# Finite- and infinite-time cluster formation for alignment dynamics on the real line

Trevor M. Leslie and Changhui Tan

*Abstract.* We show that the locations where finite- and infinite-time clustering occur for the 1D Euler-alignment system can be determined using only the initial data. Our present work provides the first results on the structure of the finite-time singularity set and asymptotic clusters associated with a weak solution. In many cases, the eventual size of the cluster can be read off directly from the flux associated with a scalar balance law formulation of the system.

## Contents

# 1. Introduction

## 1.1. The Euler-alignment system

This paper focuses on the 1-dimensional Euler-alignment system, which describes collective behavior among agents. It is a hydrodynamic analog of the celebrated Cucker–Smale system of ODEs, which we discuss below. The system is represented by the following equations:

$$\begin{cases} \partial_t \rho + \partial_x(\rho u) = 0, & (x,t) \in \mathbb{R} \times \mathbb{R}_+, \\ \partial_t(\rho u) + \partial_x(\rho u^2) = \displaystyle\int_{\mathbb{R}} \rho(x,t)\rho(y,t)\phi(x-y)(u(y,t) - u(x,t))\,\mathrm{d}y. \end{cases} \tag{1}$$

The two functions $u$ and $\rho$ represent the velocity and nonnegative density profile of a group of agents. The communication protocol $\phi : \mathbb{R} \to \mathbb{R}$ describes the strength of the nonlocal alignment interactions. It is assumed to be nonnegative and even. The system (1) is accompanied by initial data:

$$\rho(x,0) = \rho^0(x), \quad u(x,0) = u^0(x).$$

The last decade has seen a rapid development in the theory for the well-posedness and asymptotic behavior of the Euler-alignment system (1). As the nature of (1) depends strongly on the behavior of the communication protocol $\phi$ near the origin, this theory necessarily breaks into several different cases. When the communication protocol $\phi$ is *strongly singular*, meaning that it is non-integrable at the origin, the alignment force on the right-hand side of (1)$_2$ is known to exhibit dissipative properties and has a nonlinear regularizing effect. Consequently, the system takes on a parabolic character. Notably, research presented in [15,33–35] has demonstrated that the solution remains globally regular for any smooth initial data with inf $\rho_0 > 0$. In the case where the initial data contains vacuum, we refer to [1,37] for discussions on singularity formation and [16] for existence of weak solutions. Researchers have explored various extensions of the system, incorporating features such as pressure [12], misalignment

[30], and attraction–repulsion forces [23]. While progress has been made in the context of the multi-dimensional system, it is less well-studied in comparison with the one-dimensional case. Some partial results have been presented in [2,8,14,25,26,32].

We are interested in the case where the communication protocol $\phi$ is less singular, specifically when it is integrable at the origin. In such instances, the alignment force exhibits a nonlocal damping effect, and the system adopts a hyperbolic character. A significant observation known as the *critical threshold phenomenon* was first reported in [36]. This phenomenon highlights that the global regularity of solutions is contingent on the initial data: 'subcritical' initial data lead to global well-posedness, while 'supercritical' initial data result in finite-time singularity formations. Numerous research works have been dedicated to determining the critical threshold conditions for the Euler-alignment system and related systems [3,6,10,11,18,21,27,38]. In the next subsection, we will conduct a brief survey on the findings specifically related to the one-dimensional Euler-alignment system (1).

Another celebrated feature of the Euler-alignment system is its *asymptotic flocking behavior*. This behavior emerges under the assumption that the communication kernel has a *heavy tail*, indicated by the condition:

$$\int_1^\infty \phi(r)\,\mathrm{d}r = \infty. \tag{2}$$

In such cases, solutions of the system converge to a flocking state as $t \to \infty$: the velocity $u$ aligns with its average value, while the density $\rho$ stabilizes into a traveling wave form:

$$u(x,t) \to \bar{u} := \frac{\int_\mathbb{R} (\rho^0 u^0)(x)\mathrm{d}x}{\int_\mathbb{R} \rho^0 \mathrm{d}x}, \qquad \rho(x + \bar{u}t, t) \to \rho_\infty(x).$$

Here $\rho_\infty$ is the asymptotic density profile, which carries important information about the emergent flocking phenomenon. However, the structure of $\rho_\infty$ is relatively less understood. Relevant discussions on the flocking phenomenon can be found in works such as [20,28,29,34,36].

## 1.2. Regular solutions and the critical threshold conditions

As mentioned above, the existence or non-existence of a global smooth solution $(\rho, u)$ to the 1-dimensional Euler-alignment system (1) with smooth initial data $(\rho^0, u^0)$ can be determined from the *critical threshold condition* (CTC). The conditions can be precisely characterized using an auxiliary quantity introduced in [6]:

$$e^0(x) = \partial_x u^0(x) + \phi * \rho^0(x), \qquad x \in \mathbb{R}.$$

Here '$*$' denotes convolution in the spatial variable. Critical threshold conditions are available for two types of communication protocols:

- *Bounded* communication.

- *Weakly singular* communication: $\phi$ is unbounded but integrable at the origin. More precisely, we say $\phi$ is weakly singular with order $\beta \in (0, 1)$ if there exist positive constants $R > 0$ and $c > 0$ such that the following lower bound holds:

$$\phi(r) \geq c\,r^{-\beta}, \quad \forall\, r \in (0, R). \tag{3}$$

A sharp CTC has been established in [6] for bounded communication protocols, which says

  I. If $e^0(x) \geq 0$ for all $x \in \mathbb{R}$, then the solution is globally regular.

  II. If $e^0(x_0) < 0$ for some $x_0 \in \mathbb{R}$, then the solution develops a singularity at $x(x_0, T_*)$ for some finite time $T_*$.

In this context, we use the notation $x(a, t)$ for the characteristic path originating from $a$, which satisfies the following ordinary differential equation:

$$\dot{x}(a, t) = u(x(a, t), t), \quad x(a, 0) = a.$$

For bounded protocols, these conditions are often referred to as 'subcritical' (I) and 'supercritical' (II).

When the communication protocol is weakly singular, another type of finite time singularity formation was discovered in [38]. Under the subcritical CTC (I), we have the following.

 III. If $e^0(x_0) = 0$ for some $x_0 \in \mathbb{R}$, then the behavior depends on the type of communication protocol:

   (i) If $\phi$ is bounded, then the solution is globally regular.

  (ii) If $\phi$ is weakly singular, then the solution might develop a singularity at $x(x_0, T_*)$ for some finite time $T_*$.

An example was provided in [38] to illustrate the singularity formation, assuming $\phi$ has a heavy tail. Note that this is the only scenario where the two types of protocols can lead to different behaviors. We may refer it as the critical case.

The nature of the singularity in II and III(ii) is commonly known as a *singular shock*, arising when two characteristic paths collide and lead to a shock discontinuity in velocity. Additionally, there is a concentration of mass, which we refer to as *clustering* in this paper.

In [27], the author proved a refinement of the above results that renders the CTC more meaningful in the presence of vacuum. When $x \notin \mathrm{supp}\,\rho^0$, the physical velocity $u^0(x)$ is undefined, making $e^0(x)$ ill-defined. To address this, the author introduced an anti-derivative of $e^0$, denoted as:

$$\psi^0 = u^0 + \Phi * \rho^0, \qquad \Phi(x) = \int_0^x \phi(r)\,\mathrm{d}r. \tag{4}$$

The CTC are then expressed in terms of the monotonicity of $\psi^0$ inside the support of $\rho^0$. For instance, if $\phi$ is bounded, the solution is globally regular if and only if $\psi^0$ is nondecreasing in $\mathrm{supp}(\rho^0)$. The theory presented in [27] was further developed in

[24] through a comprehensive study of the evolution of characteristic paths; the latter will inform the following discussion.

The characteristics are simplest for the degenerate protocol $\phi \equiv 0$, for which the Euler-alignment system (1) reduces to the well-studied pressureless Euler system. For pressureless Euler, the characteristic paths of classical solutions are always straight lines, leading to three possible scenarios: (I) separation linearly in time, (II) collision in finite time, or (III) running parallel for all time. However, the introduction of the alignment force leads to intriguing new asymptotic behaviors.

For bounded and heavy-tailed communication protocols, assuming $\psi^0$ is nondecreasing in supp($\rho^0$), it was demonstrated in [24] that for any $a, b \in$ supp($\rho^0$) with $a < b$, the characteristic paths $x(a, t)$ and $x(b, t)$ are globally well-defined, and their distance satisfies the following quantitative bounds (for some positive constants $c$ and $C$ that do not depend on $a$, $b$, or $t$):

$$c(\psi^0(b) - \psi^0(a)) \leq \lim_{t \to \infty} \left( x(b, t) - x(a, t) \right) \leq C(\psi^0(b) - \psi^0(a)). \qquad (5)$$

The comparison (5) was then utilized to study the structure of the asymptotic density profile $\rho_\infty$. One especially interesting situation occurs when $\psi^0(a) = \psi^0(b)$, in which case the distance between characteristic paths tends to zero as $t \to +\infty$:

$$\lim_{t \to \infty} \left( x(b, t) - x(a, t) \right) = 0.$$

If there is any mass trapped between the two converging characteristics, a concentration of mass will develop in the asymptotic density profile $\rho_\infty$. This phenomenon is referred to as *infinite-time clustering*, and it represents a distinctive feature of the alignment interaction. A discrete version of this phenomenon was previously investigated in [19] for the classical Cucker–Smale system (which we note often exhibits substantially different behavior than its sticky particle version considered below, c.f. Remark 1.1). A particular bi-cluster formation was studied in [9]. The theory is further developed in [39] for the singular Cucker–Smale system, allowing particles to stick with each other.

We summarize the key results cited above in a simplified form as follows:

**Proposition 1.1.** *Suppose $\rho^0 \in C(\mathbb{R})$ is compactly supported and $u^0 \in C^1(\mathbb{R})$. Let $\phi$ be nonnegative, even, and locally integrable. Assume $a < b$.*

  I. *Suppose $\psi^0$ is nondecreasing and $\psi^0(a) < \psi^0(b)$. Then there exists $c > 0$ such that*

$$x(b, t) - x(a, t) \geq c > 0 \qquad \text{for all } t \geq 0.$$

  II. *Suppose $\psi^0(a) > \psi^0(b)$. Then $u$ loses regularity in finite time.*
  III. *Suppose $\psi^0$ is nondecreasing and $\psi^0(a) = \psi^0(b)$.*
    (i) *If $\phi$ is bounded, then $x(b, t) > x(a, t)$ for all $t \in [0, \infty)$.*
    (ii) *If $\phi$ is bounded and heavy-tailed, then we have $x(b, t) - x(a, t) \to 0$ as $t \to \infty$.*

(iii) *If $\phi$ is weakly singular and heavy-tailed, and if $\rho^0\big|_{[a,b]} \not\equiv 0$, then finite-time blowup occurs.*

The primary objective of this study is to investigate the phenomena of finite- and infinite-time clustering and to derive predictions from the initial data. We seek to extend the current theory by addressing the following gaps in Proposition 1.1:

- *For finite-time clustering* The current results only demonstrate finite-time singularity formation in cases II and III(iii). However, the behavior after the clusters form remains unknown as the classical solution ceases to exist. We aim to understand the subsequent evolution of these clusters once they have formed.
- *For infinite-time clustering* The existing results treat the infinite-time clustering phenomenon only when the solution is globally regular, meaning that there is no finite-time clustering. Moreover, the simple characterization of the infinite-time clusters in terms of the monotonicity properties of $\psi^0$ is not expected to survive in the presence of finite-time clusters; we seek a suitable generalization of this characterization.

To achieve our objective, we need to consider an appropriate class of weak solutions. For this purpose, we will build upon the theory recently established in [29], which provides a valuable framework for our study. It will allow us to explore the intricate dynamics of the Euler-alignment system and investigate the finite- and infinite-time clustering phenomena in a comprehensive manner.

### 1.3. Weak solutions and the scalar balance law

The well-posedness theory for weak solutions to systems of conservation laws poses significant challenges, particularly concerning uniqueness. In [29], the authors establish the well-posedness theory for weak solutions to (1) by employing the approach of Brenier and Grenier [5] on the 1D pressureless Euler system ($\phi \equiv 0$) to address this issue. The key idea is to reduce the system to a single scalar balance law:

$$\partial_t M + \partial_x(A(M)) = (\Phi * \partial_x M)\partial_x M, \qquad \Phi(x) = \int_0^x \phi(r)\,\mathrm{d}r, \qquad (6)$$

supplemented with initial conditions $M(\cdot, 0) = M^0$. Here $M(t) : \mathbb{R} \to \left[-\frac{1}{2}, \frac{1}{2}\right]$ is the cumulative distribution function for the density $\rho(t)$ (shifted by a constant for technical reasons) and the flux $A$ is determined from $\rho^0$ and $u^0$. Unlike the situation for (1), it is fairly straightforward to establish (though not trivial to justify) entropy conditions for (6) that are sufficient to guarantee uniqueness (c.f. [29]). We therefore refer to the solution $(\rho, u)$ of (1) that we recover from (6) as the *entropy solution* of (1).

Let us give an extremely brief description of the construction of an entropy solution to (1). Starting with initial data $(\rho^0, u^0) \in \mathcal{P}_c(\mathbb{R}) \times L^\infty(\mathrm{d}\rho^0)$ (where $\mathcal{P}_c(\mathbb{R})$ denotes the space of compactly supported probability measures on $\mathbb{R}$), we define the corresponding

cumulative distribution function $M^0$ and its generalized inverse $X^0$ as follows:

$$M^0(x) = -\frac{1}{2} + \rho^0((-\infty, x]), \quad X^0(m) = \inf \left\{ x : M^0(x) \geq m \right\}. \qquad (7)$$

We define the flux $A : \left[-\frac{1}{2}, \frac{1}{2}\right] \to \mathbb{R}$ of the scalar balance law (6) as

$$A(m) = \int_{-\frac{1}{2}}^{m} \psi^0 \circ X^0(\widetilde{m}) \, d\widetilde{m}, \quad \psi^0 = u^0 + \Phi * \rho^0. \qquad (8)$$

Note that $A$ is Lipschitz, with $A\left(-\frac{1}{2}\right) = 0$. Having determined $(M^0, A)$ from $(\rho^0, u^0)$, we evolve the scalar balance law (6). There is a unique entropy solution, which we denote by $M = M(x, t)$, associated with the initial data $M^0$ and flux $A$. We generate this entropy solution through a front-tracking approximation scheme; remarkably, the positions and magnitudes of our fronts can be encoded using the Cucker–Smale dynamics, supplemented with completely inelastic collision rules. (We will describe this *sticky particle Cucker–Smale* system in detail in Sect. 2, and it will play a crucial role in the proof of our main theorem.) Finally, we recover the solution to (1) via $\rho = \partial_x M$ and $P = -\partial_t M$. The velocity $u$ is then the Radon–Nikodym derivative of the measure $P$ with respect to $\rho$. It is shown in [29] that this pair $(\rho, u)$ solves the Euler-alignment system (1) in the sense of distributions and satisfies the initial data in an appropriate sense.

We also set notation for the generalized inverse $X(t)$ of $M(t)$, which we will use extensively below.

$$X(m, t) = \inf \left\{ x \in \mathbb{R} : M(x, t) \geq m \right\}. \qquad (9)$$

Note that $M^0$ and $M(t)$ are nondecreasing, right-continuous, and defined on $\mathbb{R}$, while $X^0$ and $X(t)$ are nondecreasing, left-continuous, and defined on $\left(-\frac{1}{2}, \frac{1}{2}\right]$. We refer to elements of $\left(-\frac{1}{2}, \frac{1}{2}\right]$ as *mass labels*.

### 1.4. Terminology and notation for cluster formation

The focus of this paper is on cluster formation of mass labels in solutions of the Euler-alignment system (1). In particular, we would like to describe the phenomena of finite- and infinite-time clustering by examining the initial conditions.

We begin by defining our clusters in terms of the function $X(\cdot, t)$, which encodes the 'location' of each mass label.

**Definition 1.2.** (Clusters) Given $m \in \left(-\frac{1}{2}, \frac{1}{2}\right]$, we say that there is a *t-cluster at $m$* (or a *finite-time cluster at $m$*, if the value of $t$ is not important) if there exists $m' < m$ such that $X(m', t) = X(m, t)$. In this case, the $t$-cluster at $m$ is defined to be the largest interval of the form $(m', m'']$, containing $m$, such that $X(\cdot, t)$ is constant on $(m', m'']$. We refer to 0-clusters as *initial clusters*.

We say there is an *infinite-time cluster at $m$* if there exists $m' < m$ such that $X(m, t) - X(m', t) \to 0$ as $t \to \infty$. In this case, the infinite-time cluster at $m$ is defined to be the largest interval of the form $I = (m', m'']$ or $I = (m', m'')$, containing $m$, such that $\operatorname{diam} X(I, t) \to 0$ as $t \to \infty$.

It is worth noting that if $(m', m'']$ is the $t$-cluster at $m$, then $X(\cdot, t)$ is constant on $(m', m'']$ by Definition 1.2, but either of the possibilities $X(m', t) = X(m'', t)$ or $X(m', t) \neq X(m'', t)$ may occur. Nevertheless, we feel that the left-continuity of $X(\cdot, t)$ provides us with a compelling reason to define our finite-time clusters as half-open intervals.

Our definition above, and the analysis of [29], guarantee that existing clusters cannot 'unstick.' Hence, the size of a cluster can only grow in time, and if there is a finite-time cluster at $m$, then there is an infinite-time cluster at $m$ as well, with the infinite-time cluster containing the finite-time cluster.

We say that there is *no finite-time clustering at* $m$ if there is no $t$-cluster at $m$ for all $t \in [0, \infty)$, and we say that there is *no infinite-time clustering at* $m$ if there is no infinite-time cluster at $m$.

In what follows, we use $A^{**}$ to denote the lower convex envelope of the flux $A$. We will give a brief reminder of some definitions related to convexity in Sect. 2.1.

**Definition 1.3.** We break up the interval $\left(-\frac{1}{2}, \frac{1}{2}\right]$ into disjoint regions, on which we will observe different clustering behavior, as follows:

- We define the *subcritical region* $\Sigma_+$ by

$$\Sigma_+ = \{m \in \left(-\frac{1}{2}, \frac{1}{2}\right] : A^{**} \text{ is not linear on any interval of the form } (m', m]\}.$$

- We define the *critical region* $\Sigma_0$ by

$$\Sigma_0 = \bigcup_{\substack{A \text{ is linear and} \\ \text{equal to } A^{**} \text{ on } (m', m'']}} (m', m''].$$

- We define the *supercritical region* $\Sigma_-$ as the (open) set on which $A > A^{**}$:

$$\Sigma_- = \{m \in (\tfrac{1}{2}, \tfrac{1}{2}) : A(m) > A^{**}(m)\}.$$

Let us provide some brief commentary on the definition above. Neglecting a set of measure zero, we can understand the three regions in terms of the table below, which provides a more intuitive picture. (Note that primes on $A$ and $A^{**}$ will always represent derivatives below.) We stress, however, that the formulation of Definition 1.3 is better suited for our analysis, as will become clear below.

|  | $A(m) = A^{**}(m)$ | $A(m) > A^{**}(m)$ |
|---|---|---|
| $(A^{**})''(m) > 0$ | $\Sigma_+$ | $\emptyset$ |
| $(A^{**})''(m) = 0$ | $\Sigma_0$ | $\Sigma_-$ |

See Fig. 1 for an illustration of the three regions.

Let us also note that in the 'typical' case where each of $\Sigma_-$, $\Sigma_0$, and $\Sigma_+$ has finitely many connected components, both $\Sigma_+$ and $\Sigma_0$ are unions of half-open intervals.
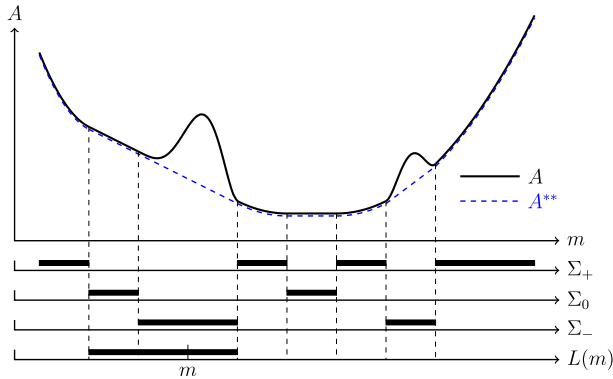
Figure 1. An illustration of the three regions of Definition 1.3, and an example of a set $L(m)$ that contains multiple connected intervals in $\Sigma_0$ and $\Sigma_-$

Consequently, the disjoint union $\Sigma_+ \cup \Sigma_0 \cup \Sigma_-$ contains all points in $\left(-\frac{1}{2}, \frac{1}{2}\right]$ *except* for the right endpoints of the intervals that constitute the connected components of $\Sigma_-$. This 'missing' set necessarily has Lebesgue measure zero (and one of our assumptions below will actually force it to be finite for the cases we consider).

We set one more piece of notation before moving on.

**Definition 1.4.** For $m \notin \Sigma_+$, we define the set $L(m)$ to be the largest half-open interval $(m', m'']$ containing $m$, such that $A^{**}$ is linear on $L(m)$. For $m \in \Sigma_+$, we define $L(m) = \{m\}$.

Ignoring endpoints, we may simply view $L(m)$ (for $m \notin \Sigma_+$) as the largest interval containing $m$ on which $(A^{**})'' = 0$. It may contain connected components of both $\Sigma_0$ and $\Sigma_-$, as illustrated in Fig. 1.

Let us briefly discuss how some of the results of Sect. 1.2 translate to the present framework, and (at an informal level) how we plan to extend them. For simplicity, we assume in this discussion that our initial data is sufficiently regular, i.e., $\psi^0$ and $X^0$ are differentiable and $X^0$ is strictly increasing. In this context, we have

$$A''(m) = \frac{d}{dm}\psi^0(X^0(m)) = e^0(X^0(m)) \cdot (X^0)'(m).$$

Then, we have following equivalent representation of CTC:

I.  $e^0(x) \geq 0$ for all $x \in \mathbb{R}$ $\iff$ $A''(m) \geq 0$ for all $m \in (-\frac{1}{2}, \frac{1}{2}]$ $\iff$ $\Sigma_- = \emptyset$,

II.  $e^0(x_0) < 0$ for some $x_0 \in \mathbb{R}$ $\iff$ $A''(m_0) < 0$ for some $m_0 \in (-\frac{1}{2}, \frac{1}{2}]$ $\iff$ $\Sigma_- \neq \emptyset$.

Furthermore, under the subcritical CTC (I), or equivalently, $\psi^0$ being nondecreasing, for $m' < m''$, we have the following equivalences for the assumptions in Proposition

1.1:

$$\text{I.} \quad \psi^0(X^0(m')) < \psi^0(X^0(m'')) \iff A'(m') < A'(m'')$$
$$\iff A \text{ is not linear on } [m', m'']$$
$$\iff L(m') \neq L(m''),$$
$$\text{III.} \quad \psi^0(X^0(m')) = \psi^0(X^0(m'')) \iff A'(m') = A'(m'')$$
$$\iff A \text{ is linear on } [m', m'']$$
$$\iff L(m') = L(m'').$$

Comparing the above discussion with Proposition 1.1, we arrive at the following conjectures:

**Conjecture 1.5.** *Mass labels from distinct $L(m)$'s never belong to the same cluster.*

**Conjecture 1.6.** *If $\phi$ is heavy-tailed, then mass labels from the same $L(m)$ belong to the same infinite-time cluster.*

We will demonstrate later that these conjectures are indeed correct. More remarkably, we show they hold even when $\Sigma_- \neq \emptyset$. Indeed, the equivalences above hold without the assumption $\Sigma_- = \emptyset$, if we replace $A$ by $A^{**}$ and we assume that $m', m'' \notin \Sigma_-$.

Let us also provide commentary on part II of Proposition 1.1. The condition $\psi^0(a) > \psi^0(b)$, $a < b$ implies that finite-time cluster formation occurs 'somewhere between $a$ and $b$.' However, the monotonicity of $\psi^0$ itself does not provide complete information about the clusters and their evolution over time. (For instance, $\psi^0(X^0(m')) > \psi^0(X^0(m''))$, $m' < m''$ implies nothing about whether $m'$ and $m''$ eventually belong to the same cluster.) Our framework allows us to obtain more detailed information about finite-time clustering. One new finding is:
*Any two mass labels from the same connected component of $\Sigma_-$ cluster together in finite time.*

We present precise versions of these statements in Theorem 1.7, along with other features that extend Proposition 1.1.

### 1.5. Global assumptions and summary of results

Throughout our manuscript, we fix initial data $(\rho^0, u^0)$ and make use of the following global assumptions, except where explicitly stated otherwise:

(A1) The communication protocol $\phi$ is locally integrable, even, and radially nonincreasing.
(A2) $(\rho^0, u^0) \in \mathcal{P}_c(\mathbb{R}) \times L^\infty(\mathrm{d}\rho^0)$, and $(\rho, u)$ is the associated entropy solution.
(A3) $M^0$, $X^0$, $A$, and $X$ are defined as in (7), (8), (9), and the sets $\Sigma_-$, $\Sigma_0$, $\Sigma_+$, and $L(m)$ are defined accordingly as in the previous subsection.
(A4) $A$ is convex in a neighborhood of every point $m$ of the boundary $S := \partial \Sigma_-$ of $\Sigma_-$.

The technical assumption (A4) is the only one we have not previously mentioned. It will simplify the structure of the set $\Sigma_-$, as will be spelled out below in Lemma 2.4; it will also play a key role in our analysis of the clustering behavior in $\Sigma_-$.

We now state our main result.

**Theorem 1.7.** *Fix* $m \in \left(-\frac{1}{2}, \frac{1}{2}\right]$. *The following statements describe the clustering behavior at* $m$.

I. *If* $m \in \Sigma_+$, *then there is no finite- or infinite-time clustering at* $m$.

II. *If* $m \in \Sigma_-$, *then there is a $t$-cluster at $m$ for all sufficiently large $t \geq 0$. Moreover, if $(m_-, m_+)$ is the connected component of $\Sigma_-$ containing $m$, then for any $\widetilde{m} \in (m_-, m_+)$ there exists a time $T \geq 0$ such that $m$ and $\widetilde{m}$ lie in the same $t$-cluster for all $t \geq T$.*

III.(i) *Suppose $\phi$ is bounded. If $(m', m'']$ is a finite-time cluster at $m$, then either $m \in \Sigma_0$ and $(m', m'']$ is an initial cluster, or $m \in \Sigma_-$ and $(m', m''] \subseteq (m_-, m_+]$, where $(m_-, m_+)$ is the connected component of $\Sigma_-$ containing $m$. No other finite-time clusters are possible.*

(ii) *Suppose $\phi$ is heavy-tailed. If $m \notin \Sigma_+$, then there is an infinite-time cluster at $m$, and it is equal to $L(m)$.*

(iii) *Suppose $\phi$ is heavy-tailed and weakly singular. If $m \notin \Sigma_+$, then there exists a finite time $T$ such that $L(m)$ is a $t$-cluster at $m$ for all $t \geq T$.*

*Remark 1.1.* The theorem above is clearly a significant upgrade over Proposition 1.1, which summarized the most relevant results on classical solutions. Indeed, our new theorem satisfactorily addresses the gaps outlined in Sect. 1.2 and applies to very general initial conditions. Let us clarify how it relates to two other lines of research that we have not yet emphasized.

We briefly mentioned above the paper [19], where the authors gave a comprehensive study of the limiting configurations associated with the 1D Cucker–Smale system with bounded communication. Their analysis is concerned with the 'free-flow' dynamics, where agents follow the Cucker–Smale ODE's for all time, without modification for the occurrence of collisions. Their analysis, comprehensive though it is for the classical Cucker–Smale system, has no hydrodynamic analog except for the case when no collisions occur. (The previously discussed work [28] is in this direction.) Once collisions occur, the proper way to pass to a hydrodynamic limit is through the *sticky particle* Cucker–Smale dynamics, as the present authors showed in our previous paper [29]. As soon as collisions are allowed, the sticky particle Cucker–Smale dynamics can differ drastically from those of their classical counterpart.

The one case where the free-flow dynamics *do* have substantial bearing on the sticky particle dynamics is in the setting of degenerate communication $\phi \equiv 0$, where (1) reduces to the 1D pressureless Euler equations. In this degenerate case, Theorem 1.7 can be recovered from already-existing theory; see for example [4,7,22,31]. (The paper [4] considers a more general system which includes pressureless Euler and Euler–Poisson as special cases.) These works all rely implicitly or explicitly on the fact that the 'sticky

particle' dynamics in that setting can be recovered from the 'free-flow' dynamics using a certain $L^2$ projection onto the convex cone of nondecreasing functions. (This is essentially the key observation of [31].) However, simple counterexamples show that the sticky particle Cucker–Smale dynamics, discussed below, cannot be recovered as such a projection. Consequently, the techniques of the above-cited works do not appear to apply to our problem. In particular, even at the discrete level, the result of applying the $L^2$ projection to the analysis of Ha et al. [19] is in general unrelated to the cluster formation described in Theorem 1.7.

### 1.6. Outline of the paper

Our proof relies on a discretization procedure from [29] involving the so-called sticky particle Cucker–Smale dynamics; in Sect. 2, we review the salient parts of this procedure for the convenience of the reader. We also establish conventions and review some standard facts related to convex functions. The remaining Sects. 3, 4 and 5 contain all the new analysis. These sections essentially track the statements I, II, and III, respectively, of Theorem 1.7, with some caveats spelled out below. We give a slightly more detailed summary presently.

In Sect. 3, we prove Conjecture 1.5: mass labels from distinct $L(m)$'s cannot belong to the same finite- or infinite-time cluster. Since $L(m)$ is a singleton for every $m \in \Sigma_+$, this implies statement I of Theorem 1.7 as a special case. The greater generality of this framework will pay dividends in the proofs of statement II and (especially) statement III.

Section 4 concerns the supercritical region $\Sigma_-$. We give the proof of part II of Theorem 1.7, which relies strongly on assumption (A4). We also prove that if $m$ lies inside the connected component $(m_-, m_+)$ of $\Sigma_-$, then no $t$-cluster at $m$ can extend beyond $(m_-, m_+]$ unless it contains all of $(m_-, m_+)$. Roughly speaking, this tells us that the interval $(m_-, m_+]$ can be treated as an indivisible unit for purposes of the larger-scale clustering analysis.

In Sect. 5, we prove part III of Theorem 1.7 (including Conjecture 1.6, and more). This is the only place in the paper where we specialize our assumptions on $\phi$ beyond (A1). The previously mentioned statement on the 'indivisibility' of the connected components of $\Sigma_-$ becomes extremely useful for proving III(i); in particular, it is essential when dealing with the situation where a connected component of $\Sigma_-$ borders an initial cluster or another connected component of $\Sigma_-$. Finally, the framework we have developed by the time we reach the proofs of statements III(ii) and III(iii) allows us to write the latter as adaptations of known arguments.

## 2. Preliminaries

In this section, we collect some preliminary results on the 1D Euler-alignment system (1) and the corresponding role of the 'sticky particle Cucker–Smale' dynamics.

First, however, we recall a few standard definitions and facts related to convex functions of a single variable, and we set a few conventions. We also justify our previous claim that $\Sigma_-$ has only finitely many connected components, as a consequence of the convexity assumption (A4).

## 2.1. Convex functions

**Definition 2.1.** (Convexity) Let $\Omega$ be an interval in $\mathbb{R}$. We say that $A : \Omega \to \mathbb{R}$ is *convex* if for every $m', m'' \in \Omega$ and every $\lambda \in (0, 1)$, we have

$$A((1 - \lambda)m' + \lambda m'') \leq (1 - \lambda)A(m') + \lambda A(m'')$$

We say $A$ is *strictly convex* if the above inequality is strict for all $m', m'' \in \Omega$ and all $\lambda \in (0, 1)$. If $I \subseteq \Omega$ is a subinterval of $\Omega$, then we say that $A$ is (strictly) convex on $I$ if the restriction of $A$ to $I$ is (strictly) convex. Finally, we say that $A$ is (strictly) convex in a neighborhood of $m \in \Omega$ if there exists an interval $I$, open relative to $\Omega$, such that $m \in I$ and $A$ is (strictly) convex on $I$.

The *lower convex envelope* of $A : \Omega \to \mathbb{R}$ is the largest convex function $A^{**} : \Omega \to \mathbb{R}$ whose graph lies below that of $A$:

$$A^{**}(m) = \sup \left\{ \widetilde{A}(m) \mid \widetilde{A} \text{ is convex and } \widetilde{A} \leq A \text{ on all of } [-\tfrac{1}{2}, \tfrac{1}{2}] \right\}.$$

We also recall the following very useful Lemma, which is a direct consequence of the definition.

**Lemma 2.2.** *Suppose $A$ is convex in an interval $I$. Let $[\widetilde{m}', \widetilde{m}'']$ and $[m', m'']$ be two sub-intervals of $I$ such that $\widetilde{m}' \leq m'$ and $\widetilde{m}'' \leq m''$. Then,*

$$\frac{A(\widetilde{m}'') - A(\widetilde{m}')}{\widetilde{m}'' - \widetilde{m}'} \leq \frac{A(m'') - A(m')}{m'' - m'}. \tag{10}$$

The following elementary lemma will be used several times in Sect. 3.

**Lemma 2.3.** *Let $A$ be a real-valued function defined on an interval containing $[m', m'']$. Suppose that for some $\theta \in (m', m'')$, it holds that*

$$(m', A(m')), \quad (\theta, A^{**}(\theta)), \quad and \quad (m'', A(m'')) \text{ are collinear.} \tag{11}$$

*Then $A(m') = A^{**}(m')$ and $A(m'') = A^{**}(m'')$, and consequently, $A^{**}$ is linear on $[m', m'']$.*

*Proof.* Pick $\lambda \in (0, 1)$ such that $\theta = (1 - \lambda)m' + \lambda m''$. Then

$$(1 - \lambda)A(m') + \lambda A(m'') = A^{**}(\theta)$$
$$\leq (1 - \lambda)A^{**}(m') + \lambda A^{**}(m'')$$
$$\leq (1 - \lambda)A(m') + \lambda A(m'').$$

Since the left and right sides are equal, this forces

$$(1 - \lambda)(A(m') - A^{**}(m')) + \lambda(A(m'') - A^{**}(m'')) = 0.$$

Since $\lambda \in (0, 1)$ and $A - A^{**} \geq 0$, we must have $A(m') = A^{**}(m')$ and $A(m'') = A^{**}(m'')$. Linearity of $A^{**}$ on $[m', m'']$ then follows from Lemma 2.2. □

The convexity assumption (A4) guarantees that the supercritical region $\Sigma_-$ has only finitely many connected components, as we establish presently.

**Lemma 2.4.** *The set $\Sigma_-$ has only finitely many connected components.*

*Proof.* Suppose not; then the set $S_r$ consisting of the right endpoints of the connected components of $\Sigma_-$ must have a limit point $m$ in $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Let $(m_j)_{j=1}^{\infty}$ be a sequence in $S_r$ that converges to $m$; we may assume without loss of generality that $(m_j)_{j=1}^{\infty}$ is strictly increasing. By assumption (A4), there exists an interval $I$ containing $m$ on which $A$ is convex. This interval contains $[m_{j-1}, m_j]$ for large enough $j$. We know that $A(m_{j-1}) = A^{**}(m_{j-1})$, $A(m_j) = A^{**}(m_j)$, and $A$ is convex on $[m_{j-1}, m_j]$; therefore, $A = A^{**}$ on $[m_{j-1}, m_j]$, whence $m_j \in \Sigma_0$. This is impossible, since $\Sigma_0 \cap S_r = \emptyset$. □

This lemma has the following obvious but useful consequence:

**Corollary 2.5.** *The set $\Sigma_+ \cup \Sigma_0$ is a union of half-open intervals of the form $(m', m'']$. In particular, if $m \in \Sigma_+ \cup \Sigma_0$, then there exists $m' < m$ such that $A = A^{**}$ on $[m', m]$.*

### 2.2. The sticky particle Cucker–Smale dynamics

The entropic solution to the Euler-alignment system (1) is compatible with the Cucker–Smale system of ODEs [13]

$$\frac{dx_i}{dt} = v_i, \quad \frac{dv_i}{dt} = \sum_{j=1}^{N} m_j \phi(x_j - x_i)(v_j - v_i), \quad i = 1, \ldots, N. \quad (12)$$

subject to a *sticky particle collision rule* (described below). The masses $m_i$ are fixed; the positions $x_i$ and velocities $v_i$ satisfy some initial conditions

$$x_i(0) = x_i^0, \quad v_i(0) = v_i^0, \quad i = 1, \ldots, N,$$

with

$$x_1^0 \leq x_2^0 \leq \cdots \leq x_N^0.$$

We do not require the $x_i^0$'s to be distinct, but we insist that

$$v_i^0 = v_j^0 \text{ whenever } x_i^0 = x_j^0. \quad (13)$$

We now specify the collision rules. Define the index cluster $J_i(t)$ to be the collection of indices associated with the agents which are stuck to agent $i$ at time $t$; let $i_*(t)$ and $i^*(t)$ denote the minimum and maximum of $J_i(t)$, respectively:

$$J_i(t) = \{j : x_j(t) = x_i(t)\} = \{i_*(t), i_*(t) + 1, \ldots, i^*(t)\}.$$

A *collision* occurs when $J_i(t)$ changes cardinality. We impose 'sticky particle' collision rules as follows:

- Each collision is *completely inelastic*, and agents stick to each other after collisions:

$$J_i(t) \supseteq J_i(s), \quad \text{whenever } t \geq s \geq 0; \tag{14}$$

- Collisions conserve momentum:

$$v_i(t) = \frac{\sum_{j \in J_i(t)} m_j v_j(t-)}{\sum_{j \in J_i(t)} m_j}. \tag{15}$$

For convenience, we also assume that the velocities are right continuous, i.e., $v_i(t) = v_i(t+)$. We also frequently use standard notation for time derivatives to indicate derivatives *from the right*, which always exist under our conventions, even though $v_i(t)$ may experience jump discontinuities.

Next, we introduce an important quantity

$$\psi_i(t) = v_i(t) + \sum_{j=1}^{N} m_j \Phi(x_i(t) - x_j(t)), \quad i = 1, \ldots, N. \tag{16}$$

It is conserved in time, in the sense described by the following Lemma.

**Lemma 2.6.** (Conservation of $\psi_i$) *For any non-collision time t, we have*

$$\frac{\mathrm{d}}{\mathrm{d}t} \psi_i(t) = 0, \quad \forall i = \{1, \ldots, N\}. \tag{17}$$

*For any collision time t (and in fact, any time), we have*

$$\psi_i(t) = \frac{\sum_{j \in J_i(t)} m_j \psi_j(t-)}{\sum_{j \in J_i(t)} m_j} = \frac{\sum_{j \in J_i(t)} m_j \psi_j^0}{\sum_{j \in J_i(t)} m_j}, \quad \forall i = \{1, \ldots, N\}. \tag{18}$$

The following lemma describes the behavior at a collision.

**Lemma 2.7.** (Barycentric lemma) *For any $i \in \{1, \ldots, N\}$ and any $t \geq 0$, we have*

$$\frac{\sum_{\ell=i_*(t)}^{j} m_\ell \psi_\ell(t-)}{\sum_{\ell=i_*(t)}^{j} m_\ell} \geq \psi_i(t) \geq \frac{\sum_{\ell=j+1}^{i^*(t)} m_\ell \psi_\ell(t-)}{\sum_{\ell=j+1}^{i^*(t)} m_\ell}, \quad \forall j = \{i_*(t), \cdots, i^*(t)\}. \tag{19}$$

A different version of the barycentric Lemma was previously used by Brenier and Grenier [5,17] to analyze the 1D pressureless Euler equations. The latter can be recovered as a special case of (1) when $\phi \equiv 0$, in which case we also have $\psi_i = v_i$. The inequality (19) then means that when collision occurs, the average velocity of the left group of particles has to be larger than the average velocity of the right group of particles. As noticed by the authors in [29], the barycentric lemma extends to the case of general (locally integrable) communication protocols $\phi$; this extension is the statement recorded in Lemma 2.7.

The properties detailed in Lemmas 2.6 and 2.7 endow the quantities $\psi_i$ with crucial information about collisions and cluster formation for the discrete sticky particle Cucker–Smale system. The consequences of these properties will be thoroughly investigated in this paper.

Before moving on, we pause to record a simple identity we will use repeatedly:

$$
\frac{\mathrm{d}}{\mathrm{d}t}(x_j(t) - x_i(t)) = \psi_j(t) - \psi_i(t)
$$

$$
- \sum_{\ell=1}^{N} m_\ell \int_{x_i(t)}^{x_j(t)} \phi(y - x_\ell(t)) \, \mathrm{d}y, \qquad \forall i, j \in \{1, \ldots, N\}.
$$

$$(20)$$

### 2.3. Atomic solutions of the Euler-alignment system

We recall the following connection between the Euler-alignment system (1) and the sticky particle Cucker–Smale dynamics (12).

**Proposition 2.8.** ( [29]) *Consider the Euler-alignment system* (1) *with atomic initial data*

$$
\rho_N^0(x) = \sum_{i=1}^{N} m_{i,N} \delta(x - x_{i,N}^0), \qquad P_N^0(x) := \rho_N^0 u_N^0(x) = \sum_{i=1}^{N} m_{i,N} v_{i,N}^0 \delta(x - x_{i,N}^0).
$$

$$(21)$$

*There exists a unique entropic solution*

$$
\rho_N(x, t) = \sum_{i=1}^{N} m_{i,N} \delta(x - x_{i,N}(t)),
$$

$$
P_N(x, t) = \rho_N u_N(x, t) = \sum_{i=1}^{N} m_{i,N} v_{i,N}(t) \delta(x - x_{i,N}(t)), \qquad (22)
$$

*where* $(x_{i,N}(t), v_{i,N}(t))_{i=1}^{N}$ *is the solution to the sticky particle Cucker–Smale dynamics* (12) *with initial data* $(m_{i,N}, x_{i,N}^0, v_{i,N}^0)_{i=1}^{N}$.

For initial data of the form (21), the cumulative distribution function $M_N^0$ and its generalized inverse $X_N^0$ (defined as in (7)) are piecewise constant functions. We write

out their formulas presently. Define

$$\theta_{i,N} = -\frac{1}{2} + \sum_{j=1}^{i} m_{j,N}, \quad i = 0, \dots, N, \tag{23}$$

so that

$$m_{i,N} = \theta_{i,N} - \theta_{i-1,N}, \quad i = 1, \dots, N. \tag{24}$$

Then, the corresponding initial data for the scalar balance law (6), and its generalized inverse, are

$$M_N^0(x) = -\frac{1}{2} + \sum_{i=1}^{N} m_{i,N} \mathbb{1}_{[0,\infty)}(x - x_{i,N}^0)$$

$$= -\frac{1}{2} \mathbb{1}_{(-\infty, x_{1,N}^0)}(x) + \sum_{i=1}^{N} \theta_{i,N} \mathbb{1}_{[x_{i,N}^0, x_{i+1,N}^0)}(x), \tag{25}$$

$$X_N^0(m) = \sum_{i=1}^{N} x_{i,N}^0 \mathbb{1}_{(\theta_{i-1,N}, \theta_{i,N}]}(m). \tag{26}$$

Here, we use the convention $x_{N+1,N}^0 = +\infty$. From Definition 1.2, the initial data include a cluster of size $\sum_{j \in J_i(0)} m_{j,N}$ located at each $x_{i,N}^0$. The blue step function in the left subplot of Fig. 2 shows a typical scenario.

Using the atomic initial data (21) in the formula (8) for the flux yields a piecewise linear function $A_N$ with the formula

$$A_N(m) = \sum_{j=1}^{i-1} m_{j,N} \psi_{j,N}^0 + (m - \theta_{i-1,N}) \psi_{i,N}^0,$$

$$\theta_{i-1,N} < m \leq \theta_{i,N}, \quad i = 1, \dots, N, \tag{27}$$

for $m \in \left(-\frac{1}{2}, \frac{1}{2}\right]$, and $A_N(-\frac{1}{2}) = 0$. Here the $\psi_{i,N}^0$'s are defined by:

$$\psi_{i,N}^0 = v_{i,N}^0 + \sum_{j=1}^{N} m_{j,N} \Phi(x_{i,N}^0 - x_{j,N}^0), \quad i = 1, \dots, N. \tag{28}$$

We also define $\psi_{i,N}(t)$ analogously to (16):

$$\psi_{i,N}(t) := v_{i,N}(t) + \sum_{j=1}^{N} m_{j,N} \Phi(x_{i,N}(t) - x_{j,N}(t)), \quad i = 1, \dots, N. \tag{29}$$

Note that we have

$$m_{i,N} \psi_{i,N}^0 = A_N(\theta_{i,N}) - A_N(\theta_{i-1,N}), \quad \text{i.e.,} \quad \psi_{i,N}^0 = \frac{A_N(\theta_{i,N}) - A_N(\theta_{i-1,N})}{\theta_{i,N} - \theta_{i-1,N}}. \tag{30}$$

Hence, $\psi_{i,N}^0$ is the slope of the line segment comprising the graph of $A_N$ in $[\theta_{i-1,N}, \theta_{i,N}]$. For discrete initial data, the flux $A_N$ (together with the masses $(m_{i,N})_{i=1}^N$) therefore carries the information of $(\psi_{i,N}^0)_{i=1}^N$. See the blue piecewise linear curve in the right subplot of Fig. 2 for an illustration. We also note that this process yields initial veloc-ities which satisfy (13), as can be easily checked.

When collisions occur, we apply (18) and (30) to get

$$\psi_{i,N}(t) = \frac{\sum_{j\in J_i(t)} m_{j,N}\psi_{j,N}^0}{\sum_{j\in J_i(t)} m_{j,N}} = \frac{A_N(\theta_{i^*(t),N}) - A_N(\theta_{i_*(t)-1,N})}{\theta_{i^*(t),N} - \theta_{i_*(t)-1,N}}. \qquad (31)$$

In other words, $\psi_{i,N}(t)$ is the slope of secant line through the graph of $A_N$ between the two points $(\theta_{i_*(t)-1,N}, A_N(\theta_{i_*(t)-1,N}))$ and $(\theta_{i^*(t),N}, A_N(\theta_{i^*(t),N}))$. For instance, as illustrated in Fig. 2, if agents 2 and 3 stick at time $t$ (and are not stuck to any other agents at time $t$), then

$$\psi_{2,N}(t) = \psi_{3,N}(t) = \frac{m_{2,N}\psi_{2,N}^0 + m_{3,N}\psi_{3,N}^0}{m_{2,N} + m_{3,N}}$$

is the slope of the red secant line through $(\theta_{1,N}, A_N(\theta_{1,N}))$ and $(\theta_{3,N}, A_N(\theta_{3,N}))$.

## 2.4. The sticky particle approximation

We now reinstate assumptions (A1)–(A4) and consider a *sequence* of atomic solu-tions $(\rho_N, u_N)_{N=1}^\infty$ that approximate the solution $(\rho, u)$ of interest. As outlined in the previous subsection, generating an atomic solution $(\rho_N, u_N)$ for any given $N$ amounts to choosing initial data $(m_{i,N}, x_{i,N}^0, v_{i,N}^0)_{i=1}^N$ and running the dynamics for the sticky particle Cucker–Smale system. We now present our choice of sticky particle initial data and make precise the sense in which the corresponding $(\rho_N, u_N)$ approximates $(\rho, u)$.

We choose our $m_{i,N}$'s in such a way that

$$\lim_{N\to\infty} \max_{1\le i\le N} m_{i,N} = 0. \qquad (D1)$$

Then, we define $\theta_{i,N}$ as in (23). For large enough $N$ (say $N \geq N_0$), we may always choose the $m_{i,N}$'s to satisfy the following additional hypothesis:

$$S = \partial\Sigma_- \subset \{\theta_{i,N}\}_{i=0}^N, \qquad N \geq N_0. \qquad (D2)$$

We define the remaining sticky particle data in terms of $(M^0, A)$ (which we recall are defined in (7)–(8)), and the $m_{i,N}$'s:

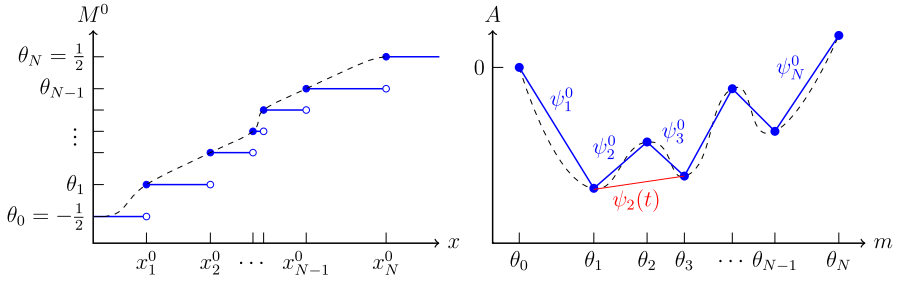$$x_{i,N}^0 = \inf\{x \in \mathbb{R} : M^0(x) \geq \theta_{i,N}\} = X^0(\theta_{i,N}), \qquad (D3)$$

Figure 2. An illustration of the discretization and the flux, with $N = 6$. We have dropped a subscript $N$ on most of the discretized quantities. Left: $M^0$ (dashed) and its piecewise constant discretization $M_N$ (solid). Right: The flux $A$ (dashed) and its piecewise linear discretization $A_N$ (solid). The slope of $A$ between the breakpoints $\theta_{j-1}$ and $\theta_j$ determines the value of $\psi_j$. If agents $j$ and $j + 1$ collide, the values of $\psi_j(t)$ and $\psi_{j+1}(t)$ are adjusted accordingly; for example, the figure shows the value of $\psi_2(t) = \psi_3(t)$ after agents 2 and 3 have collided

and finally

$$\psi_{i,N}^0 = \frac{A(\theta_{i,N}) - A(\theta_{i-1,N})}{\theta_{i,N} - \theta_{i-1,N}}, \tag{D4}$$

$$v_{i,N}^0 = \psi_{i,N}^0 - \sum_{j=1}^{N} m_{j,N} \Phi(x_{i,N}^0 - x_{j,N}^0), \qquad i = 1, \ldots, N.$$

This gives us all the information we need in order to define $M_N^0$, $X_N^0$, and $A_N$, for each $N \in \mathbb{N}$, using exactly the formulas of the previous subsection.

Our discretization scheme has several crucial properties. First of all, (D1) and (D3) guarantee that

$$\|M_N^0 - M^0\|_{L^1(\mathbb{R})} = \|X_N^0 - X^0\|_{L^1(-\frac{1}{2}, \frac{1}{2})} \to 0, \qquad \text{as } N \to \infty. \tag{32}$$

Next, (D4) guarantees that $A_N$ (defined in (27)) agrees with $A$ at the breakpoints $\theta_{i,N}$ and is linear in between:

$$A_N(\theta_{i,N}) = A(\theta_{i,N}), \; i = 0, \ldots, N;$$
$$\text{and } A_N \text{ is linear on } [\theta_{i-1}, \theta_i], \; \forall i \in \{1, \ldots, N\}. \tag{33}$$

Therefore, $A_N \to A$ uniformly as $N \to \infty$. See Fig. 2 for an illustration of the approximation: For initial data $(M^0, A)$ indicated by the dashed curves, the solid blue graphs of $(M_N^0, A_N)$ serve as approximations.

As the authors proved in [29], the entropy solutions generated by discretized initial data satisfying (D1), (D3), (D4) are a good approximation of the true solution. Letting

$(x_{i,N}(t), v_{i,N}(t))_{i=1}^{N}$ denote the sticky particle Cucker–Smale dynamics associated with the initial data $(m_{i,N}, x_{i,N}^0, v_{i,N}^0)_{i=1}^{N}$, the entropy solution of (6) associated with the initial data $M_N^0$ and flux $A_N$ defined above is

$$M_N(x, t) = -\frac{1}{2} + \sum_{i=1}^{N} m_{i,N} \mathbb{1}_{[0,\infty)}(x - x_{i,N}(t)).$$

Its generalized inverse is

$$X_N(m, t) = \sum_{i=1}^{N} x_{i,N}(t) \mathbb{1}_{(\theta_{i-1,N}, \theta_{i,N}]}(m).$$

We have the following statement from [29]:

**Proposition 2.9.** *For any fixed $t \geq 0$, the solutions $M_N(\cdot, t)$ approximate $M(\cdot, t)$ in the following sense:*

$$M_N(\cdot, t) - M(\cdot, t) \to 0 \;\; in \; L^1(\mathbb{R}). \tag{34}$$

The convergence (34) is equivalent to saying that $(\rho_N(t))_{N=1}^{\infty}$ converges to $\rho(t)$ in the Wasserstein-1 metric, and it was also proved in [29] that $(\rho_N u_N(t))_{N=1}^{\infty}$ converges weak-$*$ in the sense of measures to $\rho u(t)$. The convergence (34) is what we use below. More precisely, we note that since

$$\|X_N(\cdot, t) - X(\cdot, t)\|_{L^1(-\frac{1}{2}, \frac{1}{2})} = \|M_N(\cdot, t) - M(\cdot, t)\|_{L^1(\mathbb{R})},$$

the convergence (34) implies the existence of a subsequence $(X_{N_k}(\cdot, t))_{k=1}^{\infty}$ such that

$$X_{N_k}(m, t) \to X(m, t) \text{ as } k \to \infty, \quad \text{for almost every } m \in (-\tfrac{1}{2}, \tfrac{1}{2}]. \tag{35}$$

We will be able to leverage this almost-everywhere convergence, combined with the monotonicity of $X(\cdot, t)$, to analyze the formation of clusters in the sense of Definition 1.2.

We end this subsection by discussing the significance of the requirement (D2) in our approximation scheme. It guarantees that

$$\{m : A(m) > A^{**}(m)\} = \{m : A_N(m) > A_N^{**}(m)\}, \tag{36}$$

so that the supercritical region of mass labels associated with $(\rho_N, u_N)$ is the same as the one for $(\rho, u)$.

On account of (33) and (36), it becomes superfluous for our purposes to set additional notation for the discretized flux; we write our analysis below in terms of $A$ only.

2.5. A maximum principle

We record one final preliminary statement here; it is a uniform-in-$N$ maximum principle for the velocities associated with our discretization.

**Lemma 2.10.** *Given a sequence of discretizations satisfying* (D1)–(D4)*, there exist $N$-independent constants $u^0_{max}$ and $u^0_{min}$ such that*

$$u^0_{min} \le \min_{1 \le i \le N} v^0_{i,N} \le v_{i,N}(t) \le \max_{1 \le i \le N} v^0_{i,N} \le u^0_{max}, \qquad \forall i \in \{1, \dots, \mathbb{N}\}, \quad t \ge 0. \tag{37}$$

We refer to [29] for the details of the proof. Furthermore, we denote the initial velocity variation

$$\mathcal{V}^0 := u^0_{max} - u^0_{min}. \tag{38}$$

## 3. The subcritical regime

In this section, we prove part I of Theorem 1.7 by establishing a more general statement, namely that mass labels from distinct $L(m)$'s can never belong to the same finite- or infinite-time cluster. Now, it may of course happen that $L(m') \neq L(m'')$ without either of $m'$ or $m''$ belonging to $\Sigma_+$, but such labels are 'related subcritically,' so to speak, in that their images remain separated under $X(\cdot, t)$ for essentially the same reasons as do different mass labels in $\Sigma_+$. As already noted above, $L(m)$ is a singleton for each $m \in \Sigma_+$, so establishing the statement mentioned above proves in particular that there can never be *any* finite- or infinite-time clusters in $\Sigma_+$.

We begin by proving a discrete version of this more general statement. The finite-time part (Proposition 3.1) can be deduced from the barycentric lemma only—no explicit reference to the equations governing the dynamics is required. We bootstrap our finite-time statement into an infinite-time one with the aid of (20), proving that the distance between agents associated to different $L(m)$'s does not tend to zero, and thus that agents from different $L(m)$'s cannot be part of the same *infinite*-time cluster either. Lemma 3.4 gives a time-independent lower bound on the distance between such agents; we use it in the proof of the continuum version of our statement, which we establish in Sect. 3.2.

In this section, we work with a fixed entropically selected solution $(\rho, u)$ of (1) associated with initial data $(\rho^0, u^0)$, and we assume throughout that (A1)–(A4) hold.

3.1. The discrete setting

In this subsection, we fix $N \in \mathbb{N}$ large and a discretization $(m_{i,N}, x^0_{i,N}, v^0_{i,N})^N_{i=1}$ of $(\rho^0, u^0)$ satisfying (D2)–(D4). We let $(x_{i,N}(t), v_{i,N}(t))^N_{i=1}$ denote the associated sticky particle Cucker–Smale dynamics.

Our first main statement of this subsection is the following proposition.

**Proposition 3.1.** *Fix $i \in \{1, \dots, N\}$ and $t \ge 0$. If $i_*(t) < i^*(t)$, then $A^{**}$ is linear on $[\theta_{i_*(t)-1,N}, \theta_{i^*(t),N}]$.*

This proposition says that, at the discrete level, finite-time collisions between agents are confined to a single $L(m)$. In particular, no discrete finite-time clustering can

occur anywhere in $\Sigma_+$. Most of the work involved in proving the proposition goes into establishing the case considered in the following lemma. The slightly stronger conclusion available in this case will also be useful later in Sect. 4.

**Lemma 3.2.** *If $i < i^*(t)$ and $\theta_{i,N} \notin \Sigma_-$, then $\theta_{i_*(t)-1,N}, \theta_{i^*(t),N} \notin \Sigma_-$ as well, i.e.,*

$$A(\theta_{i_*(t)-1,N}) = A^{**}(\theta_{i_*(t)-1,N}) \qquad and \qquad A(\theta_{i^*(t),N}) = A^{**}(\theta_{i^*(t),N}). \qquad (39)$$

*Consequently, we have that*

$$A^{**} \text{ is linear on } [\theta_{i_*(t)-1,N}, \theta_{i^*(t),N}]. \qquad (40)$$

The 'consequently' claim here follows from Lemma 2.3. Note that $i < i^*(t)$ implies that $X_N(\cdot, t)$ is constant on some interval containing $(\theta_{i-1,N}, \theta_{i+1,N}]$. In particular, $\theta_{i,N}$ lies in the interior of this interval. This is why we assume $i < i^*(t)$ rather than $i_*(t) < i^*(t)$ here.

Using Lemma 3.2, we give the short proof of the rest of Proposition 3.1. Then we prove the lemma.

*Proof of Proposition 3.1.* If $\theta_{j,N} \in \Sigma_-$ for all $j \in \{i_*(t), \ldots, i^*(t)-1\}$, then assumption (D2) implies that $(\theta_{i_*(t)-1,N}, \theta_{i^*(t),N})$ must lie in a single connected component of $\Sigma_-$, on which $A^{**}$ is linear, so we are done. Otherwise, we have $\theta_{j,N} \notin \Sigma_-$ for some $j \in \{i_*(t), \ldots, i^*(t) - 1\}$. Then since $j < i^*(t) = j^*(t)$, the conclusion of Lemma 3.2 holds, with $i$ replaced by $j$. But since agents $i$ and $j$ have collided at time $t$, the intervals $[\theta_{i_*(t)-1,N}, \theta_{i^*(t),N}]$ and $[\theta_{j_*(t)-1,N}, \theta_{j^*(t),N}]$ are the same. This completes the proof. $\qquad \square$

*Proof of Lemma 3.2.* Let $t_1$ denote the first time satisfying $i < i^*(t_1)$, and let $t_2, \ldots, t_m$ denote any subsequent times (if any) where agent $i$ is involved in a collision. We prove that (39) holds at each time $t_n$, $n = 1, \ldots, m$, which will prove that (39) is valid on $[t_1, t_2), [t_2, t_3), \ldots, [t_m, +\infty)$ and thus at any time $t$ such that $i < i^*(t)$. We argue inductively on $n \in \{1, \ldots, m\}$, splitting our 'base case' into two subcases, namely $t_1 = 0$ and $t_1 > 0$.

If $t_1 = 0$, then our discretization procedure ensures that $\psi^0_{i_*(0),N} = \cdots = \psi^0_{i^*(0),N}$, which guarantees that the points $(\theta_{i_*(0)-1,N}, A(\theta_{i_*(0)-1,N}), (\theta_{i,N}, A(\theta_{i,N}))$, and $(\theta_{i^*(0),N}, A(\theta_{i^*(0),N})$ are collinear. Remembering that $A(\theta_{i,N}) = A^{**}(\theta_{i,N})$ and applying Lemma 2.3 implies that (39) holds at time $t_1 = 0$.

If $t_1 > 0$, then

$$\frac{\sum_{j=i+1}^{i^*(t_1)} m_{j,N} \psi_{j,N}(t_1-)}{\sum_{j=i+1}^{i^*(t_1)} m_{j,N}} = \frac{\sum_{j=i+1}^{i^*(t_1)} m_{j,N} \psi^0_{j,N}}{\sum_{j=i+1}^{i^*(t_1)} m_{j,N}} = \frac{A(\theta_{i^*(t_1),N}) - A(\theta_{i,N})}{\theta_{i^*(t_1),N} - \theta_{i,N}}$$

and similarly

$$\frac{\sum_{j=i_*(t_1)}^{i} m_{j,N} \psi_{j,N}(t_1-)}{\sum_{j=i_*(t_1)}^{i} m_{j,N}} = \frac{A(\theta_{i,N}) - A(\theta_{i_*(t_1)-1,N})}{\theta_{i,N} - \theta_{i_*(t_1)-1,N}}.$$

The barycentric lemma then tells us that

$$\frac{A(\theta_{i^*(t_1),N}) - A(\theta_{i,N})}{\theta_{i^*(t_1),N} - \theta_{i,N}} \leq \frac{A(\theta_{i,N}) - A(\theta_{i_*(t_1)-1,N})}{\theta_{i,N} - \theta_{i_*(t_1)-1,N}}.$$

On the other hand, our assumption that $A(\theta_{i,N}) = A^{**}(\theta_{i,N})$, the fact that $A \geq A^{**}$ in general, and Lemma 2.2 combine to give us

$$\frac{A(\theta_{i^*(t_1),N}) - A(\theta_{i,N})}{\theta_{i^*(t_1),N} - \theta_{i,N}} \geq \frac{A^{**}(\theta_{i^*(t_1),N}) - A^{**}(\theta_{i,N})}{\theta_{i^*(t_1),N} - \theta_{i,N}}$$

$$\geq \frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i_*(t_1)-1,N})}{\theta_{i,N} - \theta_{i_*(t_1)-1,N}} \geq \frac{A(\theta_{i,N}) - A(\theta_{i_*(t_1)-1,N})}{\theta_{i,N} - \theta_{i_*(t_1)-1,N}}.$$

The inequalities here thus must actually be equalities. This forces (39) to hold at time $t_1$.

Assume inductively that (39) holds at time $t_n$ for some $n < m$. Denote $j = i_*(t_n)-1$. If agent $i$ collides with an agent on its left at time $t_{n+1}$, then $t_{n+1}$ is the first time when agent $j$ collides with agent $j + 1 = i_*(t_n)$, and furthermore, we have $A(\theta_{j,N}) = A^{**}(\theta_{j,N})$ by our inductive assumption.

Applying the conclusion from our 'base case' to agent $j$, we conclude that $A^{**}$ is linear on the interval $[\theta_{j_*(t_{n+1})-1,N}, \theta_{j^*(t_{n+1}),N}]$, which is the same as $[\theta_{i_*(t_{n+1})-1,N}, \theta_{i^*(t_{n+1}),N}]$, since agents $i$ and $j$ collide at time $t_{n+1}$. Furthermore,

$$A(\theta_{i_*(t_{n+1})-1,N}) = A(\theta_{j_*(t_{n+1})-1,N}) = A^{**}(\theta_{j_*(t_{n+1})-1,N}) = A^{**}(\theta_{i_*(t_{n+1})-1,N}),$$

and similarly $A(\theta_{i^*(t_{n+1}),N}) = A^{**}(\theta_{i^*(t_{n+1}),N})$. We have thus proved the inductive step in the case where agent $i$ experiences a collision from its left at time $t_{n+1}$. The case of a collision from the right is entirely similar. □

We now turn our attention to (the absence of) infinite-time clustering, for which we must rely on equation (20). The following corollary of Lemma 3.2 will help us control the difference $\psi_{j,N}(t) - \psi_{i,N}(t)$ when $\theta_{i,N}$ and $\theta_{j,N}$ belong to different $L(m)$'s (and in particular when they lie in $\Sigma_+$).

**Corollary 3.3.** *Assume that $\theta_{i,N} \notin \Sigma_-$. Then*

$$\psi_{i,N}(t) \leq \frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i-1,N})}{\theta_{i,N} - \theta_{i-1,N}}, \qquad \forall t \geq 0, \ \ if \ i \in \{1, \ldots, N\}; \tag{41}$$

$$\psi_{i+1,N}(t) \geq \frac{A^{**}(\theta_{i+1,N}) - A^{**}(\theta_{i,N})}{\theta_{i+1,N} - \theta_{i,N}} \qquad \forall t \geq 0, \ \ if \ i \in \{0, \ldots, N-1\}. \tag{42}$$

*Proof.* We prove only (41); the proof of (42) is similar. Choose $t \geq 0$. If $i^*(t) = i$, then

$$\psi_{i,N}(t) = \frac{A(\theta_{i,N}) - A(\theta_{i_*(t)-1,N})}{\theta_{i,N} - \theta_{i_*(t)-1,N}} \leq \frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i_*(t)-1,N})}{\theta_{i,N} - \theta_{i_*(t)-1,N}}$$

$$\leq \frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i-1,N})}{\theta_{i,N} - \theta_{i-1,N}}.$$

If $i^*(t) > i$, then the previous lemma implies that $A^{**}$ is linear on $[\theta_{i_*(t)-1,N}, \theta_{i^*(t),N}]$, which of course contains $[\theta_{i-1,N}, \theta_{i,N}]$. Therefore,

$$\psi_{i,N}(t) = \frac{A(\theta_{i^*(t),N}) - A(\theta_{i_*(t)-1,N})}{\theta_{i^*(t),N} - \theta_{i_*(t)-1,N}} \leq \frac{A^{**}(\theta_{i^*(t),N}) - A^{**}(\theta_{i_*(t)-1,N})}{\theta_{i^*(t),N} - \theta_{i_*(t)-1,N}}$$

$$= \frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i-1,N})}{\theta_{i,N} - \theta_{i-1,N}}.$$

This completes the proof. $\qquad\square$

The following lemma is our second main statement of this subsection. It gives a time-independent lower bound on the distance between agents from different $L(m)$'s and in particular shows that such agents cannot belong to the same infinite-time cluster.

**Lemma 3.4.** *Fix $i, j \in \{1, \ldots, N\}$ with $i \leq j$. Assume that there exists $\sigma > 0$ such that*

$$\frac{A^{**}(\theta_{j+1,N}) - A^{**}(\theta_{j,N})}{\theta_{j+1,N} - \theta_{j,N}} - \frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i-1,N})}{\theta_{i,N} - \theta_{i-1,N}} \geq 2\sigma > 0. \qquad (43)$$

*Then agents $i$ and $j + 1$ never collide. In fact, we have the lower bound*

$$x_{j+1,N}(t) - x_{i,N}(t) \geq \max\left\{x_{j+1,N}^0 - x_{i,N}^0 - t\mathcal{V}^0, \min\{t\sigma, \eta\}\right\} \geq c > 0, \qquad \forall t \geq 0, \qquad (44)$$

*where $\mathcal{V}^0$ is defined in (38) and $\eta > 0$ is chosen so that $\left|\int_z^w \phi(r)\, dr\right| < \sigma$ whenever $|z - w| < \eta$.*

*Proof of Lemma 3.4.* The fact that agents $i$ and $j + 1$ do not collide in finite time is a consequence of Proposition 3.1, since (43) implies that $A^{**}$ is not linear on any interval containing both $\theta_{i-1,N}$ and $\theta_{j,N}$. We need only to prove the lower bound. In fact, it is clear that

$$x_{j+1,N}(t) - x_{i,N}(t) \geq x_{j+1,N}^0 - x_{i,N}^0 - t\mathcal{V}^0.$$

We therefore concentrate on proving that

$$x_{j+1,N}(t) - x_{i,N}(t) \geq \min\{t\sigma, \eta\}. \qquad (45)$$

We first give the proof under the additional assumption that $\theta_{i,N}, \theta_{j,N} \notin \Sigma_-$. In this case, Corollary 3.3 tells us that

$$\psi_{j+1,N}(t) - \psi_{i,N}(t) \geq 2\sigma > 0, \qquad \forall t \geq 0.$$

Therefore, if $\tau$ is any time such that $x_{j+1,N}(\tau) - x_{i,N}(\tau) < \eta$, then the identity (20) and the previous step tell us that

$$\frac{d}{ds}(x_{j+1,N}(s) - x_{i,N}(s))\bigg|_{s=\tau} = \psi_{j+1,N}(\tau) - \psi_{i,N}(\tau)$$

$$- \sum_{\ell=1}^N m_{\ell,N} \int_{x_{i,N}(\tau)}^{x_{j+1,N}(\tau)} \phi(y - x_{\ell,N}(\tau))\, dy \geq \sigma.$$

It follows immediately that

$$x_{j+1,N}(t) - x_{i,N}(t) \geq \min\{x^0_{j+1,N} - x^0_{i,N} + t\sigma, \eta\}, \qquad \forall t \geq 0, \qquad (46)$$

which in particular implies (45).

In the more general setting where one or both of $\theta_{i,N}, \theta_{j,N}$ may belong to $\Sigma_-$, we replace $i$ and $j$ with the closest indices $I$ and $J$ satisfying $i \leq I \leq J \leq j$ and $\theta_{I,N}, \theta_{J,N} \notin \Sigma_-$. This is possible since $\theta_{i,N}$ and $\theta_{j,N}$ cannot belong to the same connected component of $\Sigma_-$; otherwise $A^{**}$ would be linear on $[\theta_{i-1,N}, \theta_{j,N}]$, contrary to our assumption.

More specifically, if $\theta_{i,N} \in \Sigma_-$, we choose $I$ so that $\theta_{I,N}$ is the right endpoint of the connected component of $\Sigma_-$ to which $\theta_{i,N}$ belongs. If $\theta_{j,N} \in \Sigma_-$, then we choose $J$ so that $\theta_{J,N}$ is the corresponding left endpoint. The point is that now we have $\theta_{I,N}, \theta_{J,N} \notin \Sigma_-$ and

$$\frac{A^{**}(\theta_{J+1,N}) - A^{**}(\theta_{J,N})}{\theta_{J+1,N} - \theta_{J,N}} - \frac{A^{**}(\theta_{I,N}) - A^{**}(\theta_{I-1,N})}{\theta_{I,N} - \theta_{I-1,N}}$$
$$= \frac{A^{**}(\theta_{j+1,N}) - A^{**}(\theta_{j,N})}{\theta_{j+1,N} - \theta_{j,N}} - \frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i-1,N})}{\theta_{i,N} - \theta_{i-1,N}},$$

so that, applying the logic of the previous case, we have

$$x_{j+1,N}(t) - x_{i,N}(t) \geq x_{J+1,N}(t) - x_{I,N}(t) \geq \min\{x^0_{J+1,N} - x^0_{I,N} + t\sigma, \eta\}$$
$$\geq \min\{t\sigma, \eta\},$$

as needed. This completes the proof. □

*Remark 3.1.* If $\theta_{i,N}, \theta_{j,N} \in \Sigma_+$ and $i \leq j$, then the hypotheses of Lemma 3.4 are satisfied automatically for some $\sigma > 0$. Indeed, if the left side of (43) were equal to zero, it would force $A^{**}$ to be linear on $[\theta_{i-1,N}, \theta_{j+1,N}]$ contradicting the definition of $\Sigma_+$.

## 3.2. The continuum setting

We are now ready to prove the full continuum version of the statement that mass labels from distinct $L(m)$'s never belong to the same cluster. We first show in Lemma 3.5 that this is true at time zero; then, we give the full statement in Theorem 3.6.

**Lemma 3.5.** *Let $(\rho, u)$ and all related notation be defined as in assumptions (A1)– (A4). Fix $m'' \in (-\frac{1}{2}, \frac{1}{2}]$ and define $m''_L = \inf L(m'')$. Then*
   (i) $A(m''_L) = A^{**}(m''_L)$. *(I.e., $m''_L \notin \Sigma_-$.)*
   (ii) *If $m' < m''_L$, then $X^0(m') < X^0(m'')$.*

*Proof.*   (i) If $A(m''_L) > A^{**}(m''_L)$, then $m''_L \in \Sigma_-$. Let $(m_-, m_+)$ denote the connected component of $\Sigma_-$ containing $m''_L$. Then $A^{**}$ is linear on the nontrivially overlapping intervals $[m_-, m_+]$ and $[m''_L, m'']$, hence on their union $[m_-, m'']$. Since $m_- < m''_L$, this contradicts the definition of $m''_L$.

(ii) We again argue by contradiction. Suppose $X^0(m') = X^0(m'')$; then, the definition (8) of $A$ implies that $A$ is linear on $[m', m'']$. In particular, $(m''_L, A(m''_L)) = (m''_L, A^{**}(m''_L))$ lies on the segment joining $(m', A(m'))$ and $(m'', A(m''))$; by Lemma 2.3, it follows that $A^{**}$ is linear on $[m', m'']$, contradicting the definition of $m''_L$.

$\square$

**Theorem 3.6.** *Let $(\rho, u)$ and all related notation be defined as in assumptions (A1)–(A4). Fix $m'' \in (-\frac{1}{2}, \frac{1}{2}]$. If $m' < \inf L(m'')$, then there exists a time-independent constant $c > 0$ such that*

$$X(m'', t) - X(m', t) \geq c > 0, \quad \forall\, t \geq 0. \tag{47}$$

*Proof.* Define

$$m''_L := \inf L(m'').$$

There are two cases to consider, namely $m''_L \in \Sigma_+$ and $m''_L \notin \Sigma_+$. In either case, we choose a discretization satisfying (D1)–(D4) *and additionally,* $m''_L \in \{\theta_{i,N}\}_{i=0}^N$ for sufficiently large $N$. (This will only be actually used in the second case.) For each such $N$, we choose $\ell$ such that $\theta_{\ell,N} = m''_L$. (This $\ell$ will of course depend on $N$, but we suppress this in the notation.) Proposition 3.1 guarantees that agents $\ell$ and $\ell + 1$ can never collide.

We split the remainder of the argument into the two cases mentioned above.

*Case 1: $m''_L \in \Sigma_+$.* In this case, we may assume without loss of generality that $m'' = m''_L \in \Sigma_+$ and that $A = A^{**}$ on $[m', m'']$.

*Step 1.1.* Fix $\widetilde{m}'$ and $\overline{m}$ such that

$$m' < \widetilde{m}' < \overline{m} < m''.$$

Then by Lemma 2.2, together with the fact that $A$ is *not* linear on $[\widetilde{m}', m'']$, we deduce that

$$\frac{A(m'') - A(\overline{m})}{m'' - \overline{m}} > \frac{A(\overline{m}) - A(\widetilde{m}')}{\overline{m} - \widetilde{m}'}.$$

Choose $\widetilde{m}'' \in (\overline{m}, m'')$ close enough to $m''$ so that

$$\frac{A(\widetilde{m}'') - A(\overline{m})}{\widetilde{m}'' - \overline{m}} > \frac{A(\overline{m}) - A(\widetilde{m}')}{\overline{m} - \widetilde{m}'},$$

and let $2\sigma$ denote the difference between the left and right sides:

$$2\sigma := \frac{A(\widetilde{m}'') - A(\overline{m})}{\widetilde{m}'' - \overline{m}} - \frac{A(\overline{m}) - A(\widetilde{m}')}{\overline{m} - \widetilde{m}'} > 0. \tag{48}$$

*Step 1.2.* For each $N \in \mathbb{N}$, let $(m_{i,N}, x^0_{i,N}, v^0_{i,N})_{i=1}^N$ be a discretization of $(\rho^0, u^0)$, and assume that these discretizations satisfy (D1)–(D4). Using the notation of Sect. 2,

let us fix a time $t > 0$ and choose a corresponding subsequence $(N_k)_{k=1}^{\infty}$ such that the a.e. convergence (35) holds at time zero and time $t$. Then we can find $\theta' \in (m', \widetilde{m}')$ and $\theta'' \in (\widetilde{m}'', m'')$, such that

$$
\begin{aligned}
X_{N_k}(\theta', t) &\to X(\theta', t), \quad X_{N_k}(\theta'', t) \to X(\theta'', t), \\
X_{N_k}^0(\theta') &\to X^0(\theta'), \quad X_{N_k}^0(\theta'') \to X^0(\theta''),
\end{aligned}
\qquad \text{as } k \to \infty. \qquad (49)
$$

For each large enough $N$, we choose indices $i$, $j$ so that

$$
\theta' \in (\theta_{i-1,N}, \theta_{i,N}] \subset (m', \widetilde{m}'), \qquad \theta'' \in (\theta_{j,N}, \theta_{j+1,N}] \subset (\widetilde{m}'', m'').
$$

Note that $i$ and $j$ depend on $N$, but we suppress this dependence in the notation. The point is that for all sufficiently large $N \in \mathbb{N}$, we have

$$
X_N(\theta', s) = x_{i,N}(s), \qquad X_N(\theta'', s) = x_{j+1,N}(s), \qquad \forall s \geq 0, \qquad (50)
$$

so we can track the approximate positions of mass labels $\theta'$ and $\theta''$ at time $t$ by looking at $x_{i,N}(t)$ and $x_{j+1,N}(t)$, which are simpler to analyze.

*Step 1.3.* Since $A = A^{**}$ is convex on $[m', m'']$, it follows from Lemma 2.2 and our choice of $i$, $j$ that

$$
\frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i-1}, N)}{\theta_{i,N} - \theta_{i-1,N}} \leq \frac{A(\overline{m}) - A(\widetilde{m}')}{\overline{m} - \widetilde{m}'}, \qquad \text{since } \theta_{i-1,N} < \theta_{i,N} < \widetilde{m}' \leq \overline{m},
$$

$$
\frac{A^{**}(\theta_{j+1,N}) - A^{**}(\theta_j, N)}{\theta_{j+1,N} - \theta_{j,N}} \geq \frac{A(\widetilde{m}'') - A(\overline{m})}{\widetilde{m}'' - \overline{m}}, \qquad \text{since } \overline{m} < \widetilde{m}'' < \theta_{j,N} < \theta_{j+1,N}.
$$

Taking the difference of these inequalities and recalling our definition of $\sigma$ from (48), we see that the hypotheses of Lemma 3.4 are satisfied. Then, in accordance with Lemma 3.4, we have the following lower bound at time $t$:

$$
x_{j+1,N}(t) - x_{i,N}(t) \geq \max\left\{ x_{j+1,N}^0 - x_{i,N}^0 - t\mathcal{V}^0, \min\{t\sigma, \eta\} \right\}.
$$

(Actually, a slightly better lower bound holds; c.f. the proof of Lemma 3.4, but we use the more general version above so that this step generalizes to Case 2.) Substituting (50) into this lower bound yields

$$
X_N(\theta'', t) - X_N(\theta', t) \geq \max\left\{ X_N^0(\theta'') - X_N^0(\theta') - t\mathcal{V}^0, \min\{t\sigma, \eta\} \right\}.
$$

Then taking $N \to \infty$ along the subsequence $(N_k)_{k=1}^{\infty}$, we conclude that

$$
X(\theta'', t) - X(\theta', t) \geq \max\left\{ X^0(\theta'') - X^0(\theta') - t\mathcal{V}^0, \min\{t\sigma, \eta\} \right\}.
$$

*Step 1.4.* Recalling that

$$
m' < \theta' < \widetilde{m}' < \widetilde{m}'' < \theta'' < m'',
$$

we may conclude from the previous step that

$$X(m'', t) - X(m', t) \geq \max \left\{ X^0(\widetilde{m}'') - X^0(\widetilde{m}') - t\mathcal{V}^0, \min\{t\sigma, \eta\} \right\}. \tag{51}$$

Since $A^{**}$ is not linear on $[\widetilde{m}', \widetilde{m}'']$ (by (48)), we have $\widetilde{m}' < \inf L(\widetilde{m}'')$. Lemma 3.5 then guarantees that $X^0(\widetilde{m}') < X^0(\widetilde{m}'')$, so that the quantity on the right side of (51) is strictly positive. The proof of Case 1 is now complete.

*Case 2:* $m_L'' \notin \Sigma_+$. In this case, we may assume without loss of generality that $A^{**}$ is linear on $[m', m_L'']$. We note also that we must have $m_L'' < m''$, with $A^{**}$ linear on $[m_L'', m'']$. (Indeed, the only way for $m_L''$ and $m''$ to be equal is if $L(m'') = \{m''\}$, in which case $m_L'' = m'' \in \Sigma_+$, contrary to the case we are considering.) Of course, the slopes of $A^{**}$ on $[m', m_L'']$ and $[m_L'', m'']$ are different.

*Step 2.1.* As in Case 1, choose $t > 0$ and a corresponding subsequence $(N_k)_{k=1}^{\infty}$; choose $\widetilde{m}' \in (m', m_L'')$, $\widetilde{m}'' \in (m_L'', m'')$, and then $\theta' \in (m', \widetilde{m}')$, $\theta'' \in (\widetilde{m}'', m'')$ such that (49) holds.

For each large enough $N$, we choose indices $i$, $j$ so that

$$\theta' \in (\theta_{i-1,N}, \theta_{i,N}] \subset (m', \widetilde{m}'), \qquad \theta'' \in (\theta_{j,N}, \theta_{j+1,N}] \subset (\widetilde{m}'', m'').$$

*Step 2.2.* Recall that we have chosen $\ell \in \mathbb{N}$ to satisfy $\theta_{\ell,N} = m_L''$. Whenever $N$ is chosen large enough so that $m' \leq \theta_{\ell-1,N}$ and $\theta_{\ell+1,N} \leq m''$, we must have

$$\frac{A^{**}(\theta_{j+1,N}) - A^{**}(\theta_{j,N})}{\theta_{j+1,N} - \theta_{j,N}} - \frac{A^{**}(\theta_{i,N}) - A^{**}(\theta_{i-1,N})}{\theta_{i,N} - \theta_{i-1,N}}$$
$$= \frac{A^{**}(m'') - A^{**}(m_L'')}{m'' - m_L''} - \frac{A^{**}(m_L'') - A^{**}(m')}{m_L'' - m'} =: 2\sigma > 0.$$

Thus, Lemma 3.4 applies. From this point onward, the argument is identical to that of Case 1. □

## 4. Finite-time clustering in the supercritical region

In this section, we turn to the *supercritical region* $\Sigma_-$, where the flux $A$ is detached from its convex envelope $A^{**}$. We prove part II of Theorem 1.7, which guarantees that any compact subinterval of $\Sigma_-$ becomes part of a finite-time cluster; this phenomenon is completely new in the study of (1). We also show that clusters inside a given connected component $(m_-, m_+)$ of $\Sigma_-$ cannot protrude from $(m_-, m_+]$ unless they contain all of $(m_-, m_+]$. This will be useful in the proof of part III(i) of Theorem 1.7, which we give in Sect. 5. We begin by establishing discrete versions of these facts before bootstrapping to the full continuum versions of the statements.

As before, we fix a solution $(\rho, u)$ of (1) and assume (A1)–(A4) hold.

4.1. The discrete supercritical setting

**Proposition 4.1.** *Let* $(m_{i,N}, x_{i,N}^0, v_{i,N}^0)_{i=1}^N$ *be a discretization of* $(\rho^0, u^0)$ *satisfying* (D2)–(D4). *Assume that* $(\theta_{i,N}, \theta_{j,N})$ *is a connected component of* $\Sigma_-$, *so that* $A > A^{**}$ *on* $(\theta_{i,N}, \theta_{j,N})$ *and* $A(\theta_{\ell,N}) = A^{**}(\theta_{\ell,N})$ *for* $\ell = i, j$. *Then, the following statements hold:*

(i) *Agents* $i + 1, \ldots, j$ *will collide in finite time.*

(ii) *Let* $T_N$ *be the first time when agents* $i + 1$ *and* $j$ *collide. Then agents* $i$ *and* $i + 1$ *cannot collide before time* $T_N$; *similarly, agents* $j$ *and* $j + 1$ *cannot collide before time* $T_N$.

*Proof.* We start with statement (ii), which is actually a special case of something we have already proven. If agents $i$ and $i + 1$ ever collide (say at time $t \geq 0$), then the hypotheses of Lemma 3.2 are satisfied, so that in particular we must have $\theta_{i^*(t),N} \notin \Sigma_-$. As $j$ is the smallest index exceeding $i$ such that $\theta_{j,N} \notin \Sigma_-$, and since $i^*(t) > i$ by assumption, we must have $i^*(t) \geq j$. Furthermore, $T_N = \inf\{s : i^*(s) \geq j\}$, and therefore $t \geq T_N$, as claimed. The situation is the same for agents $j$ and $j + 1$.

Now we prove statement (i), arguing by contradiction. Assume that agents $i + 1$ and $j$ never collide. Then for any $t \geq 0$, we have $(i + 1)^*(t) < j$; furthermore, by statement (ii), agents $i$ and $i + 1$ cannot collide, so we also have $(i + 1)_*(t) = i + 1$. Set

$$h_N = \min_{i+1 \leq \ell \leq j-1} (A - A^{**})(\theta_{\ell,N}); \qquad c_N = \frac{h_N}{\theta_{j,N} - \theta_{i,N}}.$$

Then,

$$\begin{aligned}
\psi_{i+1,N}(t) &= \frac{A(\theta_{(i+1)^*(t),N}) - A(\theta_{i,N})}{\theta_{(i+1)^*(t),N} - \theta_{i,N}} \\
&= \frac{(A - A^{**})(\theta_{(i+1)^*(t),N}) - (A - A^{**})(\theta_{i,N})}{\theta_{(i+1)^*(t),N} - \theta_{i,N}} \\
&\quad + \frac{A^{**}(\theta_{(i+1)^*(t),N}) - A^{**}(\theta_{i,N})}{\theta_{(i+1)^*(t),N} - \theta_{i,N}} \\
&\geq c_N + \frac{A^{**}(\theta_{j,N}) - A^{**}(\theta_{i,N})}{\theta_{j,N} - \theta_{i,N}}.
\end{aligned}$$

(In the last line, we used linearity of $A^{**}$ on $[\theta_{i,N}, \theta_{j,N}]$). Similarly,

$$\psi_{j,N}(t) \leq -c_N + \frac{A^{**}(\theta_{j,N}) - A^{**}(\theta_{i,N})}{\theta_{j,N} - \theta_{i,N}}.$$

Putting the two estimates above together, we obtain a strictly positive lower bound on the difference:

$$\psi_{i+1,N}(t) - \psi_{j,N}(t) \geq 2c_N > 0, \tag{52}$$

Note in particular that $c_N$ is independent of $t$. Therefore, we can compute using (20) that for all $t \geq 0$,

$$\frac{d}{dt}\big(x_{j,N}(t) - x_{i+1,N}(t)\big) = \psi_{j,N}(t) - \psi_{i+1,N}(t)$$

$$-\sum_{\ell=1}^{N} m_{\ell,N} \int_{x_{i+1,N}(t)}^{x_{j,N}(t)} \phi(y - x_{\ell,N}(t))\, dy \leq -2c_N < 0.$$

Here, we have used (52) and the fact that $\phi$ is nonnegative. Therefore, agents $i + 1$ and $j$ must collide at some finite time $T_N \leq D^0/(2c_N)$, where $D^0 = \text{diam supp}(\rho^0)$. This finishes the proof. $\qquad\square$

### 4.2. Finite-time clustering for the Euler-alignment system

Next, we use the sticky particle approximation to extend our result to the continuum system. The difficulty is the dependence of the constant $c_N$ in (52) on $N$. In fact,

$$\lim_{N \to \infty} h_N = 0, \quad \lim_{N \to \infty} c_N = 0.$$

Consequently, our upper bound on $T_N$ tends to infinity with $N$. The argument of the previous subsection therefore does *not* prove that a connected component $(m_-, m_+)$ of $\Sigma_-$ (or more precisely, its image under $X(\cdot, t)$) collapses to a point in finite time, and in fact, this is not true in general. However, as we will show below, any compact subset $K$ of $(m_-, m_+)$ *will* collapse to a point in a finite time $T$ that we can bound from above. In order to streamline the proof of this theorem, we first state an elementary but technical lemma that will help us obtain a uniform-in-$N$ adaptation of the statement and argument of Proposition 4.1.

**Lemma 4.2.** *Let $f : [0, 1] \to \mathbb{R}$ be a Lipschitz function, and let $\widetilde{K}$ be a compact subset of $(0, 1)$. Assume that*

(a) *$f(0) = f(1) = 0$, and $f > 0$ on $(0, 1)$,*
(b) *$f$ is convex in a neighborhood of 0 and a neighborhood of 1.*

*Then there exists $h^0 > 0$ such that whenever $0 < h \leq h^0$, the following holds:*

(i) *$f^{-1}(h)$ consists of exactly two points, call them $a_h$ and $b_h$, with $a_h < b_h$, and $\widetilde{K} \subseteq [a_h, b_h]$.*
(ii) *$f > h$ on $(a_h, b_h)$ and $f < h$ on $[0, a_h) \cup (b_h, 1]$.*
(iii) *Whenever*

$$0 \leq \alpha < a_h \leq \beta < \gamma < b_h \leq \delta \leq 1, \tag{53}$$

*we have the following estimate*

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha} - \frac{f(\delta) - f(\gamma)}{\delta - \gamma} \geq h > 0, \tag{54}$$

See Fig. 3 for an illustration of the statement of the Lemma. Later on, we will apply the Lemma with $f$ equal to a rescaled version of $A - A^{**}$.

*Proof.* First, we choose $a^0$ and $b^0$ such that $\widetilde{K} \subseteq [a^0, b^0] \subset (0, 1)$, and $f$ is convex in $[0, a^0]$ and $[b^0, 1]$. By assumption (a), $f$ must be strictly increasing in $[0, a^0]$ and strictly decreasing in $[b^0, 1]$. Since $f > 0$ in $[a^0, b^0]$, its minimum on this interval must be positive. We denote

$$h^0 = \frac{1}{2} \min_{x \in [a^0, b^0]} f(x) > 0. \tag{55}$$

For any $h$ in $(0, h^0]$, we define $a_h \in (0, a^0)$ by $a_h = f^{-1}(h)$. Note that $a_h$ is well-defined, as $f$ is strictly increasing on $[0, a^0]$ and $0 = f(0) < h < f(a^0)$. Similarly, we define $b_h \in (b^0, 1)$ by $b_h = f^{-1}(h)$. It is clear that (i) and (ii) hold, and that furthermore, for $\alpha, \beta, \gamma, \delta$ satisfying (53), we must have

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha} > 0 > \frac{f(\delta) - f(\gamma)}{\delta - \gamma}, \qquad \text{which implies}$$

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha} - \frac{f(\delta) - f(\gamma)}{\delta - \gamma} > 0.$$

We need to improve this to a positive lower bound that is uniform in $\alpha, \beta, \gamma, \delta$. We do this by considering the following three cases.

*Case 1:* $\beta < a^0$. Using the convexity of $f$ on $[0, a^0]$, we have

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha} - \frac{f(\delta) - f(\gamma)}{\delta - \gamma} \geq \frac{f(a_h) - f(0)}{a_h - 0} - 0 = \frac{h}{a_h} > h.$$

*Case 2:* $a^0 \leq \beta < \gamma \leq b^0$. From the definition of $h^0$ in (55), we have $f(\beta) \geq 2h_0$ and $f(\gamma) \geq 2h^0$. Hence,

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha} - \frac{f(\delta) - f(\gamma)}{\delta - \gamma} \geq \frac{2h^0 - h}{1 - 0} - \frac{h - 2h^0}{1 - 0} \geq 2h.$$

*Case 3:* $\gamma > b^0$. Using the convexity of $f$ on $[b^0, 1]$, we obtain

$$\frac{f(\beta) - f(\alpha)}{\beta - \alpha} - \frac{f(\delta) - f(\gamma)}{\delta - \gamma} \geq 0 - \frac{f(1) - f(b_h)}{1 - b_h} = \frac{h}{1 - b_h} > h.$$

This completes the proof. $\qquad\square$

*Remark 4.1.* The technical assumption (b) on the convexity near the boundary is to make sure that there exists a small $h$ such that (ii) holds. We want to eliminate the possibility of highly oscillatory functions like

$$f(x) = x \left(2 + \sin \tfrac{1}{x}\right)$$

near 0. We will not attempt to treat fluxes that exhibit this sort of pathological behavior.

We now prove part II of Theorem 1.7, restated here in slightly modified but equivalent form.
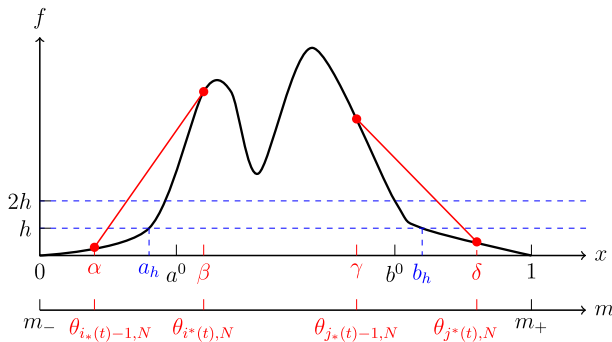
Figure 3. Illustration of Lemma 4.2 (with $h = h^0$) and its (rescaled) application in the proof of Theorem 4.3

**Theorem 4.3.** *Let $(\rho, u)$ and the accompanying functions and sets be as in (A1)–(A4). Assume $(m_-, m_+)$ is a connected component of $\Sigma_-$, i.e.,*

$$A(m) > A^{**}(m), \quad \forall\, m \in (m_-, m_+), \quad and$$
$$A(m_-) = A^{**}(m_-), \quad A(m_+) = A^{**}(m_+).$$

*We also recall that (A4) requires A to be convex in a neighborhood of $m_-$ and $m_+$. Let K be a compact subset of $(m_-, m_+)$. Then there exists a finite time $T > 0$ such that*

$$X(m, t) = X(m', t) \quad \forall\, m, m' \in K \quad and \quad \forall\, t \geq T.$$

*Proof.* Our plan is to approximate the system by the discrete dynamics, apply Lemma 4.2 to obtain uniform bounds, and then pass to the limit.

*Step 1:* Define the linear bijection $m : [0, 1] \to [m_-, m_+]$ via

$$m(x) = m_- + (m_+ - m_-)x.$$

Then define $f : [0, 1] \to \mathbb{R}$ by

$$f(x) = A(m(x)) - A^{**}(m(x)), \tag{56}$$

and define $\widetilde{K} \subset (0, 1)$ by $\widetilde{K} = m^{-1}(K)$. It is easy to check that $f$ satisfies all the assumptions in Lemma 4.2, as a consequence of assumption (A4). Choose $h^0$ as in the conclusion of the Lemma. Then for any $h \in (0, h^0]$, we have that

$$K \subseteq [m(a_h), m(b_h)],$$

and whenever we have

$$m_- < \alpha < m(a_h) \leq \beta < \gamma < m(b_h) \leq \delta \leq m_+,$$

then the following estimate holds:

$$\frac{(A - A^{**})(\beta) - (A - A^{**})(\alpha)}{\beta - \alpha} - \frac{(A - A^{**})(\delta) - (A - A^{**})(\gamma)}{\delta - \gamma}$$

$$\geq \frac{h}{m_+ - m_-} =: c(h) > 0. \tag{57}$$

Simplifying the left side of (57) using the linearity of $A^{**}$ on $[m_-, m_+]$, we obtain

$$\frac{A(\beta) - A(\alpha)}{\beta - \alpha} - \frac{A(\delta) - A(\gamma)}{\delta - \gamma} \geq c(h) > 0. \tag{58}$$

*Step 2:* Form a sequence of approximating systems $(m_{i,N}, x_{i,N}(\cdot), v_{i,N}(\cdot))_{i=1}^N$ via the procedure in Sect. 2. (In particular, assume (D1)–(D4) are satisfied). For $N$ large enough, assumption (D2) says that $S \subset \{\theta_{i,N}\}_{i=0}^N$, so that in particular $m_-$ and $m_+$ are breakpoints.

Define

$$T := \frac{X^0(m_+) - X^0(m_-)}{c(\frac{h_0}{2})}, \tag{59}$$

where $c(h) = \frac{h}{m_+ - m_-}$, as defined in (57). Fix any $t \geq T$, and choose a subsequence $(N_k)_{k=1}^\infty$ such that the a.e. convergence (35) holds at time $t$. Then we may in particular choose $h \in (\frac{h_0}{2}, h^0]$ so that, putting

$$\theta' = m(a_h), \qquad \theta'' = m(b_h), \tag{60}$$

we have

$$X_{N_k}(\theta', t) \to X(\theta', t) \quad\text{and}\quad X_{N_k}(\theta'', t) \to X(\theta'', t), \quad\text{as } k \to \infty. \tag{61}$$

Note that the subsequence $(N_k)_{k=1}^\infty$, and hence, our choice of $h$ depends on $t$. However, $h^0$ is of course time-independent.

Our goal is to show that

$$X(\theta', t) = X(\theta'', t). \tag{62}$$

Since $K \subseteq [\theta', \theta'']$ and $t \geq T$ is arbitrary, establishing (62) will prove the theorem.

*Step 3:* For each $N \in \mathbb{N}$, choose $i, j \in \{1, \dots, N\}$ such that $\theta' \in (\theta_{i-1,N}, \theta_{i,N}]$ and $\theta'' \in (\theta_{j-1,N}, \theta_{j,N}]$. Using (60) and (58) (with $\alpha = \theta_{i_*(s)-1,N}$, $\beta = \theta_{i_*(s),N}$, $\gamma = \theta_{j_*(s)-1,N}$, and $\delta = \theta_{j_*(s),N}$), we have the following uniform lower bound for all $s \geq 0$:

$$\psi_{i,N}(s) - \psi_{j,N}(s) = \frac{A(\theta_{i_*(s),N}) - A(\theta_{i_*(s)-1,N})}{\theta_{i_*(s),N} - \theta_{i_*(s)-1,N}} - \frac{A(\theta_{j_*(s),N}) - A(\theta_{j_*(s)-1,N})}{\theta_{j_*(s),N} - \theta_{j_*(s)-1,N}}$$

$$\geq c(h) \geq c(\tfrac{h_0}{2}) > 0.$$

It follows that for each $N \in \mathbb{N}$ and each time $s$ prior to the collision of agents $i$ and $j$, we have

$$\frac{d}{ds}(x_{j,N}(s) - x_{i,N}(s)) \leq -\big(\psi_{i,N}(s) - \psi_{j,N}(s)\big) \leq -c(\tfrac{h_0}{2}). \tag{63}$$

Integrating this equation and recalling the definition of $T$, as well as the fact that $X_N(\theta'', \cdot) - X_N(\theta', \cdot) = x_{j,N} - x_{i,N}$, we deduce that

$$X_N(\theta'', s) - X_N(\theta', s) = 0, \qquad \forall s \geq T.$$

Then, setting $s = t$ and taking $N \to \infty$ along the subsequence $(N_k)_{k=1}^{\infty}$ finishes the proof. □

### 4.3. Confinement of 'incomplete' clusters in supercritical intervals

We now upgrade Proposition 4.1(ii) to the full continuum version, as previewed at the beginning of this section.

**Theorem 4.4.** *Suppose* $(m', m'']$ *is a* $T$-*cluster containing* $m \in \Sigma_-$, *and let* $(m_-, m_+)$ *be the connected component of* $\Sigma_-$ *containing* $m$. *If* $(m', m'']$ *contains points outside of* $(m_-, m_+]$ *then* $(m', m'']$ *must also contain all of* $(m_-, m_+]$.

*Proof.* We assume, for purposes of obtaining a contradiction, that $m_- < m' < m_+ < m''$. (The case where $m' < m_- < m'' < m_+$ is similar.) Since $A$ is not linear on $(m', m'']$, the function $X^0$ cannot be constant on this interval, so we may assume without loss of generality that $T > 0$.

*Step 1.* We define the relevant parameters as follows. Choose $a^0, b^0 \in (m_-, m_+)$ such that $A$ is convex on $[m_-, a^0]$ and $[b^0, m_+]$. Put $h^0 = \frac{1}{2} \min_{[a^0, b^0]}(A - A^{**})$; for any $h \in (0, 2h^0)$, let $a_h \in (m_-, a^0)$ and $b_h \in (b^0, m_+)$ denote the two points in the inverse image of $h$ under $A - A^{**}$ that lie in $(m_-, m_+)$. Take $h^1 \in (0, h^0)$ small enough so that $a_{2h^1} < m' < b_{h^1}$.

As usual, consider a sequence of discretized systems satisfying (D1)–(D4), and additionally $m_-, m_+ \in \{\theta_{i,N}\}_{i=1}^{N}$ for sufficiently large $N$. Let $(N_k)_{k=1}^{\infty}$ be a subsequence such that the a.e. convergence (35) holds at time $t = T$. Choose $\theta' \in (b_{h^1}, m_+)$ and $\theta'' \in (m_+, m'')$ such that

$$X_{N_k}(\theta', T) \to X(\theta', T) \quad \text{and} \quad X_{N_k}(\theta'', T) \to X(\theta'', T), \qquad \text{as } k \to \infty.$$

Fix $h \in (0, h^1)$ such that $b_h = \theta'$. And finally, for each sufficiently large $N \in \mathbb{N}$, choose indices $I < i < J$ such that

$$m_- = \theta_{I,N}, \qquad \theta' \in (\theta_{i-1,N}, \theta_{i,N}] \subset (m', m_+), \qquad m_+ = \theta_{J,N}.$$

For convenience, we review the ordering of the relevant quantities before proceeding:

$$m_- = \theta_{I,N} < a_{2h} < m' < \theta_{i-1,N} < b_h = \theta' < \theta_{i,N} < m_+ = \theta_{J,N} < \theta'' < m''.$$

*Step 2.* We will now reason similarly to the proof of Lemma 4.2 and prove that

$$\frac{(A - A^{**})(\theta_{i^*(t), N_k}) - (A - A^{**})(\theta_{i_*(t)-1, N_k})}{\theta_{i^*(t), N_k} - \theta_{i_*(t)-1, N_k}} < -\frac{h}{m_+ - m_-}. \tag{64}$$

Since $\theta'$ lies in the $T$-cluster $(m', m'']$ at $m$ but $a_{2h} < m'$ does not, we have $X(\theta', T) > X(a_{2h}, T)$; consequently, we may assume that $X_{N_k}(\theta', T) > X_{N_k}(a_{2h}, T)$ for all $k$, so that

$$\theta_{i_*(t)-1,N_k} \geq a_{2h} \qquad \forall t \in [0, T], \ \forall k \in \mathbb{N}. \tag{65}$$

There are two cases to consider. If $\theta_{i_*(t)-1,N_k} \leq b_{2h}$, then

$$(A - A^{**})(\theta_{i_*(t)-1,N_k}) \geq 2h \qquad \text{and} \qquad (A - A^{**})(\theta_{i^*(t),N_k}) < h.$$

The inequality (64) follows immediately. On the other hand, if $b_{2h} < \theta_{i_*(t)-1,N_k} < b_h$, then by convexity of $A - A^{**}$ on $[b^0, m_+]$ we get

$$\frac{(A - A^{**})(\theta_{i^*(t),N_k}) - (A - A^{**})(\theta_{i_*(t)-1,N_k})}{\theta_{i^*(t),N_k} - \theta_{i_*(t)-1,N_k}} \leq \frac{(A - A^{**})(m_+) - (A - A^{**})(b_h)}{m_+ - b_h}$$

$$< -\frac{h}{m_+ - m_-},$$

as claimed.

*Step 3.* We will use (64), together with Proposition 4.1(ii), to prove the following lower bound:

$$\psi_{J+1,N}(t) - \psi_{i,N}(t) \geq \frac{h}{m_+ - m_-} =: 2\sigma > 0.$$

This is, in a sense, the key step of the proof, since it is here that we leverage the discrete version of the statement we want to prove.

Recalling (65), we know that

$$\theta_{I,N_k} < a_{2h} \leq \theta_{i_*(t)-1,N_k} \leq \theta_{J_*(t)-1,N_k}, \qquad \forall t \in [0, T],$$

which implies that $J_*(t) > I + 1$, i.e., agents $I + 1$ and $J$ do not collide on the time interval $[0, T]$. Consequently, Proposition 4.1(ii) tells us that agents $J$ and $J + 1$ also do not collide on $[0, T]$. Therefore, for any time $t \in [0, T]$, we thus have the following:

$$\psi_{i,N_k}(t) = \frac{A(\theta_{i^*(t),N_k}) - A(\theta_{i_*(t)-1,N_k})}{\theta_{i^*(t),N_k} - \theta_{i_*(t)-1,N_k}}$$

$$= \frac{(A - A^{**})(\theta_{i^*(t),N_k}) - (A - A^{**})(\theta_{i_*(t)-1,N_k})}{\theta_{i^*(t),N_k} - \theta_{i_*(t)-1,N_k}}$$

$$+ \frac{A^{**}(\theta_{i^*(t),N_k}) - A^{**}(\theta_{i_*(t)-1,N_k})}{\theta_{i^*(t),N_k} - \theta_{i_*(t)-1,N_k}}$$

$$\leq -\frac{h}{m_+ - m_-} + \frac{A^{**}(\theta_{(J+1)^*(t),N_k}) - A^{**}(\theta_{J,N_k})}{\theta_{(J+1)^*(t),N_k} - \theta_{J,N_k}}$$

$$\leq -\frac{h}{m_+ - m_-} + \frac{A(\theta_{(J+1)^*(t),N_k}) - A(\theta_{J,N_k})}{\theta_{(J+1)^*(t),N_k} - \theta_{J,N_k}} = -\frac{h}{m_+ - m_-} + \psi_{J+1,N_k}(t).$$

This proves the desired lower bound.

*Step 4.* We derive a contradiction. Using (20) and arguing as in (the beginning of) the proof of Lemma 3.4, we obtain

$$x_{J+1,N_k}(t) - x_{i,N_k}(t) \geq \min\{x^0_{J+1,N_k} - x^0_{i,N_k} + t\sigma, \eta\}, \qquad \forall t \geq 0,$$

where $\eta > 0$ is chosen small enough so that $|\int_z^w \phi(y)\,dr| < \sigma$ when $|z - w| < \eta$. It follows that

$$X_{N_k}(\theta'', T) - X_{N_k}(\theta', T) \geq x_{J+1, N_k}(T) - x_{i, N_k}(T) \geq \min\{T\sigma, \eta\} > 0.$$

Taking $k \to \infty$ and recalling that $m' < \theta' < \theta'' < m''$, we conclude that

$$X(m'', T) - X(m', T) \geq \min\{T\sigma, \eta\} > 0,$$

which contradicts our initial assumption that $(m', m'']$ is a $T$-cluster and therefore finishes the proof of the theorem. $\qquad\qquad\square$

## 5. Protocol-dependent behavior: clustering and non-clustering in the critical regime

So far, we have shown that $X(\cdot, t)$ exhibits two distinct behaviors in the subcritical region $\Sigma_+$ and the supercritical region $\Sigma_-$. We have also shown that clustering does not occur across different $L(m)$'s. In this section, we study the clustering behavior within a single $L(m)$. In contrast to the analysis we have presented thus far, the results of this section depend on properties of $\phi$ beyond those assumed in (A1); in that sense, the situation we consider in this section constitutes a sort of 'critical' regime (which includes in particular the behavior in the entire critical region $\Sigma_0$ but may also concern intervals in $\Sigma_-$).

The three subsections here track statements (i), (ii), and (iii) of part III of Theorem 1.7. The first statement, which concerns bounded protocols $\phi$, places limitations on the kinds of finite-time clusters that can occur within a given $L(m)$ and is the most delicate of the three results discussed in this section. The reason is that, unlike the situation in Sect. 3, it is possible for mass labels *within* a given $L(m)$ to belong to the same infinite-time cluster without belonging to the same $t$-cluster for any finite time. Ruling out finite-time clustering is therefore more subtle here than in the subcritical regime. The final two subsections show that under the heavy-tail assumption (2), each $L(m)$ becomes an infinite-time cluster, and is actually a finite-time cluster if $\phi$ is weakly singular.

As always, we assume throughout this section that $(\rho, u)$ is an entropy solution of (1) and that (A1)–(A4) hold.

### 5.1. Bounded communication

In this subsection, we assume that $\phi$ is bounded, and we prove statement III(i) of Theorem 1.7. As usual, we will start with a discrete (partial) analog of our target statement.

### 5.1.1. The discrete setting

**Lemma 5.1.** *Suppose that $\phi \in L^\infty(\mathbb{R})$. Let $(m_{i,N}, x^0_{i,N}, v^0_{i,N})^N_{i=1}$ be a discretization of $(\rho^0, u^0)$ satisfying* (D2)–(D4). *Assume that $i, j \in \{1, \ldots, N\}$ are such that $i \leq j$ and $\theta_{i,N}, \theta_{j,N} \notin \Sigma_-$.*

*Then*

$$x_{j+1,N}(t) - x_{i,N}(t) \geq (x^0_{j+1,N} - x^0_{i,N})e^{-\|\phi\|_\infty t}, \qquad \forall t \geq 0. \tag{66}$$

*Proof.* Assume without loss of generality that $x^0_{i,N} < x^0_{j+1,N}$. Since $\theta_{i,N}, \theta_{j,N} \notin \Sigma_-$, Corollary 3.3 guarantees that

$$\psi_{j+1,N}(t) - \psi_{i,N}(t) \geq 0, \qquad \forall t \geq 0.$$

Since $\phi$ is bounded, we can use an improved estimate on the sum in (20):

$$\sum_{\ell=1}^N m_{\ell,N} \int_{x_{i,N}(s)}^{x_{j+1,N}(s)} \phi(y - x_{\ell,N}(s)) \, \mathrm{d}y \leq \|\phi\|_\infty (x_{j+1,N}(s) - x_{i,N}(s)).$$

Consequently,

$$\frac{\mathrm{d}}{\mathrm{d}s}(x_{j+1,N}(s) - x_{i,N}(s)) \geq -\|\phi\|_\infty (x_{j+1,N}(s) - x_{i,N}(s)),$$

which becomes (66) after integration. $\qquad\square$

### 5.1.2. The continuum setting

We now turn to the proof of statement III(i) in Theorem 1.7. Under the assumption that $\phi \in L^\infty(\mathbb{R})$, we want to show that there are only two possible kinds of clusters. First, there might be initial clusters—intervals $(m', m'']$ on which $X^0$ is constant. And second, there might be clusters that are contained in some $(m_-, m_+]$, where $(m_-, m_+)$ is a connected component of $\Sigma_-$. We will show that no other finite-time clusters are possible. With this in mind, we make the following notation, which we will use in the proof of our theorem.

**Definition 5.2.** Fix $m \in (-\frac{1}{2}, \frac{1}{2}]$. Define the set $C(m)$ as follows.

- If there is an initial cluster at $m$, let $C(m)$ denote this initial cluster.
- If $m \in \Sigma_-$, define $C(m) = (m_-, m_+]$, where $(m_-, m_+)$ is the connected component of $\Sigma_-$ that contains $m$.
- Otherwise, put $C(m) = \{m\}$.

We also define the function $R(m, t)$ as follows.

$$R(m, t) = \begin{cases} \frac{1}{|C(m)|} \int_{C(m)} X(\widetilde{m}, t) \, \mathrm{d}\widetilde{m}, & \text{if } m \in \Sigma_- \\ X(m, t), & \text{otherwise.} \end{cases}$$

Here $|C(m)|$ denotes the Lebesgue measure of the interval $C(m)$.

The following theorem is a more precise version of statement III(i) in Theorem 1.7.

**Theorem 5.3.** *Assume $\phi \in L^\infty(\mathbb{R})$.*

(i) *If $C(m'') = \{m''\}$ is a singleton and $m' < m''$, then there exists a constant $c > 0$ such that*

$$X(m'', t) - X(m', t) \geq ce^{-\|\phi\|_\infty t}, \qquad \forall t \geq 0. \tag{67}$$

(ii) *If $C(m'')$ is an interval and $m' < \inf C(m'')$, then*

$$R(m'', t) - R(m', t) \geq ce^{-\|\phi\|_\infty t}, \qquad \forall t \geq 0. \tag{68}$$

*In either case, $m'$ and $m''$ do not belong to the same $t$-cluster for any finite time $t$.*

*Proof.* Note first of all that Theorem 4.4 is the link between (68) and the last claim, in the event that $C(m'')$ is an interval: If $m'$ and $m''$ ever belonged to the same $t$-cluster, then all of $C(m')$ and $C(m'')$ would need to belong to that $t$-cluster as well, by Theorem 4.4. This would in turn imply that $R(m', t) = R(m'', t)$, in direct contradiction with (68).

We also note that, by Theorem 3.6, we may assume without loss of generality that $m' \in L(m'')$.

**Proof of (i).** Assume that $C(m'') = \{m''\}$. Then, we must have $m'' \in \Sigma_0$ (since $m'' \in \Sigma_+$ would force $L(m') \neq L(m'')$), so we may assume without loss of generality that the interval $[m', m'']$ lies entirely in $\Sigma_0$. Since $X^0$ is left-continuous and $X^0$ is not constant on any interval of the form $[m'' - \varepsilon, m'']$ for $\varepsilon > 0$, it follows that $X^0$ must take infinitely many values on the interval $[m', m''_-]$. We may therefore choose $\tilde{m}', \tilde{m}'' \in [m', m'']$ such that

$$X^0(m') \leq X^0(\tilde{m}') < X^0(\tilde{m}'') \leq X^0(m'')$$

Consider as usual a sequence of discretizations satisfying (D1)–(D4); fix a time $t > 0$, and choose a corresponding subsequence $(N_k)_{k=1}^\infty$ such that the a.e. convergence (35) holds at time $t$. Choose $\theta' \in (m', \tilde{m}')$, $\theta'' \in (\tilde{m}'', m''_-)$ such that

$$\begin{aligned} X_{N_k}(\theta', t) \to X(\theta', t), \quad & X_{N_k}(\theta'', t) \to X(\theta'', t), \\ X^0_{N_k}(\theta') \to X^0(\theta'), \quad & X^0_{N_k}(\theta'') \to X^0(\theta''), \end{aligned} \qquad \text{as } k \to \infty. \tag{69}$$

For each sufficiently large $N \in \mathbb{N}$, choose $i, j \in \{1, \ldots, N\}$ such that

$$\theta' \in (\theta_{i-1,N}, \theta_{i,N}] \subset (m', \tilde{m}'); \qquad \theta'' \in (\theta_{j,N}, \theta_{j+1,N}] \subset (\tilde{m}'', m''_-).$$

Since $[m', m''] \subset \Sigma_0$ (and in particular $[m', m'']$ does not intersect $\Sigma_-$), Lemma 5.1 implies that

$$x_{j+1,N}(t) - x_{i,N}(t) \geq (x^0_{j+1,N} - x^0_{i,N})e^{-\|\phi\|_\infty t} \tag{70}$$

Recalling that $x_{i,N} = X_N(\theta', \cdot)$ and $x_{j+1,N} = X_N(\theta'', \cdot)$, we deduce that

$$X_N(\theta'', t) - X_N(\theta', t) \geq \left(X^0_N(\theta'') - X^0_N(\theta')\right)e^{-\|\phi\|_{L^\infty} t}.$$

Then, taking $N \to \infty$ along the subsequence $(N_k)_{k=1}^{\infty}$ and recalling that

$$m' < \theta' < \widetilde{m}' < \widetilde{m}'' < \theta'' < m'' \leq m'',$$

we obtain the desired estimate, with $c = X^0(\widetilde{m}'') - X^0(\widetilde{m}') > 0$.

**Proof of (ii).** Assume that $C(m'')$ is an interval and that $m' < \inf C(m'')$. Define

$$m''_- = \inf C(m'').$$

If $C(m''_-)$ is a singleton, i.e., $C(m''_-) = \{m''_-\}$, then $m' < \inf C(m'') = m''_-$ is upgraded to $m'_+ := \sup C(m') < m''_-$. Then statement (i) can be applied, with $m''_-$ replacing $m''$, and $m'_+$ replacing $m'$, to yield

$$R(m'', t) - R(m', t) \geq X(m''_-, t) - X(m'_+, t) \geq ce^{-\|\phi\|_{\infty}t}, \qquad \forall t \geq 0.$$

Thus, we may assume that $C(m''_-)$ and $C(m'')$ are both intervals. In this setting, we may further assume that $m'$ lies in the interval $C(m'')$, which implies that $C(m')$ and $C(m'')$ are adjacent half-open intervals. Let us therefore write

$$m'_- = \inf C(m'), \qquad m'_+ = \sup C(m') = \inf C(m'') = m''_-, \qquad m''_+ = \sup C(m'').$$

Consider a discretization satisfying (D1)–(D4), which additionally satisfies (for large enough $N$, and $i$, $j$, $\ell$ depending on $N$)

$$\theta_{i,N} = m'_-, \qquad \theta_{j,N} = m'_+ = m''_-, \qquad \theta_{\ell,N} = m''_+. \tag{71}$$

Note that $\theta_{i,N}$, $\theta_{j,N}$, and $\theta_{\ell,N}$ all belong to the complement of $\Sigma_-$, so $A(\theta_{p,N}) = A^{**}(\theta_{p,N})$ for $p = i, j, \ell$. Since $A^{**}$ is linear on $[\theta_{i,N}, \theta_{\ell,N}]$, it follows that

$$\frac{A(\theta_{j,N}) - A(\theta_{i,N})}{\theta_{j,N} - \theta_{i,N}} = \frac{A(\theta_{\ell,N}) - A(\theta_{j,N})}{\theta_{\ell,N} - \theta_{j,N}}.$$

Furthermore, Lemma 5.1 implies that no collisions occur between agents $i$ and $i + 1$, $j$ and $j + 1$, or $\ell$ and $\ell + 1$. Therefore, the following equality holds for all time:

$$\frac{\sum_{p=i+1}^{j} m_{p,N} \psi_{p,N}(t)}{\sum_{p=i+1}^{j} m_{p,N}} = \frac{A(\theta_{j,N}) - A(\theta_{i,N})}{\theta_{j,N} - \theta_{i,N}}$$

$$= \frac{A(\theta_{\ell,N}) - A(\theta_{j,N})}{\theta_{\ell,N} - \theta_{j,N}} = \frac{\sum_{q=j+1}^{\ell} m_{q,N} \psi_{q,N}(t)}{\sum_{q=j+1}^{\ell} m_{q,N}}. \tag{72}$$

The quantities on the left and right side of (72) are analogous to $\psi_i$ and $\psi_j$ in the inequality (20). The analogs of $x_i$ and $x_j$ are

$$R_N(m', t) := \frac{1}{|C(m')|} \int_{C(m')} X_N(\widetilde{m}, t) \, d\widetilde{m} = \frac{\sum_{p=i+1}^{j} m_{p,N} x_{p,N}(t)}{\sum_{p=i+1}^{j} m_{p,N}},$$

$$R_N(m'', t) := \frac{1}{|C(m'')|} \int_{C(m'')} X_N(\widetilde{m}, t) \, d\widetilde{m} = \frac{\sum_{q=j+1}^{\ell} m_{q,N} x_{q,N}(t)}{\sum_{q=j+1}^{\ell} m_{q,N}}. \tag{73}$$

Using equation (72) and performing some elementary manipulations gives us the identity

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(R_N(m'', t) - R_N(m', t)\big)$$
$$= -\frac{\sum_{r=1}^{N} \sum_{p=i+1}^{j} \sum_{q=j+1}^{\ell} m_{p,N} m_{q,N} m_{r,N} \int_{x_{p,N}(t)}^{x_{q,N}(t)} \phi(y - x_{r,N}(t))\, \mathrm{d}y}{\sum_{p=i+1}^{j} \sum_{q=j+1}^{\ell} m_{p,N} m_{q,N}}$$

Using the $L^\infty$ bound on $\phi$ and integrating then yields

$$R_N(m'', t) - R_N(m', t) \geq \big(R_N(m'', 0) - R_N(m', 0)\big)e^{-\|\phi\|_\infty t}, \qquad t \geq 0.$$

By the $L^1$-convergence of $(X_N(s))_{N=1}^{\infty}$ to $X(s)$ at $s = t$ and $s = 0$, we obtain (68), with $c = R(m'', 0) - R(m', 0)$. □

## 5.2. Heavy-tailed communication: infinite-time clustering

The lower bounds in the previous theorem are strictly positive for any finite time $t$, indicating an absence of finite-time cluster formation within $\Sigma_0$. However, the theorem is silent with regard to the possibility of infinite-time cluster formation. The infinite-time clustering phenomenon was studied extensively in [24] for the case where $\Sigma_- = \emptyset$ and the velocity field is at least $C^1$. Here we present a generalization for any entropic solution. Just as in [24], we assume that $\phi$ satisfies the heavy-tail condition (2). This assumption is sufficient to guarantee uniform global communication for all time:

$$\phi(x - y) \geq \underline{\phi} > 0, \quad \forall\, x, y \in \mathrm{supp}(\rho(\cdot, t)), \quad \forall\, t \geq 0. \tag{74}$$

Indeed, when (2) holds, it is shown in [29, Theorem 7.2] that the entropic solution to the Euler-alignment system (1) experiences *flocking*. That is, there exists a time-independent constant $\overline{D}$ such that

$$\mathrm{diam}\,\mathrm{supp}(\rho(\cdot, t)) \leq \overline{D}, \quad \forall\, t \geq 0,$$

The assertion (74) follows immediately, with $\underline{\phi} = \phi(\overline{D}) > 0$.

**Theorem 5.4.** *Assume that $\phi$ is heavy-tailed, and suppose that $L(m') = L(m'')$, with $m' < m''$. Then*

$$X(m'', t) - X(m', t) \leq D^0 e^{-\underline{\phi} t} \quad \forall\, t \geq 0, \tag{75}$$

*where $D^0 = \mathrm{diam}\,\mathrm{supp}\,\rho^0$ and $\underline{\phi}$ is defined as in (74). Consequently, $L(m') = L(m'')$ is an infinite-time cluster.*

*Proof. Step 1:* Denote

$$m_- = \inf L(m'), \qquad m_+ = \sup L(m').$$

We note first of all that it suffices to prove the bound (75) in the case $m'' < m_+$, by left-continuity of $X(\cdot, t)$.

Consider a sequence of discretizations satisfying (D1)–(D4), and additionally, $m_-, m_+ \in \{\theta_{i,N}\}_{i=0}^{N}$. For each $N$, choose $i, j \in \{0, \ldots, N\}$ such that $\theta_{i,N} = m_- \notin \Sigma_-$ and $\theta_{j,N} = m_+ \notin \Sigma_-$. Then using Corollary 3.3 and the linearity of $A^{**}$ on $L(m')$, we conclude that

$$\psi_{i+1,N}(s) \geq \frac{A^{**}(\theta_{i+1,N}) - A^{**}(\theta_{i,N})}{\theta_{i+1,N} - \theta_{i,N}} = \frac{A^{**}(\theta_{j,N}) - A^{**}(\theta_{j-1,N})}{\theta_{j,N} - \theta_{j-1,N}}$$

$$\geq \psi_{j,N}(s), \qquad \forall s \geq 0.$$

Next, we use the estimate (74) to obtain

$$\sum_{\ell=1}^{N} m_{\ell,N} \int_{x_{i+1,N}(s)}^{x_{j,N}(s)} \phi(y - x_{\ell,N}(s)) \, dy \geq \underline{\phi}\big(x_{j,N}(s) - x_{i+1,N}(s)\big). \tag{76}$$

Therefore, we have

$$\frac{d}{ds}\big(x_{j,N}(s) - x_{i+1,N}(s)\big) \leq -\underline{\phi}\big(x_{j,N}(s) - x_{i+1,N}(s)\big),$$

whence

$$x_{j,N}(s) - x_{i+1,N}(s) \leq \big(x_{j,N}^0 - x_{i+1,N}^0\big)e^{-\underline{\phi}t} \leq D^0 e^{-\underline{\phi}s}, \qquad \forall s \geq 0. \tag{77}$$

*Step 2:* Fix a time $t > 0$ and a corresponding subsequence $(N_k)_{k=1}^{\infty}$ such that the a.e. convergence (35) holds at time $t$. Choose $\theta' \in (m_-, m')$ and $\theta'' \in (m'', m_+)$ such that

$$X_{N_k}(\theta', t) \to X(\theta', t) \qquad \text{and} \qquad X_{N_k}(\theta'', t) \to X(\theta'', t), \qquad \text{as } k \to \infty.$$

For large enough $N$, we have

$$m_- = \theta_{i,N} < \theta_{i+1,N} < \theta' < m' < m'' < \theta'' < \theta_{j-1,N} < \theta_{j,N} = m_+.$$

Therefore

$$X_N(\theta'', t) - X_N(\theta', t) \leq X_N(\theta_{j,N}, t) - X_N(\theta_{i+1,N}, t)$$

$$= x_{j,N}(t) - x_{i+1,N}(t) \leq D^0 e^{-\underline{\phi}t}.$$

Taking $N \to \infty$ along the subsequence $(N_k)_{k=1}^{\infty}$ completes the proof. □

5.3. *Heavy-tailed and weakly singular communication: finite-time clustering*

As our final order of business, we consider *weakly singular* communication protocols $\phi$, i.e., those satisfying (3). These protocols are locally integrable but unbounded near the origin.

In contrast to the case of bounded communication $\phi$ (c.f. Theorem 5.3), finite-time cluster formation can occur in the critical region $\Sigma_0$ as a consequence of the weak singularity. This phenomenon was first discovered and analyzed in [38] for initial data satisfying $\Sigma_- = \emptyset$. The result is generalized in the following.

**Theorem 5.5.** *Assume $\phi$ is weakly singular (satisfying* (3)*) and heavy-tailed (in particular, satisfying* (74)*), and suppose that $L(m') = L(m'')$. Then there exists a finite time $T$ such that $L(m')$ is a $t$-cluster for all $t \geq T$, i.e., $X(\cdot, t)$ is constant on $L(m')$ for $t \geq T$.*

*Proof.* Denote $m_- = \inf L(m')$, $m_+ = \sup L(m')$, and assume without loss of generality that $m_- < m' < m'' < m_+$. We prove that there exists a time $T$, which depends on $m_-$ and $m_+$ but not on $m', m''$, such that $X(m', T) = X(m'', T)$. This is enough to establish the theorem.

Choose a time $T_1$ such that $D^0 e^{-\underline{\phi} T_1} = R$. We go through the proof of Theorem 5.4 and deduce from (77) that

$$x_{j,N}(t) - x_{i+1,N}(t) \leq R, \quad \forall t \geq T_1.$$

This allows us to apply (3) and improve the estimate in (76) by

$$\sum_{\ell=1}^{N} m_{\ell,N} \int_{x_{i+1,N}(t)}^{x_{j,N}(t)} \phi(y - x_{\ell,N}(t)) \, \mathrm{d}y \geq \sum_{\ell=i+1}^{j} m_{\ell,N} \int_{x_{i+1,N}(t)}^{x_{j,N}(t)} \phi(y - x_{\ell,N}(t)) \, \mathrm{d}y$$

$$\geq \sum_{\ell=i+1}^{j} m_{\ell,N} \cdot c \big( x_{j,N}(t) - x_{i+1,N}(t) \big)^{-\beta} \cdot \big( x_{j,N}(t) - x_{i+1,N}(t) \big)$$

$$= c(m_+ - m_-) \big( x_{j,N}(t) - x_{i+1,N}(t) \big)^{1-\beta}.$$

Hence, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \big( x_{j,N}(t) - x_{i+1,N}(t) \big) \leq -c(m_+ - m_-) \big( x_{j,N}(t) - x_{i+1,N}(t) \big)^{1-\beta}, \quad \forall t \geq T_1,$$

and therefore

$$x_{j,N}(t) - x_{i+1,N}(t) \leq \big( R^{\beta} - c\beta(m_+ - m_-)(t - T_1) \big)^{1/\beta}, \quad \forall t \geq T_1.$$

It follows that $x_{j,N}(t) - x_{i+1,N}(t)$ becomes zero no later than

$$T = T_1 + \frac{R^{\beta}}{c\beta(m_+ - m_-)},$$

where $T$ is independent of $N$.

Finally, we proceed with the same argument as Step 2 in the proof of Theorem 5.4 to pass to the limit and finish the proof. □

## Acknowledgements

**Data availability** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

## Declarations

**Conflict of interest** The authors have no competing interests to declare that are relevant to the content of this article.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

## REFERENCES

[1]   Victor Arnaiz and Ángel Castro. Singularity formation for the fractional Euler-alignment system in 1D. Transactions of the American Mathematical Society, 374(1):487–514, 2021.

[2]   Xiang Bai, Qianyun Miao, Changhui Tan, and Liutang Xue. Global well-posedness and asymptotic behavior in critical spaces for the compressible Euler system with velocity alignment. Nonlinearity, 37(2):025007, 2024.

[3]   Manas Bhatnagar, Hailiang Liu, and Changhui Tan. Critical thresholds in the Euler-Poisson-alignment system. Journal of Differential Equations, 375:82–119, 2023.

[4]   Yann Brenier, Wilfrid Gangbo, Giuseppe Savaré, and Michael Westdickenberg. Sticky particle dynamics with interactions. Journal de Mathématiques Pures et Appliquées, 99(5):577–617, 2013.

[5]   Yann Brenier and Emmanuel Grenier. Sticky particles and scalar conservation laws. SIAM journal on numerical analysis, 35(6):2317–2328, 1998.

[6]   José A Carrillo, Young-Pil Choi, Eitan Tadmor, and Changhui Tan. Critical thresholds in 1D Euler equations with non-local forces. Mathematical Models and Methods in Applied Sciences, 26(01):185–206, 2016.

[7]   Fabio Cavalletti, Marc Sedjro, and Michael Westdickenberg. A simple proof of global existence for the 1D pressureless gas dynamics equations. SIAM Journal on Mathematical Analysis, 47(1):66–79, 2015.

[8]   Li Chen, Changhui Tan, and Lining Tong. On the global classical solution to compressible Euler system with singular velocity alignment. Methods and Applications of Analysis, 28(2):155–174, 2021.

[9]   Junghee Cho, Seung-Yeal Ha, Feimin Huang, Chunyin Jin, and Dongnam Ko. Emergence of bi-cluster flocking for the Cucker–Smale model. Mathematical Models and Methods in Applied Sciences, 26(06):1191–1218, 2016.

[10]  Young-Pil Choi. The global Cauchy problem for compressible Euler equations with a nonlocal dissipation. Mathematical Models and Methods in Applied Sciences, 29(01):185–207, 2019.

[11]  Young-Pil Choi and Jan Haskovec. Hydrodynamic Cucker–Smale model with normalized communication weights and time delay. SIAM Journal on Mathematical Analysis, 51(3):2660–2685, 2019.

[12]  Peter Constantin, Theodore D Drivas, and Roman Shvydkoy. Entropy hierarchies for equations of compressible fluids and self-organized dynamics. SIAM Journal on Mathematical Analysis, 52(3):3073–3092, 2020.

[13]  Felipe Cucker and Steve Smale. Emergent behavior in flocks. Automatic Control, IEEE Transactions on, 52(5):852–862, 2007.

[14] Raphaël Danchin, Piotr B Mucha, Jan Peszek, and Bartosz Wróblewski. Regular solutions to the fractional Euler alignment system in the Besov spaces framework. Mathematical Models and Methods in Applied Sciences, 29(01):89–119, 2019.

[15] Tam Do, Alexander Kiselev, Lenya Ryzhik, and Changhui Tan. Global regularity for the fractional Euler alignment system. Archive for Rational Mechanics and Analysis, 228(1):1–37, 2018.

[16] Michal Fabisiak and Jan Peszek. Inevitable monokineticity of strongly singular alignment, 2022.

[17] Emmanuel Grenier. Existence globale pour le systeme des gaz sans pression. Comptes rendus de l'Académie des sciences. Série 1, Mathématique, 321(2):171–174, 1995.

[18] Seung-Yeal Ha, Jeongho Kim, Chanho Min, Tommaso Ruggeri, and Xiongtao Zhang. A global existence of classical solutions to the hydrodynamic Cucker–Smale model in presence of a temperature field. Analysis and Applications, 16(06):757–805, 2018.

[19] Seung-Yeal Ha, Jeongho Kim, Jinyeong Park, and Xiongtao Zhang. Complete cluster predictability of the Cucker–Smale flocking model on the real line. Archive for Rational Mechanics and Analysis, 231:319–365, 2019.

[20] Seung-Yeal Ha and Jian-Guo Liu. A simple proof of the Cucker-Smale flocking dynamics and mean-field limit. Communications in Mathematical Sciences, 7(2):297–325, 2009.

[21] Siming He and Eitan Tadmor. Global regularity of two-dimensional flocking hydrodynamics. Comptes Rendus Mathematique, 355(7):795–805, 2017.

[22] Feimin Huang and Zhen Wang. Well posedness for pressureless flow. Communications in Mathematical Physics, 222(1):117–146, 2001.

[23] Alexander Kiselev and Changhui Tan. Global regularity for 1D Eulerian dynamics with singular interaction forces. SIAM Journal on Mathematical Analysis, 50(6):6208–6229, 2018.

[24] Daniel Lear, Trevor M Leslie, Roman Shvydkoy, and Eitan Tadmor. Geometric structure of mass concentration sets for pressureless Euler alignment systems. Advances in Mathematics, 401:108290, 2022.

[25] Daniel Lear and Roman Shvydkoy. Unidirectional flocks in hydrodynamic Euler alignment system II: singular models. Communications in Mathematical Sciences, 19(3):807–828, 2021.

[26] Trevor M Leslie. Weak and strong solutions to the forced fractional Euler alignment system. Nonlinearity, 32(1):46–87, 2019.

[27] Trevor M Leslie. On the Lagrangian trajectories for the one-dimensional Euler alignment model without vacuum velocity. Comptes Rendus. Mathématique, 358(4):421–433, 2020.

[28] Trevor M Leslie and Roman Shvydkoy. On the structure of limiting flocks in hydrodynamic Euler Alignment models. Mathematical Models and Methods in Applied Sciences, 29(13):2419–2431, 2019.

[29] Trevor M Leslie and Changhui Tan. Sticky particle Cucker–Smale dynamics and the entropic selection principle for the 1D Euler-alignment system. Communications in Partial Differential Equations, 48(5):753–791, 2023.

[30] Qianyun Miao, Changhui Tan, and Liutang Xue. Global regularity for a 1D Euler-alignment system with misalignment. Mathematical Models and Methods in Applied Sciences, 31(03):473–524, 2021.

[31] Luca Natile and Giuseppe Savaré. A Wasserstein approach to the one-dimensional sticky particle system. SIAM journal on mathematical analysis, 41(4):1340–1365, 2009.

[32] Roman Shvydkoy. Global existence and stability of nearly aligned flocks. Journal of Dynamics and Differential Equations, 31(4):2165–2175, 2019.

[33] Roman Shvydkoy and Eitan Tadmor. Eulerian dynamics with a commutator forcing. Transactions of Mathematics and its Applications, 1(1):tnx001, 2017.

[34] Roman Shvydkoy and Eitan Tadmor. Eulerian dynamics with a commutator forcing II: Flocking. Discrete & Continuous Dynamical Systems, 37(11):5503–5520, 2017.

[35] Roman Shvydkoy and Eitan Tadmor. Eulerian dynamics with a commutator forcing III. fractional diffusion of order $0<\alpha<1$. Physica D: Nonlinear Phenomena, 376:131–137, 2018.

[36] Eitan Tadmor and Changhui Tan. Critical thresholds in flocking hydrodynamics with non-local alignment. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 372(2028):20130401, 2014.

[37] Changhui Tan. Singularity formation for a fluid mechanics model with nonlocal velocity. Communications in Mathematical Sciences, 17(7):1779–1794, 2019.

[38]   Changhui Tan. On the Euler-alignment system with weakly singular communication weights. Non-
       linearity, 33(4):1907, 2020.
[39]   Xiongtao Zhang and Tingting Zhu. Complete classification of the asymptotical behavior for singular
       C-S model on the real line. Journal of Differential Equations, 269(1):201–256, 2020.

*Trevor M. Leslie*
*Department of Applied Mathematics*
*Illinois Institute of Technology*
*Chicago*
*IL 60616*
*USA*
*E-mail: tleslie@iit.edu*

*Changhui Tan*
*Department of Mathematics*
*University of South Carolina*
*Columbia*
*SC 29208*
*USA*
*E-mail: tan@math.sc.edu*