SIAM J. CONTROL OPTIM.
Vol. 57, No. 4, pp. 2628–2659

SPARSE CONTROL OF HEGSELMANN–KRAUSE MODELS: BLACK HOLE AND DECLUSTERING*

BENEDETTO PICCOLI†, NASTASSIA POURADIER DUTEIL‡, AND EMMANUEL TRÉLAT§

Abstract. This paper elaborates control strategies to prevent clustering effects in opinion formation models. This is the exact opposite of numerous situations encountered in the literature where, on the contrary, one seeks controls promoting consensus. In order to promote declustering, instead of using the classical variance that does not capture well the phenomenon of dispersion, we introduce an entropy-type functional that is adapted to measuring pairwise distances between agents. We then focus on a Hegselmann–Krause-type system and design declustering sparse controls in both finite-dimensional and kinetic models. We provide general conditions characterizing whether clustering can be avoided as a function of the initial data. Such results include the description of black holes (where complete collapse to consensus is not avoidable), safety zones (where the control can keep the system far from clustering), basins of attraction (attractive zones around the clustering set), and collapse prevention (when convergence to the clustering set can be avoided).

Key words. active particles, collective behavior, control, kinetic model, declustering, black swan

AMS subject classifications. 93C15, 93C20, 91D10, 35B36, 34H05

DOI. 10.1137/18M1168911

Introduction. The term “black swan” was first used by Nassim Nicholas Taleb in 2007 in his book *The Black Swan: The Impact of the Highly Improbable* [46], in which he focuses on the extreme impact of rare and unpredictable events. The “black swan theory” has since grown to describe events that are extremely rare, have a massive impact, and are retrospectively predictable. One of the groundbreaking ideas of this recent theory is the fact that human behavior remains unpredictable. By focusing on what is known and probable, scientists tend to be surprised by major unexpected events. Taleb’s philosophy requires one to accept the fact that there will always remain unknown factors—hence, one cannot make future predictions based only on the assumption of a population’s rational behavior.

Bellomo, Herrero, and Tosin [5] have built upon this theory, applying it to the context of social competition that can lead to extreme conflicts. Their work is based upon the fact that individual behavior, whether rational or irrational, contributes in a nonlinear fashion to the global group behavior. Then, even among an initially well-distributed population, local social interactions can lead to unexpected outcomes.

Social dynamics models are particularly suited to describe these kinds of phenomena, as they focus on understanding how *self-organization* emerges from interactions of individual “agents” or “active particles” [2]. The study of collective behavior emerging from local interactions is actually of great interest to a mixed community of

*Received by the editors February 5, 2018; accepted for publication (in revised form) June 5, 2019; published electronically July 25, 2019.
https://doi.org/10.1137/18M1168911
Funding: The authors acknowledge the partial support of the NSF Project Kinetic description of emerging challenges in multiscale problems of natural sciences, DMS grant 1107444.
†Center for Computational and Integrative Biology and Department of Mathematical Sciences, Rutgers University–Camden, Camden, NJ 08102 (piccoli@camden.rutgers.edu).
‡Sorbonne Université, Inria, CNRS, Université de Paris, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France (nastassia.pouradierduteil@ljll.math.upmc.fr).
§Sorbonne Université, CNRS, Université de Paris, Inria, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France (emmanuel.trelat@sorbonne-universite.fr).

mathematicians, biologists, sociologists, economists, and engineers. These models are indeed applicable to a wide variety of fields. In biology, they are used to understand the behavior of large animal groups [3, 8, 16, 17, 37, 48, 51]. Engineering applications involve robot formation and satellite synchronization [7, 29, 34, 38, 43, 44]. Models also apply to socioeconomic problems such as population dynamics, opinion formation (including recent results on social media), voting preferences, market evolution, and emergence of cultural classes [1, 5, 27, 30, 31, 36, 45, 46, 47]. As pointed out in [49], individual behavior, especially human, is often irrational: instead of making strategic decisions, individuals tend to imitate social neighbors. This behavior leads to clustering of opinions or even consensus (agreement of all state variables). Many models reproduce this phenomenon. In [49], this is modeled in a game-theoretic setup, where agents play coordination games to improve their individual payoff. In the voter model, agents imitate the action of a randomly selected counterpart [28]. In the Hegselmann–Krause (HK) bounded confidence model, agents imitate others' behavior only if they are within a certain "confidence" radius [27]. In a competing approach, based on the so-called topological distance, agents imitate a given number of closest neighbors [3]. Another variation of the HK model consists of noticing that heterophilious dynamics enhance consensus [35]. In [24], clustering is shown to occur in the context of emergence of cultural classes, with a model where interactions between agents can be either attractive or repulsive. In [25], the authors provide an explicit computation of the number of clusters formed by the population.

Multiagent systems can be described from a microscopic point of view, by considering a system of coupled (often nonlinear) ODEs [10, 11, 18, 33]. However, as the dimension of the system increases, studying and simulating it become a harder challenge, a phenomenon known as the *curse of dimensionality.* When the number of agents tends to infinity, one can take the *mean-field limit* of the system resulting in a kinetic model, where the population is described by a density measure, and its evolution is given by a unique PDE. The mean-field limits of the HK, Vicsek, and Cucker–Smale models were respectively derived in [9, 21, 26]. Since then, kinetic formulations of social dynamics models have been the focus of many more works; see, for example, [4, 8, 15, 20, 22, 32, 41].

Self-organization has thus been extensively studied, especially focusing on the emergence of patterns such as consensus or alignment that arise from inherent properties of certain dynamics. When consensus is not reached by the system, it is natural to ask whether it can be achieved by controlling it. Such control problems have been investigated in finite-dimensional systems [10, 11, 13, 33] and in kinetic models [12, 40]. Applications involve rendez-vous problems in robotics and flock formation in animal crowd behavior. However, as seen in [7], the states of consensus or clustering are not always desirable as they can be seen as a manifestation of black swan effects. Therefore we choose to study the opposite problem: given dynamics naturally leading to consensus, we aim to, at the contrary, control the system to avoid consensus and clustering, i.e., to keep the agents as far from one another as possible. Possible motivations include keeping a market from collapsing or a crowd from converging to a localized dense conformation. While typical control problems applied to social dynamics models aim to steer the system to consensus, which is a natural feature of the dynamics, here instead we aim to drive the system against its natural behavior. The key is then to understand the interplay between the internal driving force of the system and the external applied control. It is natural to expect that the feasibility of the system will depend on the allowed strength of the control and on the nature of the interaction function. Indeed, our main results show the existence of internal attraction so strong that no control can act on the system: we will refer to such cases as "black holes."

We study a first-order opinion formation model with a positive interaction function $a(\cdot)$ and control the system via an additive term:

$$\dot{x}_i(t) = \frac{1}{N}\sum_{j\neq i} a(\|x_i(t)-x_j(t)\|)(x_j(t)-x_i(t)) + u_i(t), \quad i \in \{1,\ldots,N\}.$$

In its mean-field limit, i.e., when N tends to infinity, we study the kinetic model

$$\partial_t \mu + \operatorname{div}\left(\left(\int_{\mathbb{R}^d} a(\|x-y\|)(y-x)d\mu(y) + \chi_\omega u\right)\mu\right) = 0.$$

All terms and assumptions for both finite-dimensional and kinetic formulations are precisely defined later in sections 1 and 2. The control strategies used to steer the system away from clustering are simple explicit feedbacks depending on the state of the system at the current time. In the microscopic model, the control can be seen as an exterior force acting on the system to separate the agents from one another. For realistic purposes, we assume that we can only act on the system with finite strength, and we impose a constraint on the $\ell_1^N - \ell_2^d$ norm of the control: $\sum_i \|u_i\| \leqslant M$ for some $M > 0$. This constraint is known to promote sparsity (see [10]), so that we promote controls that act on fewer agents at a time. In the macroscopic equation, we consider the class of controls $\chi_\omega u$, where $u \in L^\infty(\mathbb{R}^+ \times \mathbb{R}^d, \mathbb{R}^d)$ and for all $t \geqslant 0$, $\omega(t)$ is a measurable subset of $\mathbb{R}^d$ (and χ is the indicator function). The control is thus constrained in two ways, as for the microscopic case. The amplitude u is constrained by the condition $\|u\|_{L^\infty(\mathbb{R}^+\times\mathbb{R}^d)} \leqslant M$. The spatial sparsity of the control is ensured by the condition $\int_{\omega(t)} dx \leqslant c$ for some $c > 0$. This particular choice of sparsity constraint means that we only allow the control to act on a given area of space. Instead, one could consider constraining the portion of the population being controlled, as done, for instance, in [41] for the control to flocking of the kinetic Cucker–Smale model.

We show that to prevent the formation of consensus, the controls may be chosen to maximize the derivative of the variance of the system. However, this strategy is not enough to achieve the stronger requirement of avoiding any degree of clustering. We show that the state of clustering may be achieved thanks to a different entropy-type functional that measures the *dispersion* of the system. Depending on the behavior of the interaction function $a(\cdot)$, several situations may arise. If $\lim_{s\to 0} sa(s) = +\infty$, there exists a *black hole* region, in which no control can prevent the system from converging. In contrast, if $\lim_{s\to 0} sa(s) = 0$, collapse to consensus can always be avoided. Far from the consensus manifold, we also observe two scenarios. If $\lim_{s\to+\infty} sa(s) = 0$, there exists a *safety zone* in which the control can always keep the system far from consensus. This safety zone does not exist if $\lim_{s\to+\infty} sa(s) = +\infty$, as the system converges to a *basin of attraction*.

We summarize these results in Table 1, giving criteria depending on $\lim_{s\to 0} sa(s)$ and $\lim_{s\to+\infty} sa(s)$.

TABLE 1
Four different configurations determined by $\lim_{s\to 0} sa(s)$ *and* $\lim_{s\to+\infty} sa(s)$.

	$s \to 0$	$s \to +\infty$
$sa(s) \to 0$	There exists a *collapse prevention* control strategy	There exists a *safety region* far from consensus
$sa(s) \to +\infty$	There exists a *black hole* (no strategy can avoid consensus for certain initial configurations)	There exists a *basin of attraction* (no safety zone far from consensus)

The paper is divided into three parts. In the first part, we consider a microscopic description of the system adapted from the HK opinion formation model. Second, we study the kinetic version of this model by taking the mean-field limit of the system, and we adequately extend the results of the microscopic description to the kinetic setting. Last, we provide numerical simulations illustrating the four cases presented in Table 1.

1. Microscopic model (finite dimension).

1.1. Generalized entropy functional for declustering control. Consider the general class of first-order differential systems:

$$\dot{x}_i = f_i(x), \quad i \in \{1, \dots, N\}, \tag{1.1}$$

where for each $i \in \{1, \dots, N\}$, $x_i(t) \in \mathbb{R}^d$. The dynamics are given by the functions $f_i \in C^1((\mathbb{R}^d)^n, \mathbb{R}^d)$.

The purpose of this work is to study the collective behavior of the system, focusing on patterns such as *consensus* and *clustering*. More specifically, we aim to design feedback control strategies to prevent the system from reaching those states. We first provide the general definitions that will be used hereafter.

DEFINITION 1.1. *The state characterized by $x_1 = \cdots = x_N$ is referred to as* consensus. *We denote by $\mathcal{M}_{\rm c}$ the* consensus manifold *defined by*

$$\mathcal{M}_{\rm c} := \{(x_i)_{i \in \{1,\dots,N\}} \mid \forall (j,k) \in \{1, \dots, N\}^2,\ x_j = x_k\}. \tag{1.2}$$

Remark 1.2. If the dynamics satisfy $f_i(x) = 0$ for every $i \in \{1, \dots, N\}$ and for all $x \in \mathcal{M}_{\rm c}$, the consensus state is an equilibrium.

Notice that if at least two agents have different states, for instance, if $x_i \neq x_j$ for some $i, j \in \{1, \dots, N\}^2$, then the system is not in consensus.

DEFINITION 1.3. *The system is said to* avoid consensus *if there exist $(i, j) \in \{1, \dots, N\}^2$ such that $x_i \neq x_j$.*

However, avoiding consensus might still leave the system in the critical state where several agents have the same state variable, which might be unwanted in some real-life situations. For instance, if each x_i represents an investor's decision, consensus might lead to a market crash. Whether or not the system is exactly in consensus state has little impact on the outcome: if one investor thinks differently than the mass (e.g., $x_j \neq x_1 = \cdots = x_{j-1} = x_{j+1} = \cdots = x_N$), it might not be enough to prevent a market collapse. With such applications in mind, we define clustering as follows.

DEFINITION 1.4. *The system is said to be* clustered *if there exist $(i, j) \in \{1, \dots, N\}^2$ such that $x_i = x_j$. We denote by $\mathcal{S}_{\rm cl}$ the* clustering set*, defined by*

$$\mathcal{S}_{\rm cl} := \{(x_i)_{i \in \{1,\dots,N\}} \mid \exists (j,k) \in \{1, \dots, N\}^2 \ \textit{s.t.}\ x_j = x_k\}. \tag{1.3}$$

Since we will focus on the avoidance of clustering, we characterize it as follows.

DEFINITION 1.5. *We say that the system is* fully declustered *if there exists $\epsilon > 0$ such that for all $(i, j) \in \{1, \dots, N\}^2$, $\|x_i - x_j\| \geqslant \epsilon$.*

Notice that the condition of avoiding consensus is weaker than the condition of declustering. The system is said to avoid consensus if it is not in a neighborhood of the consensus manifold. More constraining, the condition of declustering is satisfied if and only if the system is outside of a neighborhood of a larger manifold, which we refer to as the clustering set.

The consensus manifold $\mathcal{M}_c$ is thus contained in the clustering set $\mathcal{S}_{cl}$. More specifically, $\mathcal{M}_c$ is a d-dimensional manifold embedded in $(\mathbb{R}^d)^N$, while $\mathcal{S}_{cl}$ is a stratified set in the sense of Whitney. We recall that a set $E \subset \mathbb{R}^n$ is called *stratified* in the sense of Whitney if there exists a countable (locally finite) collection of pairwise disjoint manifolds $(\mathcal{M}_i)_{i\in\mathbb{N}}$ such that

1. $\mathcal{M}_i$ is an embedded manifold of dimension d_i,
2. if $\mathcal{M}_i \cap \partial\mathcal{M}_j \neq \emptyset$, then $\mathcal{M}_i \subset \partial\mathcal{M}_j$ and $d_i < d_j$.

Reminders on consensus achievement. Controlling a group of agents to steer it to consensus has been considered in the literature (see [10, 11, 12, 13, 40, 41, 50]). One common approach consists of modifying system (1.1) to include an additive control u:

$$\dot{x}_i = f_i(x) + u_i, \qquad i \in \{1, \dots, N\}. \tag{1.4}$$

Given $M > 0$, we impose the constraint $u \in U_M$, by defining the set of controls U_M as

$$U_M := \left\{ u : \mathbb{R}^+ \to (\mathbb{R}^d)^N \;\middle|\; u \text{ measurable, } \sum_{i=1}^N \|u_i(t)\| \leqslant M \text{ for a.e. } t \in \mathbb{R}^+ \right\}, \tag{1.5}$$

where $\|\cdot\|$ is the ℓ_2^d-Euclidean norm on $\mathbb{R}^d$. The condition $\sum_{i=1}^N \|u_i(\cdot)\| \leqslant M$, known as the $\ell_1^N - \ell_2^d$-norm constraint, promotes the componentwise sparsity of the control [10].

Previous works (see [10, 11, 40]) have proposed to construct feedback controls by minimizing the variance of the system $V : \mathbb{R}^{dN} \to \mathbb{R}$, with

$$V(x) = \frac{1}{2N^2} \sum_{i=1}^N \sum_{j=i+1}^N \|x_i - x_j\|^2 \quad \text{for all } x \in \mathbb{R}^{dN}. \tag{1.6}$$

It is easy to show that the variance characterizes the state of consensus.

Lemma 1.6. *Let $(x_i)_{i\in\{1,\dots,N\}} \in (\mathbb{R}^d)^N$, and let V be defined by* (1.6). *We say that the system $(x_i(t))_{i\in\{1,\dots,N\}}$ is in the state of consensus if and only if $V(x(t)) = 0$.*

Hence one can choose to design a feedback control strategy by minimizing the time derivative of the variance. Along any trajectory, it satisfies $\dot{V}(x(t)) = \frac{1}{2N^2}\sum_{i=1}^N \sum_{j\neq i} \langle x_i - x_j, \dot{x}_i - \dot{x}_j\rangle = \frac{1}{N}\sum_{i=1}^N \langle (x_i - \bar{x}), f_i(x) + u_i\rangle$. As done in [10] for the Cucker–Smale alignment model, one can easily define a feedback control u^c minimizing the derivative of the variance by setting $i_V := \operatorname{argmax}_{i\in\{1,\dots,N\}} \|x_i - \bar{x}\|$ so that

$$u^c_{i_V} = -M \frac{x_{i_V} - \bar{x}}{\|x_{i_V} - \bar{x}\|}\,, \text{ and } \quad u^c_i = 0 \text{ for all } i \neq i_V.$$

Searching for a functional to promote declustering. On the other hand, the purpose of this work is to design a control strategy leading the system away from clustering. Our first approach is to design a control strategy *maximizing* the time derivative of the variance (as opposed to minimizing it when controlling to consensus).

Proposition 1.7. *Let $M > 0$ and let $u \in U_M$. For all $i \in \{1, \dots, N\}$, let $R_i := x_i - \bar{x}$, where $\bar{x} := \frac{1}{N}\sum_{j=1}^N x_j$. Let $I_V := \{i \in \{1, \dots, N\} \mid$ for all $k \in \{1, \dots, N\},\ \|R_i\| \geqslant \|R_k\|\}$. Let $|I_V|$ denote the cardinality of I_V. The control u^V defined by*

$$u^V_i = \frac{M}{|I_V|}\frac{R_i}{\|R_i\|} \;\; \text{for } i \in I_V \quad \text{and} \quad u^V_i = 0 \;\; \text{for all } i \notin I_V \tag{1.7}$$

maximizes $\dot{V}$ instantaneously.

Proof. We study the time evolution of the variance. Along any trajectory, we have

$$\begin{aligned}\dot{V} &= \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j\neq i}\langle x_i - x_j, f_i(x) - f_j(x) + u_i - u_j\rangle \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\langle x_i - x_j, f_i(x) + u_i\rangle = \frac{1}{N^2}\sum_{i=1}^{N}\left\langle \sum_{j=1}^{N}(x_i - x_j), f_i(x) + u_i\right\rangle \\ &= \frac{1}{N}\sum_{i=1}^{N}\langle (x_i - \bar{x}), f_i(x) + u_i\rangle.\end{aligned}$$

Hence, $\dot{V}$ is maximized at all times by the control given by (1.7). ∎

Remark 1.8. The control defined in Proposition 1.7 requires computing $\operatorname{argmax}_{i\in\{1,\ldots,N\}}\|R_i\|$, which might not be unique. If there exists a unique index i maximizing $\|R_i\|$, i.e., $|I_V| = 1$, the control (1.7) acts only on i and is then sparse.

Notice that maximizing the variance V will only ensure that the system is far from the consensus manifold, and it does not guarantee declustering. Indeed, V can be very large even if almost all agents are concentrated at one point, as long as one agent is far from the group. This calls for a different functional, able to characterize the state of clustering like the variance characterizes consensus. A natural candidate for that purpose is the entropy functional $W \in C^1(\mathbb{R}^d \setminus \mathcal{S}_{\mathrm{cl}}, \mathbb{R})$, defined as follows:

$$W(x) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=i+1}^{N}\ln\|x_i - x_j\|. \tag{1.8}$$

Indeed, if the system is not in the clustering set, the entropy is bounded from below. However, the converse is not true, as we show in the following.

LEMMA 1.9. *Let $W \in C^1(\mathbb{R}^d\setminus\mathcal{S}_{\mathrm{cl}}, \mathbb{R})$ defined by* (1.8)*. If for all $(i,j)\in\{1,\ldots,N\}^2$, $\|x_i - x_j\| \geqslant \epsilon$ for some $\epsilon > 0$, then $W(x)$ is bounded below, i.e., there exists $K(\epsilon)\in\mathbb{R}$ such that $W(x) > K(\epsilon)$. However, the converse does not hold.*

For conciseness, we refer the reader to Appendix A for the proof. Lemma 1.9 shows that the functional W cannot characterize the boundedness away from the clustering set. This is due to the fact that the logarithm is unbounded both at zero and at infinity, which allows for the contribution of small pairwise distances to be compensated by that of large pairwise distances in (1.8).

A good entropy functional for declustering. The main idea then is to modify the entropy functional by replacing the logarithm by a function g bounded at infinity, in order to characterize clustering.

DEFINITION 1.10. *Let $g \in C^1(\mathbb{R}^+)$ be a strictly increasing function such that $\lim_{s\to 0} g(s) = -\infty$ and $\lim_{s\to+\infty} g(s) < \infty$. We define the generalized entropy functional W_g for system* (1.11) *by*

$$W_g(t) = \frac{1}{2N^2}\sum_{i=1}^{N}\sum_{j=i+1}^{N} g(\|x_i(t) - x_j(t)\|^2).$$

The advantage of defining such an entropy functional is that we are able to characterize completely the dispersion of the system.

THEOREM 1.11. *Let W_g be an entropy functional as defined in Definition* 1.10. *The following two statements are equivalent:*

1. *There exists $\eta > 0$ such that for all $t > 0$, $W_g(t) > \eta$.*
2. *There exists $\varepsilon > 0$ such that for all $t > 0$, for all $(i,j) \in \{1,\dots,N\}^2$, $\|x_i(t) - x_j(t)\| > \varepsilon$.*

If the conditions above are satisfied, the system is declustered at all times.

Proof. Let $c \in \mathbb{R}$. Suppose that for all $(i,j) \in \{1,\dots,N\}^2$, $\|x_i - x_j\| \geqslant c$. Then $W_g = \frac{1}{2N^2}\sum_{i=1}^N \sum_{j=i+1}^N g(\|x_i - x_j\|^2) \geqslant \frac{1}{2N^2}\frac{N(N-1)}{2} g(c^2)$.

Conversely, let $K \in \mathbb{R}$. Suppose that $W_g \geqslant K$. Let $m := \sup\{g(s),\ s > 0\}$. For all $(i,j) \in \{1,\dots,N\}^2$, $g(\|x_i - x_j\|^2) \leqslant m$. Let $(k,l) \in \{1,\dots,N\}^2$ with $k < l$. Notice that the assumptions on g given in Definition 1.10 imply that g is invertible. Let $g^{-1} : (-\infty, m) \to (0, +\infty)$ denote the inverse of g. Since $W_g = \frac{1}{2N^2}\sum_{i<j} g(\|x_i - x_j\|^2)$, we write $\|x_k - x_l\|^2 = g^{-1}(2N^2 W_g - \sum_{i<j,(i,j)\neq(k,l)} g(\|x_i - x_j\|^2))$. Since g^{-1} is an increasing function, we obtain $\|x_k - x_l\| \geqslant [g^{-1}(2N^2 K - (\frac{N(N-1)}{2} - 1)m)]^{1/2}$ and the result follows. ∎

From Theorem 1.11, maximizing W_g will ensure that the system is declustered, hence that it is far from the clustering set. We design a control strategy to keep the system in a declustered state, by maximizing $\dot{W}_g$ instantaneously.

PROPOSITION 1.12. *Let $M > 0$ and let $u \in U_M$. For all $i \in \{1,\dots,N\}$, let $S_i := \frac{1}{N}\sum_{j\neq i} g'(\|x_i - x_j\|^2)(x_i - x_j)$. Let $I_W := \{i \in \{1,\dots,N\} \mid$ for all $k \in \{1,\dots,N\},\ \|S_i\| \geqslant \|S_k\|\}$, and let $|I_W|$ denote the cardinality of I_W. The control u^W defined by*

$$u_i^W = \frac{M}{|I_W|}\frac{S_i}{\|S_i\|} \ \text{ for } i \in I_W \quad \text{ and } \quad u_i^W = 0 \ \text{ for all } i \notin I_W \tag{1.9}$$

maximizes the time derivative of the generalized entropy, $\dot{W}_g$, instantaneously.

Proof. Let us start by computing $\dot{W}_g(t)$. Since $\dot{x}_i - \dot{x}_j = f_i(v) - f_j(v) + u_i - u_j$, we get

$$\begin{aligned}\dot{W}_g &= \frac{1}{N^2}\sum_{i\neq j} g'(\|x_i - x_j\|^2)\langle x_i - x_j, f_i(x) + u_i\rangle \\ &= \frac{1}{N}\sum_{i=1}^N \left\langle \frac{1}{N}\sum_{j\neq i} g'(\|x_i - x_j\|^2)(x_i - x_j), f_i(x) + u_i \right\rangle.\end{aligned}$$

Let $S_i := \frac{1}{N}\sum_{j\neq i} g'(\|x_i - x_j\|^2)(x_i - x_j)$. Let $I_W := \{i \in \{1,\dots,N\} \mid \text{for all } k \in \{1,\dots,N\},\ \|S_i\| \geqslant \|S_k\|\}$. Then the control strategy (1.9) maximizes $\dot{W}_g$ at all times. ∎

Figure 1 illustrates the two control strategies designed in Propositions 1.7 and 1.12 on a group of 10 agents evolving in $\mathbb{R}^2$. The two control strategies are in general sparse, meaning that the control acts on only one agent at a given time (if $|I_V| = 1$, respectively if $|I_W| = 1$). However, they differ in fundamental ways. In order to maximize the variance V, one must act on the agent furthest away from the center of mass of the group, as shown in Proposition 1.7. On the other hand, to maximize the general entropy W_g, one must act on an agent which is very close to at least another agent. Not only do the different control strategies act on different sets of agents, but

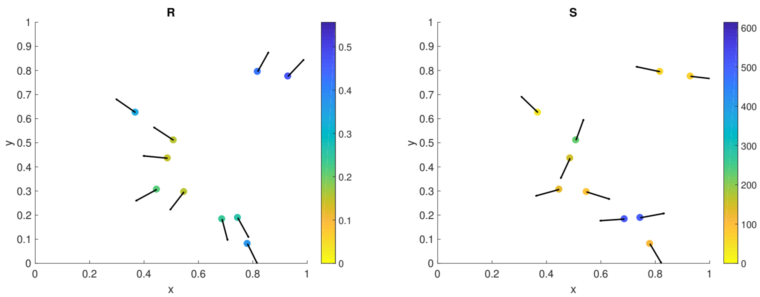

FIG. 1. *Random distribution of $N = 10$ agents in $\mathbb{R}^2$. Each agent i is colored by the value of $\|R_i\|$ (left) or $\|S_i\|$ (right), computed with $g : s \to -\frac{1}{s}$. The control strategies to minimize $\dot{V}$ and $\dot{W}_g$ consist of acting on the agent with the largest $\|R_i\|$ (respectively, $\|S_i\|$) and to drive the agent in the direction $\frac{R_i}{\|R_i\|}$ (respectively, $\frac{S_i}{\|S_i\|}$), shown in the figure by the arrows pointing from each agent.*

they also act to pull the agents in different directions. The control u_V acts on the agent furthest from the center of the group and will act to move it even further away, thus potentially acting always on the same agent. On the contrary, the control u_W acts to separate pairs of agents, hence it will potentially switch its focus once a pair has been separated enough (see Figure 1).

In particular these sparse control strategies apply to the HK system where the dynamics are given by $f : x \mapsto \frac{1}{N}\sum_{i \neq j} a(\|x_i - x_j\|)(x_j - x_i)$.

1.2. Controlling the system away from consensus and clustering. We now choose to focus our study on the well-known HK first-order consensus model:

$$\dot{x}_i = \frac{1}{N}\sum_{j \neq i} a(\|x_i - x_j\|)(x_j - x_i), \quad i \in \{1, \dots, N\}. \tag{1.10}$$

The HK model (1.10) was designed in the context of opinion dynamics and captures collective behavior such as consensus or clustering [27]. In the original "bounded confidence" model, each agent aligns its position to an average of all neighbors within a predetermined range. Here, we generalize this idea by considering that each agent x_i aligns its position to a weighted average of all other agents' positions, depending on the interaction function $a : \mathbb{R}^+ \to \mathbb{R}^+$. The HK model can be recovered in the special case of a being a step function $s \mapsto a(s) = \mathbb{1}_{s \leqslant r}$.

Let $M > 0$ and $u \in U_M$ (see (1.5)). We define the controlled evolution of the system as follows:

$$\dot{x}_i = \frac{1}{N}\sum_{j \neq i} a(\|x_i - x_j\|)(x_j - x_i) + u_i, \quad i \in \{1, \dots, N\}. \tag{1.11}$$

Let $\phi_{a,u} : (\mathbb{R}^d)^n \times \mathbb{R} \to (\mathbb{R}^d)^n$ be the flow associated with the differential equation (1.11), i.e., for all $x_0 \in (\mathbb{R}^d)^n$ for all $t \in \mathbb{R}^+$, $\phi_{a,u}(x_0, t)$ is the unique solution of (1.11) with initial condition $x(0) = x_0$.

The problem of defining the solution of (1.10) when agents collide was treated in [14, Remark 2.10]. In what follows, we allow the interaction function $a(\cdot)$ to be unbounded near zero. Less restrictive even, we allow the following: $\lim_{s \to 0} sa(s) = +\infty$. This causes the right-hand side of (1.10) to be undefined when two agents cluster. However, we can define the solution up to the time of the first clustering $\bar{t}$.

We prove that the limit of the solution of (1.10) when approaching $\bar{t}$ is unique. This will allow us to extend the solution in order to give a meaning to the system after a time of clustering.

LEMMA 1.13. *Let x denote the solution of system* (1.10)*, and let $\bar{t}$ be the first time at which a cluster occurs, i.e., for all $t < \bar{t}$, for all $(i,j) \in \{1,\ldots,N\}^2$, $x_i(t) \neq x_j(t)$, and there exist $(k,l) \in \{1,\ldots,N\}^2$ with $k \neq l$ such that $\lim_{t\to\bar{t}} \|x_k(t) - x_l(t)\| = 0$. Then there exists $x^L \in (\mathbb{R}^d)^N$ such that $\lim_{t\to\bar{t}} x(t) = x^L$.*

We refer the reader to Appendix A for a detailed proof. Since (1.10) may be undefined when two or more agents cluster, we impose $a(0) = 0$. Lemma 1.13 implies that the solution can be extended after each time of clustering. The condition $a(0) = 0$ implies that once two agents collide, they stay clustered (for the system without control).

1.2.1. Black hole. In section 1.1, we designed a control strategy in the general case of system (1.4). We now study the more specific first-order consensus model (1.11). In this section, we prove that for certain interaction functions $a(\cdot)$, there exists a *black hole*, i.e., given a certain bound M on the control (with $\sum_{i=1}^N \|u_i\| \leqslant M$), for certain initial conditions, it is impossible to avoid convergence to consensus whatever the control may be. This is a manifestation of the black swan phenomenon.

DEFINITION 1.14. *Let $M > 0$. We define the black hole region as follows:*

$$\mathcal{R}_{\mathrm{BH}}^M = \{x_0 \in (\mathbb{R}^d)^n \mid \textit{for} \quad \textit{all } u \in U_M,\ \exists T > 0,\ V(\phi_{a,u}(x_0,T)) = 0\}.$$

THEOREM 1.15. *Let a be an attraction potential such that $\lim_{s\to 0} sa(s) = +\infty$. Then for all $M > 0$, there exists $\epsilon > 0$ such that if for all $(i,j) \in \{1,\ldots,N\}^2$, $\|x_i(0) - x_j(0)\| < \epsilon$, then given any control $u \in U_M$, the system converges to consensus in finite time. In other words, for any $M > 0$, there exists $\mathcal{R}_{\mathrm{BH}}^M$ such that $\mathcal{M}_{\mathrm{c}} \subsetneq \mathcal{R}_{\mathrm{BH}}^M$.*

Proof. We study the evolution of the variance $V(t)$ Along any trajectory of x, $\dot{V}$ satisfies

$$\begin{aligned}\dot{V} = \frac{1}{2N^2}\sum_{i=1}^N \sum_{j\neq i} \Bigg\langle & x_i - x_j, \frac{1}{N}\sum_{k\neq i} a(\|x_i - x_k\|)(x_k - x_i) \\ & -\frac{1}{N}\sum_{k\neq j} a(\|x_j - x_k\|)(x_k - x_j) + u_i - u_j \Bigg\rangle.\end{aligned} \tag{1.12}$$

After calculation, the uncontrolled part of $\dot{V}$ can be written as $-\frac{1}{N^2}\sum_{i=1}^N \sum_{j=i+1}^N a(\|x_i - x_j\|)\|x_i - x_j\|^2$. Let $M > 0$. Since $\lim_{s\to 0} sa(s) = +\infty$, for all $A > 0$, there exists $\epsilon > 0$ such that for all $s < \epsilon$, $a(s) \geqslant \frac{A}{s}$. Near consensus, that is when for all i and j, $\|x_i(t) - x_j(t)\| \leqslant \epsilon$:

$$\dot{V} \leqslant -\frac{1}{N^2} A \sum_{i=1}^N \sum_{j=i+1}^N \|x_i - x_j\| + \frac{M}{N^2}\sum_{i=1}^N \sum_{j=i+1}^N \|x_i - x_j\|. \tag{1.13}$$

In particular, let $A = 2M$ and let $\epsilon > 0$ such that for all $s < \epsilon$, $a(s) \geqslant \frac{A}{s}$. Notice that $V \geqslant \frac{1}{2N^2}\max_{i,j} \|x_i - x_j\|^2$. Suppose that $V(0) = \frac{\epsilon^2}{2N^2}$. Then for all $(i,j) \in \{1,\ldots N\}^2$, $\|x_i(0) - x_j(0)\| \leqslant \sqrt{2N^2 V(0)} = \epsilon$. Then while $\|x_i - x_j\| \leqslant \epsilon$, $\dot{V} \leqslant -\frac{M}{N^2}\sum_{i\leqslant j} \|x_i - x_j\|$. By equivalence of the norms, $\dot{V} \leqslant -\frac{M}{N^2}\sqrt{\sum\sum_{i\leqslant j} \|x_i - x_j\|^2} = -\frac{M}{N^2}\sqrt{2N^2 V} = -\frac{\sqrt{2}M}{N}\sqrt{V}$. Then V decreases, which ensures that the condition $\|x_i - x_j\| \leqslant \epsilon$ holds. Hence V tends to 0 in finite time. ∎

Remark 1.16. The condition $\lim_{s\to 0} sa(s) = +\infty$ does not generalize to the integral condition given in [10, 23]: $\int_0^{s_0} a(s)ds = +\infty$. Take, for instance, $a(s) = \frac{1}{s}$. Then $\int_0^{s_0} a(s)ds = +\infty$, but $\lim_{s\to 0} sa(s) = 1$. Indeed, going back to the proof above, the derivative of the variance satisfies $\dot{V} = -\frac{1}{N^2}\sum_{i<j}\|x_i - x_j\| + \frac{1}{N^2}\sum_{i<j}\langle x_i - x_j, u_i - u_j\rangle \leqslant \frac{-1+M}{N^2}\sum_{i<j}\|x_i - x_j\|$. If $M < 1$, then convergence to consensus is unavoidable, but for bigger values of M the possibility of acting on the system to prevent consensus remains.

Remark 1.17. We can generalize Theorem 1.15 for functions $s \mapsto sa(s)$ that are bounded below for small values of s. We can prove that the existence of a black hole depends on the value of the bound M on the control, unlike in the case of Theorem 1.15, where a black hole exists no matter how strong the control is allowed to be. Similar generalizations hold for the results of sections 1.2.2, 1.2.3, and 1.2.4.

Theorem 1.15 shows that if the interaction between agents is very strong when they are close to each other (as characterized by the condition $\lim_{s\to 0} sa(s) = +\infty$), then for every bound M on the control, there exists a zone close to the consensus manifold such that no control in U_M can prevent consensus. We call this phenomenon the *black hole*. We now look at the behavior of the system far from the clustering set, that is, when each pair of agents is sufficiently separated. We show in sections 1.2.2 and 1.2.3 that depending on the strength of the decrease of a near infinity, there may or may not exist a *safety region* far from the consensus manifold, that is, a stable zone (given appropriate control).

1.2.2. Safety region. Here we give sufficient conditions on the potential for the existence of a *safety region*. Given a bound M on the control, there exist initial conditions such that the control can always keep the system away far from clustering.

DEFINITION 1.18. *Let* $M > 0$. *Construct* W_g *as in Definition* 1.10. *We define the safety region as follows:*

$$\mathcal{R}_{\mathrm{S}}^M = \{x_0 \in (\mathbb{R}^d)^n \mid \exists u \in U_M,\ \exists K \in \mathbb{R},\ \textit{for}\ \ \textit{all}\ t \geqslant 0,\ W_g(\phi_{a,u}(x_0,t)) \geqslant K\}.$$

THEOREM 1.19. *Let* a *be an attraction potential such that* $\lim_{s\to+\infty} sa(s) = 0$. *Then for all bound* $M > 0$ *on the control, there exists a safety region* $\mathcal{R}_{\mathrm{S}}^M \neq \emptyset$. *Furthermore, confinement to the safety region can be obtained with the sparse control* $u^W \in U_M$ *given in* (1.9).

Proof. From Proposition 1.12, $\dot{W}_g$ is maximized instantaneously by the sparse control (1.9), and we have

$$\max_u \dot{W}_g = \frac{1}{N}\sum_{i=1}^N \left\langle S_i, \frac{1}{N}\sum_{k=1}^N a(\|x_i - x_k\|)(x_k - x_i)\right\rangle + \frac{M}{N}\|S_{i_W}\|. \tag{1.14}$$

Let $\epsilon < \frac{M}{N}$. Since $\lim_{s\to+\infty} sa(s) = 0$, there exists $\mu > 0$ such that if for all i, j, $\|x_i - x_j\| \geqslant \mu$, then $\frac{1}{N}\sum_{k=1}^N a(\|x_i - x_k\|)\|x_i - x_k\| \leqslant \epsilon$. Suppose that at $t = 0$, the initial conditions give $W_g(0) \geqslant \frac{1}{2N^2}g(\mu^2) + \frac{m}{2N^2}(\frac{N(N-1)}{2} - 1)$. Then from the proof of Theorem 1.11, for all $(i,j) \in \{1,\ldots,N\}^2$, $\|x_i(0) - x_j(0)\|^2 \geqslant g^{-1}(g(\mu^2) + m(\frac{N(N-1)}{2} - 1) - m(\frac{N(N-1)}{2} - 1)) = \mu^2$. Consequently, $\max_u \dot{W}_g \geqslant \|S_{i_W}\|(\frac{M}{N} - \epsilon) \geqslant 0$. Then, choosing the control u^W given by (1.9) that maximizes $\dot{W}_g$ at all times, we ensure that for all $t > 0$, $W_g(t) \geqslant W_g(0)$, so that for all i, j, $\|x_i(t) - x_j(t)\| \geqslant \mu$. Furthermore, u^W is sparse. $\square$

This first theorem covers a wide range of interaction potentials. Given that the interaction potential $a(\cdot)$ decreases enough at infinity, we ensure the existence of a safety zone far from the clustering set. This, for instance, applies to potentials $a(\cdot)$ with compact support. However, notice that the interaction potential $a(s) = \frac{1}{s}$ does not meet the required conditions of Theorem 1.19. As stated in Remark 1.17, we can generalize the results of Theorem 1.19 to functions $s \mapsto sa(s)$ bounded above when $s \to +\infty$, although we do not go into details for conciseness.

Black hole horizon. In sections 1.2.1 and 1.2.2, we showed the existence of a *black hole* in a neighborhood of the consensus manifold if $\lim_{s\to 0} sa(s) = +\infty$ and the existence of a *safety region* far from the clustering set if $\lim_{s\to+\infty} sa(s) = 0$. This suggests the existence of a "horizon" between safety and attraction to the black hole for interaction potentials that meet both conditions. The question remains of clarifying this horizon.

DEFINITION 1.20. *We define the* black hole horizon $\mathcal{H}^M_{\mathrm{BH}}$ *as the subset of* $(\mathbb{R}^d)^N$ *given by* $\mathcal{H}^M_{\mathrm{BH}} := (\mathbb{R}^d)^N \setminus (\mathcal{R}^M_{\mathrm{BH}} \cup \mathcal{R}^M_{\mathrm{S}})$. *If there is no* safety region *and* $\mathcal{R}^M_{\mathrm{BH}} = (\mathbb{R}^d)^N$, *we say that the* black hole horizon *is* infinite.

If $\mathcal{H}^M_{\mathrm{BH}} = \emptyset$ *while* $\mathcal{R}^M_{\mathrm{BH}} \neq \emptyset$ *and* $\mathcal{R}^M_{\mathrm{S}} \neq \emptyset$, *then the state space* $(\mathbb{R}^d)^N$ *is divided between the black hole and the safety region, and we say that the* black hole horizon *is* sharp.

The schematic of Figure 2 illustrates the black hole horizon enclosed between the safety region and the black hole.

If the attraction potential does not satisfy the hypotheses of Theorem 1.19, we cannot ensure the existence of a *safety region*. In fact, we show that in certain cases the *safety region* does not exist and the whole space is a *black hole*, i.e., the *black hole horizon* is infinite.

LEMMA 1.21. *If* $a(s) = 1 + \frac{1}{s^2}$, *there exists* $M > 0$ *such that the* black hole horizon *is infinite.*

Proof. Let $M \leqslant \frac{\alpha}{\sqrt{2}}$ for some $\alpha < 1$. Notice that with the interaction function $a(s) = 1 + \frac{1}{s^2}$, we have from (1.12) $\dot{V} = -2V - \frac{N-1}{2N} + \frac{1}{N^2}\sum_{i<j}\langle x_i - x_j, u_i - u_j\rangle$. The controlled part of $\dot{V}$ is related to V by equivalence of the norms:

$$\begin{aligned}\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\langle x_i - x_j, u_i - u_j\rangle &\leqslant M\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\|x_i - x_j\| \\ &\leqslant M\frac{1}{N^2}N\sqrt{\sum_{i=1}^N\sum_{j=1}^N\|x_i - x_j\|^2} = M\sqrt{2V}.\end{aligned}$$

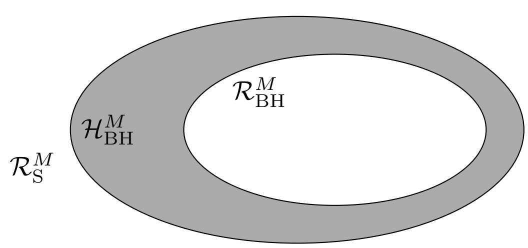

FIG. 2. *Schematic representation of the blach hole* $\mathcal{R}^M_{\mathrm{BH}}$, *the safety region* $\mathcal{R}^M_{\mathrm{S}}$, *and the black hole horizon* $\mathcal{H}^M_{\mathrm{BH}}$.

First assume that initially $\sqrt{V(0)} \leqslant \frac{\sqrt{2}}{2}M$ (i.e., pairwise distances are small relative to M). Since $a(s) \geqslant \frac{1}{s^2}$, while $\sqrt{V} \leqslant \frac{\sqrt{2}}{2}M$ we have $\dot{V} \leqslant -\frac{1}{4}+\frac{M}{2}\sqrt{2V} \leqslant -\frac{1}{4}+M^2 \leqslant \frac{\alpha-1}{4}$, so V converges to 0 in finite time.

Now suppose that $\sqrt{V(0)} > \frac{\sqrt{2}}{2}M$, so the initial conformation is far from the consensus manifold. While this condition is satisfied, since $a(s) \geqslant 1$, we write $\dot{V} \leqslant -2V + \frac{M}{2}\sqrt{2V} = \sqrt{V}(-2\sqrt{V} + \frac{M}{2}\sqrt{2}) \leqslant -\sqrt{2V}\frac{M}{2} \leqslant -\frac{M^2}{2}$. So V decreases and reaches $\sqrt{V} = \sqrt{2}M$ in finite time. Then we are brought back to the first case. ∎

1.2.3. Basin of attraction. In Theorem 1.19, we saw that if $a(\cdot)$ decreases fast enough to 0 at infinity, then there exists a "safety" zone near infinity (i.e., when the agents are far from each other). Here we show that this safety zone does not always exist.

THEOREM 1.22. *Let $M > 0$. If $\lim_{s\to+\infty} sa(s) = +\infty$, then there exists a real number $\mu > 0$ and a set $B := \{(x_i)_{i\in\{1,\dots,N\}} \mid \min_{i\neq j} \|x_i - x_j\| \leqslant \mu\}$ such that for any control $u \in U_M$ and any initial condition $x(0)$, there exists a sequence $(t_n)_{n\in\mathbb{N}}$ with $\lim_{n\to+\infty} t_n = +\infty$ such that $x(t_n) \in B$. We call B the* basin of attraction.

Proof. Let $A > M$. There exists $\mu > 0$ such that if $s > \mu$, then $sa(s) > A$. Let $B := \{(x_i)_{i\in\{1,\dots,N\}} \in \mathbb{R}^{dN} \mid \min_{i\neq j} \|x_i - x_j\| \leqslant \mu\}$ and $B^c := \mathbb{R}^{dN} \setminus B = \{(x_i)_{i\in\{1,\dots,N\}} \in \mathbb{R}^{dN} \mid \text{for all } (i,j) \in \{1,\dots,N\}^2,\ i \neq j \implies \|x_i - x_j\| > \mu\}$. Suppose that there exists $T \geqslant 0$ such that for all time $t \geqslant T$, $x(t) \in B^c$. Then the variance V decreases as a quadratic function of time. Indeed, for all $t \geqslant T$,

$$\begin{aligned}\dot{V} &= -\frac{1}{N^2}\sum_{i<j} a(\|x_i - x_j\|)\|x_i - x_j\|^2 + \frac{1}{N^2}\sum_{i<j}\langle x_i - x_j, u_i - u_j\rangle \\ &\leqslant \frac{M-A}{N^2}\sum_{i<j}\|x_i - x_j\| \leqslant \frac{\sqrt{2}(M-A)}{N}\sqrt{V}.\end{aligned}$$

Then there exists $\tau \geqslant T$ such that $V(\tau) \leqslant \frac{N-1}{4N}\mu^2$. This contradicts the statement $x(\tau) \in B^c$, as it would imply $V(\tau) = \frac{1}{2N^2}\sum_{i=1}^N \sum_{j=i+1}^N \|x_i - x_j\|^2 > \frac{N-1}{4N}\mu^2$. Hence for all $T > 0$, there exists $t \geqslant T$ such that $x(t) \in B$.

Remark 1.23. Theorem 1.22 does not ensure that $x(t)$ stays in the basin of attraction B for all time $t > T$. When $x \in B$, it might become possible to act on the system again. If the control allows us to obtain again $x \in B^c$, again V becomes strictly decreasing until $x \in B$.

1.2.4. Collapse prevention. We saw in section 1.2.1 that if $\lim_{s\to 0} sa(s) = +\infty$, there exists a *black hole* in which no control allows us to avoid clustering. On the other hand, we will show that if $\lim_{s\to 0} sa(s) = 0$, then consensus can always be avoided, in particular with the sparse control u^V strategy defined in section 1.1.

THEOREM 1.24. *Suppose that $\lim_{s\to 0} sa(s) = 0$. Let $M > 0$. Let $R_i := x_i - \bar{x}$, and define $i_V := \operatorname{argmax}_{i\in\{1,\dots,N\}} \|R_i\|$. Then the sparse control strategy u^V defined by*

$$u_i^V = M\frac{R_i}{\|R_i\|} \quad \text{for } i = i_V\ ; \quad u_i^V = 0 \quad \text{for all } i \neq i_V \tag{1.15}$$

prevents consensus. More specifically, there exists $\delta > 0$ such that if $V(0) \leqslant \delta$, then $V(t) > V(0)$ for all $t > 0$.

Proof. Let $M > 0$. Let $\epsilon = \frac{M}{4(N-1)}$. Since $\lim_{s\to 0} sa(s) = 0$, there exists $\eta > 0$ such that if for $i, j \in \{1, \dots, N\}$, $\|x_i - x_j\| \leqslant \eta$, then $a(\|x_i - x_j\|)\|x_i - x_j\| \leqslant \epsilon$. Suppose that $V(0) \leqslant \delta := \frac{\eta^2}{2N^2}$. Since $V = \frac{1}{2N^2}\sum_{i=1}^N \sum_{j=i+1}^N \|x_i(t) - x_j(t)\|^2$, we have $\max_{(i,j)\in\{1,\dots,N\}^2} \|x_i(0) - x_j(0)\| \leqslant \sqrt{2\delta}\, N = \eta$. Then using the control u^V, we compute

$$
\begin{aligned}
\dot V &= -\frac{1}{N^2}\sum_{i=1}^N \sum_{j=i+1}^N a(\|x_i - x_j\|)\|x_i - x_j\|^2 + \frac{1}{2N^2}\sum_{j\neq i_V}\left\langle x_{i_V} - x_j, M\frac{x_{i_V} - \bar x}{\|x_{i_V} - \bar x\|}\right\rangle \\
&\geqslant -\frac{1}{N^2}\epsilon \sum_{i=1}^N \sum_{j=i+1}^N \|x_i - x_j\| + \frac{1}{2N^2} MN\|x_{i_V} - \bar x\| \\
&\geqslant -\frac{1}{N^2}\epsilon\frac{N(N-1)}{2}\eta + \frac{1}{2N}M\|x_{i_V} - \bar x\|.
\end{aligned}
$$

By definition of u^V and i_V, the distance $\|x_{i_V} - \bar x\|$ represents the maximal distance between an agent and the average position of the group. Hence, $\eta \leqslant 2\|x_{i_V} - \bar x\|$. Then, $\dot V \geqslant -\frac{N-1}{2N}\epsilon\eta + \frac{M}{2N}\frac{\eta}{2} = \frac{\eta}{2N}(-\epsilon(N-1) + \frac{M}{2}) \geqslant \frac{\eta M}{8N} > 0$ by definition of ϵ. Hence $\dot V > 0$, which ensures that $V(t) > V(0)$ for all $t > 0$. ∎

Theorem 1.24 states that no matter the given strength of the control M, if initially the system is not at consensus, it can be kept from converging to consensus. We extend this result with a more general theorem that shows that if the system starts away from the clustering set, then it can be controlled to remain bounded away from the clustering set.

THEOREM 1.25. *Let $M > 0$. Suppose that $\lim_{s\to 0} sa(s) = 0$. If for every $(i, j) \in \{1, \dots, N\}^2$, $\|x_i - x_j\| > 0$, there exists a sparse feedback control strategy that prevents clustering. More specifically, let $\epsilon < \frac{M}{N}$ and $\delta > 0$ such that $sa(s) \leqslant \epsilon$ for all $s \leqslant \delta$. Consider a function $g \in C^1((0, +\infty))$ satisfying the conditions of Definition 1.10, with the additional assumptions $m := \sup g(s) > 0$ and $g(\delta^2) = 0$. We define the control u^δ by*

$$
u_i^\delta = M\frac{S_i^\delta}{\|S_i^\delta\|} \text{ for } i = i_\delta \quad \text{and} \quad u_i^\delta = 0 \text{ for all } i \neq i_\delta, \tag{1.16}
$$

where $S_i^\delta := \sum_{j\neq i,\ \|x_i - x_j\|\geqslant\delta} g'(\|x_i - x_j\|^2)\langle x_i - x_j\rangle$ and $i_\delta := \arg\max_i \|S_i^\delta\|$. Then the solution of (1.11) with control u^δ satisfies $\|x_i(t) - x_j(t)\| \geqslant \kappa := (g^{-1}(2N^2W_g^\delta(0) - (\frac{N(N-1)}{2} - 1)m))^{1/2}$ for all $t \geqslant 0$, and $i \neq j$.

The proof can be found in Appendix A.

2. Macroscopic model (kinetic equation). The second part of this paper focuses on the kinetic limit of system (1.4) and establishes the kinetic version of the results presented in section 1. We refer the reader to Appendix C for a reminder on the mean-field limit of the microscopic model (1.11) when the number of agents tends to infinity.

For every time $t \geqslant 0$, let $\mu(t, \cdot) \in \mathcal{P}(\mathbb{R}^d)$ be a probability measure representing the density of agents at time t. We suppose that we are allowed to control a part of the state space denoted by $\omega \subset \mathbb{R}^d$. Denoting by χ_ω the characteristic function of ω, we write the control as $\chi_\omega u$, with $u : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ representing the control strength.

Given M, $c > 0$, We set the following constraints on ω and u:

$$\begin{cases} u \in \mathcal{U}_M := \{u : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d \text{ measurable } \mid \|u(t,\cdot)\|_{L^\infty(\mathbb{R}^d)} \leqslant M\}, \\ \omega \in \Omega_c := \{\omega \subset \mathbb{R}^d \mid \omega \text{ measurable, of measure } |\omega| \leqslant c\}, \end{cases} \tag{2.1}$$

where $|\omega| = \int_\omega dx$ is the Lebesgue measure of ω. As in [41], the condition $|\omega| \leqslant c$ allows us to extend the idea of sparse control to the kinetic setting. Instead of acting on a single agent as in the discrete case, we limit the size of the state space region that the control can act on. Another possibility, not explored in this paper, would be to limit the mass of agents that can be controlled, with a condition such as $\int_\omega d\mu_x \leqslant c < 1$, as in [2, 41]. Let ξ denote the interaction kernel. In the case of the HK interaction, $\xi[\mu](x) = \int_{\mathbb{R}^d} a(\|x-y\|)(y-x)d\mu(y)$. Then the kinetic version of (1.4) can be written as

$$\partial_t \mu + \operatorname{div}((\xi[\mu] + \chi_\omega u)\mu) = 0. \tag{2.2}$$

The well-posedness of the kinetic equation (2.2) has been established in the case where $s \mapsto sa(s)$ and $x \mapsto \chi_{\omega(t)} u(t,x)$ are Lipschitz functions (see Appendix C and [41]). We denote equivalently by $\mu(t)$ and $\mu(t,\cdot)$ its solution at time t. Here, we want to allow possible blow-ups of the interaction function a, which in turn causes finite-time blow-up of initially regular solutions. The existence and uniqueness of weak measure solutions for kinetic equations with aggregation phenomena were studied in a number of papers, for various classes of interaction potentials; see, for instance, [14], where well-posedness was proven in the case of semiconvex potentials, and [6], where the authors establish an L^p theory for the multidimensional aggregation equation (and see all the references within).

2.1. Kinetic generalized entropy functional.

2.1.1. Controlling the kinetic system. From here onward we will suppose that μ is of compact support, i.e., $\mu \in \mathcal{P}_c(\mathbb{R}^d)$. We will define the state of kinetic clustering by the presence of one or more point masses in the measure μ. Then our control will aim to avoid the formation of point masses. In particular, if the initial measure is absolutely continuous, we aim to maintain it absolutely continuous. More precisely, let us begin by defining consensus in the context of the kinetic formulation. Let δ_x denote the Dirac mass centered at x.

DEFINITION 2.1. *For any $x_0 \in \mathbb{R}^d$, the state $\mu = \delta_{x_0}$ is referred to as* consensus.

As in the discrete case, we define the variance of the system which characterizes consensus:

$$\mathcal{V}(\mu(t,\cdot)) = \iint \|x-y\|^2 d\mu_x(t) d\mu_y(t),$$

where the indices x and y indicate the integration variable.

LEMMA 2.2. *$\mathcal{V}(\mu) = 0$ if and only if $\mu = \delta_{x_0}$ for some $x_0 \in \mathbb{R}^d$.*

Going further, if the variance is small, we can indeed ensure that the system is concentrated around its center of mass in the following sense.

LEMMA 2.3. *For all $r > 0$, $\iint_{\|x-y\|\geqslant r} d\mu_x d\mu_y \leqslant \frac{1}{r^2}\mathcal{V}(\mu)$.*

Proof. Observe that $\mathcal{V} \geqslant \iint_{\|x-y\|\geqslant r} \|x-y\|^2 d\mu_x d\mu_y \geqslant r^2 \iint_{\|x-y\|\geqslant r} d\mu_x d\mu_y$ and the claim follows. ∎

Then it is clear that maximizing $\mathcal{V}$ ensures that consensus is avoided. Like in section 1.1, we design a feedback control strategy maximizing the time derivative of the variance.

PROPOSITION 2.4. *Let $\bar{x}(t) := \int_{\mathbb{R}^d} x d\mu_x(t)$ denote the center of mass of $\mu(t,\cdot) \in \mathcal{P}_c(\mathbb{R}^d)$. Let $\mathcal{R}(x,t) := \int (x-y) d\mu_y(t) = x - \bar{x}(t)$. The control $\chi_\omega u$ such that $\omega(t) := \text{argmax}_{\omega' \in \Omega_c} \int_{\omega'} \|\mathcal{R}(x,t)\| d\mu_x(t)$ and $u(x,t) = M \frac{\mathcal{R}(x,t)}{\|\mathcal{R}(x,t)\|}$ for all $x \in \omega \backslash \{\bar{x}\}$ maximizes $\dot{V}$ instantaneously with the constraints* (2.1).

Proof. We calculate the time derivative of the kinetic entropy:

$$\begin{aligned}\frac{d}{dt}\mathcal{V}(\mu(t,\cdot)) &= 2\iint \|x-y\|^2 d\left(-\text{div}_x((\xi[\mu](x,t) + \chi_{\omega(t)} u(t,x))\mu_x(t)\right) d\mu_y(t) \\ &= 4\iint (x-y)\cdot \xi[\mu](x,t) d\mu_x(t) d\mu_y(t) + 4\int (x - \bar{x}(t)) \cdot \chi_{\omega(t)} u(t,x) d\mu_x(t).\end{aligned}$$

Hence the control given above maximizes $\dot{\mathcal{V}}$, given that for all t, the functional $F_t : \tilde{\Omega}_c \to \mathbb{R}^+$ defined by $F_t(\gamma) = \int_{\mathbb{R}^d} \|x - \bar{x}(t)\| \gamma(x) d\mu_x(t)$ admits a maximum on $\tilde{\Omega}_c := \{\gamma \in L^\infty(\mathbb{R}^d, \{0,1\}) \mid \int_{\mathbb{R}^d} \gamma(x) dx \leqslant c\}$, the set of all characteristic functions or measurable subsets of measure less than c. As done in [42], let us relax the problem to the convex closure (in the L^∞ weak-$*$ topology) of $\tilde{\Omega}_c$, defined as $\bar{\text{co}}(\tilde{\Omega}_c) = \{\gamma \in L^\infty(\mathbb{R}^d, [0,1]) \mid \int_{\mathbb{R}^d} \gamma(x) dx \leqslant c\}$. The functional $F_t : \bar{\text{co}}(\tilde{\Omega}_c) \mapsto \mathbb{R}^+$ is continuous for the weak-$*$ topology, is a linear functional defined on a closed convex set $\bar{\text{co}}(\tilde{\Omega}_c)$, compact for the weak-$*$ topology. Hence it attains its maximum on the extremal points of $\bar{\text{co}}(\tilde{\Omega}_c)$. By definition of $\bar{\text{co}}(\tilde{\Omega}_c)$, these extremal points are precisely the elements of $\tilde{\Omega}_c$. This implies that for all t, there exists a (not necessarily unique) function $\gamma_t \in \tilde{\Omega}_c$ such that $\gamma_t = \text{argmax}_{\gamma' \in \tilde{\Omega}_c} F_t(\gamma')$, and there exists a set $\omega(t)$ such that $\omega(t) = \text{argmax}_{\omega' \in \Omega_c} \int_{\omega'} \|x - \bar{x}(t)\| d\mu_x(t) = \{x \in \mathbb{R}^d \mid \gamma(x) = 1\}$. ∎

As for the finite-dimensional model of section 1, we aim to prevent not only convergence of the system to consensus but also any clustering. We define kinetic clustering as follows.

DEFINITION 2.5. *If $\mu \in \mathcal{P}(\mathbb{R}^d)$ contains at least one point mass, we say that μ is in clustered state.*

Remark 2.6. In Definition 2.5, the number of Dirac masses contained in μ represents the number of clusters. The absolutely continuous part of μ represents the nonclustered agents. Notice that consensus is a special case of clustering, with μ exactly equal to a point mass.

As in the discrete case, the strict positivity of $\mathcal{V}$ is not enough to ensure avoidance of clustering. Indeed, let $\mu = \frac{1}{2}(\delta_{x_1} + \delta_{x_2})$, with $(x_1, x_2) \in (\mathbb{R}^d)^2$, $x_1 \neq x_2$. Then μ is in a clustered state and yet $\mathcal{V}(\mu) = \frac{1}{2}\|x_1 - x_2\|^2 > 0$. As in section 1.1, we define the kinetic generalized entropy: given a function $g \in C^1(\mathbb{R}^d)$ satisfying the hypotheses of Definition 1.10,

$$\mathcal{W}_g(\mu(t,\cdot)) = \iint g(\|x-y\|^2) d\mu_x(t) d\mu_y(t).$$

The kinetic counterpart of Theorem 1.11 shows that maximizing the kinetic generalized entropy prevents clustering.

LEMMA 2.7. *Suppose that $\mathcal{W}_g(\mu) \geqslant K \in \mathbb{R}$ and $\sup_{s>0} g(s) = m$. Then for all r small enough,*

$$\iint_{\|x-y\|\leqslant r} d\mu_x d\mu_y \leqslant \frac{m-K}{-g(r^2)}. \tag{2.3}$$

Proof. We have

$$\mathcal{W}_g = \iint_{\|x-y\|\leqslant r} g(\|x-y\|^2)d\mu_x d\mu_y + \iint_{\|x-y\|>r} g(\|x-y\|^2)d\mu_x d\mu_y$$
$$\leqslant g(r^2)\iint_{\|x-y\|\leqslant r} d\mu_x d\mu_y + m.$$

For r small enough, $g(r^2) < 0$, so dividing by $g(r^2)$ yields (2.3). ◻

Lemma 2.7 implies that if $\mathcal{W}_g > K > 0$, $\lim_{r\to 0}\iint_{\|x-y\|\leqslant r} d\mu_x d\mu_y = 0$. Hence μ cannot be in a clustered state. Indeed, if $\mu = \alpha\delta_{x_0} + \tilde{\mu}$, we have $\iint_{\|x-y\|=0} d\mu_x d\mu_y \geqslant \iint_{\|x-y\|=0} \alpha^2 d\delta_{x_0} d\delta_{x_0} = \alpha^2 > 0$.

As in section 1.1, we design a control strategy maximizing $\dot{\mathcal{W}}_g$, in order to steer the system away from clustering.

PROPOSITION 2.8. *Consider $\mu \in \mathcal{P}_c(\mathbb{R}^d)$ and let $\mathcal{S}(x,t) = \int g'(\|x-y\|^2)(x-y)d\mu(t,y)$. Let $M, c > 0$ and let u and ω satisfying the conditions* (2.1). *The control $\chi_\omega u$ such that*

$$\omega(t) := \underset{\omega\in\Omega_c}{\operatorname{argmax}} \int_\omega \mathcal{S}(x,t)d\mu(t,x); \quad u(x,t) = M\frac{\mathcal{S}(x,t)}{\|\mathcal{S}(x,t)\|} \tag{2.4}$$

maximizes $\dot{\mathcal{W}}_g$ instantaneously.

Proof. The time derivative of $\mathcal{W}_g$ can be computed as

$$\frac{d}{dt}\mathcal{W}_g(t) = 2\iint g(\|x-y\|^2)d(-\mathrm{div}((\xi[\mu](x,t) + \chi_\omega(t)u(x,t))\mu_x(t)))d\mu_y(t)$$
$$= 4\int\left(\int g'(\|x-y\|^2)(x-y)d\mu_y(t)\right)\cdot(\xi[\mu](x,t) + \chi_\omega(t)u(x,t))d\mu_x(t).$$

Let $\mathcal{S}(x,t) := \int g'(\|x-y\|^2)(x-y)d\mu_y(t)$. Then $\dot{\mathcal{W}}_g(t) = 4\int_{\mathbb{R}^d}\mathcal{S}(x,t)\cdot X[\mu](x)d\mu_x(t) + 4\int_\omega \mathcal{S}(x,t)\cdot u\, d\mu_x(t)$. One can prove in the same way as in Proposition 2.4 that for all t, the functional $F_t' :\ \omega\in\Omega_c \mapsto \int_{\mathbb{R}^d}\mathcal{S}(x,t)\chi_w d\mu_x(t)$ has a maximum in Ω_c. Hence the derivative of $\mathcal{W}_g$ is maximized by the control $\chi_\omega u$ defined by (2.4). ◻

Remark 2.9. The controls designed in Propositions 2.4 and 2.8 do not satisfy the condition $x \mapsto \chi_\omega(t)u(t,x) \in \mathrm{Lip}(\mathbb{R}^d)$ that ensures existence and uniqueness of the solution to the kinetic equation (2.2), as shown in Proposition C.2. For this reason, in order to derive the specific results of Theorems 2.14, 2.17, and 2.18, we will regularize the controls.

2.1.2. Behavior of the kinetic HK system without control. Before studying the controlled system, we look at its behavior without control in two specific cases. First, we can prove that if $a(\cdot)$ is bounded below, then the system tends to consensus exponentially. Second, if $s\mapsto sa(s)$ is bounded at 0, and if μ_0 is of compact support, then the system tends to consensus in finite time.

THEOREM 2.10. *Let $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$ and μ satisfy the kinetic HK kinetic PDE*

$$\partial_t\mu + \mathrm{div}(\xi[\mu]\mu) = 0\ ; \quad \mu(0) = \mu_0, \tag{2.5}$$

where the convolution kernel is $\xi[\mu](x) = \int_{\mathbb{R}^d} a(\|x-y\|)(y-x)d\mu(y)$, for a continuous function $a \in C(\mathbb{R}^+,\mathbb{R}^+)$. If a is bounded away from zero, then $\mathcal{V}(\mu(t,\cdot))$ tends to zero exponentially.

Proof. Since $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$, $\mathcal{V}(\mu_0) \leqslant \sup_{(x,y)\in\mathrm{supp}(\mu_0)^2} \|x-y\|^2 < \infty$. We compute the time derivative of the kinetic variance:

$$\begin{aligned}\frac{d}{dt}\mathcal{V}(\mu) &= \iint \|x-y\|^2 d(-\mathrm{div}_x(\xi[\mu_x]\mu_x))d\mu_y + \iint \|x-y\|^2 d\mu_x d(-\mathrm{div}_y(\xi[\mu_y]\mu_y))\\ &= -2\iint \|x-y\|^2 a(\|x-y\|)d\mu_x d\mu_y.\end{aligned}$$

If there exists $C>0$ such that $a(s)>C$ for all $s\in\mathbb{R}^+$, then $\frac{d}{dt}\mathcal{V} \leqslant -2C\mathcal{V}$, and thus $\mathcal{V}(\mu(t,\cdot)) \leqslant \mathcal{V}(\mu_0)\exp(-2Ct)$. ∎

We now look at the behavior of measures with initially compact support, i.e., $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$. As in [41], we define the size X of the support of μ as the radius of the smallest ball centered at $\bar{x}$ and containing the support of μ. More precisely, given a time-evolving measure $\mu(t)\in\mathcal{P}_c(\mathbb{R}^d)$,

$$X(t) = \inf\{X \geqslant 0 \mid \mathrm{supp}(\mu(t)) \subseteq B(\bar{x}(t), X)\}. \tag{2.6}$$

THEOREM 2.11. *Let $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$ and suppose that $\mu(t)$ is the solution of* (2.5). *Suppose that $a \in \mathrm{Lip}(\mathbb{R}^+,\mathbb{R}^+)$ and that $s\mapsto a(s)$ is bounded below by $C>0$. Let $X(t)$ represent the size of the support of $\mu(t)$, as given by* (2.6). *Then X converges exponentially to zero.*

The proof of Theorem 2.11 can be found in Appendix B. With Theorems 2.10 and 2.11, we showed that under certain conditions on a, the system converges to consensus. We now examine under what conditions convergence to consensus can be avoided and, even further, under what conditions clustering can be avoided.

2.2. Control of the kinetic dynamics. We provide results equivalent to those of section 1, but in the case of the kinetic dynamics. We adapt the concepts of *black hole*, *safety region*, *basin of attraction*, and *collapse prevention* to the kinetic formulation. Let $\mu_0(x) = \mu(0,x)$ and $\mu(t) = \Phi_{\omega,u}(t)\#\mu_0$ denote the push-forward at time t of μ_0 by the controlled particle flow defined in Definition C.1 (see Appendix C).

2.2.1. Black hole. As in the discrete case, we start by defining the *black hole* region

$$\mathcal{R}^M_{\mathrm{BH}} = \{\mu_0 \in \mathcal{P}_c(\mathbb{R}^d) \mid \text{for all } u,\ \omega,\ \exists T>0,\ \mathcal{V}(\mu(T)) = 0\}.$$

We begin by pointing out the following.

LEMMA 2.12. *For all $r>0$, $\iint_{\|x-y\|\geqslant r} \|x-y\|d\mu_x d\mu_y \leqslant \frac{1}{r}\mathcal{V}(\mu)$.*

Proof. The proof follows the same argument as that of Lemma 2.3. ∎

THEOREM 2.13. *Suppose that $\lim_{s\to 0} sa(s) = +\infty$. Then for all $M>0$, there exists a* black hole region *$\mathcal{R}^M_{\mathrm{BH}} \neq \emptyset$. More specifically, for all $\mu_0 \in \mathcal{R}^M_{\mathrm{BH}}$, $\mathcal{V}(\mu(t))$ tends to* 0 *in finite time, and this for any controlled trajectory with $u\in\mathcal{U}_M$, $\omega\in\Omega_c$.*

Proof. Let $A = 2M$. There exists $r_0>0$ such that $sa(s)\geqslant A$ for all $s\leqslant r_0$. Suppose that $\mathcal{V}(0) \leqslant \frac{r_0^2}{4}$,

$$\begin{aligned}\dot{\mathcal{V}} &\leqslant -2\iint_{\|x-y\|\leqslant r_0} \|x-y\|^2 a(\|x-y\|)d\mu_x d\mu_y + M\iint \|x-y\|d\mu_x d\mu_y\\ &\leqslant -2A\iint \|x-y\|d\mu_x d\mu_y + 2A\iint_{\|x-y\|>r_0} \|x-y\|d\mu_x d\mu_y + M\iint \|x-y\|d\mu_x d\mu_y\\ &\leqslant (M-2A)\iint \|x-y\|d\mu_x d\mu_y + 2A\frac{\mathcal{V}}{r_0},\end{aligned}$$

where the last inequality is a consequence of Lemma 2.12. Notice that $\iint \|x - y\| d\mu_x d\mu_y \leqslant ((\iint \|x-y\|^2 d\mu_x d\mu_y)^{1/2} (\iint 1 d\mu_x d\mu_y)^{1/2} = \sqrt{\mathcal{V}}$. Then we have $\dot{\mathcal{V}} \leqslant (M - 2A + \frac{2A}{r_0}\sqrt{\mathcal{V}})\sqrt{\mathcal{V}}$. While $\sqrt{\mathcal{V}} \leqslant \frac{r_0}{2}$, $\dot{\mathcal{V}} \leqslant -M\sqrt{\mathcal{V}}$. Hence $\mathcal{V}$ decreases, which ensures that the condition $\sqrt{\mathcal{V}} \leqslant \frac{r_0}{2}$ holds. In conclusion, $\mathcal{V}$ tends to 0 in finite time. ◻

2.2.2. Safety region. The behavior of the interaction function at infinity determines the existence of either a *safety region* or a *basin of attraction*. For the discrete system, the *safety region* was defined by bounding below the smallest pairwise distance $\min_{i\neq j} \|x_i - x_j\|$. In the kinetic case, we replace this condition by requiring that the population density stays split into distinct measures concentrated around points that are far enough from one another. More precisely, when the interaction function $a(\cdot)$ decreases enough at infinity, we can prove the following.

THEOREM 2.14. *Suppose that* $\lim_{s\to+\infty} sa(s) = 0$ *and that* $s \mapsto sa(s) \in \mathrm{Lip}(\mathbb{R}^+, \mathbb{R}^+)$. *Let* $N \in \mathbb{N}$, $N \geqslant 2$, *and* $r > 0$ *small enough that* $N|B(0,r)| \leqslant c$. *Let* $\epsilon < M$ *and let* $R > 2r$ *be large enough that for all* $s \geqslant R-2r$, $sa(s) \leqslant \epsilon$. *Let* $(x_1, \dots, x_N) \in (\mathbb{R}^d)^N$ *be such that for all* $i \neq j$, $\|x_i - x_j\| \geqslant R$. *Let* $\delta \in (0, r)$.

If μ_0 *satisfies* $\mathrm{supp}(\mu_0) \subset \bigcup_{i=1}^N B(x_i, r-\delta)$, *then there exists a unique solution* $\mu \in C(\mathbb{R}^+, \mathcal{P}_c(\mathbb{R}^d))$ *to the controlled kinetic equation* (2.2), *with the control* $\chi_\omega u$ *defined by* $\omega := \bigcup_{i=1}^N B(x_i, r) \subset \Omega_c$, *and* $u \in \mathcal{U}_M$ *constant in time, such that for all* $t \in \mathbb{R}^+$, $u(t,\cdot) \in \mathrm{Lip}(\mathbb{R}^d)$ *and*

$$u(t,x) = \begin{cases} 0 & \text{for all } x \in \mathbb{R}^d \setminus \bigcup_{i=1}^N B(x_i, r), \\ -M\frac{x-x_i}{\|x-x_i\|} & \text{for all } x \in B(x_i, r-\delta) \setminus B(x_i, \frac{r-\delta}{2}), \text{ for all } i \in \{1,\dots,N\}. \end{cases} \tag{2.7}$$

Furthermore, for all $t \geqslant 0$, $\mathrm{supp}(\mu(t,\cdot)) \subset \bigcup_{i=1}^N B(x_i, r)$.

Proof. The existence and uniqueness of the solution to the controlled PDE (2.2)–(2.7) are a consequence of the Lipschitz properties of the interaction function a and the control $\chi_\omega u$, as a direct application of Proposition C.2. Suppose that $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$ and satisfies $\mathrm{supp}(\mu_0) \subset \bigcup_{i=1}^N B(x_i, r-\delta)$, with $(x_i)_{i\in\{1,\dots,N\}} \in (\mathbb{R}^d)^N$, N, r, and δ satisfying the conditions above. As in the proof of Theorem 2.11, we denote by $x(t,x_0)$ the particle trajectory of (2.5) with $x(0,x_0) = x_0$. For each $i \in \{1,\dots,N\}$, let $X_i(t) := \max\{\|x(t,x_0) - x_i\|\ ,\ x_0 \in B(x_i,r) \cap \mathrm{supp}(\mu_0)\}$. Let $K_t^i = \mathrm{argmax}\{\|x(t,x_0) - x_i\| \mid x_0 \in B(x_i,r) \cap \mathrm{supp}(\mu_0)\}$. Then for all $t \geqslant 0$, for all $x_0 \in K_t^i$, $X_i(t)^2 = \|x(t,x_0) - x_i\|^2$. From Danskin's theorem, $\frac{d}{dt}(X_i(t)^2) = \max_{x_0 \in K_t^i}\{\frac{d}{dt}\|x(t,x_0) - x_i\|^2\}$. We have

$$\begin{aligned} \frac{d}{dt}\|x(t,x_0) - x_i\|^2 = 2(x(t,x_0) - x_i) \cdot \Bigg(& \int_{B(x_i,r)} a(\|x(t,x_0) - y\|)(y - x(t,x_0)) d\mu_y \\ & + \int_{\mathbb{R}^d \setminus B(x_i,r)} a(\|x(t,x_0) - y\|)(y - x(t,x_0)) d\mu_y + \chi_\omega u(t, x(t,x_0)) \Bigg). \end{aligned}$$

Notice that for all $y \in B(x_i,r) \cap \mathrm{supp}(\mu_0)$, $(x(t,x_0) - x_i) \cdot (y - x(t,x_0)) \leqslant 0$. On the other hand, while $\mathrm{supp}(\mu(t,\cdot)) \subset \bigcup_{i=1}^N B(x_i,r)$, for all $y \in (\mathbb{R}^d \setminus B(x_i,r)) \cap \mathrm{supp}(\mu(t,\cdot))$, $\|x(t,x_0) - y\| \geqslant R - 2r$, so $a(\|x(t,x_0) - y\|)\|y - x(t,x_0)\| \leqslant \epsilon$. Then we can write $\frac{d}{dt}\|x(t,x_0) - x_i\|^2 \leqslant 2[\|x(t,x_0) - x_i\|\epsilon + (x(t,x_0) - x_i) \cdot \chi_\omega u(t, x(t,x_0))]$. We design the control u by (2.7) and the control set by $\omega := \bigcup_{i=1}^N B(x_i,r) \subset \Omega_c$.

With this control, if $x(t,x_0) \in B(x_i, r-\delta) \setminus B(x_i, \frac{r-\delta}{2})$, $\frac{d}{dt}(X_i(t)^2) = 2X_i(t)\dot{X}_i(t) \leqslant 2(X_i(t)\epsilon - MX_i(t))$, from which we get $\dot{X}_i(t) < 0$. Hence with the designed control, μ, satisfies $\mathrm{supp}(\mu(t,\cdot)) \subset \bigcup_{i=1}^N B(x_i, r)$ for all $t \geqslant 0$. ◻

Remark 2.15. Theorem 2.14 adapts the results obtained in the microscopic setting to the kinetic case, by ensuring that the population density stays confined to balls whose centers are far apart, which prevents consensus. However, this does not prevent the measure from converging to a clustering state, which can happen if the radii of the balls containing its support converge to zero.

2.2.3. Basin of attraction. We showed that there exists a *safety region* if a decreases fast enough at infinity. On the other hand, when $\lim_{s\to+\infty} sa(s) = +\infty$, no control can prevent the convergence of μ to an attractive region that we name *basin of attraction*. In the discrete case, the basin of attraction consists of all states in which at least one pairwise distance is small (see section 1.2.3). In the kinetic setting, the basin of attraction consists of measures with a large concentration around their center of mass.

THEOREM 2.16. *Suppose that* $\lim_{s\to+\infty} sa(s) = +\infty$. *Then there exists a diameter* $d > 0$, *a constant* $\delta \in (0, \frac{1}{2})$, *and a time* $T > 0$ *such that for any* $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$, *for any controlled trajectory* μ *of* (2.2)–(2.1), $\iint_{\|x-y\|\leqslant d} d\mu_x(T) d\mu_y(T) \geqslant \delta$.

Proof. Let $d > 0$ such that for all $s \geqslant d$, $sa(s) \geqslant M$. We reason by contradiction. Let $\delta \in (0, \frac{1}{2})$ and suppose that for all $t > 0$, $\iint_{\|x-y\|\leqslant d} d\mu_x(t) d\mu_y(t) < \delta$. The derivative of the variance was computed earlier and can be written as

$$
\begin{aligned}
\dot{\mathcal{V}} &\leqslant -2 \iint_{\|x-y\|>d} \|x-y\|^2 a(\|x-y\|) d\mu_x d\mu_y + M \iint \|x-y\| d\mu_x d\mu_y \\
&\leqslant -2M \iint_{\|x-y\|>d} \|x-y\| d\mu_x d\mu_y + M \iint_{\|x-y\|\geqslant d} \|x-y\| d\mu_x d\mu_y \\
&\quad + M \iint_{\|x-y\|<d} \|x-y\| d\mu_x d\mu_y \\
&\leqslant -Md \iint_{\|x-y\|>d} d\mu_x d\mu_y + Md \iint_{\|x-y\|\leqslant d} d\mu_x d\mu_y.
\end{aligned}
$$

We also have $-\iint_{\|x-y\|>d} d\mu_x d\mu_y < \delta - 1$. We obtain $\dot{\mathcal{V}} \leqslant Md(2\delta-1) < 0$ since $\delta < \frac{1}{2}$. Hence $\mathcal{V}$ converges to 0 in finite time, and the system reaches consensus, so there exists $T > 0$ such that $\iint_{\|x-y\|\leqslant d} d\mu_x(t) d\mu_y(t) = 1$. This contradicts the hypothesis. ◻

2.2.4. Collapse prevention. Last, if the interaction potential is not too big near the origin, we aim to prove that, as in the case of the discrete dynamics, there exists a control keeping the system away from consensus.

We first show that the control u constructed in Proposition 2.4 to maximize the time derivative of the variance can maintain the size of the support of $\mu(t)$ above a certain size.

THEOREM 2.17. *Suppose that* $a \in \mathrm{Lip}(\mathbb{R}^+, \mathbb{R}^+)$, *which implies that* $\lim_{s\to 0} sa(s)=0$. *Let* X *represent the size of* $\mathrm{supp}(\mu)$ *as defined by* (2.6). *Then there exists* $\eta > 0$, $\tau > 0$, *and a control* $\chi_\omega u$ *with* $(\omega, u) \in \Omega_c \times \mathcal{U}_M$ *such that there is a unique solution* μ *to* (2.2) *and for all* $t \geqslant \tau$, $X(t) \geqslant \eta$.

Proof. From the proof of Theorem 2.11 (see Appendix B), the evolution of X^2 is given by

$$\frac{d}{dt}(X(t)^2) = 2\max_{x_0\in K_t}\left\{(x(t,x_0)-\bar{x}(t))\cdot\left(\int a(\|x(t,x_0)-y\|)(y-x(t,x_0))d\mu(t,y)\right.\right.$$
$$\left.\left.+\ \chi_{w(t)}u(t,x(t,x_0))\right)\right\},$$

where $\bar{x}(t) := \int_{\mathbb{R}^d} x d\mu_x(t)$ and $K_t = \operatorname{argmax}\{\|x(t,x_0)-\bar{x}(t)\| \mid x_0 \in \operatorname{supp}(\mu_0)\}$. Let $\epsilon < M$, and let $r > 0$ be such that for all $s \leqslant r$, $sa(s) \leqslant \epsilon$. Let R_c be such that $|B(0,R_c)| \leqslant c$, and let $R < R_c$. Then we set $w(t) = B(\bar{x}(t),R) \in \Omega_c$. Suppose $X(t) \leqslant \min(\frac{r}{2},R)$. Then $\|x-y\| \leqslant r$ for all $(x,y) \in \operatorname{supp}(\mu)^2$. We then construct u such that for all $t\in\mathbb{R}^+$, $x \mapsto \chi_\omega(t)u(t,x) \in \operatorname{Lip}(\mathbb{R}^d)$ and for all $t \in \mathbb{R}^+$, $u(t,\cdot)$ satisfies

$$u(t,x(t,x_0)) = M\frac{x(t,x_0)-\bar{x}(t)}{\|x(t,x_0)-\bar{x}(t)\|} \text{ for all } x_0 \in K_t.$$

Then the well-posedness of (2.2) is given by Proposition C.2. Furthermore, $X(t)\dot{X}(t) \geqslant -\epsilon X(t) + (x(t,x_0)-\bar{x}(t))\cdot\chi_{w(t)}u(t,x(t,x_0)) = (M-\epsilon)X(t)$. Hence $\dot{X}(t) \geqslant M-\epsilon > 0$, so while $X \leqslant \min(\frac{r}{2},R)$, X increases. Consequently there exists τ such that for all $t \geqslant \tau$, $X(t) \geqslant \frac{1}{2}\min(\frac{r}{2},R)$. ∎

We proved that if $\lim_{s\to 0} sa(s) = 0$, there exists a control that keeps the support of μ from being too small. This implies that with this control, consensus cannot be reached in finite time. However, μ could still converge to a Dirac mass asymptotically.

In the finite-dimensional system, if $\lim_{s\to 0} sa(s) = 0$, we can find a control that maintains pairwise distances $\|x_i-x_j\|$ above a certain positive threshold (see Theorem 1.25). Similarly, in the kinetic system, we will prove that if initially the diameter of the support of μ_0 is small enough, and μ_0 is contained in nonoverlapping balls of small radii, then there exists a control that keeps μ_0 in its initial support, preventing consensus.

THEOREM 2.18. *Suppose that* $a \in \operatorname{Lip}(\mathbb{R}^+,\mathbb{R}^+)$, *which implies that* $\lim_{s\to 0} sa(s) = 0$. *Let* $N \in \mathbb{N}$, $N \geqslant 2$. *Let* $r_c > 0$ *small enough that* $N|B(0,r_c)| \leqslant c$, *and let* $r < r_c$. *Let* $R > 4r_c$ *be small enough that for all* $s \leqslant R$, $sa(s) \leqslant \epsilon < M$. *Let* $(x_1,\dots,x_N) \in (\mathbb{R}^d)^N$ *be such that for all* $i \neq j$, $2r_c \leqslant \|x_i - x_j\| \leqslant \frac{R}{2}$.
If $\operatorname{supp}(\mu_0) \subset \bigcup_{i=1}^N B(x_i,r)$, *then there exists a constant-in-time control* $(\omega,u) \in \Omega_c \times \mathcal{U}_M$ *constructed so that* $x \mapsto \chi_\omega u(x) \in \operatorname{Lip}(\mathbb{R}^d)$ *and*

$$\omega := \bigcup_{i=1}^N B(x_i,r_c); \quad u(x) = \begin{cases} -M\frac{x-x_i}{\|x-x_i\|} & \text{if } \|x-x_i\| = r, \\ 0 & \text{for all } x \in \mathbb{R}^d \setminus \bigcup_{i=1}^N B(x_i,r_c) \end{cases} \tag{2.8}$$

such that there is a unique solution to the controlled kinetic equation (2.2) *that satisfies for all* $t \in \mathbb{R}^+$, $\operatorname{supp}(\mu(t,\cdot)) \subset \bigcup_{i=1}^N B(x_i,r)$.

We refer the reader to Appendix B for the proof. Theorem 2.18 is a direct adaptation to the kinetic setting of Theorem 1.24, its discrete counterpart. Notice that in the discrete case, avoidance of clustering comes as a direct consequence of Theorem 1.24. In the kinetic case, Theorem 2.18 ensures that the measure stays confined to balls that are disjoint from one another. This only ensures the prevention of consensus, due to the fact that we define kinetic clustering as the presence of one or more Dirac masses (see Definitions 2.1 and 2.5). Notice, however, that here, the assumptions on $a(\cdot)$ are stronger than in the microscopic case: in order to define particle trajectories, we require $a \in \operatorname{Lip}(\mathbb{R}^+,\mathbb{R}^+)$, which would prevent a finite-time blow-up of the solution.

3. Simulations. We illustrate the results proven in the previous sections with numerical simulations, focusing on the finite-dimensional model.

3.1. Black hole and safety region. An interesting consequence of the results of sections 1.2.1 and 1.2.2 is the possible coexistence of two regions of the Nd-dimensional space of initial configurations of a *black hole* and a *safety region*. We define the *black hole* as the set of initial conditions for which for any control $u \in U_M$ the system tends to the clustering set in finite time. The *safety region* indicates the set of initial conditions for which there exists a control keeping the system away from the clustering set.

As an illustration of sections 1.2.1 and 1.2.2, we consider the interaction function given by $a : s \mapsto \frac{1}{s^2}$. Then indeed $sa(s) = \frac{1}{s}$, so that $\lim_{s\to 0} sa(s) = +\infty$ and $\lim_{s\to+\infty} sa(s) = 0$. This implies the existence of a *black hole* and of a *safety region*. We study the geometry of these regions. Let $g : s \to -\frac{1}{s}$ define the generalized entropy W_g. Then from the proof of Theorem 1.15, we know that if $V(0) \leqslant \frac{\epsilon^2}{2N^2}$, then the system converges to consensus in finite time, where ϵ is such that for any $A > M$, if $s \leqslant \epsilon$, then $sa(s) \geqslant A$. Let $\delta > 0$ arbitrarily close to 0. Then $\epsilon = \frac{1}{M+\delta}$ satisfies the condition. Hence the black hole region $\mathcal{R}^M_{\mathrm{BH}}$ satisfies $\{x \in \mathbb{R}^{dN} \mid \sum_{i<j} \|x_i - x_j\|^2 < \frac{1}{M^2}\} \subseteq \mathcal{R}^M_{\mathrm{BH}}$.

Similarly, from the proof of Theorem 1.19, we know that if $W_g(0) \geqslant \frac{1}{2N^2}(g(\mu^2) + m(\frac{N(N-1)}{2} - 1))$, then with the proper control, the system stays bounded away from the clustering set, where μ is such that for any $\epsilon < \frac{M}{N}$, if $s \geqslant \mu$, then $sa(s) \geqslant \epsilon$. Then for any $\delta > 0$ arbitrarily close to 0, let $\epsilon = \frac{M}{N} - \delta$. With the function $a : s \mapsto \frac{1}{s^2}$, the condition above is satisfied for $\mu = \frac{1}{\epsilon}$. Hence with the choice $g : s \mapsto -\frac{1}{s}$, we have $m = 0$ and the safety region $\mathcal{R}^M_{\mathrm{S}}$ satisfies: $\{x \in \mathbb{R}^{dN} \mid \sum_{i<j} \frac{1}{\|x_i - x_j\|^2} < \frac{M^2}{N^2}\} \subseteq \mathcal{R}^M_{\mathrm{S}}$.

Summarizing, we can compute explicitly:

- From the proof of Theorem 1.15, $x_0 \in \mathcal{R}^M_{\mathrm{BH}}$ if $V(0) < \frac{1}{2N^2M^2}$, i.e., if $M < \frac{1}{N\sqrt{2V(0)}}$.
- From the proof of Theorem 1.19, $x_0 \in \mathcal{R}^M_{\mathrm{S}}$ if $W_g(0) > -\frac{M^2}{2N^4}$, i.e., if $M > N^2\sqrt{-2W_g(0)}$.

Figure 3 shows the partition of the state space into the black hole region and the safety zone for the simple cases of 2 and 3 agents in $\mathbb{R}$. We illustrate the cases of black hole and safety zone in Figure 4, with $N = 10$ agents and $d = 1$, for the interaction

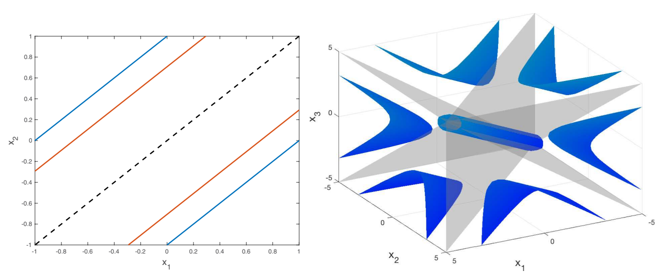

FIG. 3. *Partition of the state space into the black hole region and the safety region for $M = 1$, $a : s \mapsto \frac{1}{s^2}$ and $g : s \mapsto -\frac{1}{s}$. Left: with $(N, d) = (2, 1)$, the region enclosed by the red lines is a subset of $\mathcal{R}^M_{\mathrm{BH}}$ and the region located outside the blue lines is a subset of $\mathcal{R}^M_{\mathrm{S}}$. The dotted line represents the consensus manifold. Right: with $(N, d) = (3, 1)$, the region inside the central cylinder is a subset of $\mathcal{R}^M_{\mathrm{BH}}$ and the region outside the hyperbola branches is a subset of $\mathcal{R}^M_{\mathrm{S}}$. The gray planes represent the clustering set, and their intersection is the consensus manifold.*

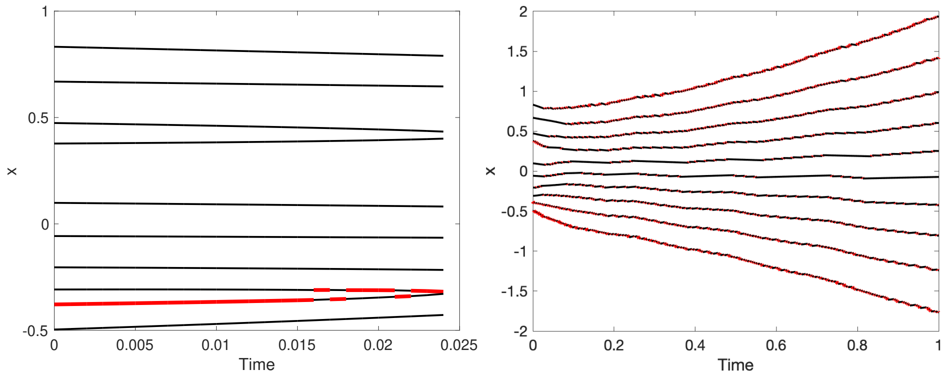

FIG. 4. *Time evolution of $N = 10$ agents' positions with the control strategy* (1.9) *respectively computed with $M = 1$ (left) and $M = 12$ (right), with the interaction function $a : s \mapsto s^{-2}$. Thick red lines correspond to controlled agents and thin black lines to uncontrolled ones. In the case $M = 1$, the system is initially in a black hole configuration: the control is unable to prevent clustering. The case $M = 12$ corresponds to a safety zone situation: the control successfully steers the system away from clustering.*

function $a : s \mapsto s^{-2}$. The agents' initial positions are selected randomly from the uniform distribution in $[0, 1]$. We compute the corresponding initial values of V and W_g (with the choice $g : s \mapsto -\frac{1}{s}$): $V(0) = 0.097$ and $W_g(0) = -4.18$. According to the computations above, if $M > 290$, the system is in a safety zone configuration, and if $M < 0.16$, the system is in a black hole configuration. Numerical simulations show that these computed thresholds are very conservative. We show simulation results for $M = 1$ and $M = 12$, with the control u_W maximizing the derivative of the generalized entropy given by (1.9). With $M = 1$ (Figure 4, left), the control u_W acts sparsely only on two agents and does not prevent convergence to clustering. The corresponding generalized entropy W_g, represented in Figure 6 (left), plunges to $-\infty$. On the other hand, with $M = 12$ (Figure 4, right), with the same initial conditions, the control manages to steer the system away from clustering. As in the case $M = 1$, it initially acts mostly on the penultimate agent but now succeeds in preventing its convergence to the neighboring agent. As time advances, the control gets distributed more evenly among all agents, eventually acting on all agents. The pairwise distances between agents are all increased as the control steers the system further still from the clustering set. As shown in Figure 6, the corresponding generalized entropy increases monotonically.

3.2. Basin of attraction and collapse prevention. To illustrate the cases of sections 1.2.3 and 1.2.4, we now consider the interaction function $a : s \mapsto \frac{1}{\sqrt{s}}$. Notice that a satisfies the condition for the existence of a *basin of attraction* ($\lim_{s\to+\infty} sa(s) = +\infty$) and the condition for the possibility of *collapse prevention* ($\lim_{s\to 0} sa(s) = 0$). We study a system with the same initial conditions as in section 3.1, for $N = 10$ and $d = 1$. We compute the control given in (1.9) respectively for $M = 1$ and $M = 12$. From the proofs of Theorems 1.22 and 1.25, we know the following:

- If $M = 1$ there exists $T > 0$ such that $x(T) \in B_1 := \{(x_i)_{i\in\{1,\dots,N\}} \mid \min_{i\neq j} \|x_i - x_j\| \leqslant 1\}$. If $M = 12$, there exists $T > 0$ such that $x(T) \in B_{12} := \{(x_i)_{i\in\{1,\dots,N\}} \mid \min_{i\neq j} \|x_i - x_j\| \leqslant 144\}$.
- There exists $\kappa > 0$ (depending on M) such that for all $t \geqslant 0$, for all $(i, j) \in \{1, \dots, N\}^2$,
 $\|x_i(t) - x_j(t)\| \geqslant \kappa$.

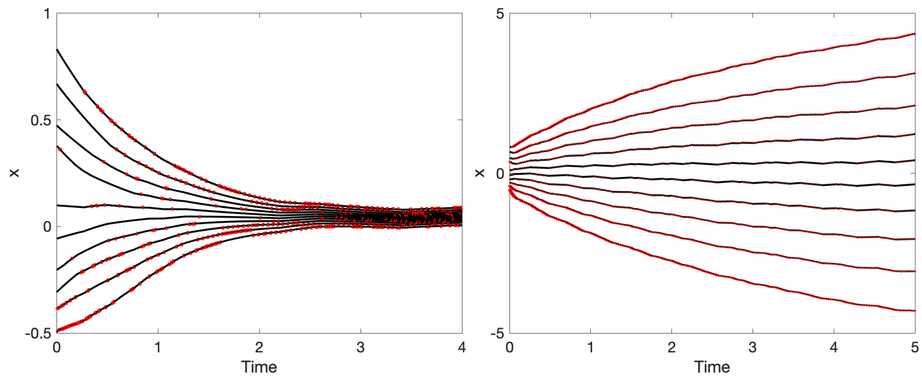

FIG. 5. *Time evolution of $N = 10$ agents' positions with the control strategy* (1.9) *respectively computed with $M = 1$ (left) and $M = 12$ (right), with the interaction function $a : s \mapsto s^{-1/2}$. Thick red lines correspond to controlled agents and thin black lines to uncontrolled ones. Left: the control u_W with the constraint $M = 1$ cannot prevent the convergence of the system to the basin of attraction. Right: the control u_W with the constraint $M = 12$ steers the system away from the clustering set.*

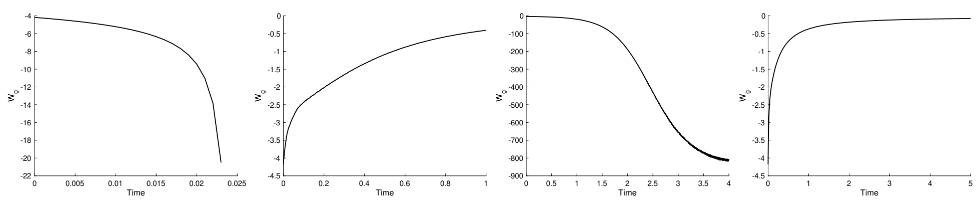

FIG. 6. *Evolution of the generalized entropy in the four cases shown in Figures* 4 *and* 5*, starting from the same initial conditions. From left to right: $a(s) = s^{-2}$ and $M = 1$; $a(s) = s^{-2}$ and $M = 12$; $a(s) = s^{-1/2}$ and $M = 1$; $a(s) = s^{-1/2}$ and $M = 12$.*

We show the behavior of the system with the interaction function $a : s \mapsto s^{1/2}$, starting with the same initial conditions as in Figure 4. Figure 5 (left) shows the evolution of the system controlled by $u_W \in U_M$ with $M = 1$. Notice that the basin of attraction appears to be much smaller than the one computed above, due to the very conservative estimates. The system is initially outside of the *basin of attraction* and converges to it. However, notice that the generalized entropy (third graph in Figure 6) does not tend to $-\infty$, as the system is able to prevent convergence to the clustering set. Figure 5 (right) shows the evolution of the same system, now controlled by $u_W \in U_M$ with $M = 12$. The control strategy defined in Theorem 1.25 is successful in preventing clustering.

Conclusion and further comments. The problem of controlling collective dynamics systems to achieve consensus or alignment has been frequently studied. In this paper, we focus on the opposite problem: controlling a system that naturally converges to consensus to achieve *declustering*. We first remark that the standard variance used to characterize consensus (with the equivalence $V = 0 \Leftrightarrow \forall (i,j) \in \{1, \cdots N\}^2,\ x_i = x_j$) does not measure declustering. Instead, to characterize the state of declustering, we introduce a generalized entropy functional W_g with the property $W_g > c \Leftrightarrow$ for all $(i,j) \in \{1, \cdots N\}^2,\ x_i \neq x_j$. With this tool in hand, we design a control strategy aiming to prevent the system from clustering by maximizing instantaneously the time derivative of W_g. The control thus constructed is *sparse*, meaning that it acts on only one agent at a time. As opposed to the problem of achieving consensus, here we fight against the system's natural tendency to form clusters. For this reason we do not expect to succeed in every situation. Indeed, the analysis of the first-order opinion formation model (1.11) with a positive interaction function $a(\cdot)$ and additive control reveals the existence of four regions of $(\mathbb{R}^d)^N$ that determine whether the system can be maintained away from consensus or clustering (see Table 1). The behavior of $a(\cdot)$ near zero determines the existence either of a black hole region or of the possibility to prevent collapse of the system. In the black hole region, no control can keep the system away from consensus, whereas in the case of collapse prevention there exist controls that can keep the system from clustering. The black hole region and the collapse prevention region can coexist with either a safety region or a basin of attraction, which are determined by the behavior of $a(\cdot)$ at infinity. In the safety region, the system can be kept far from the clustering set, given suitable initial conditions. On the other hand, in the case of a basin of attraction, the system is attracted to a neighborhood of the clustering set.

As seen in section 2, most results for the microscopic model can be extended to the kinetic equation (2.2). We define kinetic clustering as the presence of one or more Dirac masses in the population density, and we show that declustering can be characterized by the kinetic version of the generalized entropy. Similarly to the microscopic case, we design sparse control strategies maximizing the time derivative of the kinetic generalized entropy instantaneously. To extend the notion of sparse control to the PDE framework, we set a bound on the size of the support of the control. As in the microscopic setting, we show the existence of the four zones determined by the behavior of $a(\cdot)$ at zero (the black hole and the collapse prevention zones) and at infinity (the safety zone or the basin of attraction).

Last, in section 3, we present numerical simulations illustrating those four situations in the microscopic case, with two examples of interaction functions. With $a : s \mapsto s^{-2}$, we observe the coexistence of a black hole and of a safety region. Given the same initial conditions, the convergence to consensus or the avoidance of clustering are determined by the allowed strength of the control M. With the interaction function $a : s \mapsto s^{-1/2}$ and a fixed bound on the control $M = 1$, if it is initially far from the clustering set, the system converges to a basin of attraction. However, if the initial conditions are already in a neighborhood of the clustering set, we show that collapse to clustering can be avoided.

This work can be extended in many ways. We list a few of the possible future directions that stem naturally from the results presented above.

Sparse kinetic control. The crucial notion of *sparse* control in finite dimension can be extended to the kinetic setting in various ways. In this work, we made the choice of designing controls of the form $u\chi_\omega$, where we bound not only the L^∞ norm of u but also the size of the controlled region ω, with the condition $\int_\omega dx \leqslant c$ for some positive constant c. Another way to impose a notion of kinetic sparsity would be to bound the size of the controlled population, with the condition $\int_\omega d\mu(x) \leqslant c$, as done, for example, in [41].

Black hole horizon. The analysis of the model in section 1 and the numerical simulations in section 3 show the possible coexistence of a black hole and a safety region. One could go further in investigating the nature of the boundary between the two regions, and the existence of the *black hole horizon.* Two scenarios can be anticipated: either the black hole and the safety region form all of the state space, with $(\mathbb{R}^d)^N = \mathcal{R}^M_{\mathrm{BH}} \cup \mathcal{R}^M_{\mathrm{S}}$, or there exists a black hole horizon, $\mathcal{H}^M_{\mathrm{BH}} = (\mathbb{R}^d)^N \setminus (\mathcal{R}^M_{\mathrm{BH}} \cup \mathcal{R}^M_{\mathrm{S}})$.

Optimal control in finite dimension. The sparse controls designed to prevent the system from clustering minimize instantaneously the generalized entropy W_g. One could instead look for global minimizers of the functional W_g in order to design optimal control strategies.

Second-order clustering. Second-order models (like that of Cucker and Smale [18]) are commonly applied to animal groups to study coordinated collective behavior [2, 33]. Then the variable of interest is the velocity, and agreement of all agents' velocities is referred to as "alignment" [10, 11]. One could define clustering as agreement of several agents' velocities and conduct a similar control of the resulting model.

Well-posedness of the kinetic aggregation equation. The well-posedness of the kinetic equation (2.2) with aggregation phenomena is currently being investigated by several groups (see, for instance, [6] and [8] and references within). The existence and uniqueness of a solution to (2.2) for singular interaction potentials is a highly nontrivial problem. Intuitively, highly attractive interaction functions can give rise to a finite-time blow-up of the solution. This blow-up need not happen instantaneously and one could observe the coexistence of one or several Dirac masses with absolutely continuous parts of the measure in between them. A possible approach to define weak measure solutions would be through measure differential equations [39], which would allow one to prolong solutions past blow-up times. One could then define control in this new framework.

Appendix A. Proofs for the microscopic model. We provide proofs for some of the results stated in the previous sections. We start by proving that the usual entropy does not characterize clustering, as stated in Lemma 1.9.

Proof of Lemma 1.9. Let $\epsilon > 0$ and suppose that for all $(i, j) \in \{1, \dots, N\}^2$, $\|x_i - x_j\| \geqslant \epsilon$. Then

$$W(x) \geqslant \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \ln \epsilon = \frac{N(N-1)}{2N^2} \ln \epsilon.$$

We now disprove the converse. Let $K > 0$. Suppose that

$$W(x) \geqslant K \Rightarrow \exists \epsilon > 0 \text{ s.t. for all } (i, j) \in \{1, \dots, N\}^2 \text{ s.t. } i < j, \quad \|x_i - x_j\| \geqslant \epsilon, \tag{A.1}$$

where ϵ does not depend on x. Let $x \in (\mathbb{R}^d)^N$ such that $W(x) \geqslant K$. Let $\beta > 1$ and consider $\tilde{x} \in \mathbb{R}^d$ such that $\|\tilde{x}_k - \tilde{x}_l\| = \frac{1}{\beta}\|x_k - x_l\|$, $\|\tilde{x}_m - \tilde{x}_n\| = \beta\|x_m - x_n\|$, and $\|\tilde{x}_i - \tilde{x}_j\| = \|x_i - x_j\|$ for all $(i, j) \in \{1, \dots, N\}^2$ such that $i < j$, $(i, j) \neq (k, l)$ and $(i, j) \neq (m, n)$. Then

$$\begin{aligned} W(\tilde{x}) &= W(x) + \frac{1}{N^2}(-\ln(\|x_k - x_l\|) - \ln(\|x_m - x_n\|) + \ln(\|\tilde{x}_k - \tilde{x}_l\|) \\ &\qquad + \ln(\|\tilde{x}_m - \tilde{x}_n\|)) \\ &= W(x). \end{aligned}$$

Hence $W(\tilde{x}) \geqslant K$ and this result holds independently of β. Let $\beta = \frac{2\|x_k - x_l\|}{\epsilon}$. Then $\|\tilde{x}_k - \tilde{x}_l\| = \frac{\epsilon}{2}$, which contradicts (A.1). ∎

We then establish the result stated in Lemma 1.13 by proving that there exists a unique limit to the solution when approaching a time of clustering.

Proof of Lemma 1.13. Let $\bar{t}$ be the first time at which a cluster occurs, i.e., for all $t < \bar{t}$, for all $(i, j) \in \{1, \dots, N\}^2$, $x_i(t) \neq x_j(t)$, and there exists $(i, j) \in \{1, \dots, N\}^2$

such that $i \neq j$ and $\lim_{t\to\bar{t}} \|x_i(t) - x_j(t)\| = 0$. Let N_c denote the number of clusters formed at time $\bar{t}$, and let C_k denote the kth cluster, i.e., for all $k \in \{1, \dots, N_c\}$, for all $(i, j) \in C_k^2$, $\lim_{t\to\bar{t}} \|x_i(t) - x_j(t)\| = 0$, while for all $(k_1, k_2) \in \{1, \dots, N_c\}^2$ such that $k_1 \neq k_2$, for all $(i, j) \in C_{k_1} \times C_{k_2}$, for all $t \leqslant \bar{t}$, $\|x_i(\bar{t}) - x_j(\bar{t})\| > 0$. Let $|C_k|$ denote the cardinality of C_k, with $1 \leqslant |C_k| \leqslant N$ for all $k \in \{1, \dots, N_c\}$ (note that for simplicity of notation here we also refer to single agents as "clusters" of one element). Since $a(\cdot)$ takes values in $\mathbb{R}^+$, the system (1.10) is contractive and x stays bounded. Hence there exists a sequence $(t_n)_{n\in\mathbb{N}}$ converging to $\bar{t}$ such that $\lim_{n\to+\infty} x(t_n) = x^L$. Suppose that there exist two nondecreasing subsequences $(t_{2n})_{n\in\mathbb{N}}$ and $(t_{2n+1})_{n\in\mathbb{N}}$ converging to $\bar{t}$ such that $\lim_{n\to+\infty} x(t_{2n}) = \tilde{x}^L$ and $\lim_{n\to+\infty} x(t_{2n+1}) = \hat{x}^L$. For all $k \in \{1, \dots, N_c\}$, let $\overline{x}_k = \frac{1}{|C_k|}\sum_{i\in C_k} x_i$ denote the center of mass of the kth cluster. From (1.10), we have

$$\begin{aligned}\frac{d}{dt}\overline{x}_k &\leqslant \frac{1}{|C_k|}\sum_{i\in C_k}\dot{x}_i\\ &= \frac{1}{|C_k|}\sum_{i\in C_k}\left(\frac{1}{N}\sum_{j\in C_k} a(\|x_i - x_j\|)(x_j - x_i) + \frac{1}{N}\sum_{j\notin C_k} a(\|x_i - x_j\|)(x_j - x_i)\right)\\ &\leqslant \frac{1}{N|C_k|}\sum_{i\in C_k}\sum_{j\notin C_k} a(\|x_i - x_j\|)(x_j - x_i),\end{aligned}$$

where the first sum vanishes by antisymmetry of the summed coefficients. There exists $c > 0$ such that for all $i \in C_k$, for all $j \notin C_k$, $\|x_i - x_j\| \geqslant c$. Since x is bounded, there also exists $C > 0$ such that for all $j \in \{1, \dots, N\}$, $\|x_i - x_j\| \leqslant C$. This implies that there exists $A > 0$ such that for all $j \in \{1, \dots, N\}$, $a(\|x_i - x_j\|) \leqslant A$. Hence $\|\dot{\overline{x}}_k\| \leqslant \frac{|C_k|}{N} AC \leqslant AC$. Then $\|\overline{x}_k(t_{2n}) - \overline{x}_k(t_{2n+1})\| \leqslant AC|t_{2n} - t_{2n+1}| \underset{n\to+\infty}{\longrightarrow} 0$. Therefore, for all $i \in \{1, \dots, N\}$, $\tilde{x}_i^L = \hat{x}_i^L$. $\square$

Last, we prove Theorem 1.25, which shows that one can avoid clustering of the system if $\lim_{s\to 0} sa(s) = 0$.

Proof of Theorem 1.25. Let $\epsilon < \frac{M}{N}$. Since $\lim_{s\to 0} sa(s) = 0$, there exists $\eta > 0$ such that if $s \leqslant \eta$, then $sa(s) \leqslant \epsilon$. Let $\delta \geqslant \eta$. Let $g \in C^1((0, +\infty))$ satisfy the conditions of Definition 1.10 with $m := \sup g(s)$. Suppose also that $g(\delta^2) = 0$. We define the "danger" set as $\mathcal{D}^\delta(t) = \{(i, j) \in \{1, \dots, N\}^2, \mid i < j \text{ and } \|x_i(t) - x_j(t)\| \leqslant \delta\}$. We now define a partial generalized entropy using only the distances between pairs of agents belonging to $\mathcal{D}$: $W_g^\delta = \frac{1}{2N^2}\sum_{(i,j)\in\mathcal{D}^\delta} g(\|x_i - x_j\|^2)$. Notice that in the case where for all $(i, j) \in \{1, \dots, N\}^2$, $\|x_i - x_j\| \leqslant \delta$, the partial generalized entropy is equal to the generalized entropy: $W_g^\delta = W_g$. Furthermore, notice that Theorem 1.11 can be extended to W_g^δ: if $W_g^\delta \geqslant K$, then for all $(i, j) \in \{1, \dots, N\}^2$, $\|x_i - x_j\|^2 \geqslant g^{-1}(2N^2K - (\frac{N(N-1)}{2} - 1)m)$. We will prove that for all $t \geqslant 0$, for every $(i, j) \in \{1, \dots, N\}^2$, $\|x_i - x_j\|^2 \geqslant g^{-1}(2N^2 W_g^\delta(0) - (\frac{N(N-1)}{2} - 1)m)$. The derivative of the partial generalized entropy can be written as

$$\dot{W}_g^\delta = \frac{1}{N}\sum_{(i,j)\in\mathcal{D}^\delta} g'(\|x_i - x_j\|^2)\left\langle x_i - x_j, \frac{1}{N}\sum_{k\neq i} a(\|x_i - x_j\|)(x_i - x_j) + u_i\right\rangle.$$

Let $S_i^\delta := \sum_{j,\,(i,j)\in\mathcal{D}^\delta} g'(\|x_i - x_j\|^2)\langle x_i - x_j\rangle$. Let $i_\delta := \arg\max_{i\in\{1,\dots,N\}} \|S_i^\delta\|$. As in Proposition 1.12, the control defined by (1.16) maximizes W_g^δ instantaneously. With this control, the partial generalized entropy's evolution is given by $\dot{W}_g^\delta = \frac{1}{N}\sum_{i=1}^N \langle S_i^\delta, \frac{1}{N}\sum_{k\neq i} a(\|x_i - x_j\|)(x_i - x_j)\rangle + \|S_{i_\delta}^\delta\| \frac{M}{N}$.

For all $(i,j) \in \mathcal{D}^\delta$, $\|x_i - x_j\| \leqslant \delta \leqslant \eta$, so $a(\|x_i - x_j\|)\|x_i - x_j\| \leqslant \epsilon$. Let $t_\delta := \inf\{t > 0 \mid \mathcal{D}^\delta(t) \neq \mathcal{D}^\delta(0)\}$. While $t < t_\delta$, W_g^δ satisfies $\dot{W}_g^\delta \geqslant \|S_{i_\delta}^\delta\|(-\epsilon + \frac{M}{N}) > 0$. Hence for all $t < t_\delta$, $W_g^\delta(t) \geqslant W_g^\delta(0)$. This implies that for every $(i,j) \in \mathcal{D}^\delta$, for all $t < t_\delta$,

$$\|x_i(t) - x_j(t)\| \geqslant \kappa := \left(g^{-1}(2N^2 W_g^\delta(0) - \left(\frac{N(N-1)}{2} - 1\right) m)\right)^{1/2}. \tag{A.2}$$

Since $\delta > \kappa$, the inequality $\|x_i(t) - x_j(t)\| \geqslant \kappa$ holds true for every $(i,j) \in \{1,\dots,N\}^2$ for all $t < t_\delta$. Let us now consider what happens when $t > t_\delta$. Notice that even though $\mathcal{D}^\delta$ changes at t_δ, W_g^δ is continuous. Indeed, denoting $\mathcal{D}_-^\delta := \mathcal{D}^\delta(t_\delta - \theta)$ and $\mathcal{D}_+^\delta := \mathcal{D}^\delta(t_\delta + \theta)$, we have

$$\begin{aligned}
\lim_{\theta\to 0} W_g^\delta(t_\delta - \theta) &= \lim_{\theta\to 0} \sum_{(i,j)\in\mathcal{D}_-^\delta} g(\|x_i(t_\delta - \theta) - x_j(t_\delta - \theta)\|^2) \\
&= \lim_{\theta\to 0} \sum_{(i,j)\in\mathcal{D}_-^\delta\cap\mathcal{D}_+^\delta} g(\|x_i(t_\delta - \theta) - x_j(t_\delta - \theta)\|^2) \\
&\quad + \lim_{\theta\to 0} \sum_{(i,j)\in\mathcal{D}_-^\delta\setminus\mathcal{D}_+^\delta} g(\|x_i(t_\delta - \theta) - x_j(t_\delta - \theta)\|^2) \\
&= \lim_{\theta\to 0} \sum_{(i,j)\in\mathcal{D}_-^\delta\cap\mathcal{D}_+^\delta} g(\|x_i(t_\delta - \theta) - x_j(t_\delta - \theta)\|^2) + 0
\end{aligned}$$

due to the property $g(\delta^2) = 0$. Similarly,

$$\begin{aligned}
\lim_{\theta\to 0} W_g^\delta(t_\delta + \theta) &= \lim_{\theta\to 0} \sum_{(i,j)\in\mathcal{D}_+^\delta} g(\|x_i(t_\delta + \theta) - x_j(t_\delta + \theta)\|^2) \\
&= \lim_{\theta\to 0} \sum_{(i,j)\in\mathcal{D}_-^\delta\cap\mathcal{D}_+^\delta} g(\|x_i(t_\delta + \theta) - x_j(t_\delta + \theta)\|^2).
\end{aligned}$$

We can now apply the same reasoning as previously starting at $t = t_\delta$. Let $t'_\delta := \inf\{t > t_\delta \mid \mathcal{D}^\delta(t) \neq \mathcal{D}^\delta(t_\delta)\}$. We show that (A.2) becomes that for all $t_\delta < t < t'_\delta$, for every $(i,j) \in \{1,\dots,N\}^2$,

$$\|x_i(t) - x_j(t)\| \geqslant \left(g^{-1}(2N^2 W_g^\delta(t_\delta) - m\left(\frac{N(N-1)}{2} - 1\right))\right)^{1/2} = \kappa.$$

By induction, we obtain $\|x_i(t) - x_j(t)\| \geqslant \kappa$ for all $t \geqslant 0$ for every $(i,j) \in \{1,\dots,N\}^2$. $\square$

Appendix B. Proofs for the macroscopic model. We provide the proof for Theorem 2.11.

Proof of Theorem 2.11. The proof follows the same argument as in [41]. Since $a \in \mathrm{Lip}(\mathbb{R}^+, \mathbb{R}^+)$, while $\mu(t)$ has compact support, the displacement of the support has bounded velocity $\xi[\mu(t)] = \iint_{\mathrm{supp}(\mu(t))} \|x - y\|^2 a(\|x - y\|) d\mu_x(t) d\mu_y(t) < \infty$, so $X(\cdot)$ is differentiable almost everywhere. Let $x(\cdot, x_0)$ be the particle trajectory of (2.5) with $x(0, x_0) = x_0$. Hence $x(t, x_0) \in \mathrm{supp}(\mu(t))$ for all $x_0 \in \mathrm{supp}(\mu_0)$, and $X(t) =$

$\max\{\|x(t,x_0)-\bar{x}(t)\| \mid x_0 \in \mathrm{supp}(\mu_0)\}$. We remark that the maximum is reached since μ_0 is assumed to have compact support. For every $t \geqslant 0$, let K_t denote the set of points $x_0 \in \mathrm{supp}(\mu_0)$ that attain the maximum $X(t)$, i.e., $K_t = \mathrm{argmax}\{\|x(t,x_0)-\bar{x}(t)\| \mid x_0 \in \mathrm{supp}(\mu_0)\}$. By definition, $X(t)^2 = \|x(t,x_0)-\bar{x}(t)\|^2$ for all $x_0 \in K_t$. From the Danskin theorem (see [19]), $\frac{d}{dt}(X(t)^2) = \max_{x_0\in K_t}\{\frac{d}{dt}(\|x(t,x_0)-\bar{x}(t)\|^2)\}$. Then $\dot{X}(t)X(t) = \max_{x_0\in K_t}\{\int_{\mathrm{supp}(\mu(t))} a(\|x(t,x_0)-y\|)(x(t,x_0)-\bar{x}(t))\cdot(y-x(t,x_0))d\mu_y\}$. Since $x(t,x_0) \in S(\bar{x}(t),X(t))$, by convexity, $(x(t,x_0)-\bar{x}(t))\cdot(y-x(t,x_0)) \leqslant 0$ for all $y \in \mathrm{supp}(\mu(t)) \subseteq B(\bar{x}(t),X(t))$. We supposed that $a(s) \geqslant C > 0$. Hence

$$\dot{X}(t)X(t) \leqslant \max_{x_0\in K_t}\left\{\int_{\mathrm{supp}(\mu(t))} C(x(t,x_0)-\bar{x}(t))\cdot(y-x(t,x_0))d\mu_y\right\} \leqslant -CX^2(t)$$

and thus $X(t) \leqslant X(0)\exp(-Ct)$. ∎

We also prove Theorem 2.18.

Proof of Theorem 2.18. The proof follows the same argument as that of Theorem 2.14. Suppose $\mathrm{supp}(\mu_0) \subset \bigcup_{i=1}^N B(x_i,r)$ with N, r, R and $(x_1,\dots,x_N) \in (\mathbb{R}^d)^N$ satisfying the conditions listed in the Theorem. As in the proofs of Theorems 2.11 and 2.14, we denote by $x(t,x_0)$ the particle trajectory of (2.5) with $x(0,x_0)=x_0$. For each $i \in \{1,\dots,N\}$, let $X_i(t) := \max\{\|x(t,x_0)-x_i\|\ ,\ x_0 \in B(x_i,r)\cap\mathrm{supp}(\mu_0)\}$. Let $K_t^i := \mathrm{argmax}\{\|x(t,x_0)-x_i\| \mid x_0 \in B(x_i,r)\cap\mathrm{supp}(\mu_0)\}$. Then for all $t \geqslant 0$, for all $x_0 \in K_t^i$, $X_i(t)^2 = \|x(t,x_0)-x_i\|^2$. We have

$$\begin{aligned}&\frac{d}{dt}\|x(t,x_0)-x_i\|^2\\ &\quad = 2(x(t,x_0)-x_i)\cdot\left(\int_{\mathrm{supp}(\mu(t,\cdot))} a(\|x(t,x_0)-y\|)(y-x(t,x_0))d\mu_y + \chi_\omega u(t,x(t,x_0))\right).\end{aligned}$$

While $\mathrm{supp}(\mu(t,\cdot)) \subset \bigcup_{i=1}^N B(x_i,r)$, for all $(x,y) \in \mathrm{supp}(\mu(t,\cdot))^2$, $\|x-y\| \leqslant R$, so $a(\|x-y\|)\|y-x\| \leqslant \epsilon$. Then

$$\frac{d}{dt}\|x(t,x_0)-x_i\|^2 \leqslant 2(\|x(t,x_0)-x_i\|\epsilon + (x(t,x_0)-x_i)\cdot\chi_\omega u(t,x(t,x_0))).$$

As for Theorem 2.14, we design a control set $\omega := \bigcup_{i=1}^N B(x_i,r_c) \subset \Omega_c$ and a control u satisfying (2.8) with $x \mapsto \chi_\omega u(x) \in \mathrm{Lip}(\mathbb{R}^d)$. Then from Proposition C.2, there exists a unique solution to the controlled equation (2.2). Furthermore, with this control, $\frac{d}{dt}(X_i(t)^2) = 2X_i(t)\dot{X}_i(t) \leqslant 2(\epsilon-M)X_i(t) < 0$. Hence the solution μ satisfies $\mathrm{supp}(\mu(t,\cdot)) \subset \bigcup_{i=1}^N B(x_i,r)$ for all $t \geqslant 0$, and consensus is avoided. ∎

Appendix C. Reminders on mean-field limits and kinetic equations. In this section, we show that the kinetic dynamics (2.2) are the natural mean-field limit to the microscopic ones, when the number of particles N tends to infinity.

We begin by showing that there exists a unique solution to the controlled kinetic HK equation (2.2) and that, moreover, the solution can be written explicitly as the push-forward of the initial measure via the "controlled particle flow" defined hereafter. Let $V_{\omega,u}[\mu]$ denote the nonlocal vector field defined as follows: for all $\mu \in \mathcal{P}_c(\mathbb{R}^d)$, for all $(t,x) \in \mathbb{R}^+ \times \mathbb{R}^d$, $V_{\omega,u}[\mu](t,x) = \xi[\mu](t,x) + \chi_{\omega(t)}(x)u(t,x)$.

DEFINITION C.1. *We denote by "controlled particle flow" the flow $\Phi_{w,u}$ generated by the time-dependent vector field $V_{\omega,u}[\mu(t,\cdot)]$, solution to $\partial_t\Phi_{\omega,u}(t,x) = V_{\omega,u}[\mu(t,\cdot)]$ $(t,\Phi_{\omega,u}(t,x))$, with $\Phi_{\omega,u}(0,x) = x$.*

Proposition C.2. *Let $a : \mathbb{R}^+ \to \mathbb{R}^+$ be such that $s \mapsto sa(s) \in \mathrm{Lip}(\mathbb{R}^+)$. For all $\mu \in \mathcal{P}_c(\mathbb{R}^d)$, we denote by $\xi[\mu]$ the interaction kernel $\xi[\mu](x) = \int_{\mathbb{R}^d} a(\|x-y\|)(y-x)d\mu(y)$. Suppose in addition that $u\chi_w \in L^\infty(\mathbb{R}^+, \mathrm{Lip}(\mathbb{R}^d))$, where for every time $t \in \mathbb{R}^+$, $\omega(t)$ is a measurable subset of $\mathbb{R}^d$. Let $\mu^0 \in \mathcal{P}_c(\mathbb{R}^d)$. Then the Cauchy problem for the controlled kinetic HK equation*

$$\partial_t \mu + \mathrm{div}((\xi[\mu] + \chi_\omega u)\mu) = 0 \tag{C.1}$$

with initial data $\mu(0,\cdot) = \mu^0$ has a unique solution $\mu \in C(\mathbb{R}^+, \mathcal{P}_c(\mathbb{R}^d))$, and moreover μ can be obtained explicitly as the push-forward of μ^0 by the flow $\Phi_{\omega,u}$, i.e., for all $t \in \mathbb{R}^+$, $\mu(t,\cdot) = \Phi_{w,u}(t)\#\mu^0$. The solution μ is understood in the weak sense, i.e., for all $f \in C^\infty(\mathbb{R}^+ \times \mathbb{R}^d)$,

$$\int_0^T \int_{\mathbb{R}^d} \left(\partial_t f(t,x) + \nabla f(t,x) \cdot \left(\xi[\mu(t,\cdot)](x) + \chi_{\omega(t)}(x)u(t,x)\right)\right) d\mu(t,x) = 0.$$

Proof. The proof is an adaptation of the proof of the general Theorem 2.3 stated in [41]. This theorem applies to transport equations with nonlocal velocity fields $V[\mu]$ satisfying the following:

(i) $V[\cdot]$ is a Lipschitz function for the Wasserstein distance, i.e., there exists a function $K \in L^\infty_{\mathrm{loc}}(\mathbb{R}^+)$ such that for all $(\mu,\nu) \in \mathcal{P}_c(\mathbb{R}^d)^2$, $\|V[\mu] - V[\nu]\|_{L^\infty(\mathbb{R}^+, C(\mathbb{R}^d))} \leqslant K(t)W_1(\mu,\nu)$.

(ii) $V[\mu]$ is uniformly Lipschitz, i.e., there exists a function $L \in L^\infty_{\mathrm{loc}}(\mathbb{R}^+)$ such that for all $\mu \in \mathcal{P}_c(\mathbb{R}^d)$, for all $t \in \mathbb{R}^+$, for all $(x,y) \in (\mathbb{R}^d)^2$, $\|V[\mu](t,x) - V[\mu](t,y)\| \leqslant L(t)\|x-y\|$.

(iii) $V[\mu]$ has uniform sublinear growth, i.e., there exists a function $M \in L^\infty_{\mathrm{loc}}(\mathbb{R}^+)$ such that for all $\mu \in \mathcal{P}_c(\mathbb{R}^d)$, for all $t \in \mathbb{R}^+$, for all $x \in \mathbb{R}^d$, $\|V[\mu](t,x)\| \leqslant M(t)(1+\|x\|)$.

We show that assumptions (i) and (ii) hold for the velocity field $V_{\omega,u}[\mu](t,x) = \xi[\mu](t,x) + \chi_{\omega(t)}(x)u(t,x)$. Let L_a denote the Lipschitz constant of $s \mapsto sa(s)$, so that for all $(s_1,s_2) \in (\mathbb{R}^+)^2$, $\|s_1 a(s_1) - s_2 a(s_2)\| \leqslant L_a \|s_1 - s_2\|$. Let L_u denote the Lipschitz constant of $x \mapsto \chi_{\omega(t)}(x)u(t,x)$. Recall that by the Kantorovich–Rubinstein duality, the Wasserstein distance $W_1(\mu,\nu)$ can be written as

$$W_1(\mu,\nu) = \sup\left\{\int_{\mathbb{R}^d} f d(\mu-\nu),\ |\ \|f\|_{\mathrm{Lip}} \leqslant 1\right\}.$$

Then, for all $(\mu,\nu) \in \mathcal{P}_c(\mathbb{R}^d)^2$, for all $(t,x) \in \mathbb{R}^+ \times \mathbb{R}^d$,

$$\begin{aligned}\|V_{\omega,u}[\mu](t,x) - V_{\omega,u}[\nu](t,x)\| = \Bigg\| &\int_{\mathbb{R}^d} a(\|x-y\|)(y-x)d\mu(y) \\ &- \int_{\mathbb{R}^d} a(\|x-y\|)(y-x)d\nu(y)\Bigg\| \leqslant L_a W_1(\mu,\nu).\end{aligned}$$

Furthermore, for all $\mu \in \mathcal{P}_c(\mathbb{R}^d)$, $V_{\omega,u}[\mu]$ is Lipschitz with a Lipschitz constant not depending on μ. Indeed,

$$\begin{aligned}&\left\|\int_{\mathbb{R}^d}(a(\|x-z\|)(z-x) - a(\|y-z\|)(z-y))d\mu(z) + \chi_{\omega(t)}(x)u(t,x) - \chi_{\omega(t)}(y)u(t,y)\right\| \\ &\quad \leqslant (L_a + L_u)\|x-y\|.\end{aligned}$$

Last, let $\mathcal{P}_R(\mathbb{R}^d) := \{\mu \in \mathcal{P}_c(\mathbb{R}^d) \mid \mathrm{supp}(\mu) \subset B(0,R)\}$, where $B(0,R)$ denotes the ball of radius R centered at the origin. We show that although the sublinear growth condition (iii) does not hold uniformly for all $\mu \in \mathcal{P}_c(\mathbb{R}^d)$, it does hold uniformly for all $\mu \in \mathcal{P}_R(\mathbb{R}^d)$. Notice that $\|a(\|x-y\|)(y-x) - a(\|y\|)y\| \leqslant L\|x\|$. For all $\mu \in \mathcal{P}_R(\mathbb{R}^d)$,

$$\begin{aligned}\|V_{\omega,u}[\mu](t,x)\| &\leqslant \|u\|_{L^\infty(\mathbb{R}^+\times\mathbb{R}^d)} + \int_{\mathbb{R}^d} a(\|y\|)\|y\|d\mu(y) + L\|x\| \\ &\leqslant (\|u\|_{L^\infty(\mathbb{R}^+\times\mathbb{R}^d)} + M_R) + L\|x\|,\end{aligned}$$

where $M_R := \sup\{sa(s) \mid 0 \leqslant s \leqslant R\}$. Using the contractivity of the vector field $\xi[\mu]$, it is easy to prove that without control, if $\mathrm{supp}(\mu_0) \subset B(0,R)$, then for all $t \in \mathbb{R}^+$, $\mathrm{supp}(\mu(t,\cdot)) \subset B(0,R)$. Since $u \in L^\infty(\mathbb{R}^+ \times \mathbb{R}^d)$, the expansion of the support of μ can be controlled over each time interval $[0,T]$. Then the proof of Theorem 2.3 in [41] needs only to be slightly modified to prove the existence and uniqueness of the solution to (C.1), as well as the definition of the solution by the push-forward of the initial data via the particle flow $\Phi_{\omega,u}$. ∎

Proposition C.2 shows the transport property of the kinetic HK equation (C.1): its solution can be obtained as the push-forward of the initial measure via the particle flow $\Phi_{\omega,u}$.

We now recall that we can pass from the kinetic equation to the particle system by taking empirical measures. Indeed, the evolution of an empirical measure composed of Dirac masses is equivalently given by the kinetic dynamics in the space of measures and by the microscopic ones by considering each pointwise mass as an agent (as shown in [12] and [41]).

Proposition C.3. *Let $a : \mathbb{R}^+ \to \mathbb{R}^+$ be such that $s \mapsto sa(s) \in \mathrm{Lip}(\mathbb{R}^+)$ and let $u\chi_w \in L^\infty(\mathbb{R}^+, \mathrm{Lip}(\mathbb{R}^d))$, where for all $t \in \mathbb{R}^+$, $\omega(t)$ is a measurable subset of $\mathbb{R}^d$. Let $(x_i^0)_{i\in\{1,\dots,N\}} \in (\mathbb{R}^d)^N$ and let $\mu^0 = \frac{1}{N}\sum_{i=1}^N \delta_{x_i^0}$. Then, given the initial data $\mu(0,\cdot) = \mu^0$, the unique solution to*

$$\partial_t\mu + \mathrm{div}\left(\left(\int_{\mathbb{R}^d} a(\|x-y\|)(y-x)d\mu(y) + \chi_\omega u\right)\mu\right) = 0 \tag{C.2}$$

is given by $\mu = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, where $(x_i)_{i\in\{1,\dots,N\}}$ is the solution to

$$\begin{cases} \dot{x}_i(t) = \dfrac{1}{N}\displaystyle\sum_{i=1}^N a(\|x_i(t)-x_j(t)\|)(x_j(t)-x_i(t)) + \chi_{\omega(t)}(x_i(t))u(t,x_i(t)), \\ x_i(0) = x_i^0. \end{cases} \tag{C.3}$$

REFERENCES

[1] G. Ajmone Marsan, N. Bellomo, and L. Gibelli, *Stochastic evolutionary differential games toward a systems theory of behavioral social dynamics*, Math. Models Methods Appl. Sci., 26 (2016), pp. 1051–1093.

[2] A. Aydoğdu, M. Caponigro, S. McQuade, B. Piccoli, N. Pouradier Duteil, F. Rossi, and E. Trélat, *Interaction network, state space and control in social dynamics*, in Active Particles, Volume 1: Theory, Methods, and Applications, N. Bellomo, P. Degond, and E. Tadmor, eds., Springer, Berlin, 2017, pp. 99–140.

[3] 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*, Proc. Nat. Acad. Sci., 105 (2008), pp. 1232–1237.

[4] M. Bardi and F. S. Priuli, *Linear-quadratic n-person and mean-field games with ergodic cost*, SIAM J. Control Optimi., 52 (2014), pp. 3022–3052.
[5] N. Bellomo, M. A. Herrero, and A. Tosin, *On the dynamics of social conflict: Looking for the Black Swan*, Kinet. Relat. Models, 6 (2013), pp. 459–479.
[6] A. L. Bertozzi, T. Laurent, and J. Rosado, *L^p theory for the multidimensional aggregation equation*, Comm. Pure Appl. Math., 64 (2011), pp. 45–83.
[7] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks: A Mathematical Approach to Motion Coordination Algorithms*, Princeton Ser. Appl. Math., Princeton University Press, Princeton, NJ, 2009.
[8] J. A. Cañizo, J. A. Carillo, and J. Rosado, *A well-posedness theory in measures for some kinetic models of collective motion*, Math. Models Methods Appl. Sci., 21 (2011), pp. 515–539.
[9] C. Canuto, F. Fagnani, and P. Tilli, *A Eulerian approach to the analysis of rendez-vous algorithms*, in Proceedings of the IFAC World Congress, 2008.
[10] M. Caponigro, M. Fornasier, B. Piccoli, and E. Trélat, *Sparse stabilization and control of the Cucker-Smale model*, Math. Control Relat. Fields, 3 (2013), pp. 447–466.
[11] M. Caponigro, M. Fornasier, B. Piccoli, and E. Trélat, *Sparse stabilization and control of alignment models*, Math. Models Methods Appl. Sci., 25 (2015), pp. 521–564.
[12] M. Caponigro, B. Piccoli, F. Rossi, and E. Trélat, *Mean-field sparse Jurdjevic-Quinn control*, Math. Models Methods Appl. Sci., 27 (2017), pp. 1223–1253.
[13] M. Caponigro, B. Piccoli, F. Rossi, and E. Trélat, *Sparse Jurdjevic-Quinn stabilization of dissipative systems*, Automatica, 86 (2017), 110–120.
[14] J. A. Carrillo, M. Di Francesco, A. Figalli, T. Laurent, and D. Slepcev, *Global-in-time weak measure solutions and finite-time aggregation for nonlocal interaction equations*, Duke Math. J., 156 (2011), pp. 229–271.
[15] 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, Model. Simul. Sci. Eng. Technol., Springer, Berlin, 2010, pp. 297–336.
[16] A. Cavagna, A. Cimarelli, I. Giardina, G. Parisi, R. Santagati, F. Stefanini, and M. Viale, *Scale-free correlations in starling flocks*, Proc. Nat. Acad. Sci., 107 (2010), pp. 11865–11870.
[17] I. Couzin, J. Krause, N. Franks, and S. Levin, *Effective leadership and decision making in animal groups on the move*, Nature, 433 (2005), pp. 513–516.
[18] F. Cucker and S. Smale, *Emergent behavior in flocks*, IEEE Trans. Automat. Control, 52 (2007), pp. 852–862.
[19] J. Danskin, *The Theory of Max Min*, Springer, Berlin, 1967.
[20] P. Degond and S. Motsch, *Continuum limit of self-driven particles with orientation interaction*, Math. Models Methods Appl. Sci., 18 (2008), pp. 1193–1215.
[21] P. Degond and S. Motsch, *Large scale dynamics of the persistent turning walker model of fish behavior*, J. Stat. Phys., 131 (2008), pp. 989–1021.
[22] M. Fornasier and F. Solombrino, *Mean-field optimal control*, ESAIM Control Optim. Calc. Vari., 20 (2014), pp. 1123–1152.
[23] 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 (2010), pp. 1679–1683.
[24] S.-Y. Ha, E. Jeong, J.-h. Kang, and K. Kang, *Emergence of multi-cluster configurations from attractive and repulsive interactions*, Math. Models Methods Appl. Sci., 22 (2012).
[25] S.-Y. Ha, J. Kim, J. Park, and X. Zhang, *Complete cluster predictability of the Cucker–Smale flocking model on the real line*, Arch. Ration. Mech. Anal., 231 (2019), pp. 319–365.
[26] S.-Y. Ha and E. Tadmor, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinet. Relat. Models, 1 (2008), pp. 415–435.
[27] R. Hegselmann and U. Krause. *Opinion dynamics and bounded confidence models, analysis, and simulation*, J. Artificial Societies Social Simulation, 5, (2002).
[28] R. A. Holley and T. M. Liggett, *Ergodic theorems for weakly interacting infinite systems and the voter model*, Ann. Probab., 3 (1975), pp. 643–663.
[29] A. Jadbabaie, J. Lin, and A. S. Morse, *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automat. Control, 48 (2003), pp. 988–1001.
[30] J.-H. Kang, S.-Y. Ha, K. Kang, and E. Jeong, *How do cultural classes emerge from assimilation and distinction? An extension of the Cucker-Smale flocking model*, J. Math. Sociol., 38 (2014), pp. 47–71.
[31] A. P. Kirman and N. J. Vriend, *Learning to be loyal*, in A Study of the Marseille Fish Market, Springer, Berlin, 2000, pp. 33–56.

[32] A. Lachapelle and M.-T. Wolfram, *On a mean field game approach modeling congestion and aversion in pedestrian crowds*, Transportation Research Part B Methodological, 45 (2011), pp. 1572–1589.

[33] N. Leonard, *Multi-agent system dynamics: Bifurcation and behavior of animal groups*, Plenary paper, IFAC Symposium on Nonlinear Control Systems, Toulouse, France, 2013.

[34] N. Michael, D. Mellinger, Q. Lindsey, and V. Kumar, *The GRASP multiple micro-UAV testbed*, IEEE Robotics Automat. Mag., 17 (2010), pp. 56–65.

[35] S. Motsch and E. Tadmor, *Heterophilious dynamics enhances consensus*, SIAM Rev., 56 (2014), pp. 577–621.

[36] L. Pareschi, G. Toscani, A. Tosin, and M. Zanella, *Hydrodynamic Models of Preference Formation in Multi-Agent Societies*, arXiv:1901.00486, 2019.

[37] J. K. Parrish and L. Edelstein-Keshet, *Complexity, pattern, and evolutionary trade-offs in animal aggregation*, Science, 284 (1999), pp. 99–101.

[38] L. Perea, P. Elosegui, and G. Gómez, *Extension of the Cucker–Smale control law to space flight formations*, J. Guidance Control Dyn., 32 (2009), pp. 527–537.

[39] B. Piccoli, *Measure Differential Equations*, preprint, 2017.

[40] B. Piccoli, N. Pouradier Duteil, and B. Scharf, *Optimal control of a collective migration model*, Math. Models Methods Appl. Sci., 26 (2016), pp. 383–417.

[41] B. Piccoli, F. Rossi, and E. Trélat, *Control to flocking of the kinetic Cucker–Smale model*, SIAM J. Math. Anal., 47 (2015), pp. 4685–4719.

[42] Y. Privat, E. Trélat, and E. Zuazua, *Optimal observability of the multi-dimensional wave and Schrödinger equations in quantum ergodic domains*, J. Eur. Math. Soc. (JEMS), 18 (2016), pp. 1043–1111.

[43] A. Sarlette, S. Bonnabel, and R. Sepulchre, *Coordinated motion design on Lie groups*, IEEE Trans. Automat. Control, 55 (2010), pp. 1047–1058.

[44] R. Sepulchre, D. A. Paley, and N. Leonard, *Stabilization of planar collective motion: All-to-all communication*, IEEE Trans. Automat. Control, 52 (2007), pp. 811–824.

[45] H. A. Simon, *Models of Bounded Rationality*, Vol. 2, MIT Press, Cambridge, MA, 1982.

[46] N. Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House, New York, 2010.

[47] G. Toscani, A. Tosin, and M. Zanella, *Opinion modeling on social media and marketing aspects*, Phys. Rev. E, 98 (2018), 022315.

[48] 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 (1995), pp. 1226–1229.

[49] D. Vilone, J. Ramasco, A. Sánchez, and M. S. Miguel, *Social and strategic imitation: The way to consensus*, Sci. Rep., 2 (2012).

[50] S. Wongkaew, M. Caponigro, and A. Borzì, *On the control through leadership of the Hegselmann-Krause opinion formation model*, Math. Models Methods Appl. Sci., 25 (2015), pp. 565–585.

[51] C. A. Yates, R. Erban, C. Escudero, I. D. Couzin, J. Buhl, I. G. Kevrekidis, P. K. Maini, and D. J. T. Sumpter, *Inherent noise can facilitate coherence in collective swarm motion*, Proc. Nat. Acad. Sci., 106 (2009), pp. 5464–5469.