

# Symmetric Continuum Opinion Dynamics: Convergence, but sometimes only in Distribution

Julien M. Hendrickx

Alex Olshevsky

**Abstract**—This paper investigates the asymptotic behavior of some common opinion dynamic models. We show that as long as interactions in a continuum of agents are symmetric, the distribution of the agents’ opinions converges, but that there exist examples where the opinions themselves do not converge. This phenomenon is in sharp contrast with symmetric models on finite numbers of agents where convergence of opinions is always guaranteed.

However, as long as every agent in the continuum interacts with those whose opinions are close to its own (a common assumption in opinion modeling), or that the interactions are uniquely determined by their opinions, the opinions of almost all agents will in fact converge.

## I. INTRODUCTION

There has been much recent interest within the control community in the study of multi-agent systems in which the agents interact according to simple, local rules, resulting in coordinated global behavior. Unfortunately, the dynamics describing the interactions of such systems are often time-varying and nonlinear and their analysis appears to be at present impossible without making considerably simplifying assumptions. For instance, it is a common assumption in much of the literature on multi-agent control that the graphs governing the inter-agent interactions satisfy some sort of long-term connectivity condition. This assumption is necessary due to the apparent intractability of analyzing the long-term connectivity properties associated with the graph of a multi-agent system governed by time-varying and nonlinear local interactions.

Encouraging results allowing avoiding long-term connectivity conditions have however been obtained for opinion dynamics models, one of the simplest and most natural classes of multi-agent systems with time-varying interactions. These models have been recently proposed (see [7], [9], and the survey [11]) to model opinion changes resulting from repeated personal interactions among individuals and have attracted considerable attention within the control community (see [4], [5], [6], [10]). Much of this attention is due to the similarities between opinion dynamics models and various large-scale multi-agent systems arising in control applications. Opinion dynamics are nonlinear because the inter-agent interactions

change with time and depend on the states of the agents, just as many multi-agent systems of interest. Consequently, it is believed that the techniques developed to rigorously analyze the asymptotic behavior of these models will be useful in the analysis of more complex multi-agent systems whose nonlinearity arises from the dynamics of the inter-agent interactions.

It has recently been shown for opinion dynamics systems with finitely many agents that the symmetry of the inter-agent interactions (or actually a weak symmetry condition called “cut-balance”) was sufficient to guarantee the convergence of all agents, *independently of any long-term connectivity condition* [8], something which makes the analysis of the asymptotic behavior of numerous systems considerably easier. Related observations for discrete-time systems were made in [3], [10], [12], (see also references in [8]).

Such general results are lacking for systems involving infinitely many agents, or a continuous mass of agents, whose behavior is usually thought to approximate that of systems with a finite but very large number of agents. Partial results have however been obtained. For example [6] proves the convergence of the distributions of the agents (i.e. the “density” of agents converges but not necessarily of the agents themselves) under the assumptions that the interactions between any two agents are entirely determined by their difference of opinions via a smooth symmetric function with a positive value at the origin. Similarly, convergence is established in [4] for a very specific law of interactions, and under some smoothness and monotonicity assumptions on the initial condition.

Our goal in this paper is to analyze the extent to which the results obtained for finitely many agents [8] or for specific models [4], [6] remain valid for general opinion dynamics model allowing for continuous mass of agents. We will see that the results obtained for finitely many agents can only be partly generalized to systems involving continuous masses, exposing important differences between the finite and continuous cases.

### A. Problem statement

We now give a precise statement of the dynamics that we study. We consider the functions  $x_t(\alpha) : [0, 1] \rightarrow \mathbf{R}$  which are solutions of the equation

$$\dot{x}_t(\alpha) = \int_0^1 w(t, \alpha, \beta, x_t(\alpha), x_t(\beta)) (x_t(\beta) - x_t(\alpha)) d\beta, \quad (1)$$

with initial condition  $x_0(\alpha) = x_0$ . Here  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  is a non-negative function. The equation has a natural interpretation:

J. M. Hendrickx is with the ICTEAM institute, Université catholique de Louvain, Belgium. His work is supported by the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Program, initiated by the Belgian Science Policy Office. The research is also supported by the Concerted Research Action (ARC) ‘Large Graphs and Networks’ of the French Community of Belgium. Email: julien.hendrickx@uclouvain.be

Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA, (aolshev2@illinois.edu).

each agent  $\alpha$  continually adjusts its “opinion”  $x_t(\alpha)$  to move closer to the opinions of other agents, giving to each agent  $\beta$  a weight  $w(t, \alpha, \beta, x_t(\alpha), x_t(\beta))$ , i.e. a weight that may depend on the identities of the two agents, their opinions, and on time.

In particular, the function  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  is often taken to be  $1_{|x_t(\alpha) - x_t(\beta)| < r}$  for some “opinion radius  $r$ ,” which corresponds to every agent adjusting its opinion based only on the opinions of other like-minded agents, see for example [5]. We will not be making such assumption here and will instead study the more general case.

Several technical assumptions are necessary for Eq. (1) to make sense. We will not be studying questions of existence and uniqueness; rather we will assume that each  $x_t(\alpha)$  is a (Borel) measurable function such that Eq. (1) is satisfied. The function  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  will be assumed to be measurable as well. Moreover, an additional constraint on  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  is needed to ensure that the integral in Eq. (1) is finite; we will assume that  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  is bounded, i.e., there exists some constant  $W$  so that  $w(t, \alpha, \beta, x_t(\alpha), x_t(\beta)) \leq W$  for all values of  $t, \alpha, \beta, x_t(\alpha), x_t(\beta)$ . Finally, we will assume that the initial distribution of opinions is bounded, i.e.,  $x_0(\alpha) \in [0, 1]$  for all  $\alpha \in [0, 1]$ , where the choice of the specific values of the bounds on  $x_0(\alpha)$  is made without loss of generality. *We will be making these assumptions throughout the remainder of this paper without mention.*

## B. Outline and main results

In general, Eq. (1) is not guaranteed to converge in any meaningful sense. Indeed, it is not hard to see that consensus of finitely many agents is a special case of Eq. (1) and consequently all the counterexamples to convergence of finitely many agents from [1] immediately carry over to this setting. Nevertheless, we will show in Section II that under the assumption that  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  is *symmetric*, i.e.,

$$w(t, \alpha, \beta, x_t(\alpha), x_t(\beta)) = w(t, \beta, \alpha, x_t(\beta), x_t(\alpha))$$

the *distribution* of the agents is guaranteed to converge. This means that even though the agents’ opinions  $x_t(\alpha)$  do not necessarily converge, the mass of agents having their opinion in any given set (satisfying some conditions) does converge.

This result is significantly stronger than the results previously available in the literature. As already mentioned, [6] proves for example a version of Theorem 1 in a similar model under the additional assumptions that  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  depends only on  $|x_t(\beta) - x_t(\alpha)|$  (so it is a function of just one argument), is continuous, supported on some ball centered at the origin, and in addition satisfies  $w(z) = w(-z)$ . Our contribution here is to note that most of the technical conditions hitherto used to establish convergence in distribution may simply be omitted without affecting the result.

We next turn to the question of whether it is possible to improve upon the conclusion of convergence in distribution. In fact, convergence in distribution does not seem to be the most natural notion of convergence for this class of systems; the more plausible notion would be *convergence of almost*

*all agents*, i.e., that  $x_t(\alpha)$  converges for almost all  $\alpha$ . We show in Section III, however, that convergence in this sense may not occur, i.e. that there exist symmetric nonnegative  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  and functions  $x_t$  satisfying Eq. (1) for such that the set of  $\alpha$  where  $x_t(\alpha)$  does not converge has positive measure, even though  $x_t$  converges in distribution.

Nevertheless, it may be possible that a stronger form of convergence holds in many natural settings. The extent to which this is true remains an open question. But, we can prove that as long as agents with similar opinions exert a non-negligible pull on each other, almost all of the  $x_t(\alpha)$  will indeed converge. The same holds true as long as the interactions depend only on time and on the opinions of the agents via a sufficiently smooth function. These results are presented in Section IV. Their proofs, longer and more complex than those of the other results, are omitted for space reasons and will be presented in a separate publication extending this work.

We finish by offering some conclusions and describing some open questions in Section V.

## II. CONVERGENCE IN DISTRIBUTION

Before establishing the convergence in distribution of (1) under the assumption of symmetry of the interactions, we first review the relevant notion of convergence of distributions. Every function  $x_t(\alpha)$  defines a measure  $\mu_t$  on the real line in the natural way

$$\mu_t(A) = \mathcal{L}(\{\alpha \mid x_t(\alpha) \in A\}) \quad (2)$$

where  $\mathcal{L}(\cdot)$  refers to the Lebesgue measure on the real line.  $\mu_t(A)$  represents thus the “mass” of agents lying in the set  $A$  at time  $t$ . These measures  $\mu_t$  are a natural way to summarize the concentrations of the values  $x_t(\alpha)$ . The most natural definition of convergence of distributions would be that  $\lim_t \mu_t(A) = \mu_\infty(A)$  for any Borel measurable set  $A$ ; this is referred to as *strong convergence* of measures and it is usually too restrictive to be used in practice (under this definition, for example, defining  $\mu_y(A) = 1_{y \in A}$ , it is not true that  $\mu_{1/t}$  converges to  $\mu_0$  as  $t \rightarrow \infty$ , (where the functions  $1/t$  and 0 are defined on  $\mathbb{R}_0^+$ ). Consequently, the most commonly used notion is *convergence in distribution* (sometime referred to in this context as weak-\* convergence):  $\mu_t$  approaches  $\mu_\infty$  in distribution if and only if  $\lim_t \mu_t(A) = \mu_\infty(A)$  for all sets  $A$  whose boundary has measure 0 under  $\mu_\infty$ .

We can now state our first main result.

**Theorem 1** *Suppose  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  is nonnegative and symmetric and let  $x_t(\alpha)$  be a solution of Eq. (1) and  $\mu_t$  be defined as in Eq. (2). Then there exists a measure  $\mu_\infty$  on  $[0, 1]$  such that  $\mu_t$  approaches  $\mu_\infty$  in distribution.*

Our proof is quite short and rests on a novel combination of two techniques: a simple exchange of integrals to establish that *every* convex function is a Lyapunov function (see proof of Lemma 1) coupled with an appeal to some results of Hausdorff about the moment problem (see proof of Theorem 1 below).

**Definition 1** Given a Borel measurable function  $f : \mathbf{R} \rightarrow \mathbf{R}$  define

$$V_f(t) = \int_0^1 f(x_t(\alpha)) d\alpha$$

**Lemma 1** If  $f$  is convex, then for all  $t \geq 0$ ,  $\dot{V}_f(t) \leq 0$ .

*Proof:* In what follows, we use the abbreviation  $\tilde{w}_{z,\alpha}$  to denote  $w(t, z, \alpha, x_t(z), x_t(\alpha))$ . We argue as follows.  $\dot{V}_f(t) =$

$$\begin{aligned} & \int_0^1 f'(x_t(\alpha)) \dot{x}_t(\alpha) d\alpha \\ = & \int_0^1 f'(x_t(\alpha)) \int_0^1 \tilde{w}_{z,\alpha} (x_t(z) - x_t(\alpha)) dz d\alpha \\ = & \int_{[0,1]^2} f'(x_t(\alpha)) \tilde{w}_{z,\alpha} (x_t(z) - x_t(\alpha)) \\ = & \frac{1}{2} \int_{[0,1]^2} \tilde{w}_{z,\alpha} f'(x_t(\alpha)) (x_t(z) - x_t(\alpha)) \\ & + \frac{1}{2} \int_{[0,1]^2} \tilde{w}_{z,\alpha} f'(x_t(z)) (x_t(\alpha) - x_t(z)) \\ = & \frac{1}{2} \int_{[0,1]^2} \tilde{w}_{z,\alpha} (f'(x_t(\alpha)) - f'(x_t(z))) (x_t(z) - x_t(\alpha)) \end{aligned}$$

Since  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  is nonnegative and  $f'(\cdot)$  is an increasing function due to the convexity of  $f$ , the integrand is always nonpositive. ■

**Remark 1** As a consequence of Lemma 1, the functions  $m_t(k) = \int_0^1 x_t(\alpha)^k d\alpha$  are nonincreasing for each  $k \geq 1$ . Viewing  $x_t$  as a random variable with state-space  $[0, 1]$ ,  $m_t(k)$  has the interpretation that it is the  $k$ 'th moment of this random variable.

Our next lemma formally states the obvious fact that because the values  $x_0(\alpha)$  are assumed to lie in  $[0, 1]$ , we have that  $x_t(\alpha)$  remains in  $[0, 1]$  for all  $\alpha$ .

**Lemma 2** For any  $t \geq 0$  the measure of the set of  $\alpha$  with  $x_t(\alpha) \notin [0, 1]$  is zero.

*Proof:* Since  $x_0$  has range  $[0, 1]$ , we have that  $m_0(k) \leq 1$  for all  $k \geq 1$ . By Remark 1, we have that  $m_t(k) \leq 1$  for every  $t \geq 0$  and  $k \geq 1$ . But if there was a time  $t$  with a positive measure of agents  $\alpha$  such that  $x_t(\alpha) > 1$  then  $m_t(k)$  would be strictly larger than 1 for all  $k$  large enough. This is a contradiction.

By observing that  $y_t(\alpha) = 1 - x_t(\alpha)$  is a solution of Eq. (1) with initial condition  $1 - x_0$  and applying the argument of the previous paragraph, we also obtain that the measure of the set of agents  $\alpha$  with  $x_t(\alpha) < 0$  is zero for any  $t$ . ■

We can now prove the main result of this section.

*Proof:* [Proof of Theorem 1] Let us view each  $x_t$  as a random variable defined on the state-space  $\Omega = [0, 1]$ . Remark 1 implies that the moments  $m_t(k)$  are nonincreasing

for  $k \geq 1$ ; consequently,  $\lim_{t \rightarrow \infty} m_t(k)$  exists for each  $k \geq 1$ . We next argue that there exists a random variable  $x_\infty$  whose  $k$ 'th moment is  $\lim_{t \rightarrow \infty} m_t(k)$ . This follows from a result of Hausdorff, which is that a sequence  $s = s_0, s_1, s_2, \dots$  is a valid moment sequence for a random variable with values in  $[0, 1]$  if and only if a certain infinite family of linear combinations of the  $s_i$  are nonnegative; each of these combinations has only finitely many nonzero coefficients (see [14], Theorem 1.5). Clearly, the moments of each  $x_t$  satisfy this condition since each  $x_t$  defines such a random variable. Moreover, it follows from Lemma 2 that these moments are all in the compact set  $[0, 1]$ . Therefore, the limiting moments  $\lim_{t \rightarrow \infty} m_t(k)$  also satisfy the condition of Hausdorff's result (and are in  $[0, 1]$ ). We conclude that there is a random variable  $x_\infty$ , satisfying  $x_\infty(\alpha) \in [0, 1]$  for all  $\alpha \in [0, 1]$ , such that the moments of  $x_t$  converge to the moments of  $x_\infty$ .

Next, convergence of moments  $x_t$  to the moments  $x_\infty$  implies convergence in distribution of  $x_t$  to  $x_\infty$  if  $x_\infty$  is uniquely defined by its moments ([2], Theorem 30.2). However, any random variable supported on  $[0, 1]$  is uniquely defined by its moments ([2], Theorem 30.1). Thus  $x_t$  converges to  $x_\infty$  in distribution. By the Portmanteau theorem ([13], Section 7.1), this immediately implies the conclusion we are trying to prove. ■

As a final remark about Theorem 1, we note that even though it is stated for one-dimensional opinions ( $x_t(\alpha) \in \mathbf{R}$ ), it can directly be extended to opinions in  $x_t(\alpha) \in \mathbf{R}^q$  satisfying a  $q$ -dimensional version of (1) provided that the interactions weights  $w$  remain scalar or  $q$ -dimensional diagonal matrices. Indeed, one can in that case apply Theorem 1 to each component of  $x_t(\alpha)$  separately.

### III. ABSENCE OF CONVERGENCE

We now show that Theorem 1 cannot be strengthened to establish convergence for almost all  $\alpha$  without additional assumptions. Indeed, quite surprisingly, there exist systems satisfying its hypotheses and converging in distribution, but for which a significant set of agents' opinions do not converge.

**Theorem 2** There exists a symmetric, nonnegative  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  and functions  $x_t$  satisfying Eq. (1) with this  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  such that the set of  $\alpha$  where  $x_t(\alpha)$  does not converge has positive measure.

*Proof:* We first describe an evolution  $x$  for which a positive measure set of agents does not converge. We then define some symmetric weights  $w(t, \alpha, \beta)$ , and show that this evolution  $x$  is a solution of (1) for those weights.

For the clarity of exposition, the agents are separated in three distinct sets, and  $x_0$  is bounded but its image is not included in  $[0, 1]$ . The first two sets are  $C_R$  and  $C_L$ , for right and left cluster respectively, both of measure 1. The third set is  $I = [0, 2]$ . Our system can be recasted in the form (1) with agents indexed on  $[0, 1]$  by a simple change of variable. Besides, when there is no risk of ambiguity, we will

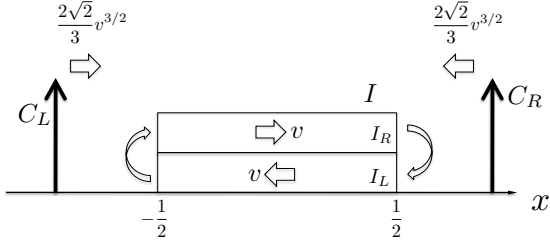


Fig. 1. Representation of the density of agents corresponding to the non-converging evolution  $x$  used in the proof of Theorem 2. The agents in  $I = I_R \cup I_L$  continuously move from left to right, with a speed  $v$  whose integral does not converge. The clusters  $C_L$  and  $C_R$  on the other hand move with a speed proportional to  $v^{3/2}$ , the integral of which does converge, so that they never reach  $[-\frac{1}{2}, \frac{1}{2}]$ .

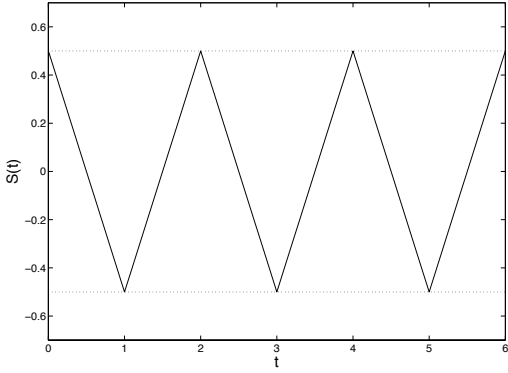


Fig. 2. Representation of the function  $S(t) = -\frac{1}{2} + |1 - (t \bmod 2)|$  that allows defining the trajectory of the agents' opinions.

sometimes refer to the agents as moving instead of as having their opinions changing, and call “velocity” the variation rate  $\dot{x}_t(\alpha)$  of their opinions.

### A. Definition of $x$

The evolution  $x$  is represented in Fig. 1. The idea is that agents in  $I$  move between  $-\frac{1}{2}$  and  $\frac{1}{2}$  without ever converging, while those in the cluster set  $C_R$  (resp.  $C_L$ ) all share the same opinion, start at some positive (resp. negative) opinion and move toward 0 without ever reaching  $1/2$  (resp.  $-1/2$ ).

Formally, we select a function  $v : \mathbf{R}^+ \rightarrow \mathbf{R}^+$  such that  $\int_0^\infty v(t)dt$  diverges, but  $\int_0^\infty v^{3/2}(t)dt < \infty$  (for example  $v(t) = (t+1)^{-3/4}$ ). Let then  $S(t) = -\frac{1}{2} + |1 - (t \bmod 2)|$ , where  $t \bmod 2$  denotes the unique value in  $[0, 2)$  equal to  $t - 2k$  for some integer  $k$ . The function  $S$  is represented in Fig. 2. Observe that  $S(t) \in [-\frac{1}{2}, \frac{1}{2}]$ , and that  $\frac{dS}{dt}(t)$  is defined everywhere except for integer  $t$ , with  $\frac{dS}{dt}(t) = 1$  when  $\lfloor t \rfloor$  is odd and  $-1$  when  $\lfloor t \rfloor$  is even. We now define  $x$ , represented in Fig. 1, in the following way:

- If  $\alpha \in I$ , then  $x_t(\alpha) = S\left(\alpha + \int_0^t v(s)ds\right)$ .
- If  $\alpha \in C_L$ , then  $x_t(\alpha) = x_t(C_L) := -C_0 + \int_0^t \frac{2\sqrt{2}}{3}v(s)^{3/2}ds$ .
- If  $\alpha \in C_R$ , then  $x_t(\alpha) = x_t(C_R) := C_0 - \int_0^t \frac{2\sqrt{2}}{3}v(s)^{3/2}ds$ ,

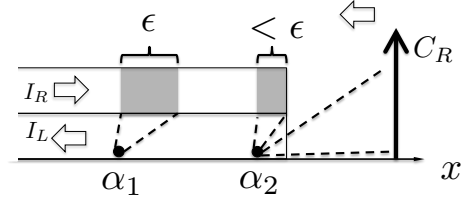


Fig. 3. Representations of the interactions in the example used in the proof of Theorem 2. The agent  $\alpha_1$  moves to the right hand side, and is attracted by all the agents moving on the left hand side that are in front of it and distant by less than  $\epsilon$  from it, guaranteeing a speed  $v$ . The agent  $\alpha_2$  is a distance smaller than  $\epsilon$  from the edge of the distribution, so there are not enough agents to attract it with a speed  $v$  (if interactions have the same intensity). This is compensated by interactions with the cluster  $C_R$ .

for some  $C_0 > 1/2 + \int_0^\infty \frac{2\sqrt{2}}{3}v(s)^{3/2}ds$ .

Observe that  $x$  does not converge, as the agents in  $I$  keep moving with speed  $v$ , the integral of which diverges. The evolution  $x$  converges however in distribution, consistently with Theorem 1. Indeed, one can verify that the density of agents on  $[-1/2, 1/2]$  remains constant at 2 for all  $t$ . At the same time, the opinions of the agents in  $C_L$  and  $C_R$  converge monotonously to some points out of the interval  $[-1/2, 1/2]$  due to our assumption on  $C_0$ .

In order to describe precisely the velocities and to later define the interaction weights, it is convenient to separate  $I$  in two time-varying disjoint sets. For every  $t$ , we let  $I_R(t)$  be the set of those  $\alpha \in [0, 2]$  for which  $\lfloor \alpha + \int_0^t v(s) \rfloor$  is odd, and  $I_L(t)$  the set of those for which this expression is even. Observe that since the shift  $\int_0^t v(s)$  is the same for all agents, which are all initially in  $[0, 2]$ ,  $I_R(t)$  and  $I_L(t)$  both have the same measure 1, and the “density” of agents of  $I_R$  and  $I_L$  on  $[-\frac{1}{2}, \frac{1}{2}]$  are both uniform and equal to 1 at all time. Besides, one can verify using the definition of  $x_t(\alpha)$  on  $I$  that  $\dot{x}_t(\alpha) = v(t)$  for every  $\alpha \in I_R(t)$  and  $\dot{x}_t(\alpha) = -v(t)$  for every  $\alpha \in I_L(t)$  (except for the two agents for which  $\alpha + \int_0^t v(s)$  is an integer, which we will neglect in this proof). Similarly,  $\dot{x}_t(\alpha) = \frac{2\sqrt{2}}{3}v(t)^{3/2}$  for every  $\alpha \in C_L$ , and  $\dot{x}_t(\alpha) = -\frac{2\sqrt{2}}{3}v(t)^{3/2}$  if  $\alpha \in C_R$ . These four equalities actually entirely characterize  $\dot{x}_t(\alpha)$  (up to the two points where it might not be defined).

### B. Interaction weights

We now define some symmetric weights  $w$  in order to later show that  $x_t$  is a solution to (1) for these weights. The idea is that agents in  $I$  will move thanks to an attraction exerted by agents moving in the other direction and lying in an interval of length  $\epsilon(t)$  ahead of them, for an appropriately chosen  $\epsilon(t)$ . Agents at distance less than  $\epsilon$  from the edge of the interval to which they are moving are in addition attracted by the clusters  $C_L$  or  $C_R$ , as represented in Fig. 3. These clusters are attracted in return, but move sufficiently slowly so that they never reach the interval  $[-\frac{1}{2}, \frac{1}{2}]$ .

The interactions between the agents of  $I$  are defined by  $w(t, \alpha, \beta) = 1$  if  $\alpha \in I_R(t), \beta \in I_L(t)$  and  $0 < x_t(\beta) - x_t(\alpha) < \epsilon(t)$ , and  $w(t, \alpha, \beta) = 0$  else, where we remind the

reader that all weights are symmetric. In addition, agents of  $I_R(t)$  with opinions above  $\frac{1}{2} - \epsilon(t)$  interact with those of  $I_R(t)$  with weights

$$w(t, \alpha, \beta) = \frac{\epsilon(t)^2 - \left(\frac{1}{2} - x_t(\alpha)\right)^2}{2(x_t(C_L) - x_t(\alpha))} \quad (3)$$

for  $\alpha \in I_R(t)$ ,  $\beta \in C_R$  and  $x_t(\alpha) \geq \frac{1}{2} - \epsilon(t)$ . Similarly,

$$w(t, \alpha, \beta) = \frac{\epsilon(t)^2 - \left(\frac{1}{2} + x_t(\alpha)\right)^2}{2(x_t(\alpha) - x_t(C_R))}$$

if  $\alpha \in I_L(t)$ ,  $\beta \in C_L$  and  $x_t(\alpha) \leq -\frac{1}{2} + \epsilon(t)$ .

C. *x* solution of (1) for these weights *w*.

To show that *x* is a solution of the system (1) for these weights, we just need to show that the right-hand side term of (1),  $\int w(t, \alpha, \beta)(x_t(\beta) - x_t(\alpha))d\beta$ , is equal to the velocities  $\dot{x}_t(\alpha)$  computed in part A. This expression should thus be equal to  $v(t)$  (resp.  $-v(t)$ ) for  $\alpha \in I_R(t)$  (resp.  $I_L(t)$ ), and to  $\frac{2\sqrt{2}}{3}v(t)^{3/2}$  (resp.  $-\frac{2\sqrt{2}}{3}v(t)^{3/2}$ ) for  $\alpha \in C_L$  (resp.  $C_R$ ), neglecting again the two agents for which  $\dot{x}_t(\alpha)$  is not defined.

We first consider an agent  $\alpha \in I_R(t)$  distant from the boundary  $\frac{1}{2}$  by more than  $\epsilon(t)$ , and interacting thus only with those agents of  $I_L(t)$  having opinions between  $x_\alpha(t)$  and  $x_\alpha(t) + \epsilon(t)$ . Its velocity  $\dot{x}_t(\alpha)$  is

$$\int_{\beta \in I_L(t): x_t(\beta) \in [x_t(\alpha), x_t(\alpha) + \epsilon(t)]} w(\alpha, \beta, t) (x_t(\beta) - x_t(\alpha)) d\beta.$$

Since the density of  $I_L(t)$  is 1 over the whole interval  $(-\frac{1}{2}, \frac{1}{2})$ , we have

$$= \int_{y=x_t(\alpha)}^{x_t(\alpha)+\epsilon(t)} (y - x_t(\alpha)) dy = \frac{1}{2}\epsilon(t)^2 = v(t), \quad (4)$$

consistently with the definition of *x*. Let us now consider an agent  $\alpha \in I_R(t)$  distant from the boundary  $\frac{1}{2}$  by less than  $\epsilon(t)$ . Such agents interact with agent in  $I_L(t)$  and with the agents in  $C_R$ . There holds thus  $\dot{x}_t(\alpha) =$

$$\int_{\beta \in I_L(t): x_t(\beta) \in [x_t(\alpha), x_t(\alpha) + \epsilon(t)]} w(\alpha, \beta, t) (x_t(\beta) - x_t(\alpha)) d\beta + \int_{\gamma \in C_R} w(\alpha, \beta, t) (x_t(\gamma) - x_t(\alpha)) d\gamma.$$

Taking into account that the density of  $I_R(t)$  is 1 in  $(-\frac{1}{2}, \frac{1}{2})$ , the fact that  $x_t(\alpha) > \frac{1}{2} - \epsilon(t)$  and the fact that all agents in  $C_R$  have the same opinion  $X_t(C_r)$ , and interact with  $\alpha$  according to (3), we obtain again

$$\begin{aligned} \dot{x}_t(\alpha) &= \int_{y=x_t(\alpha)}^{1/2} (y - x_t(\alpha)) dy \\ &+ \frac{\epsilon(t)^2 - \left(\frac{1}{2} - x_t(\alpha)\right)^2}{2(x_t(C_R) - x_t(\alpha))} (x_t(C_R) - x_t(\alpha)) \int_{\gamma \in C_r} d\gamma \\ &= \frac{1}{2} \left(\frac{1}{2} - x_t(\alpha)\right)^2 + \frac{1}{2} \left(\epsilon(t)^2 - \left(\frac{1}{2} - x_t(\alpha)\right)^2\right) \\ &= \frac{1}{2}\epsilon(t)^2 = v(t). \end{aligned}$$

We now consider an agent  $\alpha$  in the right cluster  $C_R$ . All agents in  $C_R$  have the same opinion  $x_t(C_R)$  and are interacting with agents in  $I_R(t)$  at a distance less than  $\epsilon(t)$  from  $1/2$ , with the weights (3), so that  $\dot{x}_t(\alpha) =$

$$\int_{\beta \in I_R(t): x_t(\beta) \geq \frac{1}{2} - \epsilon(t)} \frac{\epsilon^2 - \left(\frac{1}{2} - x_t(\beta)\right)^2}{2(x_t(C_R) - x_t(\alpha))} (x_t(\beta) - x_t(C_R)) d\beta$$

Using again the the density of  $I_R(t)$  is 1 over  $(-\frac{1}{2}, \frac{1}{2})$ , we obtain

$$\begin{aligned} \dot{x}_t(\alpha) &= -\frac{1}{2} \int_{y=1/2-\epsilon}^{1/2} \left(\epsilon^2 - \left(\frac{1}{2} - y\right)^2\right) dy \\ &= \frac{1}{3}\epsilon(t)^3 = -\frac{2\sqrt{2}}{3}v(t)^{3/2}. \end{aligned}$$

consistently with our definition of *x*. We have thus proved that *x* satisfies (1) with the symmetric weights that we have defined for all agents in  $I_R(t)$  and  $C_R$ . A symmetric argument applies to agents in  $I_L(t)$  and  $C_L$ , so that *x* is indeed a solution of (1), which achieves the proof, since we have already seen that  $x_t(\alpha)$  does not converge for any  $\alpha \in I = I_L(t) \cup I_R(t)$ . ■

#### IV. CONVERGENCE UNDER ADDITIONAL ASSUMPTIONS

We have seen in the previous sections that the symmetry of *w* guarantees the convergence in distribution of  $x_t$ , but not necessarily the convergence of almost all the opinions  $x_t(\alpha)$  themselves. Our construction required defining the interactions in an ad-hoc manner. In particular, two agents with the same opinions would often interact with totally disjoint sets of agents, resulting in opposite variations of opinions.

The results that we present here show that, provided that some common and reasonable assumptions on the model of opinion dynamics hold, almost all the opinions  $x_t(\alpha)$  do in fact converge. The first one concerns systems where agents with similar opinions exert a non-negligible (i.e. positive with a positive lower bound) pull on each other. Formally, let us define the set of functions  $\Gamma(r, \Delta)$  for  $r > 0, \Delta > 0$  to be the set of  $w(t, \alpha, \beta, x_t(\alpha), x_t(\beta))$  which satisfy the following property: if

$$|x_t(\alpha) - x_t(\beta)| \leq r$$

then

$$w(t, \alpha, \beta, x_t(\alpha), x_t(\beta)) \geq \Delta.$$

We then have the following theorem.

**Theorem 3** *If  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  is symmetric, nonnegative, and belongs to some  $\Gamma(r, \Delta)$  for  $r > 0, \Delta > 0$ , then any solution  $x_t(\alpha)$  of (1) converges for almost all  $\alpha$ .*

*Moreover, for almost all couples  $(\alpha, \beta)$ , either  $\lim_{t \rightarrow \infty} x_t(\alpha) = \lim_{t \rightarrow \infty} x_t(\beta)$  or  $|\lim_{t \rightarrow \infty} x_t(\alpha) - \lim_{t \rightarrow \infty} x_t(\beta)| \geq \Delta$ , so that almost all  $x_t(\alpha)$  converge to a finite set of values distant from each other by at least  $\Delta$ .*

The second result concerns systems where the interactions only depend on time and on the opinions via a sufficiently

smooth and bounded function, and states that the convergence of  $x_t(\alpha)$  for almost all  $\alpha$  is again guaranteed in that case. Intuitively, it relies on the idea that the order of the opinions is preserved in that situation (i.e. if  $x_t(\alpha) > x_t(\beta)$  for some  $t$ , this inequality remains satisfied for all further times), and that it is impossible to converge in distribution without converging for almost all  $\alpha$  while preserving the order of opinions.

**Theorem 4** *Suppose that the interaction weights  $w(\cdot, \cdot, \cdot, \cdot, \cdot)$  are symmetric, nonnegative, and depend only on time and the positions:*

$$w(t, x_t(\alpha), x_t(\beta), \alpha, \beta) = \tilde{w}(t, x_t(\alpha), x_t(\beta)).$$

If  $\tilde{w}$  is in addition

- *bounded:*  $|\tilde{w}(t, x, y)| \leq \bar{W}$  for all  $t, x, y$  for some  $\bar{W}$ ,
- *Lipschitz continuous:*  $|\tilde{w}(t, x, y) - \tilde{w}(t, x, z)| \leq L|y - z|$  for all  $t, x, y, z$ , and some  $L$ .

then any solution  $x_t(\alpha)$  of (1) converges for almost all  $\alpha$ .

We note that as a particular case of both results above, convergence of  $x_t(\alpha)$  for almost all  $\alpha$  is guaranteed for the one-dimensional versions of the systems considered in [6], where the authors proved convergence in distribution. They were however considering multi-dimensional opinions, and unlike Theorem 1, it is not clear at this stage whether Theorems 3 and 4 can be extended to such multi-dimensional cases.

## V. CONCLUSIONS

Our goal in this paper has been to analyze the asymptotic behavior of opinion dynamics. We have been able to resolve several questions implicit in the previous literature on the subject. In Theorem 1 we proved that symmetry alone appears to suffice for the convergence in distribution of such systems (in addition to the usual conditions of nonnegativity and such). Moreover, we showed in Theorems 2, 3 and 4 that while these systems converge in the sense of distributions, convergence for almost all agents is not automatic but rather crucially depends on additional assumptions, in sharp opposition with the results obtained for systems with finitely many agents [8].

Our results also cast a new light on those obtained in [8]. Convergence there is indeed first established for a sorted version of the vector containing all opinions, which is equivalent to proving the convergence of the opinions in distribution as in Theorem 1. Convergence of the opinions themselves then only follows from a continuity argument relying crucially on the discrete nature of the agents. This a priori artificial separation of the argument appears much more fundamental in view of the possibility for system with a continuous mass of agents to converge only in distribution.

Our motivation for studying these systems has been in the similarity they share with other multi-agent systems, namely the presence of a nonlinearity due to a time-varying update rule. We hope that the techniques we have developed to analyze these systems will find applications in the analysis

of other multi-agent systems. We note that Theorem 1 was proved without precise Lyapunov estimates on the decay by relying simultaneously on a large class of Lyapunov functions coupled with an appeal to results concerning the Hausdorff moment problem; to our knowledge, this is a new approach. Similarly, the proof of Theorem 3 (which will appear in a separate publication) relies on an “interval crossing” lemma bounding the mass of agents which can cross an interval infinitely often, which is new as well.

We conclude with some open questions. The contrast between Theorem 2 on the one hand and Theorems 3 and 4 on the other hand leads one to wonder whether a precise characterization of the settings in which convergence occurs for almost all agents is possible. Moreover, we wonder when it is possible to establish convergence time bounds on continuum opinion dynamics. Such bounds are likely to be useful in the analysis of convergence times in the case of finitely many agents. Other research directions could include the extension of our results to discrete-time systems, or the extensions of Theorems 3 and 4 to multi-dimensional opinions.

## REFERENCES

- [1] D.P. BERTSEKAS, J.N. TSITSIKLIS, *Parallel and Distributed Computation: Numerical Methods*, Prentice Hall, 1989.
- [2] P. BILLINGSLEY, *Probability and Measure*, John Wiley, 1986.
- [3] V.D. BLONDEL, J.M. HENDRICKX, A. OLSHEVSKY, AND J.N. TSITSIKLIS. Convergence in multiagent coordination, consensus, and flocking. In *Proceedings of the 44th IEEE Conference on Decision and Control (CDC2005)*, pp. 2996–3000, Seville, Spain, December 2005.
- [4] V.D. BLONDEL, J.M. HENDRICKX, AND J.N. TSITSIKLIS, *On Krause’s multi-agent consensus model with state-dependent connectivity*, IEEE Transactions on Automatic Control, 54(11), pp. 2586–2597, 2009.
- [5] V.D. BLONDEL, J.M. HENDRICKX, AND J.N. TSITSIKLIS, *Continuous-time average-preserving opinion dynamics with opinion-dependent communications*, SIAM Journal on Control and Optimization, 48(8), pp. 5214–5240, 2010.
- [6] C. CANUTO, F. FAGNANI, P. TILLI, *An Eulerian approach to the analysis of Krause’s Consensus Models*, SIAM Journal on Control and Optimization, 50(1), 243–265, 2012.
- [7] R. HEGSELMANN, U. KRAUSE, *Opinion dynamics and bounded confidence models, analysis, and simulations*, Journal of Artificial Societies and Social Simulation, 5(3), 2002.
- [8] J.M. HENDRICKX AND J.N. TSITSIKLIS, Convergence of type-symmetric and cut-balanced consensus seeking systems. In *IEEE Transactions on Automatic Control*, vol. 58, no. 1, pp. 214–218, 2013.
- [9] U. KRAUSE, *A discrete nonlinear and non-autonomous model of consensus formation*, in Communications in Difference Equations, S. Elaydi, G. Ladas, J. Popena, and J. Rakowski, eds., Gordon and Breach, Amsterdam, pp. 227–236, 2000.
- [10] J. LORENZ, *A stabilization theorem for continuous opinion dynamics*, Phys. A, 355, pp. 217–223, 2005.
- [11] J. LORENZ, *Continuous opinion dynamics under bounded confidence: A survey*, International Journal of Modern Physics C, 18(12), 1819–1838, 2007.
- [12] L. MOREAU, *Stability of multiagent systems with time-dependent communication links*. IEEE Transactions on Automatic Control, vol. 50, no. 2, pp. 169–182, 2005.
- [13] D. POLLARD, *A User’s Guide to Measure-Theoretic Probability*, Cambridge University Press, 2002.
- [14] J. SHOHAT, J. TAMARKIN, *The Problem of Moments*, American Mathematical Society, 1943.