tag{6.9}$$

Choose $q \in [1,+\infty)$ *so that* (6.9) *holds and let* $\mathfrak{b} \in L^q([0,T])$*. Then, for every* $N \in \mathbb{N}$ *and for every initial datum* (x_N^0, v_N^0)*, the finite dimensional optimal control Problem 1 with* $\mathrm{Adm} = \mathcal{F}_{\mathfrak{b}}([0,T])$ *admits a solution* $u_N^* \in \mathcal{F}_{\mathfrak{b}}([0,T])$*.*
Moreover, suppose there exists a compactly supported $\mu^0 \in \mathcal{P}_1(\mathbb{R}^{2d})$ *such that, by setting*

$$\mu_N^0 \triangleq \frac{1}{N}\sum_{i=1}^{N} \delta_{(x_{i,N}^0, v_{i,N}^0)},$$

then the limit $\lim_{N\to\infty} \mathcal{W}_1(\mu_N^0,\mu^0) = 0$ *holds. Then there exists a subsequence* $(u_{N_k}^*)_{k\in\mathbb{N}}$ *of the sequence of finite dimensional optimal controls* $(u_N^*)_{N\in\mathbb{N}}$ *and a function* $u^* \in \mathcal{F}_{\mathfrak{b}}([0,T])$ *such that* $u_{N_k}^*$ *converges to* u^* *in the sense of* (6.8) *and* u^* *is a solution of Problem 2 with initial datum* μ^0*.*
Hence, u^* *is a mean-field optimal control for Problem 2.*

Unfortunately, it is not clear how the choice of the control cost $\psi(\cdot) = \lambda\|\cdot\|_{\ell_1}$ enforces sparsity on the mean-field optimal control u^*: being the weak limit of controls u_N^* with very few nonzero entries, see [49], we may expect that $\operatorname{supp}(u^*(t,\cdot))$ is a "small set" too. Indeed, such concept of sparsity was successfully used in [157] 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 [65, 161, 167].

6.2 Sparsity and mean-field optimal controls

An alternative solution for a proper definition of sparse mean-field controls was proposed in [103]. Inspired by the multiscale approach explored independently in [69] and [76] to describe a mixed granular-diffuse dynamics of a crowd, let us now add to the usual Cucker-Smale system (1.6) m new agents freely interacting with the original N individuals. We assume that these new agents are far fewer than the original population ($m \ll N$) but have a great influence: for this reason they shall be called “leaders”, opposed to the N “followers”. Denoting by (y, w) the space-velocity variables of these new individuals and by

$$\rho_m(t) \triangleq \frac{1}{m}\sum_{k=1}^{m} \delta_{(y_k(t),w_k(t))} \qquad \text{for every } t \in [0,T]$$

the additional empirical measure supported on the trajectories $(y(\cdot), w(\cdot))$, then the proposed dynamics is given by

$$\begin{cases} \dot{y}_k(t) = w_k(t), & \\ \dot{w}_k(t) = (G^{[a]} * \mu_N)(y_k(t), w_k(t)) + (G^{[a]} * \rho_m)(y_k(t), w_k(t)), & k = 1, \dots, m, \\ \dot{x}_i(t) = v_i(t), & \\ \dot{v}_i(t) = (G^{[a]} * \mu_N)(x_i(t), v_i(t)) + (G^{[a]} * \rho_m)(x_i(t), v_i(t)), & i = 1, \dots, N. \end{cases}$$

The above system is similar to a standard multiagent dynamics for $N + m$ individuals, with the sole difference that the actions of leaders and followers have different weights on a single individual, $1/m$ and $1/N$, respectively. Intuitively, since leaders are much more influential than followers, a parsimonious and effective control strategy should act only on them and neglect the followers. If we now add controls on the m leaders, we obtain

$$\begin{cases} \dot{y}_k(t) = w_k(t), & \\ \dot{w}_k(t) = (G^{[a]} * \mu_N)(y_k(t), w_k(t)) + (G^{[a]} * \rho_m)(y_k(t), w_k(t)) + u_k(t), & k = 1, \dots, m, \\ \dot{x}_i(t) = v_i(t), & \\ \dot{v}_i(t) = (G^{[a]} * \mu_N)(x_i(t), v_i(t)) + (G^{[a]} * \rho_m)(x_i(t), v_i(t)), & i = 1, \dots, N, \end{cases} \tag{6.10}$$

where, as in (6.1), the functions $u_k : [0,T] \to \mathbb{R}^d$ are measurable controls for $k = 1, \dots, m$. This is the right time to stress how the sparsity of the control strategy $u = (u_1, \dots, u_m) : [0,T] \to \mathbb{R}^{dm}$ is now encoded directly in the model (6.10), since, instead of trying to influence the entire population of $N + m$ agents, the control simply focuses on the portion of m leaders. In this setting, it makes sense to choose the set Adm as $L^1([0,T];\mathcal{U})$, where $\mathcal{U}$ is a fixed nonempty compact subset of $\mathbb{R}^{dm}$, so that the Dunford-Pettis Theorem is available as a compactness result to establish the well-posedness of the control problems. As in Section 6.1.2, we shall be interested in controls u which minimize functionals of the form

$$\mathcal{J}_N(y, w, \mu_N, u) \triangleq \int_0^T \left\{ \int_{\mathbb{R}^{2d}} \ell(y(t), w(t), x, v, \mu_N(t)) d\mu_N(t, x, v) + \right.$$
$$\left. + \int_{\mathbb{R}^{2d}} \psi(u(t, y, w)) d\nu_m(t, y, w) \right\} dt.$$

Sadly, the introduction of the population of leaders has made the typographic representation of system (6.10) and of the new functional $\mathcal{J}_N$ too much cumbersome and complicated to handle. In order to overcome this shortcoming, as well as the narrowness of its representation power, we now make the drastic choice to do like in Section 4.2 and rewrite the controlled dynamics (6.10) as a first order system of the form

$$\begin{cases} \dot{y}_k(t) = (K * \mu_N(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \ldots, m, \\ \dot{x}_i(t) = (K * \mu_N(t))(x_i(t)) + g(y(t))(x_i(t)), \quad i = 1, \ldots, N, \end{cases}$$

to be coupled with the functional

$$\mathcal{J}_N(y, \mu_N, u) = \int_0^T \Big\{ L(y(t), \mu_N(t)) + \gamma(u(t)) \Big\} dt.$$

We already noticed in Chapter 4 how systems like (6.10) are basically first order, hence it is not restrictive to pass from the $2d$-dimensional phase coordinates (x, v) and (y, w) to the more compact d-dimensional variables x and y. Consequently, the empirical measures μ_N and ρ_m are now defined over the atoms $(x_1, \ldots, x_N)$ and $(y_1, \ldots, y_m)$, respectively.

Remark 6.8. The new notation allows to consider different kernels of interaction between followers and leaders, and more in general different kernels from $G^{[a]}$. It also makes possible the inclusion of *self-propulsion/friction* terms, like

$$S(x, v) \triangleq (\alpha - \beta \|v\|^2)v.$$

where α and β are nonnegative parameters. This term, introduced in [137], 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 [60, 97], and we shall use it to model pedestrian dynamics in Chapter 7.

The role of the control u on the state variables of the k-th leader is mediated by a matrix B_k; this is necessary in order to cope with the dimensionality mismatch in the dynamics (6.10) between the phase coordinates ($2d$) and the controls (only d). To avoid confusion, we denote by D the dimensionality of the control and we assume that it takes values in a *convex compact* subset $\mathcal{U}$ of $\mathbb{R}^D$. As usual, the number d is reserved for the dimensionality of the agents, hence the matrices B_k are constant $d \times D$ matrices which project the D-dimensional control onto the d-dimensional state space.

Remark 6.9. We recover system (6.10) from the above formulation of the dynamics by noticing that $x_i(t)$ stands for $(x_i(t), v_i(t))$, $y_k(t)$ stands for $(y_k(t), w_k(t))$, and choosing for any $(y, w) \in \mathbb{R}^{dm} \times \mathbb{R}^{dm}$ and $(x, v) \in \mathbb{R}^{2d}$ the functions

$$K(x, v) = \begin{pmatrix} 0 \\ G^{[a]}(x, v) \end{pmatrix}, \quad f_k(y, w) = \begin{pmatrix} w_k \\ (G^{[a]} * \rho_m)(y_k, w_k) \end{pmatrix},$$
$$g(y, w)(x, v) = \begin{pmatrix} v \\ (G^{[a]} * \rho_m)(x, v) \end{pmatrix}.$$

Notice that ρ_m is by definition a function of (y, w).

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., $K(-x) = -K(x)$ holds for all $x \in \mathbb{R}^d$ and there exists $C_K > 0$ such that for all $x \in \mathbb{R}^d$ we have

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

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

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

with $\ell \in \mathcal{C}^2(\mathbb{R}^{dm} \times \mathbb{R}^d \times \mathbb{R}^d; \mathbb{R})$, $\omega \in \mathcal{C}^2(\mathbb{R}^d; \mathbb{R}^d)$ and where $\int \omega\mu$ is intended in the sense of (4.2).

(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 $y \in \mathbb{R}^{dm}$

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

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

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

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

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

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

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

Under the above hypotheses, we study the existence of mean-field optimal controls for the following infinite dimensional optimal control problem.

Problem 3. For $T > 0$ fixed, find $u^* \in L^1([0, T]; \mathcal{U})$ minimizing the cost functional

$$\mathcal{J}(y, \mu, u) = \int_0^T \Big\{ L(y(t), \mu(t)) + \gamma(u(t)) \Big\} dt, \tag{6.11}$$

where $(y, \mu) : [0, T] \to \mathbb{R}^{dm} \times \mathcal{P}_1(\mathbb{R}^d)$ solves

$$\begin{cases} \dot{y}_k(t) = (K * \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \ldots, m, \\ \dfrac{\partial \mu}{\partial t}(t) = -\nabla_x \cdot \Big((K * \mu(t) + g(y(t)))\mu(t) \Big), \end{cases} \tag{6.12}$$

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

Notice that we have kept the number of leaders m finite, while letting the number of followers N grow to ∞: with this choice, the control strategy turns out to be sparse by default (since $m \ll N$).

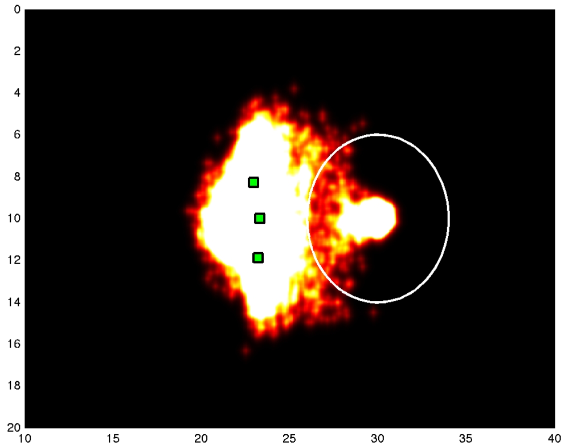


Figure 6.1: Three discrete leaders influencing a continuous density of followers. More details about this control setting in Chapter 7.

As done in [106] to prove Theorem 6.7, we shall provide a Γ-convergence result that links the above Problem 3 with the following family of finite dimensional optimal control problems, each labeled with the number of followers $N \in \mathbb{N}$. This, as a byproduct, will prove the existence of mean-fields optimal controls for Problem 3.

Problem 4. For $T > 0$ fixed, find $u^* \in L^1([0,T];\mathcal{U})$ minimizing the cost functional

$$\mathcal{J}_N(y, \mu_N, u) = \int_0^T \Big\{ L(y(t), \mu_N(t)) + \gamma(u(t)) \Big\} dt, \tag{6.13}$$

where

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

and $(y, x) : [0,T] \to \mathbb{R}^{dm} \times \mathbb{R}^{dN}$ solves

$$\begin{cases} \dot{y}_k(t) = \dfrac{1}{N} \displaystyle\sum_{j=1}^N K(y_k(t) - x_j(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \ldots, m, \\ \dot{x}_i(t) = \dfrac{1}{N} \displaystyle\sum_{j=1}^N K(x_i(t) - x_j(t)) + g(y(t))(x_i(t)), \quad i = 1, \ldots, N, \end{cases} \tag{6.14}$$

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

For the sake of completeness, we state again the notion of *mean-field optimal control* adapted to Problem 3.

Definition 6.10 (Mean-field optimal control for Problem 3)**.** Let $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$ be given. An optimal control u^* for Problem 3 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 $(\mu^0_N)_{N\in\mathbb{N}} \subset \mathcal{P}_c(\mathbb{R}^d)$ such that

(*i*) the sequence of empirical measures $(\mu^0_N)_{N\in\mathbb{N}}$ defined for every $N \in \mathbb{N}$ as

$$\mu^0_N(x) \triangleq \frac{1}{N}\sum_{i=1}^N \delta_{x^0_{i,N}}, \qquad \text{for some } x^0_{i,N} \in \operatorname{supp}(\mu^0)$$

satisfies $\mu^0_N \rightharpoonup \mu^0$ weakly* in the sense of measures;

(*ii*) for every $N \in \mathbb{N}$, u^*_N is a solution of Problem 4 with initial datum $(y^0, x^0_{1,N}, \ldots, x^0_{N,N})$;

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

6.3 Γ-convergence of optimal control problems

In this section we shall first study the coupled ODE-PDE dynamics (6.12) and then state existence and uniqueness results of solutions, together with continuous dependence on the initial data (y^0, μ^0). The proofs follow closely in the footsteps of similar results collected in Chapter 4, to which we remand for the basic ideas (see also [12, 103, 155, 156]).

6.3.1 Coupled ODE-PDE systems with nonlocal interaction

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

$$\|(y,\mu) - (y',\mu')\|_{\mathcal{X}} \triangleq \|y - y'\| + \mathcal{W}_1(\mu,\mu'), \tag{6.15}$$

where $\|y - y'\| \triangleq \sum_{k=1}^m \|y_k - y'_k\|_{\ell^d_2}$.
We can formally define the concept of solution of the controlled ODE-PDE system (6.12).

Definition 6.11. 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 system (6.12) with control u if

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

(*ii*) $(y(\cdot), \mu(\cdot))$ is continuous in time with respect to the metric (6.15) in $\mathcal{X}$;

(*iii*) $y(\cdot)$ is a Carathéodory solution of the following controlled ODE problem

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

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

(*iv*) $\mu(\cdot)$ is a solution of (4.9) for the time-varying vector field (depending on $y(\cdot)$)

$$V(t,\mu)(x) = (K * \mu)(x) + g(y(t))(x), \qquad \text{for all } (t,x,\mu) \in [0,T] \times \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d).$$

We now derive the existence of solutions of (6.12) as limits for $N \to \infty$ of the system of ODEs (6.14) following the strategy adopted in Theorem 4.11. The next result is a straightforward generalization of Proposition 4.5.

Proposition 6.12. *Fix $N \in \mathbb{N}$ and the control $u \in L^1([0,T];\mathcal{U})$. Let $(y, x_N) : [0,T] \to \mathbb{R}^{dm} \times \mathbb{R}^{dN}$ be the corresponding solution of (6.14), with $x_N(\cdot) = (x_{1,N}(\cdot), \ldots, x_{N,N}(\cdot))$. Then, the pair $(y, \mu_N) : [0,T] \to \mathcal{X}$, with $\mu_N(\cdot)$ being the empirical measure*

$$\mu_N(t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{i,N}(t)}, \qquad \text{for every } t \in [0,T],$$

is a solution of (6.12) with control u.

We are then ready to prove the controlled counterpart of Theorem 4.11.

Proposition 6.13. *Let $y^0 \in \mathbb{R}^{dm}$, $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$, and $\mu_N^0(\cdot)$ be as in Definition 6.10–(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 $(y_N, \mu_N)_{N\in\mathbb{N}} \subset \mathrm{Lip}([0,T];\mathcal{X})$ of (6.14) 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 (6.12) with initial data (y^0, μ^0) and control u. Moreover, there exists $\rho_T > 0$, 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 $k = 1, \ldots, m$ 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. First of all notice that, for every $N \in \mathbb{N}$, system (6.14) admits a unique solution $(y_N, x_N) : [0,T] \to \mathbb{R}^{dm} \times \mathbb{R}^{dN}$ by the set of hypotheses (H) and analogous considerations to those made in the proof of Proposition 4.9.
We start by fixing $N \in \mathbb{N}$ and estimating the growth in time of $\|y_{k,N}(\cdot)\|^2 + \|x_{i,N}(\cdot)\|^2$ for $k = 1, \ldots, m$ and $i = 1, \ldots N$. Let

$$\Sigma \triangleq \{1, \ldots, m\} \times \{1, \ldots, N\}.$$

By using the same estimation techniques of (4.14), from the hypotheses (H) and the compactness of $\mathcal{U}$, for any $(k,i) \in \Sigma$ it follows

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\left(\|y_{k,N}\|^2 + \|x_{i,N}\|^2\right) &= \dot{y}_{k,N} \cdot y_{k,N} + \dot{x}_{i,N} \cdot x_{i,N} \\
&= ((K * \mu_N)(y_{k,N}) + f_k(y) + B_k u) \cdot y_{k,N} + \\
&\quad + ((K * \mu_N)(x_i) + g(y)(x_{i,N})) \cdot x_{i,N} \\
&\leq \|(K * \mu_N)(y_{k,N})\| \, \|y_{k,N}\| + f_k(y_N) \cdot y_{k,N} + \|B_k u\| \|y_{k,N}\| + \\
&\quad + \|(K * \mu_N)(x_{i,N})\| \|x_{i,N}\| + g(y_N)(x_{i,N}) \cdot x_{i,N} \\
&\leq C_1 \max_{(l,j)\in\Sigma} \left\{\|y_{l,N}\|^2 + \|x_{j,N}\|^2\right\} + C_2,
\end{aligned}$$

with $C_1 \triangleq 4C_K + F_1 + G_2 + M_1$ and $C_2 \triangleq C_K + F_2 + G_3 + M_1$. We now denote for all

$t \in [0, T]$

$$\xi_{(k,i)}(t) \triangleq \|y_{k,N}(t)\|^2 + \|x_{i,N}(t)\|^2 \quad \text{and} \quad \zeta(t) \triangleq \max_{(l,j)\in\Sigma}\{\xi_{(l,j)}(t)\}.$$

Notice that, being $(y_N(\cdot), x_N(\cdot)) \in W^{1,\infty}([0,T];\mathbb{R}^{dm}\times\mathbb{R}^{dN})$, then $\zeta(\cdot) \in W^{1,\infty}([0,T];\mathbb{R}_+)$. This implies that $\zeta(\cdot)$ is a.e. differentiable, hence again by Stampacchia's Lemma, for a.e. $t \in [0, T]$ there exists $(l, j) \in \Sigma$ such that

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

which let us infer the inequality

$$\dot{\zeta}(t) \leq 2C_1\zeta(t) + 2C_2.$$

We can, therefore, apply Gronwall's Lemma A.1, which together with Definition 6.10–(i) implies that

$$\zeta(t) \leq (\zeta(0) + 2C_2 t)e^{2C_1 t} \leq (C_0 + 2C_2 t)e^{2C_1 t}, \tag{6.16}$$

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

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

which does not depend neither on t nor 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 $\|\dot{y}_{k,N}(\cdot)\|$ and $\|\dot{x}_{i,N}(\cdot)\|$ and noticing that all the functions involved are bounded by the hypotheses (H) and the fact that we are inside $B(0, \rho_T)$. Therefore

$$\|\dot{y}_{k,N}(t)\| \leq \rho'_T, \quad \|\dot{x}_{i,N}(t)\| \leq \rho'_T \tag{6.17}$$

holds for every $t \in [0, T]$ where the constant ρ'_T does not depend neither on t nor 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 $(y_N, \mu_N)_{N\in\mathbb{N}}$ converging uniformly to a limit $(y, \mu) : [0, T] \to \mathcal{X}$, 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 obtain that for all $t_1, t_2 \in [0, T]$ and for some $L_T > 0$ it holds

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

Hence, the limit trajectory $(y(\cdot), \mu(\cdot))$ belongs to $\mathrm{Lip}([0, T]; \mathcal{X})$ as well. It is now necessary to show that the limit $(y(\cdot), \mu(\cdot))$ is a solution of (6.12). We first verify that $y(\cdot)$ is a solution of the ODE part. To this end, we observe that the limit $(y_N(\cdot), \mu_N(\cdot)) \to (y(\cdot), \mu(\cdot))$ in $\mathcal{X}$ specifies into

$$y_N \to y \qquad \text{uniformly in } [0, T], \tag{6.18}$$

$$\dot{y}_N \rightharpoonup \dot{y} \qquad \text{in } L^1([0,T];\mathbb{R}^{2d}), \tag{6.19}$$

$$\lim_{N\to\infty} \mathcal{W}_1(\mu_N(t),\mu(t)) = 0 \qquad \text{uniformly with respect to } t\in[0,T]. \tag{6.20}$$

As a consequence of (6.18), (6.19), (6.20), hypothesis (K), and Lemma 4.10, we have that for all $k=1,\dots,m$

$$\begin{aligned} (K*\mu_N)(y_{k,N}) &\to (K*\mu)(y_k), \\ f_k(y_N) &\to f_k(y), \end{aligned} \qquad \text{uniformly in } [0,T] \text{ as } N\to+\infty. \tag{6.21}$$

To prove that $y(\cdot)$ is actually the Carathéodory solution of (6.14), we have only to show that for all $k=1,\dots,m$ one has

$$\dot{y}_k(t) = (K*\mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \qquad \text{for all } t\in[0,T].$$

This is clearly equivalent to the following: for every $\eta\in\mathbb{R}^d$ and every $t\in[0,T]$ it holds

$$\eta\cdot\int_0^t \dot{y}_k(s)\,ds = \eta\cdot\int_0^t \Big((K*\mu(s))(y_k(s)) + f_k(y(s)) + B_k u(s)\Big)ds, \tag{6.22}$$

which follows from (6.21) and from the weak L^1-convergence of $\dot{y}_{k,N}$ to $\dot{y}_k$ and of u_N to u for $N\to\infty$.

The proof of the fact that $\mu(\cdot)$ is a solution of (6.12) follows along the same lines of the one of Theorem 4.11 by noticing that the functions $g_N = g(y_N)$ converge uniformly to $g=g(y)$ on $[0,T]\times B(0,\rho_T)$ and satisfy the hypotheses (i)–(iii) of Proposition 4.9.

This, together with (6.22), proves that $(y(\cdot),\mu(\cdot))$ is a solution of (6.14) with initial datum (y^0,μ^0) and control u. Similarly to Remark 4.18, we can show that the sequence $(y_N,\mu_N)_{N\in\mathbb{N}}$ actually converges to (y,μ) without passing to a subsequence as soon as we show the continuous dependance on the initial data. □

Corollary 6.14. *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* (6.12) *with control u and initial datum (y^0,μ^0).*

Proof. Follows from Proposition 6.13 by taking any sequence of empirical measures μ_N^0 as in Definition 6.10–(i), and the constant sequence $u_N=u$ for all $N\in\mathbb{N}$. □

We now prove the continuous dependence on the initial data, that also grants the uniqueness of the solution for (6.12).

Proposition 6.15. *Let the hypotheses (H) hold. Let $u\in L^1([0,T];\mathcal{U})$ be given, and take two solutions $(y_1(\cdot),\mu_1(\cdot))$ and $(y_2(\cdot),\mu_2(\cdot))$ of* (6.12) *with control u and with initial data $(y_1^0,\mu_1^0),(y_2^0,\mu_2^0)\in\mathcal{X}$, respectively, where μ_1^0 and μ_2^0 have both compact support. Then there exists a constant $C_T>0$ such that*

$$\|(y_1(t),\mu_1(t)) - (y_2(t),\mu_2(t))\|_{\mathcal{X}} \le C_T\|(y_1^0,\mu_1^0)-(y_2^0,\mu_2^0)\|_{\mathcal{X}} \quad \textit{for all } t\in[0,T].$$

Proof. We start by noticing that, by the definition of a solution, we infer the existence of a $\rho_T>0$ for which $y_1(t),y_2(t)\in B(0,\rho_T)\subset\mathbb{R}^{dm}$ and $\operatorname{supp}(\mu_1(t)),\operatorname{supp}(\mu_2(t))\subseteq B(0,\rho_T)\subset\mathbb{R}^d$ hold for all $t\in[0,T]$.

As a preliminary estimate, by hypothesis (K), Lemma 4.7 and Lemma 4.10 with the choice $r = \widehat{R} = \rho_T$, we infer the existence of a constant $L^K_{\rho_T} > 0$ such that

$$\|(K * \mu_1)(x) - (K * \mu_2)(y)\| \leq L^K_{\rho_T} \left(\mathcal{W}_1(\mu_1, \mu_2) + \|x - y\|\right). \tag{6.23}$$

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

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

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

$$\begin{aligned} \|g(y_1) - g(y_2)\|_{L^\infty(B(0,\rho_T))} &\leq G\|y_1 - y_2\|, \\ \|f_k(y_1) - f_k(y_2)\| &\leq F\|y_1 - y_2\|. \end{aligned} \tag{6.24}$$

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

$$\dot{y}_k(t) = (K * \mu(t))(y_k(t)) + f_k(y(t)) + B_k u(t), \quad k = 1, \ldots, m, \tag{6.25}$$

with the one of the PDE

$$\frac{\partial \mu}{\partial t}(t) = -\nabla_x \cdot \Big((K * \mu(t) + g(y(t)))\mu(t)\Big), \tag{6.26}$$

first addressing the dependence of (6.25). By integration we have

$$\begin{aligned} \|y_{1,k}(t) - y_{2,k}(t)\| \leq \|y^0_{1,k} - y^0_{2,k}\| + \\ + \int_0^t \Big(\|(K * \mu_1(s))(y_{1,k}(s)) - (K * \mu_2(s))(y_{2,k}(s))\| + \\ + \|f_k(y_1(s)) - f_k(y_2(s))\|\Big) ds. \end{aligned} \tag{6.27}$$

For the left-hand side of (6.27) we have that (6.23), (6.24), and the uniform bound on $y_1(\cdot)$ and $y_2(\cdot)$ yield

$$\begin{aligned} \|y_{1,k}(t) - y_{2,k}(t)\| \leq \|y^0_{1,k} - y^0_{2,k}\| + \int_0^t \Big(L^K_{\rho_T} \mathcal{W}_1(\mu_1(s), \mu_2(s)) + \\ + L^K_{\rho_T}\|y_{1,k}(s) - y_{2,k}(s))\| + F\|y_1(s) - y_2(s)\|\Big) ds \end{aligned} \tag{6.28}$$

We now consider (6.26). Theorem 4.16 for the choices $\overline{R} = \rho_T$, $K_1 = K_2 = K$, $g_1 = g(y_1)$ and $g_2 = g(y_2)$, together with the estimate (6.24), yields

$$\begin{aligned} \mathcal{W}_1(\mu_1(t), \mu_2(t)) \leq \alpha \mathcal{W}_1(\mu^0_1, \mu^0_1) + \\ + \beta \int_0^t \left(L_{K,\rho_T,\rho_T} \mathcal{W}_1(\mu_1(s), \mu_2(s)) + G\|y_1(s) - y_2(s)\|\right) ds. \end{aligned} \tag{6.29}$$

for certain constants $\alpha, \beta > 0$.

We now consider the function

$$\varepsilon(t) \triangleq \|(y_1(t), \mu_1(t)) - (y_2(t), \mu_2(t))\|_{\mathcal{X}} \quad \text{for every } t \in [0, T].$$

Combining (6.28) for each $k = 1, \ldots, m$ and (6.29), we obtain

$$\begin{aligned}\varepsilon(t) \leq\ & \|y_1^0 - y_2^0\| + \int_0^t \Big(mL^K_{\rho_T} \mathcal{W}_1(\mu_1(s), \mu_2(s)) + L^K_{\rho_T} \|y_1(s) - y_2(s))\| + \\ & + mF\|y_1(s) - y_2(s)\| \Big) ds + \alpha \mathcal{W}_1(\mu_1^0, \mu_2^0) + \\ & + \beta \int_0^t \Big(L_{K,\rho_T,\rho_T} \mathcal{W}_1(\mu_1(s), \mu_2(s)) + G\|y_1(s) - y_2(s)\| \Big) ds \\ \leq\ & \alpha\varepsilon(0) + \int_0^t \Big(mL^K_{\rho_T} + mF + \beta(L_{K,\rho_T,\rho_T} + G) \Big) \varepsilon(s) ds.\end{aligned}$$

Gronwall's Lemma A.1 then implies

$$\varepsilon(t) \leq \alpha e^{T\left(mL^K_{\rho_T} + mF + \beta(L_{K,\rho_T,\rho_T} + G)\right)} \varepsilon(0),$$

for any $t \in [0, T]$. This concludes the proof. □

Remark 6.16. We can argue as in Remark 4.18 to prove that the sequence $(y_N, \mu_N)_{N\in\mathbb{N}}$ converges to (y, μ) without passing to a subsequence.

Remark 6.17. Since equicompactly supported solutions are unique, given the initial datum, by Proposition 6.15 and Proposition 6.13 we infer that the support of the unique solution can be estimated as a function of the data, namely it is contained in $B(0, \rho_T)$, where the constant ρ_T is depending only on y^0, $\mathrm{supp}(\mu^0)$, K, g, f_k, B_k, $\mathcal{U}$, and T.

6.3.2 Existence and construction of solutions of Problem 3

In this section, we prove that Problem 3 admits a solution which is a mean-field optimal control. The proof generalizes similar results in [103].

In view of the Definition 5.14 of Γ-convergence, let us fix as our domain $\mathbb{X} \triangleq L^1([0, T]; \mathcal{U})$ which, endowed with the weak L^1-topology, is actually a metrizable space. Fix now 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 6.10–(i). Consider the functional

$$\mathcal{J}(u) \triangleq \mathcal{J}(y^u(\cdot), \mu^u(\cdot), u)$$

on $\mathbb{X}$ introduced in (6.11), where the pair $(y^u(\cdot), \mu^u(\cdot))$ defines the unique solution of (6.12) with initial datum (y^0, μ^0) and control u. Similarly, consider the functional

$$\mathcal{J}_N(u) \triangleq \mathcal{J}_N(y_N^u(\cdot), \mu_N^u(\cdot), u)$$

on $\mathbb{X}$ introduced in (6.13), where the pair $(y_N^u(\cdot), \mu_N^u(\cdot))$ defines the unique solution of (6.12) with initial datum (y^0, μ_N^0) and control u. As recalled in Proposition 6.12, such solution coincides with a properly rewritten solution of the ODE system (6.14).

We now pass to prove the Γ-convergence of the sequence of functionals $(\mathcal{J}_N)_{N\in\mathbb{N}}$ on $\mathbb{X}$ to the target functional $\mathcal{J}$. Let us mention that Γ-convergence in optimal control problems has been already considered, see for instance [45], but it has been only recently employed in the context of mean-field limits in [103, 106].

Theorem 6.18. *Let the functionals* (6.11)–(6.13) *and system* (6.12) *satisfy the set of 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 6.10–(i). Then the sequence of functionals* $(\mathcal{J}_N)_{N\in\mathbb{N}}$ *on* $\mathbb{X}$ *defined in* (6.13) Γ*-converges to the functional* $\mathcal{J}$ *defined in* (6.11).

Proof. Let us start by showing the $\Gamma - \liminf$ condition. Let us fix a weakly convergent sequence of controls $u_N \rightharpoonup u$ in $\mathbb{X}$. We associate to each of these controls a sequence of solutions $(y_N(\cdot), \mu_N(\cdot))$ of (6.12) uniformly convergent to a solution $(y(\cdot), \mu(\cdot))$ with control u and initial datum (y^0, μ^0). In view of the fact that solutions $(y_N(\cdot), \mu_N(\cdot))$ and $(y(\cdot), \mu(\cdot))$ will have uniformly bounded supports with respect to N and $t \in [0, T]$ and by the uniform convergence of trajectories $y_N(\cdot) \to y(\cdot)$ in $[0, T]$, as well as the uniform convergence $\mathcal{W}_1(\mu_N(t), \mu(t)) \to 0$ for $t \in [0, T]$, it follows from the continuity of L under the hypotheses (H) that

$$\lim_{N\to\infty} \int_0^T L(y_N(t), \mu_N(t)) = \int_0^T L(y(t), \mu(t))dt. \tag{6.30}$$

By the assumed weak convergence of $(u_N)_{N\in\mathbb{N}}$ to $u \in \mathbb{X}$ and Ioffe's Theorem (see, for instance, [11, Theorem 5.8]) we obtain the lower semicontinuity of γ

$$\liminf_{N\to\infty} \int_0^T \gamma(u_N(t))dt \geq \int_0^T \gamma(u(t))dt. \tag{6.31}$$

By combining (6.30) and (6.31), we immediately obtain the $\Gamma - \liminf$ condition

$$\liminf_{N\to\infty} \mathcal{J}_N(u_N) \geq \mathcal{J}(u).$$

We now prove the $\Gamma - \limsup$ condition. We fix $u \in \mathbb{X}$ and consider the constant recovery sequence $u_N = u$ for all $N \in \mathbb{N}$. Similarly as above for the argument of the $\Gamma - \liminf$ condition, we can associate to each of these controls a sequence of solutions $(y_N(\cdot), \mu_N(\cdot))$ of (6.12) uniformly convergent to a solution $(y(\cdot), \mu(\cdot))$ with control u and initial datum (y^0, μ^0) and we can similarly conclude the limit (6.30). Additionally, since $(u_N)_{N\in\mathbb{N}}$ is a constant sequence, we have

$$\liminf_{N\to\infty} \int_0^T \gamma(u_N(t))dt = \int_0^T \gamma(u(t))dt. \tag{6.32}$$

Hence, combining (6.30) and (6.32) we can easily infer

$$\limsup_{N\to\infty} \mathcal{J}_N(u_N) = \lim_{N\to\infty} \mathcal{J}_N(u) = \mathcal{J}(u).$$

This concludes the proof of the Γ-convergence of the functionals $(\mathcal{J}_N)_{N\in\mathbb{N}}$ to $\mathcal{J}$. □

The following result shows that Problem 4 is automatically well-posed under our set of hypothesis.

Proposition 6.19 ([63, Theorem 23.11]). *Under the hypotheses (H), Problem 4 admits solutions.*

We are now ready to show the existence of mean-field optimal controls for Problem 3.

Corollary 6.20. *Let the hypotheses (H) 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 3.*

Proof. For every $N \in \mathbb{N}$, consider an empirical measure μ_N^0 as in Definition 6.10–(i), and let u_N^* be an optimal control for Problem 4 with initial datum (y^0, μ_N^0) (it exists by Proposition 6.19). Notice that the optimal controls u_N^* belong to the space $\mathbb{X} = L^1([0,T];\mathcal{U})$, which is a compact set with respect to the weak topology of L^1 by the Dunford-Pettis Theorem. Hence, the sequence $(\mathcal{J}_N)_{N\in\mathbb{N}}$ is equicoercive, and $(u_N^*)_{N\in\mathbb{N}}$ admits a subsequence, which we do not relabel, weakly convergent to some $u^* \in \mathbb{X}$.
We can associate to each of these controls u_N^* and initial data (y^0, μ_N^0) a solution $(y_N^*(\cdot), \mu_N^*(\cdot))$ of (6.12). The sequence of solutions $(y_N^*(\cdot), \mu_N^*(\cdot))$ is then uniformly convergent to a solution $(y^*(\cdot), \mu^*(\cdot))$ of (6.12) with control u^*, by Proposition 6.13. In order to conclude that u^* is an optimal control for Problem 3 (and hence, by construction, that u^* is a *mean-field optimal control*) we need to show that it is actually a minimizer of $\mathcal{J}$. To prove it we use the fact that $\mathcal{J}$ is the Γ-limit of the sequence $(\mathcal{J}_N)_{N\in\mathbb{N}}$ as proved in Theorem 6.18. Let $u \in \mathbb{X}$ be an arbitrary control and let $(u_N)_{N\in\mathbb{N}}$ be a recovery sequence given by the $\Gamma - \limsup$ condition, so that

$$\mathcal{J}(u) \geq \limsup_{N\to\infty} \mathcal{J}_N(u_N). \tag{6.33}$$

By using the optimality of the controls in the sequence $(u_N^*)_{N\in\mathbb{N}}$, we obtain

$$\limsup_{N\to\infty} \mathcal{J}_N(u_N) \geq \limsup_{N\to\infty} \mathcal{J}_N(u_N^*) \geq \liminf_{N\to\infty} \mathcal{J}_N(u_N^*). \tag{6.34}$$

Applying the $\Gamma - \liminf$ condition yields

$$\liminf_{N\to\infty} \mathcal{J}_N(u_N^*) \geq \mathcal{J}(u^*). \tag{6.35}$$

By chaining the inequalities (6.33)–(6.34)–(6.35) we conclude that

$$\mathcal{J}(u) \geq \mathcal{J}(u^*) \quad \text{for all } u \in \mathbb{X},$$

i.e., that u^* is an optimal control for Problem 3. □

Remark 6.21. 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 3 are mean-field optimal controls.

Remark 6.22. Notice that the inequality (6.31) actually holds for several non-strictly convex functionals, like the ℓ_1–norm

$$\gamma(u) = \frac{1}{m}\sum_{k=1}^{m} \|u_i\|,$$

as shown in [103, 106]. The strict convexity shall actually be a necessary property to derive the optimality conditions for mean-field optimal controls in the next section.

6.4 Mean-field Pontryagin Maximum Principle

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 [15, 43, 50, 163]. However, up to now no corresponding results have appeared in the literature for optimal control problems with mixed ODE-PDE constraints of the kind of Problem 3. This section is devoted to the proof of a Pontryagin Maximum Principle to characterize mean-field optimal controls of such problems.

The Pontryagin Maximum Principle (PMP in what follows, see [159]) is used in Optimal Control Theory to find the best possible control for steering a dynamical system from one state to another. It is a generalization of the Euler-Lagrange equation of the Calculus of Variations and, similarly to it, when satisfied along a trajectory, the PMP is a necessary condition for an optimum.

To formulate the PMP, an *adjoint variable* p for every state variable x of the system is introduced. The variables p are elements of the dual space $\mathcal{S}'$ of the state space $\mathcal{S}$, and can be interpreted as Langrange multipliers associated with the state equations of the system: the state equations represent constraints of the minimization problem, and the adjoint variables represent the marginal cost of violating those constraints. The evolution of state and adjoint variables are deeply interwined: if u^* is a control maximizing the Hamiltonian $\mathbb{H}(x, p, u)$ of the system with respect to the control variable u, and $x^* : [0, T] \to \mathcal{S}$ and $p^* : [0, T] \to \mathcal{S}'$ are the state and adjoint trajectories associated to u^*, then x^* and p^* satisfy

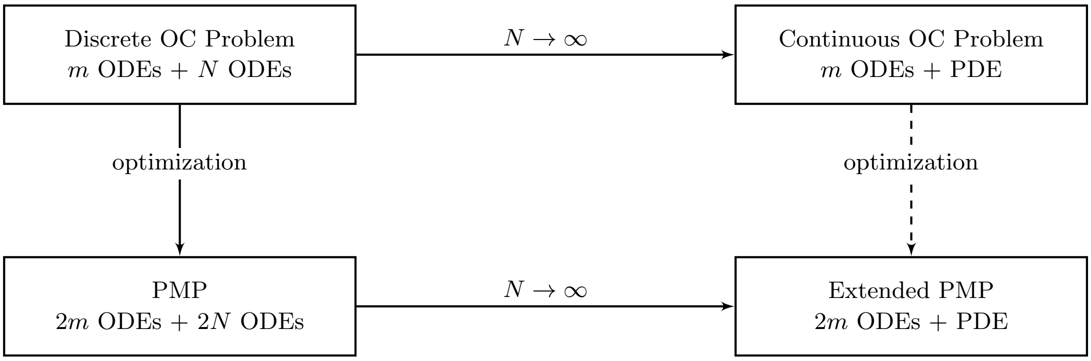
$$\begin{cases} \dot{x}^*(t) = \nabla_p \mathbb{H}(x^*(t), p^*(t), u^*(t)), \\ \dot{p}^*(t) = -\nabla_x \mathbb{H}(x^*(t), p^*(t), u^*(t)), \\ u^*(t) \in \operatorname*{arg\,max}_{u(t)\in\mathcal{U}} \mathbb{H}(x^*(t), p^*(t), u(t)). \end{cases}$$

The state equations are subject to an initial condition and are solved forwards in time. The adjoint equations must satisfy, instead, a terminal condition and are solved backwards in time, from the final time towards the beginning.

Whenever we deal with a finite dimensional state space $\mathcal{S}$, the above formulation presents no ambiguity. Since the dual of $\mathbb{R}^d$ is $\mathbb{R}^d$ itself then the state and adjoint variables are both d-dimensional vectors, and the gradients ∇_x and ∇_p are intended in the usual sense. However, when the dynamical system describes the evolution of a measure μ, things become more problematic. The first issue is that the dual space of $\mathcal{M}(\mathbb{R}^d)$ is $\mathcal{C}(\mathbb{R}^d)$: does it matter to preserve the duality between the space of state variables and that of adjoint variables, or it is more important to have them lie in the same space? Moreover, in which sense we should compute $\nabla_\mu \mathbb{H}$ when μ is a measure? Is it only a formal symbol or has it sense in some differential manifold?

We stress again that we are not interested in all possible optima of Problem 3, but mainly on mean-field optimal controls, i.e., those which arise as limits of optimal strategies of the discrete Problems 4. This already suggests us a possible strategy to provide an answer to the questions raised above. Indeed, since Problem 4 is a standard finite dimensional optimal control problem with $m + N$ ODE constraints, a PMP for it is already available (see, for instance, Theorem 6.28 below). Nothing prevents us to try to compute the limit for $N \to \infty$ of the equations of this PMP, and calling it *extended PMP*. Eventually, the fact that Problem 3 is the $\Gamma-$limit of Problem 4 for $N \to \infty$ will imply that mean-field optimal controls of Problem 3 satisfy this extended PMP.

We can summarize our strategy, borrowing a leaf from the diagram in [50], as follows:



We shall see that, under the set of hypotheses (H), the dashed line from the upper-right to the bottom-right box is valid for all mean-field optimal controls. The result of this procedure will be the following first-order condition.

Theorem 6.23. *Fix an initial datum $(y^0, \mu^0) \in \mathbb{R}^{dm} \times \mathcal{P}_c(\mathbb{R}^d)$ and assume that the set of hypotheses (H) holds. If u^* is a mean-field optimal control for Problem 3 and $(y^*(\cdot), \mu^*(\cdot))$ is the corresponding trajectory, then the triple (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, $\mathrm{supp}(\mu^0)$, d, K, g, f_k, B_k, $\mathcal{U}$, and T, such that $\mathrm{supp}(\nu^*(\cdot)) \subseteq B(0, R_T)$ and it satisfies $\nu^*(t)(E \times \mathbb{R}^d) = \mu^*(t)(E)$ for all $t \in [0,T]$ and for every Borel set $E \subseteq \mathbb{R}^d$;*
- *for every $t \in [0,T]$ it holds*

$$\begin{cases} \dot{y}_k^*(t) = \nabla_{q_k} \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t), u^*(t)), \\ \dot{q}_k^*(t) = -\nabla_{y_k} \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t), u^*(t)), \\ \dfrac{\partial \nu^*}{\partial t}(t) = -\nabla_{(x,r)} \cdot \Big((\mathbf{J} \nabla_\nu \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t), u^*(t))) \nu^*(t) \Big), \\ u^*(t) = \underset{u(t) \in \mathcal{U}}{\arg\max}\, \mathbb{H}_c(y^*(t), q^*(t), \nu^*(t), u(t)) \end{cases} \tag{6.36}$$

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

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

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*

$$\mathbb{H}_c(y,q,\nu,u) \triangleq \begin{cases} \mathbb{H}(y,q,\nu,u) & \textit{if } \operatorname{supp}(\nu) \subseteq \overline{B(0,R_T)}, \\ +\infty & \textit{elsewhere;} \end{cases}$$

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

$$\begin{aligned}\mathbb{H}(y,q,\nu,u) \triangleq \frac{1}{2}\int_{\mathbb{R}^{4d}} &(r-r')\cdot K(x-x')d\nu(x,r)d\nu(x',r')+ \\ &+ \int_{\mathbb{R}^{2d}} r\cdot g(y)(x)d\nu(x,r) + \sum_{k=1}^m \int_{\mathbb{R}^{2d}} q_k \cdot K(y_k-x)d\nu(x,r)+ \\ &+ \sum_{k=1}^m q_k\cdot(f_k(y)+B_k u) - L(y,\pi_{1\#}\nu) - \gamma(u).\end{aligned} \tag{6.37}$$

- *the following initial conditions for system* (6.36) *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 terminal conditions for system* (6.36) *hold at time* T*:* $q^*(T) = 0$ *and* $\nu^*(T)(\mathbb{R}^d\times E) = \delta_0(E)$ *for every Borel set* $E \subseteq \mathbb{R}^d$*.*

The above result answers the questions raised at the beginning of the section: first, it introduces a new measure ν^* on $\mathbb{R}^{2d}$ whose first marginal is the measure of followers μ^*. The adjoint of μ^* is given by the second marginal of ν^*, which is also an element of $\mathcal{P}_1(\mathbb{R}^d)$, as μ^* itself. The property of state variables and their adjoints to belong to the same ambient space is thus preserved in the mean-field limit. Second, we point out the difference between the usual notion of gradient in $\mathbb{R}^{2d}$ with respect to the state and the adjoint variables (x,r), denoted by $\nabla_{(x,r)}$, and the Wasserstein gradient ∇_ν of $\mathbb{H}_c$, which, as will be shown in Section 6.4.2, whenever ν has support contained in $B(0,R_T)$ can be computed explicitly as follows:

- For $l = 1,\dots,d$, it holds

$$\begin{aligned}\nabla_\nu\mathbb{H}_c(y,q,\nu,u)(x,r)\cdot e_l = \int_{\mathbb{R}^{2d}} &(r-r')\cdot(\mathbf{D}K(x-x')e_l)\,d\nu(x',r')+ \\ &+ r\cdot(\mathbf{D}_x g(y)(x)e_l) - \sum_{k=1}^m q_k\cdot(\mathbf{D}K(y_k-x)e_l)- \\ &- \nabla_\xi \ell(y,x,\textstyle\int\omega\mu)\cdot e_l - \big(\nabla_\varsigma \ell(y,x,\textstyle\int\omega\mu)\mathbf{D}\omega(x)\big)\cdot e_l.\end{aligned} \tag{6.38}$$

These are the components of $\nabla_\nu\mathbb{H}_c(y,q,\nu,u)(x,r)$ in the x_l coordinates. Notice that, by the hypothesis (L), the functions $\ell \in \mathcal{C}^2(\mathbb{R}^{dm}\times\mathbb{R}^d\times\mathbb{R}^d;\mathbb{R})$ and $\omega \in \mathcal{C}^2(\mathbb{R}^d;\mathbb{R}^d)$

appearing in (6.38) are related to the functional L in (6.11) via

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

where $\int \omega\mu = \omega\mu(\mathbb{R}^d)$ in the sense of (4.2), while $\nabla_\xi \ell$ and $\nabla_\varsigma \ell$ denote the partial derivatives of the function $\ell(\eta, \xi, \varsigma)$. The notations $\mathbf{D}K(x)$, $\mathbf{D}_x g(y)(x)$ and $\mathbf{D}\omega(x)$ stand for the Jacobian of the functions K, $g(y)$ and ω evaluated at x.

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

$$\nabla_\nu \mathbb{H}_c(y,q,\nu,u)(x,r) \cdot e_l = \int_{\mathbb{R}^{2d}} K(x-x') \cdot e_{l-d}\, d\nu(x',r') + g(y)(x) \cdot e_{l-d}. \quad (6.39)$$

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

Remark 6.24. There are two main consequences of the fact that the control does not act directly on the PDE component of (6.12). The first is that $\nabla_\nu \mathbb{H}(y,q,\nu,u)$ actually does not depend on u, since ν is not associated to the directly controlled part of the dynamics. The second is that, for every $(y(t), q(t), \nu(t)) \in \mathbb{R}^{2d} \times \mathcal{P}_1(\mathbb{R}^{2d})$ with $\mathrm{supp}(\nu(t)) \subseteq \overline{B(0,R_T)}$, the expression (6.37) immediately implies

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

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

Remark 6.25. System (6.36) shows that $(y^*(\cdot), q^*(\cdot), \nu^*(\cdot))$ 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 [12]. 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 (6.12), since both are flows in a Wasserstein space.

Remark 6.26. The notion of solution for system (6.36) can be trivially obtained by a slight modification of Definition 6.11. Indeed, the dynamics of ν^* is a Vlasov-type equation (4.9) for the choice of the vector field

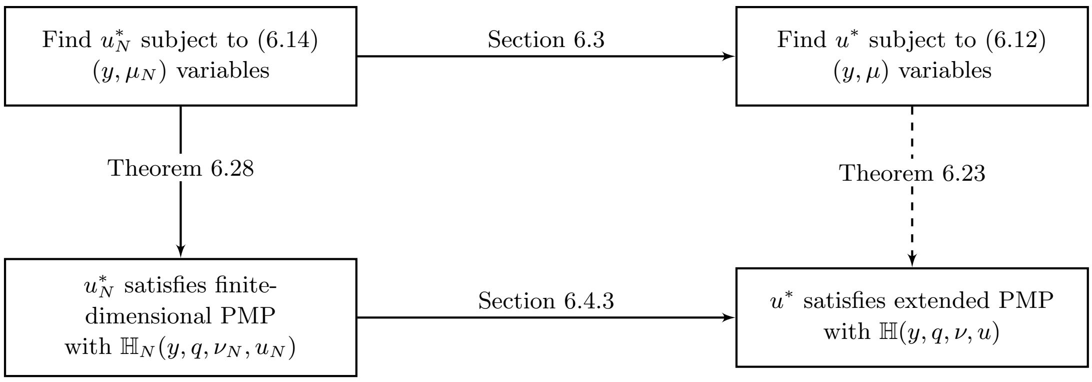
$$V(t,\nu)(x,r) = \mathbf{J}\nabla_\nu \mathbb{H}_c(y^*(t), q^*(t), \nu, u^*(t))(x,r) \qquad \text{for every } (x,r) \in \mathbb{R}^{2d}.$$

The extended PMP will be derived after reformulating the finite-dimensional PMP applied to Problem 4 in terms of the empirical measure in the phase space $\mathbb{R}^{2d}$ (to which the pairs state-adjoint variables (x_i, p_i) belongs) defined as

$$\nu_N(t) \triangleq \frac{1}{N} \sum_{i=1}^N \delta_{(x_i(t), Np_i(t))} \qquad \text{for every } t \in [0,T].$$

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 (see also Remark 6.33); within this scaling, the right-hand side of the finite-dimensional PMP is brought back to the form considered, for instance, in [84], with a different Hamiltonian.

The following diagram recollects the strategy of the proof, making use of a more precise notation and reporting the ingredients disseminated in the chapter which make the proof possible.



Remark 6.27. We briefly compare the hypotheses (H) with those of [15, 43], where optimality conditions for similar optimal control problems are considered. In [15], which deals with an SDE-constrained 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, the hypothesis (H) do not require differentiability of the Lagrangian. The authors of [43] deal, instead, with 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 γ to be strictly convex.

6.4.1 Mean-field of the finite-dimensional optimality conditions

We now introduce the adjoints of the state variables x_i and y_k, denoted by p_i and q_k, respectively, and state the PMP for Problem 4.

Theorem 6.28 ([63, Theorem 22.2])**.** *Let u_N^* be a solution of Problem 4 with initial datum $(y(0), x(0)) = (y^0, x^0) \in \mathbb{R}^{dm} \times \mathbb{R}^{dN}$, and denote with $(y^*(\cdot), x^*(\cdot)) : [0,T] \to \mathbb{R}^{dm} \times \mathbb{R}^{dN}$ the corresponding trajectory. Then there exists a curve $(y^*(\cdot), q^*(\cdot), x^*(\cdot), p^*(\cdot)) \in \mathrm{Lip}([0,T]; \mathbb{R}^{dm} \times \mathbb{R}^{dN})$ solving the system*

$$\begin{cases} \dot{y}_k^*(t) = \nabla_{q_k} \mathbb{H}_N(y^*(t), q^*(t), x^*(t), p^*(t), u^*(t)), & \multirow{2}{*}{$k = 1, \dots, m,$} \\ \dot{q}_k^*(t) = -\nabla_{y_k} \mathbb{H}_N(y^*(t), q^*(t), x^*(t), p^*(t), u^*(t)), & \\ \dot{x}_i^*(t) = \nabla_{p_i} \mathbb{H}_N(y^*(t), q^*(t), x^*(t), p^*(t), u^*(t)), & \multirow{2}{*}{$i = 1, \dots, N,$} \\ \dot{p}_i^*(t) = -\nabla_{x_i} \mathbb{H}_N(y^*(t), q^*(t), x^*(t), p^*(t), u^*(t)), & \\ u_N^*(t) = \operatorname*{arg\,max}_{u(t) \in \mathcal{U}} \mathbb{H}_N(y^*(t), q^*(t), x^*(t), p^*(t), u(t)), & \end{cases} \tag{6.40}$$

for every $t \in [0,T]$ *with initial and terminal data given by* $(y(0), x(0)) = (y^0, x^0)$ *and* $(q(T), p(T)) = 0$. *The finite dimensional Hamiltonian* $\mathbb{H}_N : \mathbb{R}^{2dm} \times \mathbb{R}^{2dN} \to \mathbb{R}$ *is*

$$\begin{aligned}\mathbb{H}_N(y, q, x, p, u) \triangleq \sum_{i=1}^{N} p_i \cdot \left(\frac{1}{N} \sum_{j=1}^{N} K(x_i - x_j) + g(y)(x_i) \right) + \\ + \sum_{k=1}^{m} q_k \cdot \left(\frac{1}{N} \sum_{j=1}^{N} K(y_k - x_j) + f_k(y) + B_k u \right) - L(y, \mu_N) - \gamma(u),\end{aligned} \tag{6.41}$$

with $\mu_N = \sum_{i=1}^{N} \delta_{x_i}/N$.

Remark 6.29. 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. [63, Corollary 22.3].

Remark 6.30. The uniqueness of the maximizer of $\mathbb{H}_N$ follows from the same motivations reported in Remark 6.24. Indeed, the form of the Hamiltonian implies that for each $(y(t), q(t), x(t), p(t)) \in \mathbb{R}^{4d}$ it holds

$$\begin{aligned}\overline{u}(t) = \operatorname*{arg\,max}_{u(t) \in \mathcal{U}} \mathbb{H}_N(y(t), q(t), x(t), p(t), u(t)) \iff \\ \iff \overline{u}(t) = \operatorname*{arg\,max}_{u(t) \in \mathcal{U}} \left(\sum_{k=1}^{m} q_k(t) \cdot B_k u(t) - \gamma(u(t)) \right).\end{aligned}$$

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 4 as solutions of the extended PMP for Problem 3. As a first step, we prove that pairs control-trajectories $(u_N^*, (y_N^*, q_N^*, x_N^*, p_N^*))$ satisfying system (6.40) have support uniformly bounded in time and in the number of followers $N \in \mathbb{N}$. To this end, for every $N \in \mathbb{N}$ we define the mapping $\Pi_N : \mathbb{R}^{2dN} \to \mathcal{P}_1(\mathbb{R}^{2d})$ as follows

$$\Pi_N : (x_1, p_1, \dots, x_N, p_N) \mapsto \frac{1}{N} \sum_{i=1}^{N} \delta_{(x_i, N p_i)}. \tag{6.42}$$

We also introduce the auxiliary space $\mathcal{Y} \triangleq \mathbb{R}^{2dm} \times \mathcal{P}_1(\mathbb{R}^{2d})$, endowed with the distance

$$\|(y, q, \nu) - (y', q', \nu')\|_{\mathcal{Y}} \triangleq \|y - y'\| + \|q - q'\| + \mathcal{W}_1(\nu, \nu'). \tag{6.43}$$

Proposition 6.31. *Let* $y^0 \in \mathbb{R}^{dm}$, $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$, *and* μ_N^0 *be as in Definition 6.10–(i). Let* u_N^* *be a solution of Problem 4 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 4 with initial datum* (y^0, μ_N^0) *and control* u_N^* *given by Theorem 6.28.*

Then, by defining $\nu_N^*(t) \triangleq \Pi_N(x_N^*(t), p_N^*(t), \dots, x_{N,N}^*(t), p_{N,N}^*(t))$ *for every* $t \in [0,T]$*, it follows that the trajectories* $(y_N^*, q_N^*, \nu_N^*) : [0,T] \to \mathcal{Y}$ *are equibounded and equi-Lipschitz continuous.*

Furthermore, there exists $R_T > 0$*, depending only on* y^0*,* $\mathrm{supp}(\mu^0)$*,* d*,* 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*

$$\mathbb{H}(y_N^*(t), q_N^*(t), \nu_N^*(t), u_N^*(t)) = \mathbb{H}_c(y_N^*(t), q_N^*(t), \nu_N^*(t), u_N^*(t)) \quad \textit{for every } t \in [0,T].$$

Proof. As a first step, notice that the pair $(y_N^*(\cdot), x_N^*(\cdot))$ solves the system (6.14). It then follows from (6.16) and (6.17) that there exist two positive constants ρ_T and ρ_T', not depending on N such that it holds

$$\|y_{k,N}^*(t)\| \le \rho_T\,, \quad \|x_{i,N}^*(t)\| \le \rho_T, \quad \|\dot{y}_{k,N}^*(t)\| \le \rho_T'\,, \quad \|\dot{x}_{i,N}^*(t)\| \le \rho_T', \tag{6.44}$$

for all $i = 1, \dots, N$, for all $k = 1, \dots, m$, and for a.e. $t \in [0,T]$. It follows in particular that there exists a uniform constant W_T such that for all $t \in [0,T]$

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

We now observe that, for any $(y, q, x, p, u) \in \mathbb{R}^{4d} \times L^1([0,T]; \mathcal{U})$, by an explicit computation follows

$$\begin{aligned}
&N \nabla_{x_i} \mathbb{H}_N(y, q, x, p, u) \cdot e_l = \\
&\quad = N \frac{r_i}{N} \cdot \frac{1}{N} \sum_{j=1}^N \mathbf{D}K(x_i - x_j) e_l - N \sum_{j=1}^N \frac{r_j}{N} \cdot \frac{1}{N} \mathbf{D}K(x_j - x_i) e_l + \\
&\qquad + N \frac{r_i}{N} \cdot \nabla g(y)(x_i) - N \sum_{k=1}^m q_k \cdot \frac{1}{N} \mathbf{D}K(y_k - x_i) e_l - \\
&\qquad - N \frac{1}{N} \Bigg(\nabla_\xi \ell \left(y, x_i, \tfrac{1}{N} \textstyle\sum_{j=1}^N \omega(x_j) \right) \cdot e_l + \\
&\qquad + \nabla_\varsigma \ell \left(y, x_i, \tfrac{1}{N} \textstyle\sum_{j=1}^N \omega(x_j) \right) \mathbf{D}\omega(x_i) \cdot e_l \Bigg) \\
&\quad = \frac{1}{N} \sum_{j=1}^N (r_i - r_j) \cdot (\mathbf{D}K(x_i - x_j) e_l) + r_i \cdot \mathbf{D}_x g(y)(x_i) - \\
&\qquad - \sum_{k=1}^m q_k \cdot (\mathbf{D}K(y_k - x_i) e_l) - \nabla_\xi \ell \left(y, x_i, \tfrac{1}{N} \textstyle\sum_{j=1}^N \omega(x_j) \right) \cdot e_l - \\
&\qquad - \left(\nabla_\varsigma \ell \left(y, x_i, \tfrac{1}{N} \textstyle\sum_{j=1}^N \omega(x_j) \right) \mathbf{D}\omega(x_i) \right) \cdot e_l,
\end{aligned} \tag{6.45}$$

for each $i = 1, \dots, N$ and each $l = 1, \dots, d$ (where we have used that $\mathbf{D}K$ is even and we merged the first two terms).

Therefore, since $\dot{p}^*_{i,N}(\cdot)$ solves (6.40), by setting $\dot{r}^*_{i,N}(\cdot) \triangleq N\dot{p}^*_{i,N}(\cdot)$, we get

$$
\begin{aligned}
-\dot{r}^*_{i,N}(t)\cdot e_l = \frac{1}{N}\sum_{j=1}^{N}&(r^*_{i,N}(t)-r^*_{j,N}(t))\cdot(\mathbf{D}K(x^*_{i,N}(t)-x^*_{j,N}(t))e_l)+\\
&+r^*_{i,N}(t)\cdot(\mathbf{D}_y g(y^*_N(t))(x^*_{i,N}(t))e_l)-\sum_{k=1}^{m}q^*_{k,N}(t)\cdot(\mathbf{D}K(y^*_{k,N}(t)-x^*_{i,N}(t))e_l)-\\
&-\nabla_\xi \ell\left(y^*_N(t),x^*_{i,N}(t),\tfrac{1}{N}\textstyle\sum_{j=1}^{N}\omega(x^*_{j,N}(t))\right)\cdot e_l-\\
&-\left(\nabla_\varsigma \ell\left(y^*_N(t),x^*_{i,N}(t),\tfrac{1}{N}\textstyle\sum_{j=1}^{N}\omega(x^*_{j,N}(t))\right)\mathbf{D}\omega(x^*_{i,N}(t))\right)\cdot e_l,
\end{aligned}
$$

for each $i=1,\dots,N$, each $l=1,\dots,d$, and for a.e. $t\in[0,T]$, where we have used the fact that $\mathbf{D}K$ is even. We now denote with L_T a uniform constant satisfying

$$
\begin{gathered}
\|\mathbf{D}K\|_{L^\infty(B(0,\rho_T);\mathbb{R}^{d\times d})}\le L_T\,,\qquad \sup_{|y|\le\sqrt{m}R_T}\|\mathbf{D}_y g(y)(\cdot)\|_{L^\infty(B(0,\rho_T);\mathbb{R}^{d\times d})}\le L_T,\\
\|\nabla_\xi \ell\|_{L^\infty(B(0,\sqrt{m}\rho_T)\times B(0,\rho_T)\times B(0,W_T);\mathbb{R}^d)}\le L_T,\\
\|\nabla_\varsigma \ell\|_{L^\infty(B(0,\sqrt{m}\rho_T)\times B(0,\rho_T)\times B(0,W_T);\mathbb{R}^d)}\le L_T,\quad \|\mathbf{D}\omega\|_{L^\infty(B(0,\rho_T);\mathbb{R}^{d\times d})}\le L_T\,,
\end{gathered}
$$

and we easily get the estimate

$$
\|\dot{r}^*_{i,N}(t)\|\le\sqrt{d}L_T\left(2\|r^*_{i,N}(t)\|+\frac{1}{N}\sum_{j=1}^{N}\|r^*_{j,N}(t)\|+\sum_{k=1}^{m}\|q^*_{k,N}(t)\|+1+L_T\right)\tag{6.46}
$$

for each $i=1,\dots,N$ and for a.e. $t\in[0,T]$. An explicit computation of $\nabla_{y_k}\mathbb{H}_N$ and a similar argument, possibly with another constant L_T, show the estimate

$$
\|\dot{q}^*_{k,N}(t)\|\le\sqrt{d}L_T\left(\frac{1}{N}\sum_{i=1}^{N}\|r^*_{i,N}(t)\|+2\|q^*_{k,N}(t)\|+L_T\right)\tag{6.47}
$$

for each $k=1,\dots,m$ and for a.e. $t\in[0,T]$. We now set

$$
\varepsilon_N(t)\triangleq\sum_{k=1}^{m}\|q^*_{k,N}(t)\|+\frac{1}{N}\sum_{i=1}^{N}\|r^*_{i,N}(t)\|\qquad\text{for every } t\in[0,T],
$$

and observe that it holds

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

Therefore, (6.46) and (6.47) yield

$$
|\dot{\varepsilon}_N(t)|\le\sqrt{d}L_T\left(4\varepsilon_N(t)+1+2L_T\right).\tag{6.48}
$$

Defining then the increasing functions $\eta_N(t)$ through

$$
\eta_N(t)\triangleq\sup_{\tau\in[0,t]}\varepsilon_N(T-\tau)\qquad\text{for every } t\in[0,T],
$$

and observing that it holds $\eta_N(0) = 0$ for the boundary conditions in Theorem 6.28, from (6.48) and Gronwall's Lemma A.1 we obtain

$$\eta_N(\tau) \le \sqrt{d} L_T \tau (1 + 2L_T) e^{(4\sqrt{d} L_T)\tau},$$

and with this

$$\varepsilon_N(t) \le \eta_N(T) \le \sqrt{d} L_T T (1 + 2L_T) e^{(4\sqrt{d} L_T)T} \triangleq C_T \tag{6.49}$$

for all $t \in [0, T]$. Plugging (6.49) into (6.48), again by Gronwall's Lemma we get the existence of a constant C'_T such that

$$|\dot{\varepsilon}_N(t)| \le C'_T \tag{6.50}$$

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

$$\begin{aligned} \mathcal{W}_1(\nu^*_N(t+\tau), \nu^*_N(t)) \le \sqrt{2} \Bigg(\frac{1}{N} \sum_{i=1}^{N} \|x^*_{i,N}(t+\tau) - x^*_{i,N}(t)\| + \\ + \frac{1}{N} \sum_{i=1}^{N} \|r^*_{i,N}(t+\tau) - r^*_{i,N}(t))\| \Bigg), \end{aligned}$$

from the previous inequality, (6.44), (6.48), (6.49), and (6.50) we obtain that $y^*_N(\cdot)$ and $q^*_N(\cdot)$ are equibounded, that there exist a constant denoted by R_T for which $\mathrm{supp}(\nu^*_N(t)) \subset B(0, R_T)$ for all $t \in [0, T]$ and that $(y^*_N, q^*_N, \nu^*_N) : [0, T] \to \mathcal{Y}$ is Lipschitz continuous with Lipschitz constant uniform in N. □

Proposition 6.32. *Let $N \in \mathbb{N}$ and $u^*_N \in L^1([0, T]; \mathcal{U})$ be an optimal control for Problem 4 with given initial datum $(y^0_N, x^0_N) \in \mathbb{R}^{dm} \times \mathbb{R}^{dN}$, and let $(y^*_N(\cdot), q^*_N(\cdot), x^*_N(\cdot), p^*_N(\cdot)) \in \mathrm{Lip}([0, T], \mathbb{R}^{2dm} \times \mathbb{R}^{2dN})$ denote the corresponding trajectory of the PMP.*

Set $\nu^*_N(t) \triangleq \Pi_N(x^*_N(t), p^*_N(t), \ldots, x^*_{N,N}(t), p^*_{N,N}(t))$ for every $t \in [0, T]$. Then, the control u^*_N is optimal for Problem 3 and the pair control-trajectory $(u^*_N, (y^*_N, q^*_N, \nu^*_N))$ satisfies the extended PMP.

Proof. First observe that, by Proposition 6.31,

$$\mathbb{H}_c(y^*_N(t), q^*_N(t), \nu^*_N(t), u^*_N(t)) = \mathbb{H}(y^*_N(t), q^*_N(t), \nu^*_N(t), u^*_N(t))$$

holds for every $t \in [0, T]$. Moreover, we have

$$\begin{aligned} u^*_N(t) = \arg\max_{u(t) \in \mathcal{U}} \mathbb{H}_N(y^*_N(t), q^*_N(t), x^*_N(t), p^*_N(t), u(t)) \iff \\ \iff u^*_N(t) = \arg\max_{u(t) \in \mathcal{U}} \mathbb{H}(y^*_N(t), q^*_N(t), \nu^*_N(t), u(t)), \end{aligned}$$

for every $t \in [0, T]$, due to the specific form of the Hamiltonian $\mathbb{H}_N$ and $\mathbb{H}$, see Remarks 6.24 and 6.30.

We now prove that (for space reasons, we omit the time dependency)

$$\begin{cases} \dot{y}^*_{k,N} = \nabla_{q_k}\mathbb{H}_N(y^*_N, q^*_N, x^*_N, p^*_N, u^*_N), \\ \dot{q}^*_{k,N} = -\nabla_{y_k}\mathbb{H}_N(y^*_N, q^*_N, x^*_N, p^*_N, u^*_N), \end{cases} \implies \begin{cases} \dot{y}^*_{k,N} = \nabla_{q_k}\mathbb{H}(y^*_N, q^*_N, \nu^*_N, u^*_N), \\ \dot{q}^*_{k,N} = -\nabla_{y_k}\mathbb{H}(y^*_N, q^*_N, \nu^*_N, u^*_N), \end{cases}$$

i.e., that if the (y, q) variables satisfy the PMP for Problem 4 then they satisfy the extended PMP for Problem 3. It is sufficient to observe that x^*_N and p^*_N in $\mathbb{H}_N$ can be substituted by ν^*_N as follows

$$\begin{aligned} \mathbb{H}_N(y^*_N(t), q^*_N(t), \nu^*_N(t), u^*_N(t)) = &\int_{\mathbb{R}^{4d}} r \cdot (K * \pi_{1\#}\nu^*_N(t))(x)\, d\nu^*_N(t, x, r) + \\ &+ \int_{\mathbb{R}^{2d}} r \cdot g(y^*_N(t))(x)\, d\nu^*_N(t, x, r) + \\ &+ \sum_{k=1}^{m} \int_{\mathbb{R}^{2d}} q^*_{k,N}(t) \cdot K(y^*_{k,N}(t) - x)\, d\nu^*_N(t, x, r) + \\ &+ \sum_{k=1}^{m} q^*_{k,N}(t) \cdot (f_k(y^*_{k,N}(t)) + B_k u^*_N(t)) - \\ &- L(y^*_N(t), \pi_{1\#}\nu^*_N(t)) - \gamma(u^*_N(t)), \end{aligned}$$

where we used the rescaled variable $r = Np$.

Comparing the above expression with that of $\mathbb{H}(y^*_N(t), q^*_N(t), \nu^*_N(t), u^*_N(t))$, one has that their expressions coincide up to the first term. Since such first term is independent of y_k and q_k, then $\nabla_{y_k}\mathbb{H}_N = \nabla_{y_k}\mathbb{H}$ and $\nabla_{q_k}\mathbb{H}_N = \nabla_{q_k}\mathbb{H}$, hence the equations for $\dot{y}^*_{k,N}$ and $\dot{q}^*_{k,N}$ in the PMP for Problem 4 and in the extended PMP for Problem 3 coincide.

We now prove a similar result for the $x^*_{i,N}(\cdot)$ and $r^*_{i,N}(\cdot)$, with $r^*_{i,N}(\cdot) = Np^*_{i,N}(\cdot)$. After this change of variable, the third and the fourth equation in (6.40) become

$$\begin{cases} \dot{x}^*_{i,N}(t) = N\nabla_{r_i}\mathbb{H}_N(y^*_N(t), q^*_N(t), x^*_N(t), p^*_N(t), u^*_N(t)), \\ \dot{r}^*_{i,N}(t) = -N\nabla_{x_i}\mathbb{H}_N(y^*_N(t), q^*_N(t), x^*_N(t), p^*_N(t), u^*_N(t)). \end{cases}$$

We want to prove that the following identity holds (again omitting the time dependency)

$$\mathbf{J}(\nabla_\nu \mathbb{H}_c(y^*_N, q^*_N, \nu^*_N, u^*_N))(x^*_{i,N}, r^*_{i,N}) = \begin{pmatrix} N\nabla_{r_i}\mathbb{H}_N(y^*_N, q^*_N, x^*_N, p^*_N, u^*_N) \\ -N\nabla_{x_i}\mathbb{H}_N(y^*_N, q^*_N, x^*_N, p^*_N, u^*_N) \end{pmatrix}, \tag{6.51}$$

i.e., that the Hamiltonian vector fields generated by $\mathbb{H}$ and $\mathbb{H}_N$ coincide in each point of the trajectory $(x^*_{i,N}(\cdot), r^*_{i,N})(\cdot)$. The presence of the constant N in the right-hand side is due to the change of variables $r = Np$. By applying $\mathbf{J}^{-1}$ on both sides of (6.51), we need to prove

$$\begin{aligned} \nabla_\nu \mathbb{H}_c(y^*_N, q^*_N, &\nu^*_N, u^*_N)(x^*_{i,N}, r^*_{i,N}) \cdot e_l = \\ &= N\nabla_{x_i}\mathbb{H}_N(y^*_N, q^*_N, x^*_N, p^*_N, u^*_N) \cdot e_l, \end{aligned} \qquad \text{for } l = 1, \ldots, d, \tag{6.52}$$

$$\begin{aligned} \nabla_\nu \mathbb{H}_c(y^*_N, q^*_N, &\nu^*_N, u^*_N)(x^*_{i,N}, r^*_{i,N}) \cdot e_l = \\ &= N\nabla_{r_i}\mathbb{H}_N(y^*_N, q^*_N, x^*_N, p^*_N, u^*_N) \cdot e_{l-d}, \end{aligned} \qquad \text{for } l = d+1, \ldots, 2d. \tag{6.53}$$

By writing explicitly the left hand sides of (6.52) and (6.53), by using the expressions (6.38)–(6.39) and evaluating them in $(x^*_{i,N}, r^*_{i,N})$, we have

$$\begin{aligned}\nabla_\nu \mathbb{H}_c(y^*_N, q^*_N, \nu^*_N, u^*_N)(x^*_{i,N}, r^*_{i,N}) \cdot e_l = \\ = \frac{1}{N}\sum_{j=1}^{N}(r^*_{i,N} - r^*_{j,N}) \cdot \left(\mathbf{D}K(x^*_{i,N} - x^*_{j,N})e_l\right) + r^*_{i,N} \cdot \left(\mathbf{D}_x g(y^*_N)(x^*_{i,N})e_l\right) - \\ - \sum_{k=1}^{m} q^*_{k,N} \cdot \left(\mathbf{D}K(y^*_{k,N} - x^*_{i,N})e_l\right) - \nabla_\xi \ell\left(y^*_N, x^*_{i,N}, \tfrac{1}{N}\textstyle\sum_{j=1}^N \omega(x^*_{j,N})\right) \cdot e_l - \\ - \left(\nabla_\varsigma \ell\left(y^*_N, x^*_{i,N}, \tfrac{1}{N}\textstyle\sum_{j=1}^N \omega(x^*_{j,N})\right)\mathbf{D}\omega(x^*_{i,N})\right) \cdot e_l,\end{aligned}$$

for $l = 1, \dots, d$, so that (6.52) follows immediately from (6.45). Similarly, we have

$$\nabla_\nu \mathbb{H}_c(y^*_N, q^*_N, \nu^*_N, u^*_N)(x^*_{i,N}, r^*_{i,N}) \cdot e_l = \frac{1}{N}\sum_{j=1}^{N} K(x^*_{i,N} - x^*_{j,N}) \cdot e_{l-d} + g(y^*_N)(x^*_{i,N}) \cdot e_{l-d},$$

for $l = d+1, \dots, 2d$, which coincides with the right hand side of (6.53) by an explicit computation. After the identification $\nu^*_N = \Pi_N(x^*_{1,N}, p^*_{1,N}, \dots, x^*_{N,N}, p^*_{N,N})$, the boundary conditions of Problem 4 and Problem 3 coincide too, hence the result follows now by (6.52)–(6.53) arguing, for instance, as in [106, Lemma 4.3]. □

Remark 6.33. It is interesting to observe that the process of rewriting a trajectory of the PMP for Problem 4 in the empirical measure formulation depends on the number N of followers, see the definition of Π_N in (6.42). This is a consequence of the fact that the Hamiltonian of the PMP for Problem 4 actually depends on N. Indeed, consider a population of followers composed of a unique agent with state and adjoint variables (x_1, p_1), for which the first term of the Hamiltonian reads as $p_1 \cdot g(y)(x_1)$. Consider now another population of followers composed of two agents (x_1, p_1) and (x_2, p_2) satisfying $x_1 = x_2$ and $p_1 = p_2$, for which the first term of the Hamiltonian reads as $2p_1 \cdot g(y)(x_1)$. Clearly, in both cases the empirical measure in the state variables is $\mu_1 = \mu_2 = \delta_{x_1}$, while the definition of Π_N gives two different empirical measures for the cotangent bundle: $\nu_1 = \delta_{(x_1,p_1)}$ and $\nu_2 = \delta_{(x_1,2p_1)}/2$. This difference is needed to compensate the dependence of the Hamiltonian of the PMP for Problem 4 on N.

6.4.2 The Wasserstein gradient

At the beginning of this section, we anticipated that the dynamics of $\nu^*(\cdot)$ in (6.36) is an Hamiltonian flow in the Wasserstein space of probability measures, in the sense of [12]. This means that, for every $t \in [0, T]$, the vector field $\nabla_\nu \mathbb{H}_c(\nu^*(t))$ is an element with minimal norm in the Fréchet subdifferential at the point $\nu^*(t)$ of the maximized Hamiltonian $\mathbb{H}_c$ introduced in the statement of Theorem 6.23 (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 [13, Chapter 10], which however cannot be applied verbatim to our case due to the peculiar nature of our operators. 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 have already proven in Proposition 6.31

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.
We start with some basic definitions and general results on functionals defined on $\mathcal{P}_2(\mathbb{R}^{2d})$: the following one is motivated by Definition 10.3.1 and Remark 10.3.3 in [13].

Definition 6.34. Let $\psi : \mathcal{P}_2(\mathbb{R}^{2d}) \to (-\infty, +\infty]$ be a proper and lower semicontinuous functional, and let

$$\nu_0 \in D(\psi) \triangleq \left\{ \nu \in \mathcal{P}_2(\mathbb{R}^{2d}) : \psi(\nu) < +\infty \right\}.$$

We say that $w \in L^2(\mathbb{R}^{2d}, \nu_0)$ 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

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

It can be seen [12] that whenever $\partial\psi(\nu_0)$ is nonempty, it has an element with minimal $L^2(\mathbb{R}^{2d}; \nu_0)$-norm, which we call the *Wasserstein gradient* $\nabla_\nu \psi(\nu_0)$ of ψ at ν_0.

Proposition 6.35 ([13, Theorem 10.3.10])**.** *Fix the functional $\psi : \mathcal{P}_2(\mathbb{R}^{2d}) \to (-\infty, +\infty]$. Then, for every $\nu_0 \in D(\psi)$, the metric slope*

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

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

The following property shall be used to prove that the subdifferential of $\mathbb{H}_c$ is nonempty.

Definition 6.36. A proper, lower semicontinuous functional $\psi : \mathcal{P}_2(\mathbb{R}^n) \to (-\infty, +\infty]$ is *semiconvex along geodesics* whenever, for every $\nu_0, \nu_1 \in \mathcal{P}_2(\mathbb{R}^n)$ and $\rho \in \Gamma_o(\nu_0, \nu_1)$ there exists $C \in \mathbb{R}$ for which for every $s \in [0, 1]$ it holds

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

In what follows, we shall fix $y, q \in \mathbb{R}^{dm}$ and $u \in L^1([0, T]; \mathcal{U})$, omit the time dependency, and write for the sake of compactness $\mathbb{H}_c(\nu)$ in place of $\mathbb{H}_c(y, q, \nu, u)$. Moreover, Ω shall denote a convex, compact subset of $\mathbb{R}^{2d}$ and $z = (x, r)$ a variable in $\mathbb{R}^{2d}$.
Whenever $\text{supp}(\nu) \subseteq \overline{B(0, R_T)}$, $\mathbb{H}_c(\nu)$ can be rewritten as

$$\mathbb{H}_c(\nu) = \frac{1}{2} \int_{\mathbb{R}^{4d}} \mathcal{F}(z - z') d\nu(z)\nu(z') + \int_{\mathbb{R}^d} \mathcal{G}(z) d\nu(z) - \int_{\mathbb{R}^d} \ell(\pi_1(z), \textstyle\int \omega \pi_{1\#}\nu) d\nu(z) + Q,$$

where we have denoted by

$$\mathcal{F}(x, r) \triangleq r \cdot K(x) \quad \text{ and } \quad \mathcal{G}(x, r) \triangleq r \cdot g(y)(x) + \sum_{k=1}^m q_k \cdot K(y_k - x),$$

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

In order to prove the semiconvexity of $\mathbb{H}_c$, we shall establish the semiconvexity of the following functionals for any ν satisfying $\mathrm{supp}(\nu) \subseteq \Omega$:

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

where $\hat{\mathcal{F}}$, $\hat{\mathcal{G}}$, $\hat{\ell}$, and $\hat{\omega}$ are $\mathcal{C}^2$ functions. The desired result follows 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 $\Omega = \overline{B(0, R_T)}$.
The following simple property is needed to prove semiconvexity of the above functionals.

Lemma 6.37. *Let $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$ with support contained in Ω. Let $\rho \in \Gamma(\nu_0, \nu_1)$ and denote by*

$$\nu_s \triangleq ((1-s)\pi_1 + s\pi_2)_{\#}\rho \quad \text{for every } s \in [0,1]. \tag{6.54}$$

Then, it holds

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

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

$$\mathrm{supp}(\rho) \subseteq \Omega \times \Omega. \tag{6.55}$$

This follows from the identity

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

and from the fact that both $\mathbb{R}^{2d} \times (\mathbb{R}^{2d} \backslash \Omega))$ and $(\mathbb{R}^{2d} \backslash \Omega) \times \mathbb{R}^{2d}$ are ρ-null sets by hypothesis. To prove the statement, it suffices to show that for all $f \in \mathcal{C}(\mathbb{R}^{2d}; \mathbb{R})$ satisfying $f|_{\Omega} \equiv 0$ it holds

$$\int_{\mathbb{R}^{2d}} f \; d\nu_s = 0, \qquad \text{for every } s \in [0,1]. \tag{6.56}$$

Indeed we have

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

since, by (6.55), $\mathrm{supp}(\rho) \subseteq \Omega \times \Omega$. From the convexity of Ω follows that $(1-s)z_0 + sz_1 \in \Omega$ for every $s \in [0,1]$, which, together with the assumption $f|_{\Omega} \equiv 0$, yield (6.56), as desired. □

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

Remark 6.38. Let Ω be a convex, compact subset of $\mathbb{R}^{2d}$ and let $f \in \mathcal{C}^2(\mathbb{R}^{2d}; \mathbb{R})$. Then there exists $C_{\Omega,f} \in \mathbb{R}$ depending only on Ω and f such that

$$f((1-s)x_0 + sx_1) \leq (1-s)f(x_0) + sf(x_1) + C_{\Omega,f}s(1-s)\|x_0 - x_1\|^2, \tag{6.57}$$

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

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

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

Proof. We may assume $\mathrm{supp}(\nu_0), \mathrm{supp}(\nu_1) \subseteq \Omega$ for some convex and compact set $\Omega \subset \mathbb{R}^{2d}$, otherwise the inequality is trivial. Hence, from Lemma 6.37, it follows $\mathrm{supp}(\nu_s) \subseteq \Omega$ 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 [13, Proposition 9.3.2, Proposition 9.3.5]. ☐

Corollary 6.40. *Let $\hat{\omega} \in \mathcal{C}^2(\mathbb{R}^{2d}; \mathbb{R}^d)$, $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$, $\rho \in \Gamma(\nu_0, \nu_1)$ and define ν_s as in* (6.54) *for $s \in [0,1]$. If we set*

$$\xi_s \triangleq \int_{\mathbb{R}^{2d}} \hat{\omega} d\nu_s \quad \text{for every } s \in [0,1]. \tag{6.58}$$

for every $s \in [0,1]$, then

$$\|\xi_s - (1-s)\xi_0 - s\xi_1\| \leq Cs(1-s)\mathcal{W}_2(\nu_0,\nu_1)^2 \quad \text{for every } s \in [0,1],$$

where C is independent of ν_0 and ν_1.

Proof. Follows from Lemma 6.39 applied first to the functions $\hat{\mathcal{F}} \equiv 0$ and $\hat{\mathcal{G}} = \hat{\omega}$, and then to $\hat{\mathcal{F}} \equiv 0$ and $\hat{\mathcal{G}} = -\hat{\omega}$. ☐

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

Lemma 6.41. *Suppose that $\hat{\ell} \in \mathcal{C}^2(\mathbb{R}^{2d} \times \mathbb{R}^d; \mathbb{R})$, let $z_0, z_1 \in \Omega$ and set $z_s \triangleq (1-s)z_0 + sz_1$ for all $s \in [0,1]$. Furthermore, let $\nu_0, \nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$, $\rho \in \Gamma(\nu_0, \nu_1)$ and define ν_s and ξ_s as in* (6.54) *and* (6.58) *for $s \in [0,1]$. Then, for all $s \in [0,1]$, it holds*

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

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

Proof. Since Ω is convex, $z_s \in \Omega$ for all $s \in [0,1]$. Moreover, $(1-s)\xi_0 + s\xi_1 \in \widehat{\Omega}$ for all $s \in [0,1]$, for some convex and compact set $\widehat{\Omega} \subset \mathbb{R}^d$. Notice that from (6.57) follows

$$\begin{aligned}\hat{\ell}(z_s, (1-s)\xi_0 + s\xi_1) \leq{}& (1-s)\hat{\ell}(z_0,\xi_0) + s\hat{\ell}(z_1,\xi_1) + \\ &+ C_{\Omega,\widehat{\Omega}}\, s(1-s)\left(\|z_0 - z_1\|^2 + \|\xi_0 - \xi_1\|^2\right),\end{aligned} \tag{6.59}$$

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

$$\|\xi_0 - \xi_1\|^2 \leq \mathrm{Lip}_\Omega(\omega)\mathcal{W}_1(\nu_0,\nu_1)^2 \leq \mathrm{Lip}_\Omega(\omega)\mathcal{W}_2(\nu_0,\nu_1)^2. \tag{6.60}$$

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

$$\begin{aligned}\|\hat{\ell}(z_s,\xi_s) - \hat{\ell}(z_s,(1-s)\xi_0 + s\xi_1)\| &\leq \mathrm{Lip}_{\Omega\times\widehat{\Omega}} \|\xi_s - (1-s)\xi_0 - s\xi_1\| \\ &\leq \mathrm{Lip}_{\Omega\times\widehat{\Omega}} s(1-s)C\mathcal{W}_2(\nu_0,\nu_1)^2.\end{aligned} \tag{6.61}$$

Hence, for every $s \in [0,1]$, using (6.59), (6.60) and (6.61), we get

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

This concludes the proof. □

Corollary 6.42. *Let $\nu_0,\nu_1 \in \mathcal{P}_c(\mathbb{R}^{2d})$ and $\rho \in \Gamma_o(\nu_0,\nu_1)$. Then, there exists $C \in \mathbb{R}$ independent of ν_0 and ν_1 for which*

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

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

Proof. Notice that, by Lemma 6.37, $\hat{\mathbb{H}}^2_c(\nu_s)$ can be rewritten as

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

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

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

The statement then follows from Lemma 6.41. □

Proposition 6.43. *The functional $\mathbb{H}_c$ is semiconvex along geodesics.*

Proof. Follows directly from Lemma 6.39 and Corollary 6.42, by noticing that $\mathbb{H}_c(\nu) = \hat{\mathbb{H}}^1_c(\nu) + \hat{\mathbb{H}}^2_c(\nu)$ for the choices $\hat{\mathcal{F}} = \mathcal{F}$, $\hat{\mathcal{G}} = \mathcal{G}$, $\hat{\ell} = -\ell \circ (\pi_1, \mathrm{Id})$, $\hat{\omega} = \omega \circ \pi_1$ and $\Omega = \overline{B(0,R_T)}$. □

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

$$\nabla_\nu \mathcal{L}(z) \triangleq \begin{pmatrix} \nabla_\xi \ell(y,\pi_1(z), \textstyle\int \omega \pi_{1\#}\nu) + \nabla_\varsigma \ell(y,\pi_1(z), \textstyle\int \omega \pi_{1\#}\nu)\mathbf{D}\omega(\pi_1(z)) \\ 0 \end{pmatrix}$$

for every $z \in \mathbb{R}^{2d}$. This notation is reminiscent of the fact that this vector field will eventually turn out to be the 2-Wasserstein gradient of the functional L, as it will follow from Theorem 6.45 in the case $\mathcal{F} = \mathcal{G} \equiv 0$. We can thus define our candidate vector field for the Wasserstein gradient $\nabla_\nu \mathbb{H}_c(\nu_0)$ in the case that $\mathrm{supp}(\nu_0) \subseteq B(0, R_T)$:

$$w \overset{\Delta}{=} (\nabla \mathcal{F}) * \nu + \nabla \mathcal{G} - \nabla_\nu \mathcal{L}. \tag{6.62}$$

Notice that, by the set of hypotheses (H), w is a continuous function in z, and hence it is well-defined ν-a.e.

Lemma 6.44. *Let $\nu \in \mathcal{P}_c(\mathbb{R}^{2d})$. Then w defined by* (6.62) *belongs to $L^p(\mathbb{R}^{2d}, \nu)$ for every $p \in [1, +\infty]$ and it satisfies*

$$\begin{aligned}\int_{\mathbb{R}^{4d}} w(z_0) \cdot (z_1 - z_0) d\rho(z_0, z_1) = \int_{\mathbb{R}^{6d}} \Big(\nabla \mathcal{F}(z_0 - z_2) + \nabla \mathcal{G}(z_0) - \\ - \nabla_\nu \mathcal{L}(z_0)\Big) \cdot (z_1 - z_0) d\rho(z_0, z_1) d\nu(z_2),\end{aligned} \tag{6.63}$$

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

Proof. Since w is continuous, the fact that w is $L^p(\mathbb{R}^{2d}, \nu)$-integrable follows the fact that ν has compact support. Equation (6.63) then follows by Fubini-Tonelli and from the fact that ρ is compactly supported too by Remark 6.37. □

Theorem 6.45. *Let $\nu \in \mathcal{P}_2(\mathbb{R}^{2d})$ be such that* $\mathrm{supp}(\nu) \subseteq B(0, R_T)$. *Then $\nu \in D(|\partial \mathbb{H}_c|)$ if and only if w as in* (6.62) *belongs to $L^2(\mathbb{R}^{2d}, \nu)$. In this case, $\|w\|_{L^2(\mathbb{R}^{2d}, \nu)} = |\partial \mathbb{H}_c|(\nu)$, i.e., w is an element with minimal norm in $\partial \mathbb{H}_c(\nu)$.*

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 $L^2(\mathbb{R}^{2d}, \nu)$ and that $\|w\|_{L^2(\mathbb{R}^{2d}, \nu)} \leq |\partial \mathbb{H}_c|(\nu)$. We compute the directional derivative of $\mathbb{H}_c$ along a direction induced by the transport map $Id + \xi$, where ξ is a smooth function with compact support such that $\mathrm{supp}((Id + s\xi)_{\#}\nu) \subseteq \overline{B(0, R_T)}$ for any sufficiently small $s > 0$. If we denote by

$$\begin{aligned}\mathcal{L}_1(s) &\overset{\Delta}{=} \ell(y, \pi_1(z_0) + s(\pi_1 \circ \xi)(z_1), \textstyle\int \omega d(\pi_1 \circ (Id + s\xi))_{\#}\nu), \\ \mathcal{L}_2(s) &\overset{\Delta}{=} \ell(y, \pi_1(z_0), \textstyle\int \omega d(\pi_1 \circ (Id + s\xi))_{\#}\nu),\end{aligned}$$

then the map

$$\begin{aligned}s \mapsto \frac{\mathcal{F}((z_0 - z_1) + s(\xi(z_0) - \xi(z_1))) - \mathcal{F}(z_0 - z_1)}{s} + \\ + \frac{\mathcal{G}(z_0 + s\xi(z_0)) - \mathcal{G}(z_0)}{s} - \frac{\mathcal{L}_1(s) - \mathcal{L}_2(s)}{s} - \frac{\mathcal{L}_2(s) - \mathcal{L}_2(0)}{s},\end{aligned}$$

as $s \to 0$ converges to then the map

$$\nabla \mathcal{F}(z_0 - z_1) \cdot (\xi(z_0) - \xi(z_1)) + (\nabla \mathcal{G}(z_0) - \nabla_\nu \mathcal{L}(z_0)) \cdot \xi(z_0).$$

Since ν has compact support and $\nabla \mathcal{F}$ is odd, the dominated convergence theorem and the identity (6.63) imply

$$+\infty > \lim_{s \to 0} \frac{\mathbb{H}_c((Id + s\xi)_{\#}\nu) - \mathbb{H}_c(\nu)}{s}$$

$$= \frac{1}{2}\int_{\mathbb{R}^{4d}} \nabla\mathcal{F}(z_0 - z_1)\cdot(\xi(z_0)-\xi(z_1))d\nu(z_0)d\nu(z_1)+ \\ + \int_{\mathbb{R}^{2d}} (\nabla\mathcal{G}(z_0) - \nabla_\nu\mathcal{L}(z_0))\cdot\xi(z_0)d\nu(z_0) \\ = \int_{\mathbb{R}^{2d}} w(z_0)\cdot\xi(z_0)d\nu(z_0).$$

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

$$\mathcal{W}_2((Id + s\xi)_\#\nu, \nu) \le s\|\xi\|_{L^2(\mathbb{R}^{2d},\nu)},$$

we get

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

and finally

$$\left|\int_{\mathbb{R}^{2d}} w(z_0)\cdot\xi(z_0)d\nu(z_0)\right| \le |\partial\mathbb{H}_c|(\nu)\|\xi\|_{L^2(\mathbb{R}^{2d},\nu)}.$$

This proves that $w \in L^2(\mathbb{R}^{2d},\nu)$ and that $\|w\|_{L^2(\mathbb{R}^{2d},\nu)} \le |\partial\mathbb{H}_c|(\nu)$.
We now prove that the vector w belongs to $\partial\mathbb{H}_c$; 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 6.35.
For proving the claim, we start by remarking that by Proposition 6.35, the vector $w \in L^2_\nu(\mathbb{R}^{2d})$. We further consider a test measure $\overline{\nu}$, a plan $\rho \in \Gamma_o(\nu,\overline{\nu})$, and we compute the directional derivative of $\mathbb{H}_c$ along the direction induced by ρ. Denoting by

$$\mathcal{L}_1(s) \triangleq \ell(y, (1-s)z_0 + sz_1, \textstyle\int \omega d((1-s)\pi_1 + s\pi_2)_\#\rho), \\ \mathcal{L}_2(s) \triangleq \ell(y, z_0, \textstyle\int \omega d((1-s)\pi_1 + s\pi_2)_\#\rho),$$

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

$$s \mapsto \frac{\mathcal{F}((1-s)(z_0-\overline{z}_0) + s(z_1-\overline{z}_1)) - \mathcal{F}(z_0-\overline{z}_0)}{s} + \\ + \frac{\mathcal{G}((1-s)z_0+sz_1) - \mathcal{G}(z_0)}{s} - \frac{\mathcal{L}_1(s)-\mathcal{L}_2(s)}{s} - \frac{\mathcal{L}_2(s)-\mathcal{L}_2(0)}{s},$$

as $s \to 0$ converges to

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

Hence, since $\nabla\mathcal{F}$ is odd, we get from Proposition 6.43, the dominated convergence theorem and the identity (6.63), the following inequality

$$\mathbb{H}_c(\overline{\nu}) - \mathbb{H}_c(\nu) \ge \lim_{s\to 0}\frac{\mathbb{H}_c(((1-s)\pi_1+s\pi_2)_\#\rho) - \mathbb{H}_c(\nu)}{s} + o(\mathcal{W}_2(\overline{\nu},\nu)) \\ = \frac{1}{2}\int_{\mathbb{R}^{8d}} \nabla\mathcal{F}(z_0-\overline{z}_0)\cdot((z_1-z_0)-(\overline{z}_1-\overline{z}_0))\,d\rho(z_0,z_1)d\rho(\overline{z}_0,\overline{z}_1)$$

$$+\int_{\mathbb{R}^{4d}} (\nabla\mathcal{G}(z_0) - \nabla_\nu \mathcal{L}(z_0)) \cdot (z_1 - z_0) d\rho(z_0, z_1) + o(\mathcal{W}_2(\overline{\nu}, \nu))$$
$$= \int_{\mathbb{R}^{4d}} w(z_0) \cdot (z_1 - z_0) d\rho(z_0, z_1) + o(\mathcal{W}_2(\overline{\nu}, \nu)).$$

This proves that $w \in \partial\mathbb{H}_c(\nu)$. □

6.4.3 Proof of Theorem 6.23

In this section, we prove Theorem 6.23. We first recall that we already shown in Corollary 6.20 that there exists a mean-field optimal control for Problem 3. 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 3 with initial datum (y^0, μ^0). Fix μ_N^0 as in Definition 6.10–(i), and consider a sequence $(u_N^*)_{N\in\mathbb{N}}$ of optimal controls of Problem 4 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 by $(y_N^*(\cdot), x_N^*(\cdot))$ the trajectory of (6.14) corresponding to the control u_N^* and the initial datum (y^0, μ_N^0) of Problem 4. Compute the corresponding pair control-trajectory $(u_N^*, (y_N^*, q_N^*, x_N^*, p_N^*))$ satisfying the PMP for Problem 4, that exists due to Theorem 6.28. Set $\nu_N^*(\cdot) \triangleq \Pi_N(x_N^*(\cdot), p_N^*(\cdot))$ and $r_N^*(\cdot) \triangleq Np_N^*(\cdot)$. By Proposition 6.31, the trajectories $(y_N^*(\cdot), q_N^*(\cdot), \nu_N^*(\cdot))$ are equibounded and equi-Lipschitz from $[0,T]$ to the space $\mathcal{Y}$ endowed with the distance (6.43), and the empirical measure $\nu_N^*(\cdot)$ have equibounded support. Moreover, the pair $(u_N^*, (y_N^*, q_N^*, \nu_N^*))$ satisfies the extended PMP by Proposition 6.32.

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 $(y^*, q^*, \nu^*) : [0,T] \to \mathcal{Y}$ 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^* proven in Proposition 6.13 we get $\pi_{1\#}\nu^* = \mu^*$. Observe that $(y^*(\cdot), q^*(\cdot), \nu^*(\cdot))$ is a Lipschitz function with respect to time and $\nu^*(\cdot)$ 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 $q^*(T) = 0$, $\pi_{2\#}(\nu^*(T)) = \delta_0$.

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

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

Recall that by (6.40) and by Remark 6.30, $u_N^*(t)$ satisfies

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

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 the variable $u \in \mathcal{U}$. 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

$$v_t = \operatorname*{arg\,max}_{u(t)\in\mathcal{U}} E(q^*(t), u(t)) \tag{6.64}$$

and is therefore uniquely determined. This shows that the sequence $(u_N^*)_{N\in\mathbb{N}}$ is pointwise converging in $[0,T]$ to the function $v(t) \triangleq v_t$. 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 (6.64) implies that

$$u_N^* \to u^* \text{ strongly in } L^1([0,T];\mathcal{U}) \tag{6.65}$$

and that

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

for a.e. $t \in [0,T]$. Due to the explicit expression of $\mathbb{H}(y,q,\nu,u)$ in (6.37), this is equivalent to say that

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

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

We finally prove that $(y^*(\cdot), q^*(\cdot), \nu^*(\cdot))$ satisfies the Hamiltonian system (6.36) with control u^*. Due to equi-Lipschitz continuity, we have that the derivatives $(\dot{y}_N^*, \dot{q}_N^*)$, and $\partial\nu_N^*/\partial t$ converge to $(\dot{y}^*, \dot{q}^*)$, and $\partial\nu^*/\partial t$ weakly in $L^1([0,T];\mathbb{R}^{2dm})$ and in the sense of distributions, respectively. Observe now that by (6.38) and (6.39) the vector field $\nabla_\nu \mathbb{H}_c(y,q,\nu)(\cdot,\cdot)$ – which is independent of u, so we can omit u – is continuously depending on the variables (y,q,ν). By the uniform convergence of (y_N^*, q_N^*, ν_N^*) and since $\operatorname{supp}(\nu_N^*(t)) \subseteq B(0,R_T)$ for all $t \in [0,T]$ we get that

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

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 $\operatorname{supp}(\nu_N^*(t)) \subseteq B(0,R_T)$, we then get the uniform bound

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

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

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

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 properties above are enough to show that

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

in the sense of distributions, so that $\nu^*(\cdot)$ solves the third equation in (6.36).

For all $k = 1, \ldots, m$, taking derivatives in the explicit expression in (6.37) 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 $q^*(\cdot)$ solves the second equation in (6.36).
A similar argument, also using the strong L^1 convergence of u_N^* to u^* proved in (6.65), 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 $y^*(\cdot)$ solves the first equation in (6.36). This concludes the proof of Theorem 6.23.

6.4.4 An example

We now show the application of the extended Pontryagin Maximum Principle to the control problem (6.10). For simplicity of notation, we study the $d = 1$ dimensional problem. Given a positive function $a \in \mathcal{C}^2(\mathbb{R}_+;\mathbb{R}_+)$ (like the Cucker-Smale kernel (1.7)), the mean-field limit for $N \to \infty$ is given by (see Proposition 6.13)

$$\begin{cases} \dot{y}_k = w_k, & \\ \dot{w}_k = (G^{[a]} * \mu)(y_k, w_k) + \dfrac{1}{m}\displaystyle\sum_{j=1}^m a(|y_k - y_j|)(w_j - w_k) + u_k, & k = 1, \dots, m, \\ \dfrac{\partial \mu}{\partial t} = -v \cdot \nabla_x \mu - \nabla_v \cdot \left[\left(G^{[a]} * \mu + \dfrac{1}{m}\displaystyle\sum_{j=1}^m a(|x - y_j|)(w_j - v)\right)\mu\right], & \end{cases} \tag{6.66}$$

where $\mu = \mu(x,v)$ is the density of followers and $G^{[a]}(x,v) = -a(|x|)v$. The time dependency is omitted to keep the notation compact.
Notice that this is a particular case of system (6.12), where the state variables for the leaders and the followers are

$$\mathbf{y}_k = \begin{pmatrix} y_k \\ w_k \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x \\ v \end{pmatrix},$$

one sets

$$\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_m), \qquad K(\mathbf{x}) = \begin{pmatrix} 0 \\ G^{[a]}(\mathbf{x}) \end{pmatrix},$$

$$f_k(\mathbf{y}) = \begin{pmatrix} w_k \\ \dfrac{1}{m}\displaystyle\sum_{j=1}^m a(|y_k - y_j|)(w_j - w_k) \end{pmatrix}, \qquad g(\mathbf{y})(\mathbf{x}) = \begin{pmatrix} v \\ \dfrac{1}{m}\displaystyle\sum_{j=1}^m a(|x - y_j|)(w_j - v) \end{pmatrix},$$

and, for every $k = 1, \dots, m$, B_k is the $2 \times m$ matrix defined as

$$B_k : \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \in \mathbb{R}^m \mapsto \begin{pmatrix} 0 \\ u_k \end{pmatrix} \in \mathbb{R}^2.$$

Notice that since $a(|\cdot|)$ is a radial function, the function $G^{[a]}$, and thus K, is odd.

As seen since Chapter 1, a standard problem in the study of Cucker-Smale systems is to find conditions to ensure flocking, i.e., alignment of the whole crowd towards the same velocity. For this reason, it is interesting in our case to study the minimization of the variance of the crowd, by choosing

$$\begin{aligned} L(\mathbf{y},\mu) &= \int_{\mathbb{R}^{2d}} \left(\frac{2}{m}\sum_{k=1}^m |w_k|^2 + 2|v|^2\right) d\mu(x,v) - \left|\frac{1}{m}\sum_{k=1}^m w_k + \int_{\mathbb{R}^{2d}} v d\mu(x,v)\right|^2 \\ &= \int_{\mathbb{R}^{2d}} \left(\frac{2}{m}\sum_{k=1}^m |w_k|^2 + 2|v|^2 - \left(\frac{1}{m}\sum_{k=1}^m w_k + \int v' d\mu(x',v')\right)\times \right. \\ &\quad \left. \times \left(\frac{1}{m}\sum_{k=1}^m w_k + v\right)\right) d\mu(x,v). \end{aligned} \tag{6.67}$$

(For simplicity of computation, we have considered the minimization of 4 times the variance.) The Lagrangian L falls into hypothesis (L) since, by choosing $\omega(\mathbf{x}) = v$ and

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

it is of the form $\int_{\mathbb{R}^{2d}} \ell(\mathbf{y},\mathbf{x},\int \omega\mu) d\mu(\mathbf{x})$. For the control constraints, we set $\mathcal{U} = [-1,1]^m$ and we choose to penalize the L^2-norm of the control, i.e., $\gamma(u) = |u|^2$.

Remark 6.46. Other forms for the Lagrangian L can be of interest. For example, one may want to drive the crowd to a given fixed velocity $\widehat{v}$. In this case, one can consider the minimization of the functional

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

that is again of the form $\int_{\mathbb{R}^{2d}} \ell(\mathbf{y},\mathbf{x},\int \omega\mu) d\mu(\mathbf{x})$, with ℓ not depending on its third variable, this time.

Since the set of hypotheses (H) are clearly satisfied by the above choice of the functionals, we can apply the extended Pontryagin Maximum Principle to the optimal control problem with cost functional (6.67) constrained by system (6.66). We introduce the dual variables of $\mathbf{y}_k$ and $\mathbf{x}$, denoted by $\mathbf{q}_k \triangleq (q_k, z_k)$ and $\mathbf{r} \triangleq (r,s)$, respectively. The Hamiltonian $\mathbb{H}$ in (6.37) can be found by direct substitution:

$$\begin{aligned} \mathbb{H}(\mathbf{y},\mathbf{q},\nu,u) = \frac{1}{2}\int_{\mathbb{R}^8} & (s-s')a(|x-x'|)(v'-v) d\nu(x',v',r',s') d\nu(x,v,r,s) + \\ & + \int_{\mathbb{R}^4} \left(rv + s\frac{1}{m}\sum_{j=1}^m a(|x-y_j|)(w_j - v)\right) d\nu(x,v,r,s) + \\ & + \sum_{k=1}^m \left(q_k w_k + z_k \int_{\mathbb{R}^4} a(|y_k - x|)(v-w_k) d\nu(x,v,r,s)\right) + \end{aligned}$$

$$+\sum_{k=1}^{m}\left(z_k\frac{1}{m}\sum_{j=1}^{m}a(|y_k-y_j|)(w_j-w_k)+z_ku_k\right)-$$
$$-\int_{\mathbb{R}^2}\ell(\mathbf{y},\mathbf{x},\textstyle\int\omega\mu)\,d\mu(\mathbf{x})-|u|^2.$$

The components of the vector field $\nabla_\nu \mathbb{H}_c(\mathbf{y},\mathbf{q},\nu,u^*)$ can be computed at every point $(\mathbf{x},\mathbf{r})=(x,v,r,s)\in\mathbb{R}^4$ and are given by

$$\begin{aligned}
\nabla_\nu\mathbb{H}_c\cdot e_1 &= \int_{\mathbb{R}^4}(s-s')\mathbf{D}a(|x-x'|)(v'-v)d\nu(x',v',r',s')+\\
&\quad + s\frac{1}{m}\sum_{j=1}^{m}\mathbf{D}a(|x-y_j|)(w_j-v)-\sum_{k=1}^{m}z_k\mathbf{D}a(|y_k-x|)(v-w_k),\\
\nabla_\nu\mathbb{H}_c\cdot e_2 &= -\int_{\mathbb{R}^4}(s-s')a(|x-x'|)d\nu(x',v',r',s')+r-s\frac{1}{m}\sum_{j=1}^{m}a(|x-y_j|)+\\
&\quad +\sum_{k=1}^{m}z_k\,a(|y_k-x|)-3v+\frac{2}{m}\sum_{k=1}^{m}w_k+\int_{\mathbb{R}^2}v d\mu(x,v),\\
\nabla_\nu\mathbb{H}_c\cdot e_3 &= v,\\
\nabla_\nu\mathbb{H}_c\cdot e_4 &= (G^{[a]}*\mu)(x,v)+\frac{1}{m}\sum_{j=1}^{m}a(|x-y_j|)(w_j-v).
\end{aligned}$$

In the above expressions, $\mathbf{D}a(|\cdot|)$ reduces to the derivative of the function $a(|\cdot|)$. The optimal control can be explicitly computed by (6.41) as follows

$$u_k^*(\mathbf{y},\mathbf{q})=\begin{cases}\dfrac{z_k}{2} & \text{if } z_k\in[-2,2],\\ \operatorname{sgn}(z_k) & \text{otherwise.}\end{cases}$$

Since μ is the first marginal of ν, the PMP dynamics of the state and adjoint variables is

$$\begin{cases}
\dot y_k = w_k,\\
\dot w_k = (G^{[a]}*\mu)(y_k,w_k)+\dfrac{1}{m}\displaystyle\sum_{k=1}^{m}a(|y_k-y_j|)(v_j-v_k)+u_k^*(\mathbf{y},\mathbf{q}),\\
\dot q_k = \dfrac{1}{m}\displaystyle\int_{\mathbb{R}^4}s\,\mathbf{D}a(|x-y_k|)(w_k-v)d\nu(x,v,r,s)-\\
\qquad -z_k\displaystyle\int_{\mathbb{R}^4}\mathbf{D}a(|y_k-x|)(v-w_k)d\nu(x,v,r,s)-\dfrac{1}{m}\sum_{j\neq k}z_j\mathbf{D}a(|y_k-y_j|)(w_j-w_k),\\
\dot z_k = -\displaystyle\int_{\mathbb{R}^4}\left(\frac{1}{m}s-z_k\right)a(|x-y_k|)d\nu(x,v,r,s)-q_k+\sum_{j\neq k}z_j\,a(|y_k-y_j|)-\\
\qquad -\dfrac{4}{m}w_k+\dfrac{2}{m^2}\displaystyle\sum_{j=1}^{m}w_j+\frac{2}{m}\int_{\mathbb{R}^2}v\,d\mu(x,v),\\
\dfrac{\partial\nu}{\partial t} = -\nabla_{(x,v,r,s)}\cdot\left((\mathbf{J}\nabla_\nu\mathbb{H}_c(\mathbf{y},\mathbf{q},\nu,u^*))\nu\right).
\end{cases}$$

We remark that, as it happens for the standard PMP, the explicit computation of the third and the fourth components gives the vector field determining the dynamics of μ (6.66).

CHAPTER 7

Invisible sparse control

We now employ the indirect sparse control strategy developed in Chapter 6 to address the problem of the evacuation of a crowd of individuals from an unknown environment under limited visibility. First, we present a discrete model for pedestrians in such situations, where the alignment dynamics (typical of uncertain behavior) is coupled with a random walk (modeling the need to explore the unknown area). We then introduce few controlled unrecognizable leaders to improve the evacuation process. To treat situations where the number of followers makes the discrete model numerically unfeasible, we obtain a mean-field approximation of the discrete model by means of the grazing interaction limit, outlined in Chapter 4. Numerical simulations show the efficiency of the control strategy in both the discrete and the continuous setting, which is then implemented in a real experiment with people.

7.1 Bottom-up and top-down control strategies

In this chapter we are concerned with the multiscale modeling and control of a human crowd which leaves an unknown area when the visibility is limited. In such scenarios, we claim that people exhibit two opposite tendencies: on one hand, since agents are assumed to be not informed about the positions of the exits, they tend to spread out in order to explore the environment efficiently; on the other hand they follow the group mates in the hope that they have already found a way out (*herding effect*). We assume that agents are selfish and do not intend to help the others: hence, they do not communicate with other agents or share information directly with them.

Controlling these natural behaviors is not easy, especially in emergency or panic situations. Typically, the emergency management actors (e.g., police, stewards) adopt a *top-down* approach: they receive information about the current situation, update the evacuation strategies, and finally inform people communicating directives. It is useful to stress here that this approach can be very inefficient in large environments (where communications to people are difficult) and in panic situations (since "instinctive" behavior prevails over the rational one). Let us also mention that in some cases people are simply not prone to follow authority's directives because they are opposing it (e.g., in public demonstrations). For such reasons, we explore the possibility of controlling crowds adopting a *bottom-up* approach. Control is obtained by means of special agents (the leaders) who are *hidden in the crowd, and are not recognized by the mass as special.* From the modeling point of view, this is translated by the fact that the individuals of the crowd (the followers) interact in the same way with both the other crowd mates and the leaders.

This control technique, known in the literature as *invisible* or *soft control*, was first studied in [73] in connection with a repulsion-alignment-attraction model: the authors show that a little percentage of informed agents pointing towards a target is sufficient to steer the whole population to it. Invisible control was successfully implemented in real experiments involving animals, where leaders come in the form of disguised robots; see, e.g., [117], treating the case of cockroaches and [44], which study the response of zebrafish. A Vicsek-like model with a single invisible leader is considered in [118, 119].

7.2 Model guidelines

In order to deal with the curse of dimensionality, followers shall be described below both at the *microscopic* level (i.e., when they are seen as discrete) and at the *mesoscopic* one (when they are considered as a continuum). The general features of the model discussed here are valid at any scale of observation.

The model for followers is characterized by an *exploration phase* and an *evacuation phase*. This is achieved by a combination of repulsion, alignment, self-propulsion and random (social) force, together with the introduction of an *exit's visibility area* (modeling the fact that exits can be seen only from specific locations). We also consider both *metrical* and *topological* interactions so to avoid unnatural all-to-all interactions. A few leaders, not recognized as such by the crowd and implementing the bottom-up control strategy, are added to the model with a special, controlled dynamics.

First- vs. second-order model One can notice that walking people, animals and robots are in general able to adjust their velocity almost instantaneously, reducing to a negligible duration the acceleration/deceleration phase. For this reason, in principle a second-order (inertia-based) framework does not appear to be the most natural setting to describe their dynamics. However, as argued in Section 1.2, if we include in the model the tendency of the agents to move together and to align with group mates, we implicitly assume that agents can perceive the velocity of the others: this makes unavoidable the use of a second-order model, where both position and velocity are state variables, but only velocity is the *consensus parameter* (consensus here is intended in the sense of Definition 1.4). See [77, Section 4.2] for a general discussion on this point. In our model we adopt a mixed approach, describing leaders by a first-order model (since they do not need to align) and followers by a second-order one. In the latter case, the small inertia is obtained by means of a fast relaxation towards the target velocity.

Metrical vs. topological interactions The interaction of one agent with group mates is said to be *metrical* if it involves only mates within a predefined *sensory region*, regardless of the number of individuals which actually fall in it. Interaction is instead called *topological* if it involves a predefined number of group mates regardless of their distance from the considered agent. Again, in our model we adopt a mixed approach, assuming short-range interactions to be metrical and long-range ones to be topological.

Isotropic vs. anisotropic interactions Living agents (people, animals) are generally asymmetric, in the sense that they better perceive stimuli coming from their "front", rather than their "back". For instance, in pedestrian models interactions are usually restricted to the half-space in front of the person, since the human visual field is approximately 180°. Nevertheless, humans can easily see all around them simply turning the head. In the case we are interested in, people have no idea of the location of their target, hence we expect that they often look around to explore the environment and see the behavior of the others. This is why we prefer to adhere to isotropic interactions.

Let us describe the specific social forces acting on the agents.

- *Leaders*
  - Leaders are subject to an isotropic metrical short-range *repulsion force* directed against all the others, translating the fact that they want to avoid collisions and that a maximal density exists.
  - Leaders are assumed to know the environment and the self-organizing features of the crowd. They respond to an *optimal force* which is the result of an off-line optimization procedure, defined as to minimizing some cost functional (see Section 7.5).
- *Followers*
  - Similarly to leaders, followers respond to an isotropic metrical short-range *repulsion force* directed against all the others.
  - Followers tend to a *desired velocity* which corresponds to the velocity they would have if they were alone in the domain. This term takes into account the fact that the environment is unknown. Followers describe a *random walk* if the exit is not visible (exploration phase) or a sharp motion toward the exit if the exit is visible (evacuation phase). In addition, we include a self-propulsion term which translates the tendency to reach a given characteristic speed (modulus of the velocity), see Remark 6.8.
  - If the exit is not visible, we also prescribe that followers adhere to a "stick-to-your-neighbors" strategy, in the hope to stay with agents who already know where to go. As already discussed in Chapter 1, the behavior to move coherently with the other members of the group can be mathematically described by an *isotropic alignment force*. Motivated by empirical evidences, such force will be *topological*, so that agents will only align with their closest neighbors.

Concerning pedestrian modeling, virtually any kind of models have been investigated so far: for a quick introduction, we refer the reader to the reviews [25, 123] and the books [77, 132]. Some papers deal specifically with *evacuation problems*: a very good source of references is the recent paper [1], where evacuation models both with and without top-down optimal planning search are discussed. Evacuation problems were studied by means of lattice models [62, 113], social force models [152], cellular automata models [1, 179], mesoscopic models [2], and macroscopic models [55]. Limited visibility issues were also considered in [55, 62, 113].

7.3 The microscopic model

Formally, we consider N followers and m leaders in dimension d (even if, for practical reasons, we shall pay particular attention to the case $d = 2$), and as in Section 6.2 we assume that $m \ll N$. We denote by $\Omega \subseteq \mathbb{R}^d$ the walking area and by $x^\tau \in \Omega$ the target point (i.e., the exit). To define the target's visibility area, we consider the set Σ, with $x^\tau \in \Sigma \subset \Omega$, and we assume that the target is completely visible from any point belonging to Σ and completely invisible from any point belonging to $\Omega \backslash \Sigma$.

For every $i = 1, \ldots, N$, let $(x_i(t), v_i(t)) \in \mathbb{R}^{2d}$ denote position and velocity of the agents belonging to the population of followers at time $t \geq 0$ and, for every $k = 1, \ldots, m$, let $(y_k(t), w_k(t)) \in \mathbb{R}^{2d}$ denote position and velocity of the agents among the population of leaders at time $t \geq 0$.

Let us also define $\mathbf{x} \triangleq (x_1, \ldots, x_N)$ and $\mathbf{y} \triangleq (y_1, \ldots, y_m)$. By $\mathcal{B}_\mathcal{N}(x; \mathbf{x}, \mathbf{y})$ we denote the *minimal* ball centered on x encompassing at least $\mathcal{N}$ agents among followers and leaders, and by $\mathcal{N}^*$ the actual number of agents in $\mathcal{B}_\mathcal{N}(x; \mathbf{x}, \mathbf{y})$. Note that $\mathcal{N}^* \geq \mathcal{N}$.

Remark 7.1. The computation of $\mathcal{B}_\mathcal{N}(x; \mathbf{x}, \mathbf{y})$ requires the knowledge of the positions of all the agents, since all the distances $\|x_i - x\|$, $i = 1, \ldots, N$, and $\|y_k - x\|$, $k = 1, \ldots, m$ must be evaluated in order to find the $\mathcal{N}$ closest agents to x. This motivates the explicit dependency on the vectors of positions $\mathbf{x}$ and $\mathbf{y}$.

The microscopic dynamics described by the two populations is given by the following set of ODEs: for $i = 1, \ldots, N$ and $k = 1, \ldots, m$,

$$\begin{cases} \dot{x}_i(t) = v_i(t), \\ \dot{v}_i(t) = A(x_i(t), v_i(t)) + \sum_{j=1}^{N} K^{\mathrm{F}}(x_i(t), v_i(t), x_j(t), v_j(t); \mathbf{x}(t), \mathbf{y}(t)) + \\ \qquad + \sum_{l=1}^{m} H^{\mathrm{F}}(x_i(t), v_i(t), y_l(t), w_l(t); \mathbf{x}(t), \mathbf{y}(t)), \\ \dot{y}_k(t) = w_k(t) = \sum_{j=1}^{N} K^{\mathrm{L}}(y_k(t), x_j(t)) + \sum_{l=1}^{m} H^{\mathrm{L}}(y_k(t), y_l(t)) + u_k(t). \end{cases} \tag{7.1}$$

We will consider the case where

- A is a self-propulsion term, given by the relaxation toward a random direction or the relaxation toward a unit vector pointing to the target (the choice depends on the position), plus a term which translates the tendency to reach a given characteristic speed $s \geq 0$ (the one described in Remark 6.8), i.e.,

$$A(x, v) \triangleq \theta(x) C_z (z - v) + (1 - \theta(x)) C_\tau \left(\frac{x^\tau - x}{\|x^\tau - x\|} - v \right) + C_s (s^2 - \|v\|^2) v, \tag{7.2}$$

where $\theta : \mathbb{R}^d \to [0,1]$ is the characteristic function of $\Omega \backslash \Sigma$, i.e., $\theta(x) = \chi_{\Omega\backslash\Sigma}(x)$, the vector z is a d-dimensional random vector with normal distribution $\mathcal{N}(0,\sigma^2)$, and C_z, C_τ, C_s are positive constants.

- The interactions follower-follower and follower-leader coincide and are equal to

$$\begin{aligned} K^{\text{F}}(x,v,y,w;\mathbf{x},\mathbf{y}) &\triangleq -C_r^{\text{F}} R_{\gamma,r}(x,y) + \theta(x)\frac{C_a}{\mathcal{N}^*}(w-v)\,\chi_{B_{\mathcal{N}}(x;\mathbf{x},\mathbf{y})}(y), \\ H^{\text{F}}(x,v,y,w;\mathbf{x},\mathbf{y}) &\triangleq K^{\text{F}}(x,v,y,w;\mathbf{x},\mathbf{y}), \end{aligned} \tag{7.3}$$

for given positive constants C_r^{F}, C_a, r, and γ, where

$$R_{\gamma,r}(x,y) \triangleq \begin{cases} e^{-\|y-x\|^\gamma}\dfrac{y-x}{\|y-x\|} & \text{if } y \in B(x,r)\backslash\{x\}, \\ 1 & \text{otherwise,} \end{cases}$$

models a (metrical) repulsive force, and the second term accounts for the (topological) alignment force, which vanishes inside Σ. Notice that, once the summations over the agents $\sum_j K^{\text{F}}$ and $\sum_l H^{\text{F}}$ are done, the *topological alignment term* models the tendency of the followers to relax toward the average velocity of the $\mathcal{N}$ closest agents. With the choice $K^{\text{F}} = H^{\text{F}}$ the leaders are not recognized by the followers as special. This feature opens a wide range of new applications, including the control of crowds not prone to follow authority's directives.

- The interactions leader-follower and leader-leader reduce to a mere (metrical) repulsion, i.e.,

$$K^{\text{L}}(x,y) = H^{\text{L}}(x,y) \triangleq C_r^{\text{L}} R_{\zeta,r}(x,y),$$

where $C_r^{\text{L}} > 0$ and $\zeta > 0$ are in general different from C_r^{F} and γ, respectively. Note that here the repulsion force is interpreted as a velocity field, while for followers it was an acceleration field.

- u_k is the control variable belonging to the set of admissible control functions Adm.

Remark 7.2. The leader dynamics do not depend explicitly on the agents' velocities. This allows us to consider a first-order model for the leaders. However, a generalization to fully second-order model is straightforward.

Remark 7.3. The behaviour of the leaders is entirely encapsulated in the control term u but for a short-range repulsion force. The latter should be indeed interpreted as a force due to the presence of the others, thereby non controllable.

7.4 Formal derivation of a Boltzmann-type equation

The main difference between the above model and system (6.10) is given by the scaling of the force exerted on each agent: in system (7.1) we are not in the presence of the mean-field scaling, and the reason lies in the fact that we want leaders and followers to have the same weight in the followers' dynamics (in (6.10) theses weights were $1/m$ and $1/N$, respectively). To pass to the mean-field we have then to adopt the Boltzmann binary interaction approach followed by the grazing interaction limit, a procedure outlined in Section 4.4.

7.4.1 Binary interactions for system (7.1)

We consider the evolution of the distribution of followers at time $t \geq 0$, denoted by $f(t) \in \mathcal{M}_b(\mathbb{R}^{2d})$, together with the microscopic equations for the leaders (whose number is small by design). As in (4.30), we denote by $\rho[f](t)$ the total mass of followers at time $t \geq 0$, i.e.,

$$\rho[f](t) \triangleq \int_{\mathbb{R}^{2d}} f(t,x,y)\ dx\ dv,$$

We introduce, for symmetry reasons, the distribution of leaders g and its total mass

$$g(t,x,v) \triangleq \sum_{k=1}^{m} \delta_{(y_k(t),w_k(t))}(x,v) \quad \text{and} \quad \rho[g](t) \triangleq \int_{\mathbb{R}^{2d}} g(t,x,v)\ dx\ dv. \tag{7.4}$$

We shall require both masses to be constant and equal to N and m, respectively. This means that for every $t \geq 0$ it holds

$$\rho[f](t) = \rho[f] = N \quad \text{and} \quad \rho[g](t) = \rho[g] = m.$$

Similarly to what we did in Section 4.4, we shall describe the evolution of f by means of a Boltzmann-type equation obtained from binary interactions between f and itself and between f and the mass of leaders g. The binary interactions will be such that the grazing interaction limit recovers a Fokker-Planck-type equation corresponding to the dynamics of followers in (7.1), to be coupled with the previously presented ODE dynamics for leaders (see equation (7.14) below for the resulting system).

To derive the Boltzmann-like dynamics, we assume that, before interacting, each follower has at his disposal the values $\mathbf{x}$ and $\mathbf{y}$ that he needs in order to compute its trajectory: hence, in a binary interaction between two followers with state parameters (x,v) and $(\hat{x},\hat{v})$, the value of $K^{\mathrm{F}}(x,v,\hat{x},\hat{v};\mathbf{x},\mathbf{y})$ does not depend on $\mathbf{x}$ and $\mathbf{y}$. In the case of K^{F} of the form (7.3), this means that the ball $B_{\mathcal{N}}(x;\mathbf{x},\mathbf{y})$ and the value of $\mathcal{N}^*$ have been already computed before interacting.

Moreover, since we are considering the distributions f and g of followers and leaders, respectively, the vectors $\mathbf{x}$ and $\mathbf{y}$ are derived from f and g by means of their first marginals, indicated by $\pi_{1\#}f$ and $\pi_{1\#}g$, respectively, which give the spatial variables of those distribution. Hence, since no confusion arises, we write $K^{\mathrm{F}}(x,v,\hat{x},\hat{v};\pi_{1\#}f,\pi_{1\#}g)$ in place of $K^{\mathrm{F}}(x,v,\hat{x},\hat{v};\mathbf{x},\mathbf{y})$ to stress the dependency of this term on f and g.

We thus consider two followers with state parameter (x,v) and $(\hat{x},\hat{v})$ respectively, and we describe the evolution of their post-interaction velocities as

$$\begin{cases} v^* = v + \eta^{\mathrm{F}} \left[\theta(x) C_z \xi + S(x,v) + \rho[f] K^{\mathrm{F}}(x,v,\hat{x},\hat{v};\pi_{1\#}f,\pi_{1\#}g)\right], \\ \hat{v}^* = \hat{v} + \eta^{\mathrm{F}} \left[\theta(\hat{x}) C_z \xi + S(\hat{x},\hat{v}) + \rho[f] K^{\mathrm{F}}(\hat{x},\hat{v},x,v;\pi_{1\#}f,\pi_{1\#}g)\right], \end{cases} \tag{7.5}$$

where η^{F} is the strength of interaction among followers, ξ is a random variables whose entries are i.i.d. following a normal distribution with mean 0, variance ς^2 (which shall be related to the variance σ^2 of the random vector z in (7.2) as in (4.35)), taking values in a set $\mathcal{B}$, and S is defined as the deterministic part of the self-propulsion term (7.2),

$$S(x,v) = -\theta(x)C_z v + (1-\theta(x))C_\tau \left(\frac{x^\tau - x}{\|x^\tau - x\|} - v \right) + C_s(s^2 - \|v\|^2)v. \tag{7.6}$$

As in Section 4.4.1, we need to assume that the change of variables (7.5) satisfies the hypothesis (Inv) of being invertible.
We then consider the same follower as before with state parameters (x, v) and a leader agent $(\tilde{x}, \tilde{v})$; in this case the modified velocities satisfy

$$\begin{cases} v^{**} = v + \eta^{\mathrm{L}} \rho[g] H^{\mathrm{F}}(x, v, \tilde{y}, \tilde{w}; \pi_{1\#}f, \pi_{1\#}g), \\ \tilde{v}^{*} = \tilde{v}, \end{cases} \tag{7.7}$$

where η^{L} is the strength of the interaction between followers and leaders. Note that (7.7) accounts only for the change of the followers' velocities, since leaders are not evolving via binary interactions. Again, we assume that (7.7) satisfies the hypothesis (Inv)
The time evolution of f is then given by a balance between bilinear gain and loss of space and velocity terms according to the two binary interactions (7.5) and (7.7), quantitatively described by the Boltzmann-type equation

$$\frac{\partial f}{\partial t}(t) + v \cdot \nabla_x f(t) = \lambda^{\mathrm{F}} Q(f,f)(t) + \lambda^{\mathrm{L}} Q(f,g)(t), \tag{7.8}$$

where λ^{F} and λ^{L} stand for the interaction frequencies among followers and between followers and leaders, respectively. The Boltzmann-Povzner operators $Q(f,f)$ and $Q(f,g)$ are defined as

$$Q(f,f)(t,x,v) \triangleq \mathbb{E}\left[\int_{\mathbb{R}^{4d}} \left(\frac{1}{\mathfrak{J}_{\mathrm{F}}} f(t, x_*, v_*) f(t, \hat{x}_*, \hat{v}_*) - f(t,x,v) f(t,\hat{x},\hat{v}) \right) d\hat{x}\; d\hat{v} \right], \tag{7.9}$$

$$Q(f,g)(t,x,v) \triangleq \mathbb{E}\left[\int_{\mathbb{R}^{4d}} \left(\frac{1}{\mathfrak{J}_{\mathrm{L}}} f(t, x_{**}, v_{**}) g(t, \tilde{x}_*, \tilde{v}_*) - f(t,x,v) g(t,\tilde{x},\tilde{v}) \right) d\tilde{x}\; d\tilde{v} \right]. \tag{7.10}$$

In the expression above, the couples (x_*, v_*) and $(\hat{x}_*, \hat{v}_*)$ are the pre-interaction states that generates (x, v) and $(\hat{x}, \hat{v})$ via (7.5), and $\mathfrak{J}_{\mathrm{F}}$ is the Jacobian of the change of variables given by (7.5) (well-defined by (Inv)). Similarly, (x_{**}, v_{**}) and $(\tilde{x}_*, \tilde{v}_*)$ are the pre-interaction states that generates (x, v) and $(\tilde{x}, \tilde{v})$ via (7.7), and $\mathfrak{J}_{\mathrm{L}}$ is the Jacobian of the change of variables given by (7.7). As in (4.28), the expected value $\mathbb{E}$ is computed with respect to $\xi \in \mathcal{B}$.

Remark 7.4. If we would have opted for a description of agents as hard-sphere particles, the arising Boltzmann equation (7.8) would be of Enskog-type, see [170]. The relationship between the hard- and soft-sphere descriptions (i.e., where repulsive forces between point masses are considered, instead) has been deeply discussed, for instance, in [14]. In our model, the repulsive force $R_{\gamma,r}$ is not singular at the origin for computational reasons, therefore the parameters γ and r have to be chosen properly to avoid arbitrarily high density concentrations.

Equation (7.8) is then to be coupled with the ODE dynamics of the leaders from system

(7.1). The controlled ODE-PDE dynamics obtained in this way is

$$\begin{cases} \dfrac{\partial f}{\partial t}(t) + v \cdot \nabla_x f(t) = \lambda^{\mathrm{F}} Q(f,f)(t) + \lambda^{\mathrm{L}} Q(f,g)(t), \\ \dot{y}_k(t) = w_k(t) = \displaystyle\int_{\mathbb{R}^{2d}} K^{\mathrm{L}}(y_k(t), x) f(t,x,v)\; dx\; dv + \sum_{l=1}^{m} H^{\mathrm{L}}(y_k(t), y_l(t)) + u_k(t). \end{cases} \tag{7.11}$$

Since we are not limiting our study to the functions S, K^{F}, H^{F}, K^{L} and H^{L} introduced in Section 7.3 (hence, allowing for the possibility of nonsmooth and singular kernels), we say that $(f, y_1, w_1, \ldots, y_m, w_m) \in \mathcal{M}_b(\mathbb{R}^{2d}) \times \mathbb{R}^{2dm}$ is *admissible* if the following quantities exist and are finite:

(*i*) $\int_{\mathbb{R}^{2d}} S(x,v) f(x,v)\; dx\; dv$,

(*ii*) $\int_{\mathbb{R}^{4d}} K^{\mathrm{F}}(x,v,\hat{x},\hat{v}; \pi_{1\#}f, \mathbf{y}) f(x,v) f(\hat{x},\hat{v})\; dx\; dv\; d\hat{x}\; d\hat{v}$,

(*iii*) $\sum_{l=1}^{m} \int_{\mathbb{R}^{2d}} H^{\mathrm{F}}(x,v,y_l,w_l; \pi_{1\#}f, \mathbf{y}) f(x,v)\; dx\; dv$,

(*iv*) $\int_{\mathbb{R}^{2d}} K^{\mathrm{L}}(y_k, x) f(x,v)\; dx\; dv$ for every $k = 1, \ldots, m$,

(*v*) $\sum_{l=1}^{m} H^{\mathrm{L}}(y_k, y_l)$ for every $k = 1, \ldots, m$.

Consequently, we introduce our definition of solution for system (7.11).

Definition 7.5. Fix $T > 0$, $\delta > 0$, and $u : [0,T] \to \mathbb{R}^{2dm}$. By a *$\delta$-weak solution* of system (7.11) with control u and initial datum $(f^0, y_1^0, \ldots, y_m^0) \in \mathcal{M}_b(\mathbb{R}^{2d}) \times \mathbb{R}^{dm}$ in the interval $[0,T]$, we mean any $(f, y_1, \ldots, y_m) \in L^2([0,T]; \mathcal{M}_b(\mathbb{R}^{2d})) \times W^{1,\infty}([0,T]; \mathbb{R}^{dm})$ such that:

(*i*) $f(0,x,v) = f^0(x,v)$ for every $(x,v) \in \mathbb{R}^{2d}$;

(*ii*) the vector $(f(t), y_1(t), \dot{y}_1(t), \ldots, y_m(t), \dot{y}_m(t))$ is admissible for every $t \in [0,T]$;

(*iii*) there exists $R > 0$ depending only on $(f^0, y_1^0, \ldots, y_m^0)$, T, δ, u, θ, S, K^{F}, H^{F}, K^{L}, and H^{L}, such that $\mathrm{supp}(f(t)) \subseteq B(0,R)$ for every $t \in [0,T]$;

(*iv*) $\rho[f](t) = N$ for every $t \in [0,T]$;

(*v*) $f(\cdot)$ satisfies the weak form of the equation (7.8), i.e., for every $\varphi \in \mathcal{T}_\delta$ and every $t \in (0,T]$ it holds

$$\frac{\partial}{\partial t} \langle f(t), \varphi \rangle + \langle f(t), v \cdot \nabla_x \varphi \rangle = \lambda^{\mathrm{F}} \langle Q(f,f)(t), \varphi \rangle + \lambda^{\mathrm{L}} \langle Q(f,g)(t), \varphi \rangle,$$

where

$$\langle Q(f,f)(t), \varphi \rangle = \mathbb{E}\left[\int_{\mathbb{R}^{4d}} (\varphi(x,v^*) - \varphi(x,v))\, f(t,x,v) f(t,\hat{x},\hat{v})\, dx\, dv\, d\hat{x}\, d\hat{v}\right],$$
$$\langle Q(f,g)(t), \varphi \rangle = \mathbb{E}\left[\int_{\mathbb{R}^{4d}} (\varphi(x,v^{**}) - \varphi(x,v))\, f(t,x,v) g(t,\tilde{x},\tilde{v})\, dx\, dv\, d\tilde{x}\, d\tilde{v}\right],$$

and where v^* and v^{**} are given by (7.5) and (7.7), respectively;

(vi) for every $k = 1, \ldots, m$, $y_k(\cdot)$ satisfies

$$\dot{y}_k(t) = w_k(t) = \int_{\mathbb{R}^{2d}} K^{\mathrm{L}}(y_k(t), x) f(t, x, v)\ dx\ dv + \sum_{l=1}^{m} H^{\mathrm{L}}(y_k(t), y_l(t)) + u_k(t),$$

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

In what follows, we shall assume that a solution of (7.11) in the sense of Definition 7.5 exists (see also Remark 4.21), and we shall provide a mean-field approximation of the Boltzmann-like equation (7.8) via the grazing interaction limit procedure.

7.4.2 Mean-field via grazing interaction limit

Similarly to Section 4.4.2, we pass to scale η^{F}, η^{L}, λ^{F}, λ^{L} and ς^2 with respect to a common parameter $\varepsilon > 0$ in such a way that we obtain for $\varepsilon \to 0$ the Fokker-Planck operator associated to the discrete dynamics of followers in system (7.1) given by

$$FP(t, x, v) \triangleq -\nabla_v \cdot \big(\mathcal{G}[f, g](t, x, v) f(t, x, v)\big) + \frac{1}{2}\sigma^2(\theta(x)C_z)^2 \Delta_v f(t, x, v),$$

where $\mathcal{G}[f, g](t, x, v) \triangleq S(x, v) + \mathcal{H}[f](t, x, v) + \mathcal{H}[g](t, x, v)$, and

$$\mathcal{H}[f](t, x, v) \triangleq \int_{\mathbb{R}^{2d}} K^{\mathrm{F}}(x, v, \hat{x}, \hat{v}; \pi_{1\#}f(t), \pi_{1\#}g(t)) f(t, \hat{x}, \hat{v})\ d\hat{x}\ d\hat{v},$$
$$\mathcal{H}[g](t, x, v) \triangleq \int_{\mathbb{R}^{2d}} H^{\mathrm{F}}(x, v, \tilde{x}, \tilde{v}; \pi_{1\#}f(t), \pi_{1\#}g(t)) g(t, \tilde{x}, \tilde{v})\ d\tilde{x}\ d\tilde{v}.$$

Indeed, it can be seen by direct substitution that if $(x, v) : [0, T] \to \mathbb{R}^{2dN}$ is a solution of the followers' dynamics of system (7.1), then the curve $f : [0, T] \to \mathcal{M}_b(\mathbb{R}^{2d})$ defined for every $t \in [0, T]$ as $f(t) = \sum_{i=1}^{N} \delta_{(x_i(t), v_i(t))}$ satisfies

$$\frac{\partial f}{\partial t}(t) + v \cdot \nabla_x f(t) = FP(t, x, v)$$

for every $t \geq 0$. More precisely, we set

$$\eta^{\mathrm{F}} = \varepsilon, \qquad \eta^{\mathrm{L}} = \varepsilon, \qquad \lambda^{\mathrm{F}} = \frac{1}{\varepsilon \rho[f]}, \qquad \lambda^{\mathrm{L}} = \frac{1}{\varepsilon \rho[g]}, \qquad \varsigma^2 = \frac{\sigma^2}{\varepsilon}, \tag{7.12}$$

and we apply the machinery of Section 4.4.2 separately to the operators $\lambda^{\mathrm{F}} Q(f, f)$ and $\lambda^{\mathrm{F}} Q(f, g)$ in order to pass to the mean-field. Remember that we assumed $\rho[f](t) = \rho[f] = N$ and $\rho[g](t) = \rho[g] = m$ for every $t \geq 0$.

Let us fix $T > 0$, $\delta > 0$, $\varepsilon > 0$ a control $u : [0, T] \to \mathbb{R}^{2dm}$ and an initial datum $(f^0, y_1^0, \ldots, y_m^0) \in \mathcal{M}_b(\mathbb{R}^{2d}) \times \mathbb{R}^{dm}$. Consider a δ-weak solution $(f(\cdot), y_1(\cdot), \ldots, y_m(\cdot))$ of system (7.11) with control u, initial datum $(f^0, y_1^0, \ldots, y_m^0)$ and define g as in (7.4). Let the scale (7.12) holds for the chosen ε. Moreover, suppose that $\mathbb{E}\left[\|\xi\|^{2+\delta}\right] < +\infty$ and that θ, S, K^{F} and H^{F} belong to L^p_{loc} for $p = 2, 2 + \delta$.

Denote by $Q(f, f) \triangleq Q^{\varepsilon}(f, f)$ the collisional operator (7.9) associated to the scaling (7.12) for the chosen $\varepsilon > 0$. Since the binary interactions (7.5) correspond to (4.27) for the choices $\alpha(x,v)=\theta(x)C_z$, $\beta(x,v)=S(x,v)$, and $\gamma(x,v,\hat{x},\hat{v})=K^{\mathrm{F}}(x,v,\hat{x},\hat{v},\pi_{1\#}f,\pi_{1\#}g)$, then analogous computations to those employed to prove Theorem 4.23 lead to

$$\lambda^{\mathrm{F}}\langle Q(f,f),\varphi\rangle \to \left\langle f, \nabla_v\varphi\cdot(S+\mathcal{H}[f])+\frac{1}{2}\sigma^2(\theta C_z)^2\Delta_v\varphi\right\rangle$$

for $\varepsilon\to 0$. On the other hand, the binary interactions (7.7) correspond to (4.27) for the choices $\alpha=\beta\equiv 0$ and $\gamma(x,v,\hat{x},\hat{v})=H^{\mathrm{F}}(x,v,\hat{x},\hat{v},\pi_{1\#}f,\pi_{1\#}g)$. Therefore, if we denote by $Q(f,g)\overset{\Delta}{=}Q^{\varepsilon}(f,g)$ the collisional operator (7.10) associated to the scaling (7.12) for the chosen $\varepsilon>0$, the same strategy used for $\lambda^{\mathrm{F}}\langle Q(f,f)$ implies this time that

$$\lambda^{\mathrm{L}}\langle Q(f,g),\varphi\rangle \to \langle f,\nabla_v\varphi\cdot\mathcal{H}[g]\rangle$$

for $\varepsilon\to 0$. Putting together the above limits, by Remark 4.22 we obtain the weak formulation of a Fokker-Planck-type equation for the followers' dynamics as $\varepsilon\to 0$:

$$\frac{\partial}{\partial t}\langle f(t),\varphi\rangle+\langle f(t),v\cdot\nabla_x\varphi\rangle=\left\langle f(t),\nabla_v\varphi\cdot\mathcal{G}\left[f,g\right](t)+\frac{1}{2}\sigma^2(\theta C_z)^2\Delta_v\varphi\right\rangle. \tag{7.13}$$

Being φ compactly supported, equation (7.13) can be recast in strong form by means of integration by parts. Coupling the resulting PDE with the microscopic ODEs for the leaders we eventually obtain the system

$$\begin{cases}\dfrac{\partial}{\partial t}f(t)+v\cdot\nabla_x f(t)=-\nabla_v\cdot\left(\mathcal{G}\left[f,g\right](t)f(t)\right)+\dfrac{1}{2}\sigma^2(\theta C_z)^2\Delta_v f(t),\\ \dot{y}_k(t)=w_k(t)=\displaystyle\int_{\mathbb{R}^{2d}}K^{\mathrm{L}}(y_k(t),x)f(t,x,v)\;dx\;dv+\sum_{l=1}^{m}H^{\mathrm{L}}(y_k(t),y_l(t))+u_k(t).\end{cases} \tag{7.14}$$

Let us stress again that, assuming $f(t)$ to be the empirical measure $f(t)=\sum_{i=1}^{N}\delta_{(x_i(t),v_i(t))}$ concentrated on the trajectories $(x(\cdot),v(\cdot))$ of the microscopic dynamic (7.1), we recover again the original microscopic model.

We have thus obtained the following

Corollary 7.6. *Fix $T>0$, $\delta>0$ and $u:[0,T]\to\mathbb{R}^{2dm}$, and let, for every $\varepsilon>0$, $(f^{\varepsilon}(\cdot),y_1^{\varepsilon}(\cdot),\ldots,y_m^{\varepsilon}(\cdot))$ be a δ-weak solution of* (7.11) *corresponding to the control u and the initial condition $(f^0,y_1^0,\ldots,y_m^0)$, and where the quantities $\eta^{\mathrm{F}},\eta^{\mathrm{L}},\lambda^{\mathrm{F}},\lambda^{\mathrm{L}}$, and ς^2 are rescaled with respect to ε according to* (7.12). *Suppose that*

(i) $\mathbb{E}\left[\|\xi\|^{2+\delta}\right]$ *is finite, and*

(ii) the functions θ, S, K^{F}, and H^{F} are in L^p_{loc} for $p=2,2+\delta$.

Then, as $\varepsilon\to 0$, the solutions $(f^{\varepsilon}(\cdot),y_1^{\varepsilon}(\cdot),\ldots,y_m^{\varepsilon}(\cdot))$ converges pointwise, up to a subsequence, to $(f(\cdot),y_1(\cdot),\ldots,y_m(\cdot))$, where $f(\cdot)$ satisfies the Fokker-Planck-type equation (7.13) *with initial datum $(f^0,y_1^0,\ldots,y_m^0)$, and for every $k=1,\ldots,m$, $y_k(\cdot)$ satisfies*

$$\begin{cases}\dot{y}_k(t)=w_k(t)=\displaystyle\int_{\mathbb{R}^{2d}}K^{\mathrm{L}}(y_k(t),x)f(t,x,v)\;dx\;dv+\sum_{l=1}^{m}H^{\mathrm{L}}(y_k(t),y_l(t))+u_k(t),\\ y_k(0)=y_k^0.\end{cases}$$

Remark 7.7. The hypothesis of admissibility in Definition 7.5 makes sure that all the integrals considered in the proof of Corollary 7.6 exist and are finite.

Remark 7.8. The choice of the functions θ, S, K^{F} and H^{F} made in (7.6) and (7.3) clearly satisfy the hypotheses of Corollary 7.6 for any $\delta > 0$. Hence, we expect the simulation of system (7.11) with the scaling (7.12) to be a good approximation of the ones of the original microscopic model (7.1) for sufficiently small values of ε.

7.5 Optimal control of the crowd

In this section we discuss how one can (optimally) control the crowd of followers by means of "invisible" leaders, whose dynamics is given by control strategies minimizing certain cost functionals.

7.5.1 Group splitting and influence of the leaders

It is well known that, in the case of alignment-dominated models, the invisibility of leaders is not a limitation *per se*, but nothing is known about more complex models. In particular, *the coupling between random walk and alignment* gives rise to interesting phenomena, see [173]. Notice that, in order to fit real behavior (see also Section 7.7), the ratio between random walk force and alignment force should be large enough to assure a complete exploration of the domain, and small enough to catch the herding behavior. However, this also narrows the choice of parameters.

To fix the ideas, we plot a typical outcome of the microsimulator when the crowd is far away from the exit. At the initial time, agents are uniformly distributed in a square, see Figure 7.1. We refer to this situation as Setting 0 (model parameters for this and following simulations are reported in Table 7.1). With no leaders, the group splits in several subgroups, each of them having a common direction of motion (local consensus is reached in very short time), see Figure 7.1-left. With 5 leaders moving rightward, the crowd reaches immediately a consensus and align with leaders, see Figure 7.1-center. If, for some reason, leaders cease to be influential, the crowd tend to split again, see Figure 7.1-right.

Setting	m	N	$\mathcal{N}$	C_r^{F}	C_r^{L}	C_a	C_z	C_τ	C_s	s^2	$r=\zeta$	γ
0	0-5	150	10	2	1.5	2	0.25	–	1	0.5	0.4	1
1	0-3	150	10	2	1.5	3	0.2	1	1	0.5	0.4	1
2	0-2	100	10	2	1.5	3	0.2	1	1	0.5	0.4	1
3	0-6	30-70	10	1	0	–	–	1	1	0.5	0.5	1

Table 7.1: Model parameters.

This makes clear how difficult it is to control the crowd in this scenario, since leaders have to fight *continuously* against the natural tendency of the crowd to split in subgroups and move randomly, even after that a consensus is reached, see [119].

A natural question is about the minimum number of leaders needed to lead the whole crowd

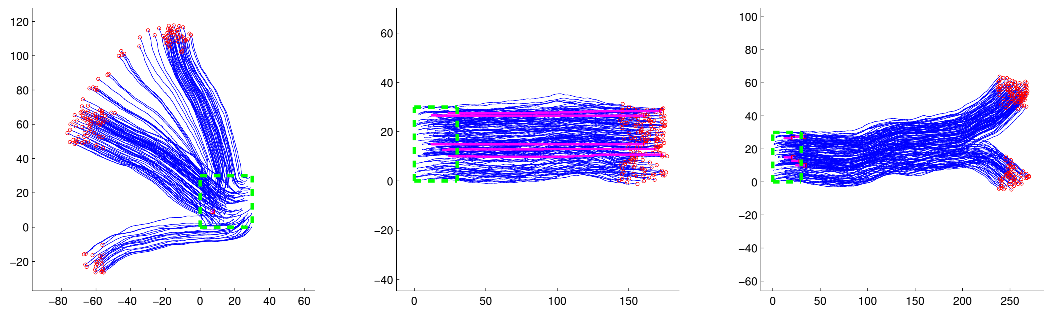


Figure 7.1: Setting 0 (no exit is visible here). Microscopic dynamics. The crowd is initially confined in the green dashed square. Leaders' trajectories are in magenta, followers' are in blue. Final positions of followers are in red. Left: no leaders. Center: 5 leaders moving rightward. Right: 5 leaders moving rightward, disappearing after a short time.

to consensus, i.e., to align all the agents along a desired direction. Numerical simulations suggests that this number strongly depends on the initial conditions. Table 7.2 gives a rough idea in the case of multiple runs with random initial data.

#leaders	1	2	3	4	5
consensus reached	0%	1%	27%	58%	100%

Table 7.2: Number of leaders versus percentage of runs where consensus is reached.

At mesoscopic level we consider the control of the Boltzmann-type dynamics (7.11), where we account for small values of the scaling parameter ε, in order to be sufficiently close to the Fokker-Planck-type model (7.14). In this case, the control through few microscopic leaders has to face the tendencies of the continuous density to spread around the domain and to locally align with the surrounding mass. Moreover, their action is weakened by the type of interaction considered. Figure 7.2 shows the results corresponding to the kinetic density in the Setting 0. Unlike the microscopic case, no splitting occurs where there is no action of the leaders. Instead, the mass smears out around the domain, due to the diffusion term.

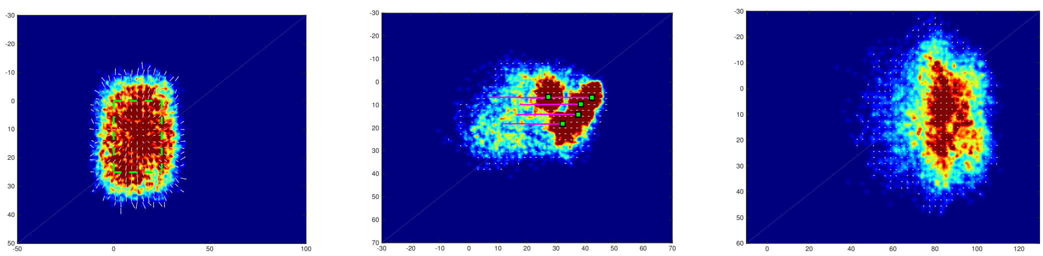


Figure 7.2: Setting 0 (no exit is visible here). Mesoscopic dynamics. The crowd density is initially confined in the green dashed square shown in the left figure. Left: no leaders. Center: 5 leaders moving rightward all the time. Right: 5 leaders moving rightward, disappearing after a short time.

7.5.2 Optimization problem

The functional to be minimized can be chosen in several ways. The effectiveness mostly depends on the optimization method which is used afterwards. The most natural functional is the *evacuation time*,

$$\min \{t > 0 : x_i(t) \notin \Omega \text{ for all } i = 1, \ldots, N\}, \tag{7.15}$$

subject to (7.1) or (7.11) and with $u \in \mathrm{Adm}$. Concerning the set of admissible controls Adm, we only impose a box constraint of the form $u(t) \in \mathcal{U}$ for every $t \in [0, T]$, where $T > 0$ is a sufficiently large time horizon and $\mathcal{U}$ is some compact subset of $\mathbb{R}^{dm}$.
Another cost functional, more affordable by standard methods [6, 39], is

$$\ell(\mathbf{x}, \mathbf{y}, u) = \omega^{\mathrm{F}} \sum_{i=1}^{N} \|x_i - x^{\tau}\|^2 + \omega^{\mathrm{L}} \sum_{i=1}^{N} \sum_{k=1}^{m} \|x_i - y_k\|^2 + \omega^{\mathrm{C}} \sum_{k=1}^{m} \|u_k\|^2, \tag{7.16}$$

for some positive constants ω^{F}, ω^{L}, and ω^{C}. The first term promotes the fact that followers have to reach the exit while the second forces leaders to keep contact with the crowd. The last term penalizes excessive velocities. This minimization is performed at every instant (instantaneous control), or along a fixed time frame $[t_i, t_f]$

$$\min_{u \in \mathrm{Adm}} \int_{t_i}^{t_f} \ell(\mathbf{x}(t), \mathbf{y}(t), u(t))\, dt, \qquad \text{subject to (7.1) or (7.11).} \tag{7.17}$$

With regards to the mesoscopic scale, both functionals (7.15) and (7.16) can be considered, however at the continuous level the presence of few invisible leaders does not assure that the whole mass of followers is evacuated. The major difficulty to reach a complete evacuation of the continuous density is mainly due to the presence of the diffusion term and to the invisible interaction with respect to the leaders. Therefore a more appropriate functional is given by the mass evacuated at the final time T,

$$\min_{u \in \mathrm{Adm}} \int_{\mathbb{R}^d} \int_{\Omega \setminus \mathcal{E}} f(T, x, v)\, dx\, dv \qquad \text{subject to (7.11),} \tag{7.18}$$

where $\mathcal{E}$ is a set containing the exit x^{τ} and representing the zone in which agents can be considered safe (one could think of x^{τ} as the center of the rendezvous zone in case of emergency and $\mathcal{E}$ as the rendezvous zone itself).

Remark 7.9. We will not require that leaders reach the exit, but only the followers.

Remark 7.10. We shall not address here the issue of the relationship between the microscopic and the mesoscopic optimal control problems. We point out that we are not under the hypotheses of Theorem 6.18 due to the lack of mean-field scaling and the presence of discontinuous interaction terms in the dynamics. Despite that, since the current situation bears several similarities with that of Chapter 6, we expect the optimal strategies of the continuous problem to be a reasonable approximation of those of the discrete problem whenever the number of followers is sufficiently high. Such intuition is partially corroborated by the numerical results we shall present in the next section, which show that some distinctive features are shared by the optimal strategies of both problems.

The optimization techniques we shall use to find numerically the minimizer of the above functionals are:

- **Model predictive control**: a computationally efficient way to address the optimal control problem (7.17) is by means of a relaxed approach known as *model predictive control* (MPC) [142]. We consider a sampling of the dynamics (7.1) at every time interval Δt, and the following minimization problem over N_{MPC} time steps, starting from the current time step $\bar{n}$,

$$\min_{u \in \text{Adm}} \sum_{n=\bar{n}}^{\bar{n}+N_{\text{MPC}}-1} \ell(\mathbf{x}(n\Delta t), \mathbf{y}(n\Delta t), u(n\Delta t)) \tag{7.19}$$

  generating an optimal sequence of controls $\{u(\bar{n}\Delta t), \ldots, u((\bar{n}+N_{\text{MPC}}-1)\Delta t)\}$, from which only the first term is taken to evolve the dynamics for a time Δt, to recast the minimization problem over an updated time frame $\bar{n} \leftarrow \bar{n}+1$. Note that for $N_{\text{MPC}} = 2$, the MPC approach recovers an instantaneous controller, whereas for $N_{\text{MPC}} = (t_f - t_i)/\Delta t$ it solves the full time frame problem (7.17). Such flexibility is complemented with a robust behavior, as the optimization is re-initialized every time step, allowing to address perturbations along the optimal trajectory.

- **Modified compass search**: when the cost functional is highly irregular and the search of local minima is particularly difficult, it could be convenient to move towards random methods as compass search (see [18] and references therein), genetic algorithms, or particle swarm optimization. In the following we describe a compass search method that works surprisingly well for our problem.

  First of all, we consider only piecewise constant trajectories, introducing suitable *switching times* for the leaders' controls. More precisely, we assume that leaders move at constant velocity for a given fixed time interval and, when the switching time is reached, a new velocity vector is chosen. Therefore, the control variables are the velocities at the switching times for each leader. Note that controlling directly the velocities rather than the acceleration makes the optimization problem much simpler because minimal control variations have an immediate impact on the dynamics.

  Starting from an initial guess, at each iteration the optimization algorithm modifies the current best control strategy found so far by means of small random variations of the current values. Then, the cost functional is evaluated. If the variation is advantageous (the cost decreases), the variation is kept, otherwise it is discarded. The method stops when the strategy cannot be improved further. As initial guess for the strategy, leaders move at constant speed along the direction joining the target with their position at initial time $t = 0$.

7.6 Numerical tests

In what follows we present some numerical tests to validate our modeling setting at the microscopic and mesoscopic level.

The microscopic model (7.1) is discretized by means of the explicit Euler method with a time step $\Delta t = 0.1$. The evolution of the kinetic density in (7.14) is approximated by means of binary interaction algorithms, which approximates the Boltzmann dynamics (7.8) with a meshless Monte-Carlo method for small values of the parameter ε, as presented in [7]. We choose $\varepsilon = 0.02$, $\Delta t = 0.01$ and a sample of $N_s = 10000$ particles to reconstruct the kinetic density. This type of approach is inspired by numerical methods for plasma physics and it allows to solve the interaction dynamics with a reduced computational cost compared with mesh-based methods, and an accuracy of $O(N_s^{-1/2})$.

Concerning optimization, in the microscopic case we adopt either the compass search with functional (7.15) or MPC with functional (7.17). In the mesoscopic case we adopt the compass search with functional (7.18).

We consider three settings for pedestrians, without and with obstacles, hereafter referred to as Setting 1, 2, and 3, respectively. In Setting 1 we set the compass search switching times every 20 time steps, and in Setting 2 every 50, having fixed the maximal random variation to 1 for each component of the velocity. In Setting 1, the inner optimization block of the MPC procedure is performed via a direct formulation, by means of the `fmincon` routine in Matlab, which solves the optimization problem via an SQP method.

7.6.1 Setting 1

To begin with, let us consider the case of a large room with no obstacles, see Figure 7.3.

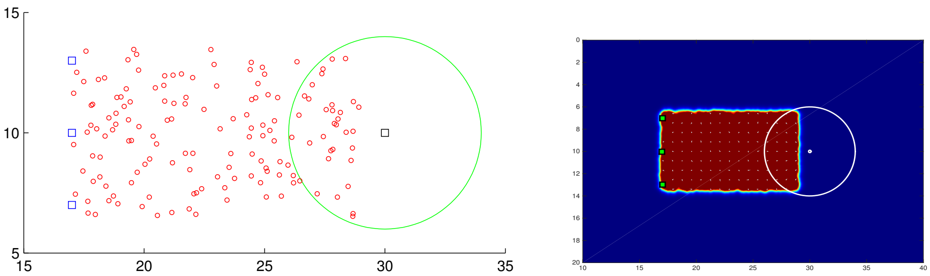


Figure 7.3: Setting 1. Left: initial positions of followers (circles) and leaders (squares). Right: uniform density of followers and the microscopic leaders (squares).

The exit is a point located at $x^\tau = (30, 10)$ which can be reached from any direction. We set $\Sigma = \left\{x \in \mathbb{R}^2 : \|x - x^\tau\| < 4\right\}$. This simple setting helps elucidating the role and the interplay of the different terms of our model. Followers are initially randomly distributed in the domain $[17, 29] \times [6.5, 13.5]$ with velocity $(0, 0)$. Leaders, if present, are located to the left of the crowd. Parameters are reported in Table 7.1.

7.6.1.1 Microscopic model

Figure 7.4-first row shows the evolution of the agents computed by the microscopic model, without leaders. Followers having a direct view of the exit immediately point towards it, and some group mates close to them follow thanks to the alignment force. On the contrary, farthest people split in several but cohesive groups with random direction and never reach the exit.

Figure 7.4-second row shows the evolution of the agents with three leaders. The leaders'

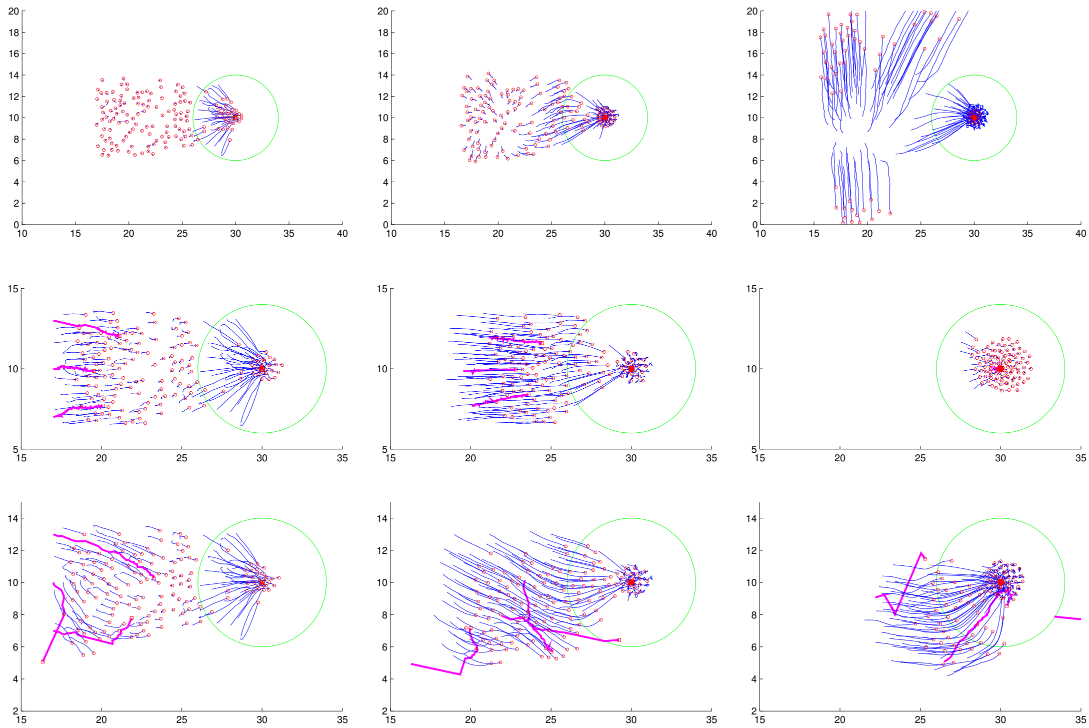


Figure 7.4: Setting 1. Microscopic dynamics. First row: no leaders. Second row: three leaders, go-to-target strategy. Third row: three leaders, optimal strategy (computed with the compass search method).

strategy is defined manually. More precisely, at any time the control is equal to the unit vector pointing towards the exit from the current position. Hereafter we refer to this strategy as “go-to-target”. Notice that the final leaders’ trajectories are not straight lines because of the additional repulsive force. As it can be seen, the crowd behavior changes completely since, this time, the whole crowd reaches the exit. However, followers form a heavy congestion around the exit which delays notably the evacuation. This suggests that the strategy of the leaders is not optimal and that it can be improved by an optimization method. The shape of the congestion is circular: this is perfectly in line with the results of other social force models as well as physical observation, which shows the formation of an “arch” near the exits. The arch is correctly substituted here by a full circle due to the absence of walls.

Figure 7.4-third row shows the evolution of the agents with three leaders and the optimal strategy obtained by the compass search algorithm. Surprisingly enough, the optimizator prescribes that leaders should *divert* some pedestrians from the right direction, in order to not steer the whole crowd to the exit at the same time. In this way, congestion is avoided and pedestrian flow through the exit is increased.

In this test we have also run the MPC optimization, including a box constraint of the type $u(t) \in \mathcal{U} = [-1,1]^{2m}$ for every $t \geq 0$. We choose $\omega^{\mathrm{F}} = 1$, $\omega^{\mathrm{L}} = 1e-5$, and $\omega^{\mathrm{C}} = 1e-5$. MPC results are consistent in the sense that for $N_{\mathrm{MPC}} = 2$, the algorithm recovers a controlled behavior similar to the application of the instantaneous controller (or go-to-target strategy). Increasing the time frame up to $N_{\mathrm{MPC}} = 6$ improves both congestion and evacuation times, but results still remain non competitive if compared to the whole time frame optimization performed with a compass search.

In Figure 7.5 we compare the occupancy of the exit's visibility zone as a function of time for go-to-target strategy and optimal strategies (compass search, 2-step, and 6-step MPC). We also show the decrease of the value function as a function of attempts (compass search) and time (MPC). Evacuation times are compared in Table 7.3. It can be seen that only the long-term optimization strategies are efficient, being able to moderate congestion and clogging around the exit. This suggests a quite unethical but effective evacuation

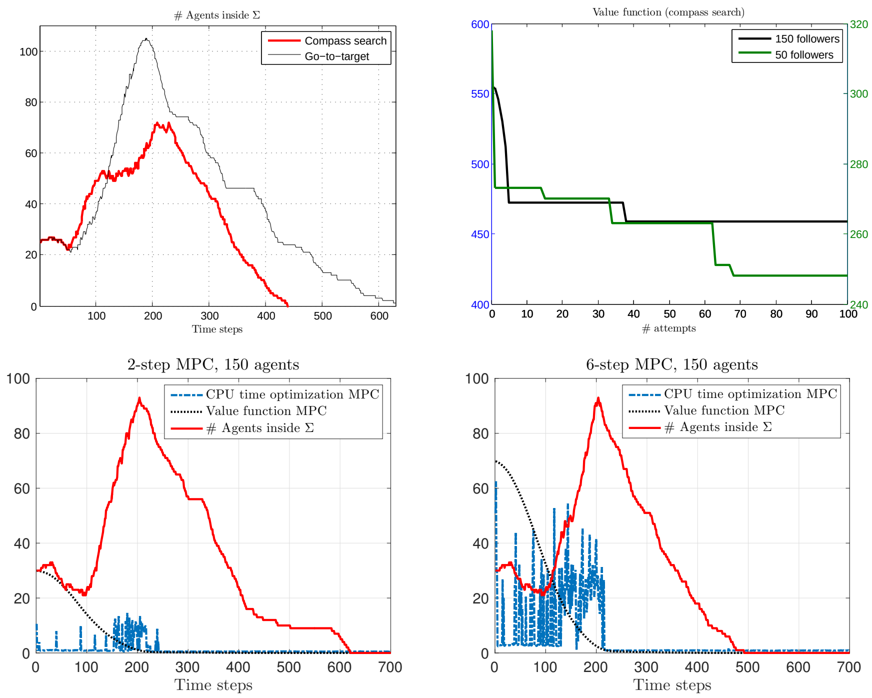


Figure 7.5: Setting 1. Optimization of the microscopic dynamics. Top-left: occupancy of the exit's visibility zone Σ as a function of time for optimal strategy (compass search) and go-to-target strategy. Top-right: decrease of the value function (7.15) as a function of the iterations of the compass search (for 50 and 150 followers). Bottom: MPC optimization. occupancy of the exit's visibility zone Σ as a function of time, CPU time of the optimization call embedded in the MPC solver, and the evolution of the corresponding value (2-step and 6-step).

	no leaders	go-to-target	2-MPC	6-MPC	CS (IG)
$N = 50$	335	297	342	278	248 (318)
$N = 150$	∞	629	619	491	459 (554)

Table 7.3: Setting 1. Evacuation times (time steps). CS=compass search, IG=initial guess.

procedure, namely misleading some people to a false target and then leading them back to the right one, when exit conditions are safer. Notice that, in real-life situations, most of the injuries are actually caused by overcompression and suffocation rather than the present danger.

7.6.1.2 Mesoscopic model

We consider here the case of a continuous density of followers. Figure 7.6-first row shows the evolution of the uncontrolled system of followers. Due to the diffusion term and the topological alignment, large part of the mass spreads around the domain and is not able to reach the target exit.

In Figure 7.6-second row we take into account the action of three leaders, driven by a go-to-target strategy defined as in the microscopic case. It is clear that also in this case the action of leaders is able to influence the system and promote the evacuation, but the presence of the diffusive term causes the dispersion of part of the continuous density. The result is that part of the mass is not able to evacuate, unlike the microscopic case.

In order to improve the go-to-target strategy, we rely on the compass search method to minimize – differently from the microscopic case – the objective functional (7.18), i.e., the total mass evacuated at final time. Figure 7.6-third row sketches the optimal strategy found in this way: on one hand, the two external leaders go directly towards the exit, evacuating part of the density; on the other hand the central leader moves slowly backward, misleading part of the density and only later it moves forward towards the exit. The efficiency of the leaders' strategy is due in particular by the latter movement of the last leader, which is able to gather the followers' density left behind by the others, and to reduce the occupancy

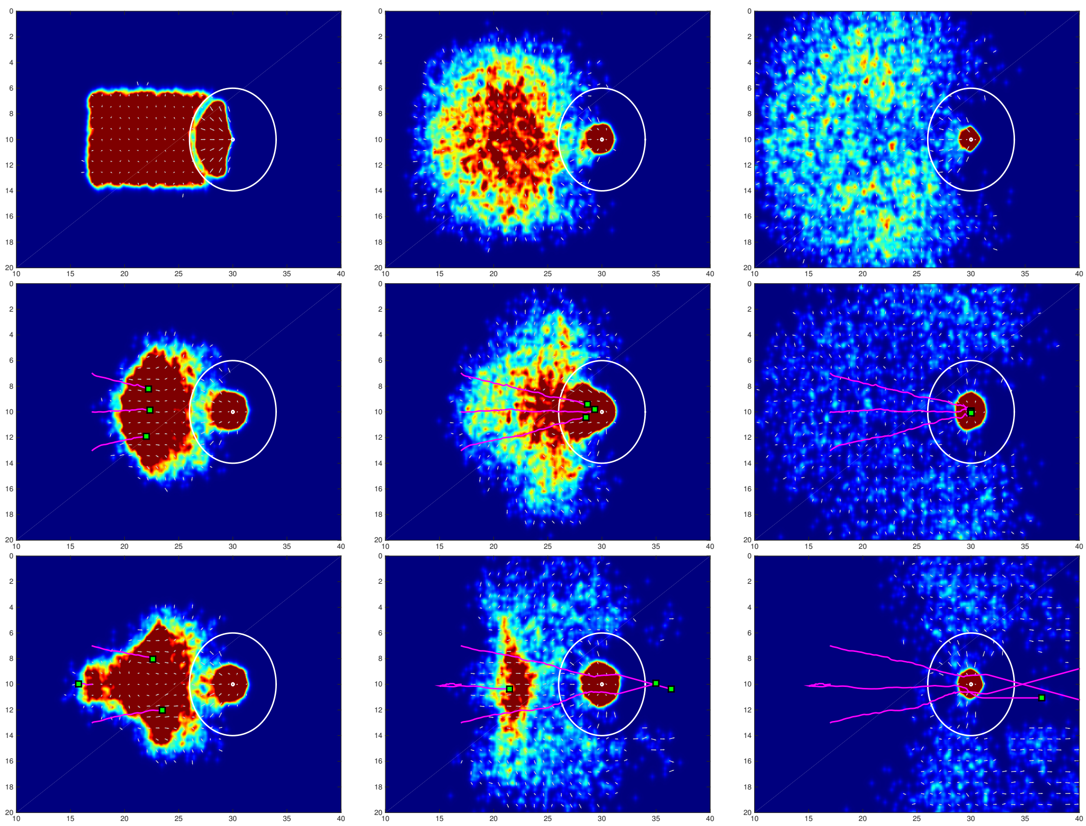


Figure 7.6: Setting 1. Mesoscopic dynamics. First row: no leaders. Second row: three leaders, go-to-target strategy. Third row: three leaders, optimal strategy (compass search).

of the exit's visibility area by delaying the arrival of part of the mass.

In Figure 7.7 we summarize, for the three numerical experiments, the evacuated mass and the occupancy of the exit's visibility area Σ as functions of time. The occupancy of the exit's visibility area shows clearly the difference between the leaders' action: for the go-to-target strategy, the amount of mass occupying Σ concentrates quickly and the evacuation is partially hindered by the clogging effect, as only 71.3% of the total mass is evacuated. The optimal strategy (obtained after 30 iterations) is able to better distribute the mass arrival in Σ, and an higher efficiency is reached, evacuating up to 85.2% of the total mass.

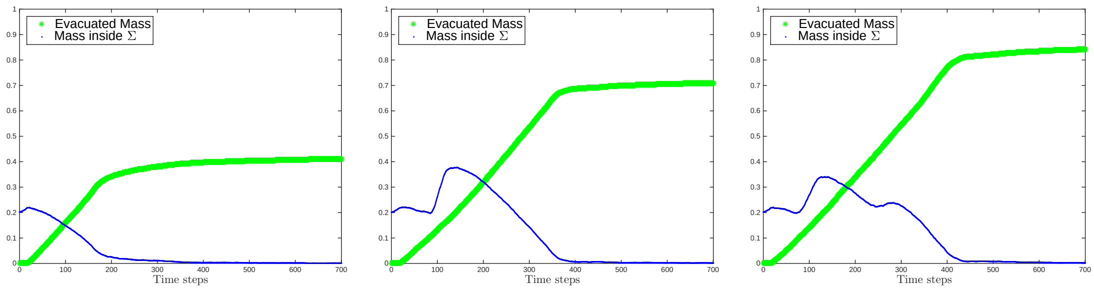


Figure 7.7: Setting 1. Occupancy of the exit's visibility zone Σ, dotted lines, and percentage of evacuated mass, star lines, as function of time. Left: histograms for the case without leaders (percentage of evacuated mass 41.2%). Center: histograms for leaders moving with the go-to-target strategy (percentage of evacuated mass 71.3%). Right: histograms for leaders with an optimal strategy (percentage of evacuated mass 85.2%).

Remark 7.11. In a realistic context it is impossible to know a priori the initial position and velocity of a crowd, but they can be expressed as a joint probability density, obtained from statistics on real data. Therefore from the simulation of the kinetic model (7.14), one can estimate the evolution of these probabilities and testing the efficiency of specific control strategies.

7.6.2 Setting 2

In this section we test the microscopic model (with compass search optimization) in a more complicated setting, which has also some similarities with the one considered in Section 7.7. The crowd is initially confined in a rectangular room with three walls. In order to evacuate, people must first leave the room and then search for the exit point. We assume that walls are not visible, i.e., people can perceive them only by physical contact. This corresponds to an evacuation in case of null visibility (except for the exit point which is still visible from within Σ). Walls are handled as in [76] (see also [75] for an overview of the techniques employed to handle obstacles in pedestrian simulations).

If no leaders are present, the crowd splits in several groups and most of the people hit the wall, see Figure 7.8-first row. After some attempts, the crowd finds the way out, and then it crashes into the right boundary of the domain. Finally, by chance people decide *en cascade* to go upward. The crowd leaves the domain in 1162 time steps.

If instead we hide in the crowd two leaders who point fast towards the exit (Figure 7.8-second row), the evacuation from the room is completed in very short time, but after that, the influence of the leaders vanishes. Unfortunately, this time people decide to go downward after hitting the right boundary, and nobody leaves the domain. Slowing down the two leaders helps keeping the leaders' influence for longer time, although it is quite difficult to find a good choice.

Compass search optimization finds (after 30 iterations) a nice strategy for the two leaders which remarkably improves the evacuation time, see Figure 7.8-third row. One leader behaves similarly to the previous case, while the other diverts the crowd pointing SE, then comes back to wait for the crowd, and finally points NE towards the exit. This strategy allows to bring everyone to the exit in 549 time steps, without bumping anyone against the boundary, and avoiding congestion near the exit.

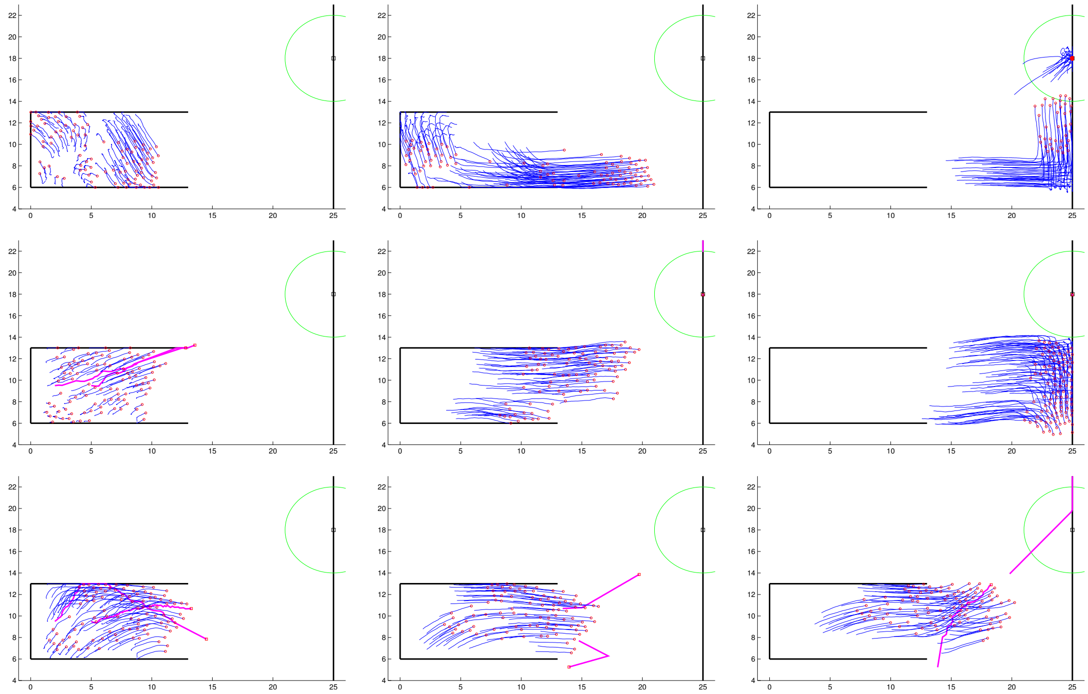


Figure 7.8: Setting 2. Microscopic simulation. First row: no leaders. Second row: two leaders and go-to-target strategy. Third row: two leaders and best strategy computed by the compass search.

7.6.3 Setting 3

In the following test we propose a different use of the leaders. Rather than steering the mass towards the exit, they can be employed to fluidify the evacuation near a door. In fact, it is observed that when several panicky pedestrians reach a bottleneck at the same time they block each other, coming to a deadlock. As a consequence, none of them can pass through the bottleneck. We want to investigate the possibility of using leaders as small *smart obstacles*, which alleviate the high friction among the bodies just moving randomly near the exit.

We consider a large room with a single exit door located on the right wall, visible from any point ($\Sigma = \mathbb{R}^2$). We nullify the repulsion force perceived by the leaders, see Table 7.1. Most important, in this test we assume that followers have nonzero size and hard

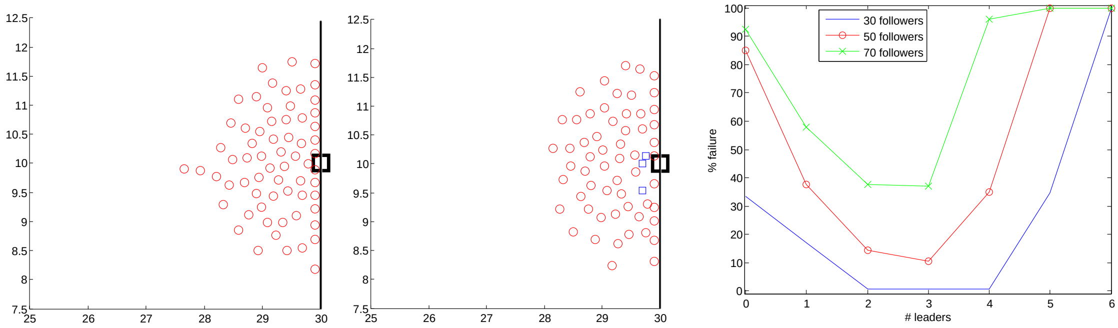


Figure 7.9: Setting 3 (exit is visible from any point). Left: evacuation with no leaders. Center: evacuation with 3 leaders. Right: percentage of evacuation failures as a function of the number of leaders, for $N = 30$, 50, 70.

collisions are not possible. This is achieved assuming that they are circular-shaped with diameter 0.25, and keeping frozen those pedestrians who aim at moving too close to another pedestrian in one time step. The width of the exit is 0.45. Leaders stay near the exit at any time and move with a random velocity, each component being uniformly distributed in the interval $[-1.3, 1.3]$. We say that the evacuation succeeded if $N - 10$ followers leave the room within 2000 time steps, otherwise we say that the evacuation failed. Average quantities are computed running the simulation 200 times.

Figure 7.9 shows a typical outcome without (left) and with (center) leaders. Figure 7.9-right shows the percentage of evacuation failures as a function of the number of leaders. It is clearly visible that in this scenario the optimal number of leaders is 3, and this number is independent form the number of evacuees. This is reasonable since leaders act at local level near the exit and do not affect people queuing far from the exit. We noted that the leaders initially act as a barrier for the incoming pedestrians and slow down the evacuation. Later on, instead, they are quite efficient in breaking symmetries and avoiding deadlocks. As a by-product, we confirmed once again the Braess's paradox, which states that reducing the freedom of choice (e.g., adding obstacles in front of the exit [78, 171]) one can actually improve the overall dynamics.

7.7 Validation of the proposed crowd control technique

The crowd control technique investigated in the previous sections relies on the fact that pedestrians actually exhibit herding behavior in special situations, so that the alignment term in the model is meaningful. To confirm this, we have organized an experiment involving volunteer pedestrians. It is not the first time that a real experiment with people is performed; for instance, one can be found in [113].

The experiment took place on October 1, 2014 at Department of Mathematics of Sapienza – University of Rome, Rome (Italy), at 1 p.m. Participants were chosen among first-year students. Since courses started just two days before the date of the experiment and the Department has a complex ring-shaped structure, we could safely assume that most of the students were unfamiliar with the environment.

7.7.1 The experiment

Students were informed to be part of a scientific experiment aiming at studying the behavior of a group of people moving in a partially unknown environment. Their task was to leave the classroom II and reach a certain target inside the Department. Students were asked to accomplish their task as fast as possible, without running and without speaking with others. Participation was completely voluntary and not remunerated (only a celebratory T-shirt was given at the end).

76 students (39 girls, 37 boys) agreed to participate in the experiment. The students were then divided randomly in two groups: group A (42 people) and group B (34 people). The two groups performed the experiment independently one after the other. The target was communicated to the students just before the beginning of the experiment. It was the Istituto Nazionale di Alta Matematica (INdAM), which is located just upstairs with respect to the classroom II. Due to the complex shape of the environment, there are many paths joining classroom II and INdAM. The shortest path requires people to leave the classroom, to go leftward in a not so frequented and unfamiliar area (even for experienced students), and climb the stairs. The target can also be reached by going rightward and climb other, more frequented, stairs. Figure 7.10-left shows the experiment setting, with classroom II and its two exits.

Figure 7.10: Left: the two exits of classroom II. Just outside the classroom, people have to decide whether to go right or left. Right: 10 students following the unaware leader in group A.

In group B there were 5 incognito students (hereafter referred to as leaders, to make a parallel with the model discussed in this chapter) of the same age of the others. Leaders were previously informed about the goal of the experiment and the location of the target. They were also trained in order to steer the crowd toward the target in minimal time. Nobody recognized them as "special" before or during the experiment. Unexpectedly, also in group A there was a girl who knew the target. Therefore, she acted as an *unaware* leader.

It is important to stress that all the other students continued their usual activities and *participants were not officially recorded by the organizers*. This choice was crucial to get *natural* behavior and meaningful results, see [77, Section 3.4.3]. The price to pay is that we had to extract participants' trajectories by low-resolution videos taken by two observers.

Moreover, the area outside classroom II was rather crowded, introducing a high level of noise in the experiment. Finally, let us mention that some participants broke the rules of the experiment speaking with others and running.

7.7.2 Results and conclusions from the experiment

In Figure 7.11 we show the history of left/right decisions taken by the students just outside the two doors of the classroom. Group A used both exits of the classroom. The 8-people

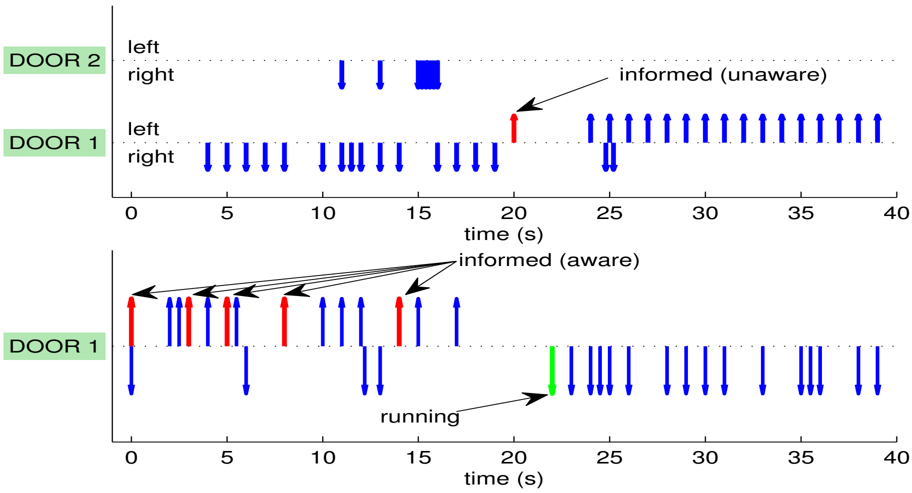


Figure 7.11: Time instants corresponding to the students' right/left decisions in front of the door for group A (top) and B (bottom).

group which used door 2 was uncertain about the direction for a while. Then, once the first two students decided to go rightward, the others decided simultaneously to follow them. Similarly, the first student leaving from door 1 was uncertain for a few seconds, then he moved rightward and triggered a clear *domino effect*. At $t = 20s$ the unaware leader moved to the left, inducing hesitation and mixed behavior in the followers. After that, another domino effect arose.

Group B used only door 1. Invisible leaders were able to trigger a domino effect but this time 4 people decided unilaterally not to follow them, although they were not informed about the destination. At $t = 20s$, after the passage of the last leader, a girl passed through the door and went leftward. Then, at $t = 22s$, she suddenly began to run toward the right. She first induced hesitation, then triggered a new rightward domino effect.

Students have shown a tendency to go rightward, being that part of the Department more familiar, but this tendency was greatly overcame by the wish to follow the group mates in front, regardless of their right/left preferences. This fact was confirmed by the students themselves after the experiment. Indeed, they admitted that they have been influenced by the leaders because of their clear direction of motion. Interestingly, no more than 5 people have been influenced by a single leader. This is compatible with the topological alignment term used in our model. The small value of $\mathcal{N}$ here is due to the fact that the space in front of the doors was rather crowded and the visibility was reduced.

It is also interesting to notice that only 4 students reached the target alone. This confirms the tendency not to remain isolated and to form clusters. Video recordings also show that (bad or well) informed people behaved in a rather recognizable manner, walking faster, overtaking the others and exhibiting a clear direction of motion. No follower has overtaken other people.

The 5 incognito leaders adopted an effective but non optimal strategy, being too close to each other, too much concentrated in the front line, and too fast (see Figure 7.8-second row). This was due to the fact that organizers expected a clearer environment in front of the doors and consequently a faster coming out of the people. A better distribution in the crowd would lead to better results, although nothing can be done against a quite noticeable bad informed person.

APPENDIX A

Gronwall's estimates and variations on the theme

Gronwall's estimates are needed in virtually every chapter of the present thesis. Besides the classical one, we need to develop a variation for piecewise continuous evolutions. Both are reported below.

Lemma A.1 (Classical Gronwall's Lemma)**.** *Let ρ, β and u be real valued functions and furthermore assume that β is non-negative and continuous, u is continuous and ρ is non-decreasing and integrable on $[a, b]$. Moreover, assume that we have*

$$u(t) \leq \rho(t) + \int_a^t \beta(s)u(s)\ ds, \quad \textit{for all } t \in [a, b].$$

Then we have

$$u(t) \leq \rho(t)e^{\int_a^t \beta(s)\ ds}, \quad \textit{for all } t \in [a, b].$$

Lemma A.2 (Discrete Gronwall's Lemma)**.** *Fix $\tau \in [0, T]$ and let ρ, β_1, β_2, and u be real functions on $[0, T]$ such that*

- *(i) ρ is non-decreasing and bounded on $[0, T]$,*
- *(ii) β_1 is non-decreasing and continuous on $[0, T]$,*
- *(iii) β_2 is non-negative and continuous on $[0, T]$ and*
- *(iv) u is non-negative and continuous on $[0, T]$.*

Assume that for every $t \in [0, T]$ the following property holds: let $n \in \mathbb{N}_0$ such that $n\tau \leq t < (n + 1)\tau$ and assume

$$u(t) \leq (\rho(t) - \rho(n\tau)) + (1 + \beta_1(t) - \beta_1(n\tau))u(n\tau) + \int_{n\tau}^t \beta_2(s)u(s)\ ds \quad \textit{for all } t \in [0, T].$$

Then it holds

$$u(t) \le u(0)e^{\beta_1(t)-\beta_1(0)+\int_0^t \beta_2(s)\ ds} + (\rho(t)-\rho(0))e^{\beta_1(t)-\beta_1(0)+\int_0^t \beta_2(s)\ ds} \quad \textit{for all } t \in [0,T].$$

Proof. The proof uses n applications of the classical Gronwall's Lemma for the intervals $[0,\tau], [\tau, 2\tau], \dots, [(n-1)\tau, n\tau]$, $[n\tau, t]$. The first application over $[n\tau, t]$ gives

$$u(t) \le [(\rho(t)-\rho(n\tau)) + (1+\beta_1(t)-\beta_1(n\tau))u(n\tau)]\, e^{\int_{n\tau}^t \beta_2(s)\ ds}.$$

The second application for the interval $[(n-1)\tau, n\tau]$ gives

$$u(n\tau) \le [(\rho(n\tau)-\rho((n-1)\tau)) + (1+\beta_1(n\tau)-\beta_1((n-1)\tau))u((n-1)\tau)] \times \\ \times e^{\int_{(n-1)\tau}^{n\tau} \beta_2(s)\ ds}.$$

Plugging the last estimate into the first one we arrive at

$$\begin{aligned} u(t) \le\ & (\rho(t)-\rho(n\tau))e^{\int_{n\tau}^t \beta_2(s)\ ds} + \\ & + [1+\beta_1(t)-\beta_1(n\tau)]\,[\rho(n\tau)-\rho((n-1)\tau)]\, e^{\int_{(n-1)\tau}^t \beta_2(s)\ ds} + \\ & + [1+\beta_1(t)-\beta_1(n\tau)]\,[1+\beta_1(n\tau)-\beta_1((n-1)\tau)]\, u((n-1)\tau) e^{\int_{(n-1)\tau}^t \beta_2(s)\ ds}. \end{aligned}$$

Now, by induction on this iterated operations we obtain

$$\begin{aligned} u(t) \le\ & (\rho(t)-\rho(n\tau))e^{\int_{n\tau}^t \beta_2(s)\ ds} + \\ & + [1+\beta_1(t)-\beta_1(n\tau)] \sum_{i=0}^{n-1} [\rho((n-i)\tau)-\rho((n-i-1)\tau)]\, e^{\int_{(n-i-1)\tau}^t \beta_2(s)\ ds} \times \\ & \times \prod_{j=n-i+1}^{n} [1+\beta_1(j\tau)-\beta_1((j-1)\tau)] + \\ & + u(0)e^{\int_0^t \beta_2(s)\ ds}\,[1+\beta_1(t)-\beta_1(n\tau)] \prod_{j=1}^{n} [1+\beta_1(j\tau)-\beta_1((j-1)\tau)]\,. \end{aligned}$$

It is a well-known fact that $1+a \le e^a$ for all $a \ge 0$. Hence it follows

$$\prod_{i=1}^{n}(1+a_n) \le e^{\sum_{i=1}^n a_n} \quad \text{for all } a_1, \dots, a_n \ge 0.$$

If we use the above inequality we get

$$\begin{aligned} u(t) \le\ & (\rho(t)-\rho(n\tau))e^{\int_{n\tau}^t \beta_2(s)\ ds} + \\ & + \sum_{i=0}^{n-1} [\rho((n-i)\tau)-\rho((n-i-1)\tau)]\, e^{\int_{(n-i-1)\tau}^t \beta_2(s)\ ds} e^{\beta_1(t)-\beta_1((n-i)\tau)} + \\ & + u(0)e^{\int_0^t \beta_2(s)\ ds} e^{\beta_1(t)-\beta_1(0)} \\ \le\ & u(0)e^{\beta_1(t)-\beta_1(0)+\int_0^t \beta_2(s)\ ds} + (\rho(t)-\rho(0))e^{\beta_1(t)-\beta_1(0)+\int_0^t \beta_2(s)\ ds}, \end{aligned}$$

which is the desired statement. □

APPENDIX B

Some facts about Carathéodory solutions

For the reader's convenience we briefly recall some general, well-known results about solutions to Carathéodory differential equations.

We fix a domain $\Omega \subset \mathbb{R}^n$, a Carathéodory function $g : [0,T] \times \Omega \to \mathbb{R}^n$, and $0 < \tau \leq T$. A function $y : [0,\tau] \to \Omega$ is called a *solution* of the Carathéodory differential equation

$$\dot{y}(t) = g(t, y(t)) \quad \text{for } t \in [0,\tau], \tag{B.1}$$

if and only if y is absolutely continuous and (B.1) is satisfied a.e. in $[0,\tau]$. The following existence result holds.

Theorem B.1. *Fix $T > 0$ and $y^0 \in \mathbb{R}^n$. Suppose that there exists a compact subset Ω of $\mathbb{R}^n$ such that $y^0 \in \mathrm{int}(\Omega)$ and there exists $m_\Omega \in L^1([0,T])$ for which it holds*

$$\|g(t,y)\| \leq m_\Omega(t), \tag{B.2}$$

for a.e. $t \in [0,T]$ and for all $y \in \Omega$. Then there exists a $\tau > 0$ and a solution $y(\cdot)$ of (B.1) *defined on the interval $[0,\tau]$ which satisfies $y(0) = y^0$. If there exists a function $\ell \in L^1([0,T])$ such that the function g also satisfies the condition*

$$g(t,y) \cdot y \leq \ell(t)(1 + \|y\|^2), \tag{B.3}$$

for a.e. $t \in [0,T]$ and every $y \in \Omega$, and it holds $B(0,R) \subseteq \Omega$ for

$$R > \left((1 + \|y^0\|^2)e^{\int_0^T \ell(t)dt} - 1\right)^{1/2}, \tag{B.4}$$

then the local solution $y(\cdot)$ of (B.1) *which satisfies $y(0) = y^0$ can be extended to the whole* *interval* $[0, T]$.

Proof. Since $y^0 \in \text{int}(\Omega)$, we can consider a ball $B(y^0, r) \subset \Omega$ for some $r > 0$. The classical result [100, Chapter 1, Theorem 1] and (B.2) yield the existence of a local solution defined on an interval $[0, \tau]$ and taking values in $B(y^0, r)$.
If (B.3) holds, then we have

$$\frac{d}{dt}\|y(t)\|^2 = 2\dot{y}(t) \cdot y(t) = 2g(t, y(t)) \cdot y(t) \leq 2\ell(t)(1 + \|y(t)\|^2),$$

for every $t \in [0, \tau]$, hence, by direct integration we can conclude

$$\|y(t)\| \leq \left((1 + \|y^0\|^2)e^{\int_0^T \ell(t)dt} - 1\right)^{1/2} \qquad \text{for every } t \in [0, \tau].$$

Choosing R as in (B.4), the graph of a solution $y(\cdot)$ cannot reach the boundary of $[0, T] \times B(0, R)$ unless $\tau = T$, therefore the continuation of the local solution to a global one on $[0, T]$ follows, for instance, from [100, Chapter 1, Theorem 4]. □

Remark B.2. The condition

$$\|g(t, y)\| \leq \ell(t)(1 + \|y\|) \tag{B.5}$$

is usually encountered in place of (B.3). However, the latter is weaker than (B.5) as the following set of inequalities shows

$$\begin{aligned} g(t, x) \cdot x &\leq \|g(t, x)\|\|x\| \\ &\leq \ell(t)(1 + \|x\|)\|x\| \\ &\leq \ell(t)(\|x\| + \|x\|^2) \\ &\leq \frac{3}{2}\ell(t)(1 + \|x\|^2). \end{aligned}$$

Therefore, if (B.5) is satisfied with $\ell \in L^1([0, T])$, then (B.3) holds with $3\ell/2 \in L^1([0, T])$.

Gronwall's Lemma A.1 easily gives us the following results on continuous dependence on the initial data.

Theorem B.3. *Let* $g_1, g_2 : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n$ *be Carathéodory functions both satisfying* (B.3) *for the same* $\ell \in L^1([0, T])$*. For every* $r > 0$ *define*

$$\rho_{r,\ell,T} \triangleq \left((1 + r^2)e^{\int_0^T \ell(t)dt} - 1\right)^{1/2}.$$

Assume in addition that there exists a function $L \in L^1([0, T])$ *satisfying*

$$\|g_i(t, y_1) - g_i(t, y_2)\| \leq L(t)\|y_1 - y_2\|$$

for every $t \in [0, T]$*, every* $i = 1, 2$*, and every* y_1*,* y_2 *such that* $\|y_i\| \leq \rho_{r,\ell,T}$*, for* $i = 1, 2$*.*

Then, if for all $i = 1, 2$ *we have*

$$\begin{cases} \dot{y}_i(t) = g_i(t, y_i(t)) & \text{for every } t \in (0, T], \\ y_i(0) = y_i^0, \end{cases}$$

with initial data satisfying $\|y_1^0\|, \|y_2^0\| \leq r$, *it holds*

$$\|y_1(t) - y_2(t)\| \leq e^{\int_0^t L(s)ds} \left(\|y_1^0 - y_2^0\| + \int_0^t \|g_1(s, \cdot) - g_2(s, \cdot)\|_{L^\infty(B(0,\rho_{r,\ell,T}))} ds \right) \quad \text{(B.6)}$$

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

Proof. By Theorem B.1, for every $i = 1, 2$ we may consider a solution $y_i : [0, T] \to \mathbb{R}^d$ of (B.1) with $g = g_i$ and initial datum y_i^0. For every $t \in [0, T]$, we can bound $\|y_1(t) - y_2(t)\|$ from above as follows:

$$\begin{aligned} \|y_1(t) - y_2(t)\| &\leq \|y_1^0 - y_2^0\| + \int_0^t \|\dot{y}_1(s) - \dot{y}_2(s)\| ds \\ &= \|y_1^0 - y_2^0\| + \\ &\quad + \int_0^t \|g_1(s, y_1(s)) - g_1(s, y_2(s)) + g_1(s, y_2(s)) - g_2(s, y_2(s))\| ds \\ &\leq \|y_1^0 - y_2^0\| + \int_0^t \|g_1(s, \cdot) - g_2(s, \cdot)\|_{L^\infty(B(0,\rho_{r,\ell,T}))} \, ds + \\ &\quad + \int_0^t L(s) \|y_1(s) - y_2(s)\| ds. \end{aligned}$$

Since the function $f(t) = \|y_1(0) - y_2(0)\| + \int_0^t \|g_1(s, \cdot) - g_2(s, \cdot)\|_{L^\infty(B(0,\rho_{r,\ell,T}))} \, ds$ is increasing, an application of Gronwall's Lemma A.1 gives (B.6), as desired. □

Remark B.4. Fix $y^0 \in \mathbb{R}^n$ and suppose that a function g satisfies the assumptions of Theorems B.1 and B.3 for every $T \geq 0$. Then, for every $T \geq 0$, we have a unique solution $y_T : [0, T] \to \mathbb{R}^n$ of (B.2) with initial condition y^0. By uniqueness, this family of solutions $(y_T)_{T \geq 0}$ satisfies

$$T_1 \leq T_2 \Rightarrow y_{T_1}(t) = y_{T_2}(t), \quad \text{for every } t \in [0, T_1].$$

We can thus define $y : \mathbb{R}_+ \to \mathbb{R}^n$ as $y(t) = y_t(t)$ for every $t \geq 0$, and obtain by construction a solution of (B.1) with initial datum y^0.

Bibliography

[1] A. Abdelghany, K. Abdelghany, H. Mahmassani, and W. Alhalabi. Modeling framework for optimal evacuation of large-scale crowded pedestrian facilities. *Eur. J. Oper. Res.*, 237(3):1105–1118, 2014.

[2] J. P. Agnelli, F. Colasuonno, and D. Knopoff. A kinetic theory approach to the dynamics of crowd evacuation from bounded domains. *Math. Models Methods Appl. Sci.*, 25(1):109–129, 2015.

[3] S. Ahn, H.-O. Bae, S.-Y. Ha, Y. Kim, and H. Lim. Application of flocking mechanism to the modeling of stochastic volatility. *Math. Models Methods Appl. Sci.*, 23(9):1603–1628, 2013.

[4] S. M. Ahn and S.-Y. Ha. Stochastic flocking dynamics of the Cucker-Smale model with multiplicative white noises. *J. Math. Phys.*, 51(10):103301, 2010.

[5] G. Albi, M. Bongini, E. Cristiani, and D. Kalise. Invisible sparse control of self-organizing agents leaving unknown environments. *To appear in SIAM J. Appl. Math.*, 2015.

[6] G. Albi, M. Herty, and L. Pareschi. Kinetic description of optimal control problems and applications to opinion consensus. *Commun. Math. Sci.*, 13(6):1407–1429, 2015.

[7] G. Albi and L. Pareschi. Binary interaction algorithms for the simulation of flocking and swarming dynamics. *Multiscale Model. Simul.*, 11(1):1–29, 2013.

[8] G. Albi, L. Pareschi, and M. Zanella. Boltzmann type control of opinion consensus through leaders. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20140138, 2014.

[9] R. Alexandre and C. Villani. On the Boltzmann equation for long-range interactions. *Commun. Pure Appl. Math.*, 55(1):30–70, 2002.

[10] L. Ambrosio. Transport equation and Cauchy problem for non-smooth vector fields. In *Calculus of variations and nonlinear partial differential equations. Lectures given at the C.I.M.E. summer school, Cetraro, Italy, June 27–July 2, 2005. With a historical overview by Elvira Mascolo*, pages 1–42. Springer, 2008.

[11] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of Bounded Variation and Free Discontinuity Problems*, volume 254. Clarendon Press Oxford, 2000.

[12] L. Ambrosio and W. Gangbo. Hamiltonian ODEs in the Wasserstein space of probability measures. *Comm. Pure Appl. Math.*, 61(1):18–53, 2008.

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

[14] H. C. Andersen, J. D. Weeks, and D. Chandler. Relationship between the hard-sphere fluid and fluids with realistic repulsion force. *Phys. Rev. A*, 4(4):1597–1607, 1971.

[15] D. Andersson and B. Djehiche. A maximum principle for SDEs of mean-field type. *Appl. Math. Comput.*, 63(3):341–356, 2011.

[16] F. Arvin, J. C. Murray, L. Shi, C. Zhang, and S. Yue. Development of an autonomous micro robot for swarm robotics. In *Proceedings of the IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 635–640. IEEE, 2014.

[17] J.-P. Aubin and A. Cellina. *Differential inclusions. Set-valued maps and viability theory.* Springer-Verlag, 1984.

[18] C. Audet, K.-C. Dang, and D. Orban. Optimization of algorithms with OPAL. *Math. Prog. Comp.*, 6(3):233–254, 2014.

[19] P. Bak. *How nature works: the science of self-organized criticality.* Springer Science & Business Media, 2013.

[20] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study. *P. Natl. Acad. Sci. USA*, 105(4):1232–1237, 2008.

[21] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, M. Viale, and V. Zdravkovic. Empirical investigation of starling flocks: a benchmark study in collective animal behaviour. *Anim. Behav.*, 76(1):201–215, 2008.

[22] R. G. Baraniuk and M. B. Wakin. Random projections of smooth manifolds. *Found. Comput. Math.*, 9(1):51–77, 2009.

[23] S. Battiston, D. Delli Gatti, M. Gallegati, B. Greenwald, and J. Stiglitz. Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *J. Econ. Dyn. Control*, 36(8):1121–1141, 2012.

[24] R. Bellman. *Dynamic programming.* Princeton University Press, 1957.

[25] N. Bellomo and C. Dogbé. On the modeling of traffic and crowds: A survey of models, speculations, and perspectives. *SIAM Rev.*, 53(3):409–463, 2011.

[26] N. Bellomo and G. Toscani. On the Cauchy problem for the nonlinear Boltzmann equation: global existence uniqueness and asymptotic stability. *J. Math. Phys.*, 26:334–338, 1985.

[27] M. Bergler. Dimensionality Reduction of the Cucker-Dong Model of Attraction-Repulsion Systems. Master's thesis, Technische Universität München, 2015.

[28] E. R. Berlekamp, J. H. Conway, and R. K. Guy. *Winning Ways for your Mathematical Plays.* New York: Academic Press, 1982.

[29] P. Binev, A. Cohen, W. Dahmen, and R. DeVore. Universal algorithms for learning theory. Part II: Piecewise polynomial functions. *Constr. Approx.*, 26(2):127–152, 2007.

[30] P. Binev, A. Cohen, W. Dahmen, R. DeVore, and V. Temlyakov. Universal algorithms for learning theory. Part I: Piecewise constant functions. *J. Mach. Learn. Res.*, 6:1297–1321, 2005.

[31] F. Bolley, J. A. Canizo, and J. A. Carrillo. Stochastic mean-field limit: non-Lipschitz forces and swarming. *Math. Models Methods Appl. Sci.*, 21(11):2179–2210, 2011.

[32] M. Bongini and M. Fornasier. Sparse stabilization of dynamical systems driven by attraction and avoidance forces. *Netw. Heterog. Media*, 9(1):1–31, 2014.

[33] M. Bongini, M. Fornasier, F. Frölich, and L. Hagverdi. Sparse control of force field dynamics. In *International Conference on NETwork Games, COntrol and OPtimization*, October 2014.

[34] M. Bongini, M. Fornasier, M. Hansen, and M. Maggioni. Inferring interaction rules from observations of evolutive systems I: The variational approach. Submitted, 2016.

[35] M. Bongini, M. Fornasier, M. Hansen, and M. Maggioni. Inferring interaction rules from observations of evolutive systems II: The universal learning approach. In preparation, 2016.

[36] M. Bongini, M. Fornasier, O. Junge, and B. Scharf. Sparse control of alignment models in high dimension. *Netw. Heterog. Media*, 10(3):647–697, 2015.

[37] M. Bongini, M. Fornasier, and D. Kalise. (Un)conditional consensus emergence under perturbed and decentralized feedback controls. *Discrete Contin. Dyn. Syst.*, 35(9):4071–4094, 2015.

[38] M. Bongini, M. Fornasier, F. Rossi, and F. Solombrino. Mean-field Pontryagin maximum principle. Submitted, 2015.

[39] A. Borzì and S. Wongkaew. Modeling and control through leadership of a refined flocking system. *Math. Models Methods Appl. Sci.*, 25(2):255–282, 2015.

[40] J. Bouvrie and M. Maggioni. Geometric multiscale reduction for autonomous and controlled nonlinear systems. In *Proceedings of the 51st IEEE Conference on Decision and Control*, pages 4320–4327, 2012.

[41] W. Braun and K. Hepp. The Vlasov dynamics and its fluctuations in the $1/N$ limit of interacting classical particles. *Comm. Math. Phys.*, 56(2):101–113, 1977.

[42] A. Bressan and B. Piccoli. *Introduction to the mathematical theory of control*, volume 2 of *AIMS Series on Applied Mathematics*. American Institute of Mathematical Sciences (AIMS), Springfield, MO, 2007.

[43] M. Burger, M. Di Francesco, P. Markowich, and M.-T. Wolfram. Mean field games with nonlinear mobilities in pedestrian dynamics. *Discrete Contin. Dyn. Syst. Ser. B*, 19:1311–1333, 2014.

[44] S. Butail, T. Bartolini, and M. Porfiri. Collective response of zebrafish shoals to a free-swimming robotic fish. *PLoS ONE*, 8(10):e76123, 2013.

[45] G. Buttazzo and G. Dal Maso. Γ-convergence and optimal control problems. *J. Optim. Theory Appl.*, 38(3):385–407, 1982.

[46] J. Cañizo, J. Carrillo, and J. Rosado. A well-posedness theory in measures for some kinetic models of collective motion. *Math. Models Methods Appl. Sci.*, 21(3):515–539, 2011.

[47] S. Camazine, J.-L. Deneubourg, N. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau. *Self-organization in biological systems.* Princeton University Press, 2002.

[48] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[49] M. Caponigro, M. Fornasier, B. Piccoli, and E. Trélat. Sparse stabilization and control of alignment models. *Math. Models Methods Appl. Sci.*, 25(03):521–564, 2015.

[50] R. Carmona, F. Delarue, and A. Lachapelle. Control of McKean–Vlasov dynamics versus mean field games. *Math. Financ. Econ.*, 7(2):131–166, 2013.

[51] J. A. Carrillo, Y.-P. Choi, and M. Hauray. The derivation of swarming models: mean-field limit and Wasserstein distances. In *Collective Dynamics from Bacteria to Crowds*, pages 1–46. Springer, 2014.

[52] J. A. Carrillo, M. R. D'Orsogna, and V. Panferov. Double milling in self-propelled swarms from kinetic theory. *Kinet. Relat. Models*, 2(2):363–378, 2009.

[53] J. A. Carrillo, M. Fornasier, J. Rosado, and G. Toscani. Asymptotic flocking dynamics for the kinetic Cucker-Smale model. *SIAM J. Math. Anal.*, 42(1):218–236, 2010.

[54] J. A. Carrillo, M. Fornasier, G. Toscani, and F. Vecil. Particle, kinetic, and hydrodynamic models of swarming. In *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, Modeling and Simulation in Science, Engineering and Technology, pages 297–336. Birkhäuser Boston, 2010.

[55] J. A. Carrillo, S. Martin, and M.-T. Wolfram. An improved version of the Hughes model for pedestrian flow. *Math. Models Methods Appl. Sci.*, 26(4):671–697, 2016.

[56] E. Casas, C. Clason, and K. Kunisch. Approximation of elliptic control problems in measure spaces with sparse solutions. *SIAM J. Control Optim.*, 50(4):1735–1752, 2012.

[57] A. Cavagna, A. Cimarelli, I. Giardina, A. Orlandi, G. Parisi, A. Procaccini, R. Santagati, and F. Stefanini. New statistical tools for analyzing the structure of animal groups. *Math. Biosci.*, 214(1-2):32–37, 2008.

[58] C. Cercignani, R. Illner, and M. Pulvirenti. *The Mathematical Theory of Dilute Gases*. Springer, 1994.

[59] K. J. Cheung, E. Gabrielson, Z. Werb, and A. J. Ewald. Collective invasion in breast cancer requires a conserved Basal epithelial program. *Cell*, 155(7):1639–1651, 2013.

[60] Y.-L. Chuang, M. R. D'Orsogna, D. Marthaler, A. L. Bertozzi, and L. S. Chayes. State transitions and the continuum limit for a 2D interacting, self-propelled particle system. *Phys. D*, 232(1):33–47, 2007.

[61] F. R. K. Chung. *Spectral graph theory*, volume 92. American Mathematical Society, 1997.

[62] E. N. M. Cirillo and A. Muntean. Dynamics of pedestrians in regions with no visibility - A lattice model without exclusion. *Physica A*, 392:3578–3588, 2013.

[63] F. Clarke. *Functional Analysis, Calculus of Variations and Optimal Control*, volume 264. Springer, 2013.

[64] F. H. Clarke, Y. S. Ledyaev, E. D. Sontag, and A. I. Subbotin. Asymptotic controllability implies feedback stabilization. *IEEE Trans. Automat. Control*, 42(10):1394–1407, 1997.

[65] C. Clason and K. Kunisch. A duality-based approach to elliptic control problems in non-reflexive Banach spaces. *ESAIM Control Optim. Calc. Var.*, 17(01):243–266, 2011.

[66] C. Clason and K. Kunisch. A measure space approach to optimal source placement. *Comput. Optim. Appl.*, 53(1):155–171, 2012.

[67] M. A. Cohen and S. Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans. Syst., Man, Cybern., Syst.*, 13(5):815–826, 1983.

[68] R. R. Coifman and M. J. Hirn. Diffusion maps for changing data. *Appl. Comput. Harmon. Anal.*, 36(1):79–107, 2014.

[69] R. Colombo and M. Lécureux-Mercier. An analytical framework to describe the interactions between individuals and a continuum. *J. Nonlinear Sci.*, 22(1):39–61, 2012.

[70] S. Cordier, L. Pareschi, and G. Toscani. On a kinetic model for a simple market economy. *J. Stat. Phys.*, 120(1-2):253–277, 2005.

[71] J. Cortés and F. Bullo. Coordination and geometric optimization via distributed dynamical systems. *SIAM J. Control Optim.*, 44(5):1543–1574, 2005.

[72] I. D. Couzin and N. R. Franks. Self-organized lane formation and optimized traffic flow in army ants. *P. Roy. Soc. Lond. B Bio.*, 270(1511):139–146, 2003.

[73] I. D. Couzin, J. Krause, N. R. Franks, and S. A. Levin. Effective leadership and decision-making in animal groups on the move. *Nature*, 433:513–516, 2005.

[74] A. J. Craig and I. Flügge-Lotz. Investigation of optimal control with a minimum-fuel consumption criterion for a fourth-order plant with two control inputs; synthesis of an efficient suboptimal control. *J. Fluids Eng.*, 87(1):39–58, 1965.

[75] E. Cristiani and D. Peri. Handling obstacles in pedestrian simulations: models and optimization. Submitted, 2015.

[76] E. Cristiani, B. Piccoli, and A. Tosin. Multiscale modeling of granular flows with application to crowd dynamics. *Multiscale Model. Simul.*, 9(1):155–182, 2011.

[77] E. Cristiani, B. Piccoli, and A. Tosin. *Multiscale Modeling of Pedestrian Dynamics*. Modeling, Simulation & Applications. Springer, 2014.

[78] E. Cristiani, F. S. Priuli, and A. Tosin. Modeling rationality to control self-organization of crowds: An environmental approach. *SIAM J. Appl. Math.*, 75(2):605–629, 2015.

[79] F. Cucker and J.-G. Dong. A general collision-avoiding flocking framework. *IEEE Trans. Automat. Control*, 56(5):1124–1129, 2011.

[80] F. Cucker and J.-G. Dong. A conditional, collision-avoiding, model for swarming. *Discrete Contin. Dynam. Systems*, 34(3):1009–1020, 2014.

[81] F. Cucker and S. Smale. Emergent behavior in flocks. *IEEE Trans. Automat. Control*, 52(5):852–862, 2007.

[82] F. Cucker and S. Smale. On the mathematics of emergence. *Jpn. J. Math.*, 2(1):197–227, 2007.

[83] F. Cucker and D. X. Zhou. *Learning Theory: An Approximation Theory Viewpoint*, volume 24. Cambridge University Press, 2007.

[84] M. Cullen, W. Gangbo, and G. Pisante. The semigeostrophic equations discretized in reference and dual variables. *Arch. Ration. Mech. Anal.*, 185(2):341–363, 2007.

[85] S. Currarini, M. O. Jackson, and P. Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

[86] G. Dal Maso. *An Introduction to* Γ*-Convergence*. Progr. Nonlinear Differential Equations Appl. Birkhäuser Boston Inc., Boston, MA, 1993.

[87] F. Dalmao and E. Mordecki. Cucker-Smale flocking under hierarchical leadership and random interactions. *SIAM J. Appl. Math.*, 71(4):1307–1316, 2011.

[88] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures Algorithms*, 22(1):60–65, 2003.

[89] C. De Boor. B(asic)-Spline Basics. Technical report, Wisconsin University–Madison Mathematics Research Center, 1986.

[90] S. Dereich, M. Scheutzow, and R. Schottstedt. Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincaré (B)*, 49(4):1183–1203, 2013.

[91] L. Desvillettes. On asymptotics of the Boltzmann equation when the collisions become grazing. *Transp. Theory Stat. Phys.*, 21(3):259–276, 1992.

[92] J. Dickinson. Animal social behavior. In *Encyclopaedia Britannica Online*. Encyclopaedia Britannica Inc., 2016.

[93] R. J. DiPerna and P.-L. Lions. On the Cauchy problem for Boltzmann equations: Global existence and weak stability. *Ann. of Math. (2)*, 130(2):321–366, 1989.

[94] S. Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *Found. Comp. Math.*, pages 1–30, 2014.

[95] R. Dobrushin. Vlasov equations. *Funct. Anal. Appl.*, 13(2):115–123, 1979.

[96] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[97] M. R. D'Orsogna, Y.-L. Chuang, A. L. Bertozzi, and L. S. Chayes. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. *Phys. Rev. Lett.*, 96(10):104302, 2006.

[98] Y. Eldar and H. Rauhut. Average case analysis of multichannel sparse recovery using convex relaxation. *IEEE Trans. Inform. Theory*, 56(1):505–519, 2010.

[99] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Trans. Automat. Control*, 49(9):1465–1476, 2004.

[100] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides.* Kluwer Academic Publishers, 1988.

[101] M. Fornasier, J. Haskovec, and J. Vybiral. Particle systems and kinetic equations modeling interacting agents in high dimension. *Multiscale Model. Simul.*, 9(4):1727–1764, 2011.

[102] M. Fornasier and J.-C. Hütter. Consistency of probability measure quantization by means of power repulsion-attraction potentials. Submitted, 2015.

[103] M. Fornasier, B. Piccoli, and F. Rossi. Mean-field sparse optimal control. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 372(2028):20130400, 2014.

[104] M. Fornasier and H. Rauhut. Recovery algorithms for vector-valued data with joint sparsity constraints. *SIAM J. Numer. Anal.*, 46(2):577–613, 2008.

[105] M. Fornasier and H. Rauhut. *Handbook of Mathematical Methods in Imaging*, chapter Compressive Sensing, pages 187–228. Springer-Verlag, 2010.

[106] M. Fornasier and F. Solombrino. Mean-field optimal control. *ESAIM Control Optim. Calc. Var.*, 20(4):1123–1152, 2014.

[107] F. Golse. The mean-field limit for the dynamics of large particle systems. In *Journées "Équations aux dérivées partielles", Forges-les-Eaux, France, 2 au 6 juin 2003. Exposés Nos. I-XV*, pages 1–47. Nantes: Université de Nantes, 2003.

[108] M. Granovetter. Threshold models of collective behavior. *Am. J. Sociol.*, pages 1420–1443, 1978.

[109] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.

[110] M. Grant and S. Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.

[111] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comp. Phys.*, 73(2):325–348, 1987.

[112] L. Greengard and V. Rokhlin. The rapid evaluation of potential fields in three dimensions. In *Vortex methods. Proceedings of the U.C.L.A. Workshop held in Los Angeles, May 20–22, 1987*, volume 1360, pages 121–141. Springer, 1988.

[113] R.-Y. Guo, H.-J. Huang, and S. C. Wong. Route choice in pedestrian evacuation under conditions of good and zero visibility: Experimental and simulation results. *Transportation Res. B*, 46(6):669–686, 2012.

[114] S.-Y. Ha, T. Ha, and J.-H. Kim. Emergent behavior of a Cucker-Smale type particle model with nonlinear velocity couplings. *IEEE Trans. Automat. Control*, 55(7):1679–1683, 2010.

[115] S.-Y. Ha, J.-G. Liu, et al. A simple proof of the Cucker-Smale flocking dynamics and mean-field limit. *Commun. Math. Sci.*, 7(2):297–325, 2009.

[116] S.-Y. Ha and E. Tadmor. From particle to kinetic and hydrodynamic description of flocking. *Kinet. Relat. Models*, 1(3):415–435, 2008.

[117] J. Halloy, G. Sempo, G. Caprari, C. Rivault, M. Asadpour, F. Tâche, I. Saïd, V. Durier, S. Canonge, J. M. Amé, C. Detrain, N. Correll, A. Martinoli, F. Mondada, R. Siegwart, and J. L. Deneubourg. Social integration of robots into groups of cockroaches to control self-organized choices. *Science*, 318(5853):1155–1158, 2007.

[118] J. Han, M. Li, and L. Guo. Soft control on collective behavior of a group of autonomous agents by a shill agent. *Jrl. Syst. Sci. & Complexity*, 19(1):54–62, 2006.

[119] J. Han and L. Wang. Nondestructive intervention to multi-agent systems through an intelligent agent. *PLoS ONE*, 8(5):e61542, 2013.

[120] G. Hardin. The tragedy of the commons. *Science*, 162(3859):1243–1248, 1968.

[121] J. Haskovec. A note on the consensus finding problem in communication networks with switching topologies. *Appl. Anal.*, 94(5):991–998, 2015.

[122] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Soc. Simulat.*, 5(3), 2002.

[123] D. Helbing. Traffic and related self-driven many-particle systems. *Rev. Mod. Phys.*, 73(4):1067–1141, 2001.

[124] J. E. Herbert-Reada, A. Pernab, R. P. Mannb, T. M. Schaerfa, D. J. T. Sumpterb, and A. J. W. Warda. Inferring the rules of interaction of shoaling fish. *PNAS*, 108(46):18726?18731, 2011.

[125] E. S. Herman and N. Chomsky. *Manufacturing Consent: The Political Economy of the Mass Media.* Pantheon Books, 1988.

[126] R. Herzog, G. Stadler, and G. Wachsmuth. Directional sparsity in optimal control of partial differential equations. *SIAM J. Control Optim.*, 50(2):943–963, 2012.

[127] H. Hildenbrandt, C. Carere, and C. Hemelrijk. Self-organized aerial displays of thousands of starlings: a model. *Behav. Ecol.*, 21(6):1349–1359, 2010.

[128] M. Huang, P. E. Caines, and R. P. Malhamé. Individual and mass behaviour in large population stochastic wireless power control problems: centralized and Nash equilibrium solutions. In *Proceedings of the 42nd IEEE Conference on Decision and Control*, volume 1, pages 98–103. IEEE, 2003.

[129] A. Ilachinski. *Artificial War: Multiagent-Based Simulation of Combat.* World Scientific, 2004.

[130] P.-E. Jabin. A review of the mean field limits for Vlasov equations. *Kinet. Relat. Models*, 7(4):661–711, 2014.

[131] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Proceedings of the Conference in modern analysis and probability, New Haven, Conn., 1982*, volume 26 of *Contemp. Math.*, pages 189–206. American Mathematical Society, Providence, RI, 1984.

[132] P. Kachroo, S. J. Al-nasur, S. A. Wadoo, and A. Shende. *Pedestrian dynamics. Feedback control of crowd evacuation.* Understanding Complex Systems. Springer-Verlag, Berlin Heidelberg, 2008.

[133] J. L. Kelley. *General Topology.* Springer-Verlag, 1955.

[134] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and their Applications.* Academic Press, New York, NY, 1980.

[135] A. Kirman, S. Markose, S. Giansante, and P. Pin. Marginal contribution, reciprocity and equity in segregated groups: Bounded rationality and self-organization in social networks. *J. Econ. Dyn. Control*, 31(6):2085–2107, 2007.

[136] J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.

[137] H. Levine, W.-J. Rappel, and I. Cohen. Self-organization in systems of self-propelled particles. *Phys. Rev. E*, 63(1):017101, 2000.

[138] E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20(2):130–141, 1963.

[139] B. Mandelbrot. The variation of certain speculative prices. *J. Bus.*, 36(4):394–419, 1963.

[140] R. Mann. Bayesian inference for identifying interaction rules in moving animal groups. *PLoS ONE*, 6(8):e22827. doi:10.1371/journal.pone.0022827, 2011.

[141] R. N. Mantegna and H. E. Stanley. *An Introduction to Econophysics: Correlations and Complexity in Finance.* Cambridge University Press, 1999.

[142] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: Stability and optimality. *Automatica*, 36(6):789–814, 2000.

[143] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.*, pages 415–444, 2001.

[144] B. Mohar. The Laplacian spectrum of graphs. In Y. Alavi, G. Chartrand, O. R. Oellermann, and A. J. Schwenk, editors, *Graph theory, Combinatorics, and Applications*, volume 2, pages 871–898. Wiley, 1991.

[145] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Trans. Automat. Control*, 50(2):169–182, 2005.

[146] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM Rev.*, 56(4):577–621, 2014.

[147] J. F. Nash. Equilibrium points in N-person games. *Proc. Natl. Acad. Sci. USA*, 36(1):48–49, 1950.

[148] H. Neunzert and J. Wick. The convergence of simulation methods in plasma physics. In *Mathematical methods of plasmaphysics (Oberwolfach, 1979)*, volume 20, pages 271–286. Lang Frankfurt, 1980.

[149] M. Nourian, P. E. Caines, and R. P. Malhamé. Mean field analysis of controlled Cucker-Smale type flocking: Linear analysis and perturbation equations. In *Proceedings of 18th IFAC World Congress Milano*, pages 4471–4476, 2011.

[150] F. Paganini, J. Doyle, and S. Low. Scalable laws for stable network congestion control. In *Proceedings of the 40th IEEE Conference on Decision and Control*, volume 1, pages 185–190. IEEE, 2001.

[151] L. Pareschi and G. Toscani. *Interacting Multiagent Systems. Kinetic Equations & Monte Carlo Methods.* Oxford University Press, USA, 2013.

[152] D. R. Parisi and C. O. Dorso. Microscopic dynamics of pedestrian evacuation. *Physica A*, 354:606–618, 2005.

[153] L. Perea, P. Elosegui, and G. Gómez. Extension of the Cucker-Smale control law to space flight formations. *J. Guid. Control Dynam.*, 32(2):527–537, 2009.

[154] L. Petrovic, M. Henne, and J. Anderson. Volumetric Methods for Simulation and Rendering of Hair. Technical report, Pixar Animation Studios, 2005.

[155] B. Piccoli and F. Rossi. Transport equation with nonlocal velocity in Wasserstein spaces: Convergence of numerical schemes. *Acta Appl. Math.*, 124(1):73–105, 2013.

[156] B. Piccoli and F. Rossi. Generalized Wasserstein distance and its application to transport equations with source. *Arch. Ration. Mech. Anal.*, 211(1):335–358, 2014.

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

[158] B. T. Polyak, M. V. Khlebnikov, and P. S. Shcherbakov. Sparse feedback in linear control systems. *Autom. Remote Control*, 75(12):2099–2111, 2014.

[159] L. Pontryagin, V. Boltyanskij, R. Gamkrelidze, and E. Mishchenko. *The Mathematical Theory of Optimal Processes.* John Wiley & Sons, 1962.

[160] A. Y. Povzner. On the Boltzmann equation in the kinetic theory of gases. *Mat. Sb.*, 100(1):65–86, 1962.

[161] Y. Privat, E. Trélat, and E. Zuazua. Optimal location of controllers for the one-dimensional wave equation. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 30(6):1097–1126, 2013.

[162] M. Pulvirenti. From Hamiltonian particle systems to Kinetic equations. Technical report, WIAS Berlin, 2012.

[163] J. P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. *Appl. Math. Optim.*, 39(2):143–177, 1999.

[164] C. W. Reynolds. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH Computer Graphics*, 21(4):25–34, 1987.

[165] H. Risken. *The Fokker-Planck equation. Methods of solutions and applications.* Berlin: Springer-Verlag, 1996.

[166] J. Shen. Cucker-Smale flocking under hierarchical leadership. *SIAM J. Appl. Math.*, 68(3):694–719, 2007.

[167] G. Stadler. 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.

[168] H. G. Tanner, A. Jadbabaie, and G. J. Pappas. Flocking in fixed and switching networks. *IEEE Trans. Automat. Control*, 52(5):863–868, 2007.

[169] G. Toscani. Kinetic models of opinion formation. *Comm. Math. Sci.*, 4(3):481–496, 2006.

[170] G. Toscani and N. Bellomo. The Enskog-Boltzmann equation in the whole space $\mathbb{R}^3$: Some global existence, uniqueness and stability results. *Comput. Math. Applic.*, 13(9–11):851–859, 1987.

[171] M. Twarogowska, P. Goatin, and R. Duvigneau. Macroscopic modeling and simulations of room evacuation. *Appl. Math. Model.*, 38:5781–5795, 2014.

[172] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. arXiv preprint arXiv:1011.3027, 2010.

[173] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.*, 75(6):1226, 1995.

[174] T. Vicsek and A. Zafeiris. Collective motion. *Phys. Rep.*, 517(3):71–140, 2012.

[175] C. Villani. A review of mathematical topics in collisional kinetic theory. *Handbook of mathematical fluid dynamics*, 1:71–74, 2002.

[176] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[177] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

[178] G. Wachsmuth and D. Wachsmuth. Convergence and regularization results for optimal control problems with sparsity functional. *ESAIM Control Optim. Calc. Var.*, 17(3):858–886, 2011.

[179] J. Wang, L. Zhang, Q. Shi, P. Yang, and X. Hu. Modeling and simulating for congestion pedestrian evacuation with panic. *Physica A*, 428:396–409, 2015.

[180] G. Weisbuch, G. Deffuant, F. Amblard, and J.-P. Nadal. Meet, discuss, and segregate! *Complexity*, 7(3):55–63, 2002.