

MEAN FIELD MODELS FOR LARGE DATA-CLUSTERING PROBLEMS

MICHAEL HERTY

RWTH Aachen University
Institut für Geometrie und Praktische Mathematik
Templergraben 55, 52062 Aachen, Germany

LORENZO PARESCHI*

University of Ferrara
Mathematics and Computer Science Department
Via Machiavelli 35, 44121 Ferrara, Italy

GIUSEPPE VISCONTI

RWTH Aachen University
Institut für Geometrie und Praktische Mathematik
Templergraben 55, 52062 Aachen, Germany

ABSTRACT. We consider mean-field models for data-clustering problems starting from a generalization of the bounded confidence model for opinion dynamics. The microscopic model includes information on the position as well as on additional features of the particles in order to develop specific clustering effects. The corresponding mean-field limit is derived and properties of the model are investigated analytically. In particular, the mean-field formulation allows the use of a random subsets algorithm for efficient computations of the clusters. Applications to shape detection and image segmentation on standard test images are presented and discussed.

1. Introduction. Particle and kinetic models for consensus and cluster formation appeared in recent literature for self-organized socio-economic dynamical systems as opinion formation, flocking of birds or fish, elections and referendums under influence of mass media, etc. See e.g. [5, 11, 18, 19, 20, 23, 25, 34, 16], the review articles [35, 4, 1] and the book [40]. A related research direction is based on using the consensus features of these models in an artificial way to solve problems of optimization or segmentation of data in large dimensions [28, 41, 15, 29].

In this paper, we aim at formulating suitable models on the microscopic (or particle) level as well as on the mean-field (or kinetic) level to describe the partition of a large set of data in clusters. This problem is also known as data clustering problem and it is widely studied in many applications like pattern recognition, shape detection and image segmentation problems. The proposed methods do not need to have fixed a-priori number of clusters and clusters are characterized by small in-group and large out-group distances.

2020 *Mathematics Subject Classification.* Primary: 82C40, 94A08; Secondary: 68U10.

Key words and phrases. Data clustering, opinion dynamic, mean field equations, image segmentation, shape detection.

* Corresponding author: Lorenzo Pareschi.

Since we are interested in modes characterizing distance and qualitative features without additional physical or socio-economical modeling background the proposed model will generalize the Hegselmann–Krause (HK) opinion dynamics model [25]. Originally, the HK model was proposed in a microscopic setting, in one-dimensional spatial and time discrete framework. Several extensions exist, in the sequel we will briefly review the basic model before discussing its extension towards the image clustering problems.

To this aim, let us consider a group of n particles with a (scalar) initial state $x_i(0) \in \mathbb{R}$, $i = 1, \dots, n$, and the state of each particle varies depending on the state of the others. The key idea of the Hegselmann–Krause (HK) model [25] is that particles with completely different opinions do not influence each other, and a sort of mediation occurs among agents whose opinions are within a bounded confidence interval described by a parameter $\epsilon \geq 0$. Let $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$ be the state of the system at time $t \geq 0$. Then the dynamic of the i -th particle is given by

$$\frac{d}{dt}x_i(t) = \sum_{j=1}^n A_{ij}(t, \epsilon) (x_j(t) - x_i(t)), \quad i = 1, \dots, n \quad (1)$$

where $\mathbf{A}(t, \epsilon) \in \mathbb{R}^{n \times n}$ is the time-varying adjacency matrix whose entries are in the form

$$A_{ij}(t, \epsilon) := \begin{cases} \frac{1}{\sigma_i}, & \text{if } j \in \mathcal{N}_i(t, \epsilon) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

with

$$\mathcal{N}_i(t, \epsilon) := \{j \in \{1, \dots, n\} : |x_i(t) - x_j(t)| \leq \epsilon\}, \quad i = 1, \dots, n \quad (3)$$

defining the neighborhood of the i -th particle at time t , and

$$\sigma_i := \begin{cases} |\mathcal{N}_i(t, \epsilon)|, \\ n \end{cases} \quad (4)$$

the type of interactions. Precisely, when $\sigma_i = n$ the interactions are symmetric since $A_{ij} = A_{ji}$, $\forall i, j$, but \mathbf{A} is not a stochastic matrix. Instead, when $\sigma_i = |\mathcal{N}_i(t, \epsilon)|$, then \mathbf{A} is a right stochastic matrix but interactions are no longer symmetric.

Several works have been proposed in the literature which analyze the properties of the HK model. For instance, for the analysis in the time discrete setting we refer to [27]. In [7] it is proven that, during the evolution of the system (1), the order of the states is preserved. Thanks to the definition of the interaction kernel (2)–(4), if $|x_i(t) - x_{i+1}(t)| > \epsilon$, at some time t , it remains true for larger times. Therefore, the HK model tends to group the initial states in a finite number of clusters as proved in [35]. Following [35], we define a cluster $\mathcal{C}(t)$ at time $t \geq 0$ a subset of particles separated from all the other particles

$$A_{ij}(t, \epsilon) \neq 0 \quad \text{for all } i, j \in \mathcal{C}(t), \quad A_{ij}(t, \epsilon) = 0 \quad \text{whenever } i \in \mathcal{C}(t), j \notin \mathcal{C}(t).$$

In [7, 22, 30] the stability of the dynamical model is investigated. In particular, the fact that the system converges to a steady profile in finite time is proved in [7]. For further results on the one-dimensional local and symmetric model we refer also to [8]. We point out that also behavior of cluster formation in the transient is of interest in the mathematical literature [21].

In [36, 10], the one-dimensional Hegselmann–Krause is generalized to the case of a multi-dimensional data-set. Subsequently, the multi-dimensional HK model has

been used as a technique to cluster a big amount of data into a small number of subsets with some common features [38] and to compare its performance with the k -means algorithm [31]. Recently, in [29, 37] new approaches to clustering problems and image segmentation have been proposed based on the Kuramoto model.

Here, we introduce a generalization of the multi-dimensional formulation of the HK model to solve data clustering problems. This amounts to take into account clustering with respect to different features. We deal with data having both time dependent and static features. The latter describe intrinsic properties of a datum, such as the measure of a trustworthy information or the color intensity of pixels in images. We derive the corresponding mean-field limit and investigate analytically the properties of the model. In particular, following [2] the mean-field formulation allows the use of a random subset algorithm for efficient computations of the clusters.

The rest of the manuscript is organized as follows. The microscopic model is introduced at the beginning of Section 2 and briefly discussed in Section 2.1. The proposed model is still a microscopic model and we describe the case of large data using a mean-field equation in Section 3. Analytical properties of the kinetic equation, such as a-priori estimation on the evolution of the moments and characterization of the limit distribution, are discussed in Section 3.1. Numerical evidence of the theoretical results is provided in Section 4.1. Further, we propose applications on detection and compression of data, such as shape detection, in Section 4.2, and image segmentation, in Section 4.3. We finally conclude with some remarks and future research directions in Section 5.

2. Microscopic models for data-clustering. Each particle $i = 1, \dots, N$ is endowed with a time-dependent state vector $\mathbf{x}_i = \mathbf{x}_i(t) \in \mathbb{R}^{d_1}$ as well as features $\mathbf{c}_i = [c_{i,1}, \dots, c_{i,d_2}] \in \mathbb{R}^{d_2}$ representing static characteristics of the system, i.e., \mathbf{c}_i is independent of time. As a motivation example consider an image segmentation problem where \mathbf{x}_i are the center point of a pixel or voxels of the image and \mathbf{c}_i the color coding at the center point.

As in the HK model we define the neighborhood of the particles by

$$\mathcal{N}_i(t, \epsilon_1, \epsilon_2) := \{j \in \{1, \dots, n\} : \|\mathbf{x}_i(t) - \mathbf{x}_j(t)\|_{\mathbb{R}^{d_1}} \leq \epsilon_1, \|\mathbf{c}_i - \mathbf{c}_j\|_{\mathbb{R}^{d_2}} \leq \epsilon_2\}, \quad (5)$$

for $i = 1, \dots, n$, and $A_{ij}(t, \epsilon_1, \epsilon_2)$ are entries of the time-varying matrix $\mathbf{A}(t, \epsilon_1, \epsilon_2) \in \mathbb{R}^{n \times n}$ defined as

$$A_{ij}(t, \epsilon_1, \epsilon_2) := \begin{cases} \frac{1}{\sigma_i}, & \text{if } j \in \mathcal{N}_i(t, \epsilon_1, \epsilon_2) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

with σ_i defined analogously to (4). Here, $\epsilon_1 \geq 0$ and $\epsilon_2 \geq 0$ are two bounded confidence levels. The two metrics $\|\cdot\|_{\mathbb{R}^{d_1}}$ and $\|\cdot\|_{\mathbb{R}^{d_2}}$ need to be properly defined according to the specific context of the problem. Then, the mathematical model for any $t \geq 0$ is given by

$$c_{i,k}(t) = c_{i,k}(0), \quad i = 1, \dots, n, \quad k = 1, \dots, d_2, \quad (7)$$

$$\frac{d}{dt} x_{i,k}(t) = \sum_{j=1}^n A_{ij}(t, \epsilon_1, \epsilon_2) (x_{j,k}(t) - x_{i,k}(t)), \quad i = 1, \dots, n, \quad k = 1, \dots, d_1 \quad (8)$$

and initial condition $c_{i,k}(0) = c_{i,k}^0$ and $x_{i,k}(0) = x_{i,k}^0$. Notice that a sufficient condition to reduce model (8) to the multi-dimensional version of the HK model (1) is $\epsilon_2 > \max_{i,j=1,\dots,n} \|\mathbf{c}_i - \mathbf{c}_j\|_{\mathbb{R}^{d_2}}$.

2.1. Properties of the microscopic model. Existence and convergence of solutions to system (8) can be established by using same techniques as in [8] where the original HK in the one-dimensional case was analyzed. In fact, system (7)- (8) belongs to the same class of state-switched systems.

In the case of symmetric interactions we can recover from the previous model similar results on the moment behavior as presented for the HK model in [8, 26]. We only record the results here since the proofs are slight variations of existing results (see for example [8]).

Define the moments $\mathbf{m}_1(t) \in \mathbb{R}^{d_1}$ and $\mathbf{m}_2(t) \in \mathbb{R}^{d_1 \times d_1}$ with respect to the time dependent feature as

$$\mathbf{m}_1(t) := \sum_{i=1}^n \mathbf{x}_i(t), \quad \mathbf{m}_2(t) := \sum_{i=1}^n \mathbf{x}_i(t) \otimes \mathbf{x}_i(t), \quad (9)$$

then the following result holds true.

Lemma 2.1. *Let $(\mathbf{x}_i(t))_{1 \leq i \leq n}$ be the solution of the dynamical system (8) with symmetric interactions, i.e. $\sigma_i = n$ in (6). Then, we obtain*

$$\mathbf{m}_1(t) = \mathbf{m}_1(0), \quad \frac{d}{dt} (\mathbf{m}_2(t))_{kk} \leq 0, \quad k = 1, \dots, d_1.$$

Corollary 1. *Let $(\mathbf{x}_i(t))_{1 \leq i \leq n}$ be the solution of the dynamical system (8). Assume that ϵ_1 and ϵ_2 are sufficiently large so that interactions are global. Then the first moment is conserved and the following decay estimate holds:*

$$\begin{aligned} \frac{d}{dt} (\mathbf{m}_2(t))_{k\ell} &= \frac{2}{n} (\mathbf{m}_1(0))_k (\mathbf{m}_1(0))_\ell - 2 (\mathbf{m}_2(t))_{k\ell}, \\ \lim_{t \rightarrow \infty} (\mathbf{m}_2(t))_{k\ell} &= \frac{(\mathbf{m}_1(0))_k (\mathbf{m}_1(0))_\ell}{n}. \end{aligned}$$

Later, we will show that similar results hold for the continuous model. Some remarks on further properties concerning the particle model (8) are in order.

Remark 1. • Formation of clusters in the large time behavior is extensively investigated in the review article [35] for general systems of the form (8). However, number of clusters cannot be a-priori predicted starting from a given initial configuration.

- In the non-symmetric case, conservation of the first moment does not hold true. Also, in general, it is not possible to show the decay of the second moment although clustering still appears, see [35].
- Extensions of the model to take into account non static features are obtained by including an interaction term on the right hand side of (7). The determination of such interaction term, however, would be rather problem dependent and in this paper we will not explore further this direction.

3. Mean-field description. In the case of many particles we derive the formal mean-field equation. Let $\Omega_1 \subseteq \mathbb{R}^{d_1}$, $\Omega_2 \subseteq \mathbb{R}^{d_2}$ be compact domains and $\Omega = \Omega_1 \times \Omega_2$. For $n \geq 1$ we denote by $f_n : \mathbb{R}^+ \times \Omega \rightarrow \mathbb{R}$ the empirical distribution on $\Omega \subset \mathbb{R}^{d_1 \times d_2}$ given by

$$f_n(t, \mathbf{x}, \mathbf{c}) := \frac{1}{n} \sum_{i=1}^n \delta(\mathbf{x} - \mathbf{x}_i(t)) \delta(\mathbf{c} - \mathbf{c}_i(t)).$$

Let us consider a test function $\varphi(\mathbf{x}, \mathbf{c}) \in C_0^1(\Omega)$, i.e. the space of continuous and compactly supported functions on Ω with continuous derivative. Denote by $\langle \cdot, \cdot \rangle$ the integration of f_n against the test function φ on Ω . We have

$$\begin{aligned} \frac{d}{dt} \langle f_n(t), \varphi \rangle &= \frac{1}{n} \sum_{i=1}^n \frac{d}{dt} \varphi(\mathbf{x}_i(t), \mathbf{c}_i(t)) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{\sigma_i} \sum_{j \in \mathcal{N}_i(t, \epsilon_1, \epsilon_2)} \nabla_{\mathbf{x}} \varphi(\mathbf{x}_i(t), \mathbf{c}_i(t)) \cdot (\mathbf{x}_j(t) - \mathbf{x}_i(t)) \\ &= \left\langle f_n(t), \frac{1}{n \sigma(t, \mathbf{x}, \mathbf{c})} \sum_{j=1}^n \chi_{\epsilon_1}(\|\mathbf{x}_j(t) - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{c}_j(t) - \mathbf{c}\|) (\mathbf{x}_j(t) - \mathbf{x}) \cdot \nabla_{\mathbf{x}} \varphi \right\rangle \end{aligned}$$

where we defined

$$\chi_{\epsilon}(x) = \begin{cases} 1, & x \leq \epsilon \\ 0, & \text{else} \end{cases}$$

and we used the fact that equation (8) can be re-written as

$$\frac{d}{dt} \mathbf{x}_i(t) = \frac{1}{n \sigma(t, \mathbf{x}_i(t), \mathbf{c}_i(t))} \sum_{j=1}^n \chi_{\epsilon_1}(\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|) \chi_{\epsilon_2}(\|\mathbf{c}_j(t) - \mathbf{c}_i(t)\|) (\mathbf{x}_j(t) - \mathbf{x}_i(t)),$$

with $\sigma_i = n \sigma(t, \mathbf{x}_i(t), \mathbf{c}_i(t))$ and

$$\sigma(t, \mathbf{x}_i(t), \mathbf{c}_i(t)) = \frac{1}{n} \sum_{j=1}^n \chi_{\epsilon_1}(\|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|) \chi_{\epsilon_2}(\|\mathbf{c}_j(t) - \mathbf{c}_i(t)\|).$$

We can easily compute

$$\begin{aligned} \sigma(t, \mathbf{x}, \mathbf{c}) &= \frac{1}{n} \sum_{j=1}^n \chi_{\epsilon_1}(\|\mathbf{x}_j(t) - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{c}_j(t) - \mathbf{c}\|) \\ &= \langle \chi_{\epsilon_1}(\|\mathbf{y} - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{z} - \mathbf{c}\|), \delta(\mathbf{y} - \mathbf{x}_j(t)) \delta(\mathbf{z} - \mathbf{c}_j(t)) \rangle \\ &= \int_{\Omega} \chi_{\epsilon_1}(\|\mathbf{y} - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{z} - \mathbf{c}\|) f_n(t, \mathbf{y}, \mathbf{z}) \, d\mathbf{z} \, d\mathbf{y}, \end{aligned}$$

and similarly

$$\begin{aligned} &\frac{1}{n \sigma(t, \mathbf{x}, \mathbf{c})} \sum_{j=1}^n \chi_{\epsilon_1}(\|\mathbf{x}_j(t) - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{c}_j(t) - \mathbf{c}\|) (\mathbf{x}_j(t) - \mathbf{x}) \\ &= \frac{1}{\sigma(t, \mathbf{x}, \mathbf{c})} \int_{\Omega} \chi_{\epsilon_1}(\|\mathbf{y} - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{z} - \mathbf{c}\|) (\mathbf{y} - \mathbf{x}) f_n(t, \mathbf{y}, \mathbf{z}) \, d\mathbf{z} \, d\mathbf{y}. \end{aligned}$$

Collecting these formal computations, after integration by part in \mathbf{x} , we obtain the weak form of the mean-field equation

$$\begin{aligned} &\frac{d}{dt} \langle f_n, \varphi \rangle + \\ &\left\langle \nabla_{\mathbf{x}} \cdot \left(f_n(t, \mathbf{x}, \mathbf{c}) \int_{\Omega} \frac{\chi_{\epsilon_1}(\|\mathbf{y} - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{z} - \mathbf{c}\|)}{\sigma(t, \mathbf{x}, \mathbf{c})} (\mathbf{y} - \mathbf{x}) f_n(t, \mathbf{y}, \mathbf{z}) \, d\mathbf{z} \, d\mathbf{y} \right), \varphi \right\rangle = 0. \end{aligned}$$

If we now define a *kernel* \mathcal{A} as continuous extension of the adjacency matrix (6)

$$\mathcal{A}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}, \mathbf{y}, \mathbf{z}) = \frac{\chi_{\epsilon_1}(\|\mathbf{y} - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{z} - \mathbf{c}\|)}{\sigma(t, \mathbf{x}, \mathbf{c})}, \quad (10)$$

and \mathbf{V} given by

$$\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}) = \int_{\Omega} \mathcal{A}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}, \mathbf{y}, \mathbf{z})(\mathbf{y} - \mathbf{x})f(t, \mathbf{y}, \mathbf{z}) \, d\mathbf{z} \, d\mathbf{y} \quad (11)$$

in the limit $n \rightarrow \infty$, assuming that the empirical measure $f_n(t, \mathbf{x}, \mathbf{c})$ converge to $f(t, \mathbf{x}, \mathbf{c})$, we formally obtain the strong form of the kinetic equation as

$$\partial_t f(t, \mathbf{x}, \mathbf{c}) + \nabla_{\mathbf{x}} \cdot (\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c})f(t, \mathbf{x}, \mathbf{c})) = 0. \quad (12)$$

Rigorous analytical results on convergence in the case of ϵ_2 very large and symmetric interactions have been already discussed, for instance in [13, 14]. These results guarantee convergence of the distribution f in (12) to a probability limit distribution f^∞ , provided the initial distribution f_0 at time $t = 0$ has finite second moment and \mathcal{A} is a non-negative, bounded, measurable and symmetric kernel. Observe that these assumptions are verified also in our framework. For a more detailed discussion on analytical results on convergence in mean-field models, we refer e.g. to [12, 16]

We briefly discuss the relation to similar kinetic models.

In [9] the analysis of a homogeneous kinetic model for opinion dynamics under bounded confidence is studied. Therein, the model is derived by using a Boltzmann-like approach with instantaneous binary interactions describing compromise. An analogous derivation of the kinetic equation (12) is also possible using a binary interaction model based on an explicit Euler discretization of the underlying particle dynamics (8) with $n = 2$ and performing a grazing collision limit. We omit this computation.

In [9] the authors prove the weak convergence of the solution to a convex combination of Dirac delta functions. In [13] a similar asymptotic distribution is found for the mean-field limit of the classical Hegselmann-Krause model.

A further symmetric clustering model with weighted interactions with respect a fixed number of closest neighbors and corresponding mean-field limit has been introduced in [3]. Moments and long-time behavior could also be studied therein. Finally, models for other applications are also able to cluster information e.g. in traffic flow modeling [42] where the physical acceleration has the role of the bounded confidence.

3.1. Properties of the mean-field model. As preliminary remark we observe that the marginal distribution $\tilde{f}^c(t, \mathbf{c}) := \int_{\Omega_1} f(t, \mathbf{x}, \mathbf{c}) \, d\mathbf{x}$ is preserved in time by the kinetic equation (12). Instead, it is easy to check that the marginal distribution $\tilde{f}^x(t, \mathbf{x}) := \int_{\Omega_2} f(t, \mathbf{x}, \mathbf{c}) \, d\mathbf{c}$ is not stationary and its behavior in time is still influenced by a \mathbf{c} dependent kernel. These considerations are direct consequence of the microscopic model (8) and (7). For this reason, in the following results and if not otherwise stated, we mainly focus on the analysis of moments with respect to the variable \mathbf{x} .

As in the discrete case we define the p -th moment of the kinetic distribution with respect to \mathbf{x} as

$$\langle \mathbf{x}^\alpha \rangle(t) = \int_{\Omega} \mathbf{x}^\alpha f(t, \mathbf{x}, \mathbf{c}) \, d\mathbf{x} \, d\mathbf{c}, \quad |\alpha| = p \in \mathbb{N}, \quad \alpha_i \in \mathbb{N}.$$

Here, we used the following multi-index notation $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_{d_1}^{\alpha_{d_1}}$ and $|\alpha| = \sum_{i=1}^{d_1} \alpha_i$. In particular, for each $k, j = 1, \dots, d_1$ we will denote

$$u_k(t) = \langle \mathbf{x}^\alpha \rangle(t), \quad |\alpha| = 1, \quad \alpha_i = \delta_{ik}, \quad i = 1, \dots, d_1$$

$$E_{kj}(t) = \langle \mathbf{x}^\alpha \rangle(t), \quad |\alpha| = 2, \quad \alpha_i = \delta_{ik} + \delta_{ij}, \quad i = 1, \dots, d_1,$$

the first moment and the second moment, respectively. Then, for a symmetric kernel \mathcal{A} we obtain by the integration of the kinetic equation

Lemma 3.1. *Let $f(t, \mathbf{x}, \mathbf{c})$ be the solution of the model (12) with symmetric kernel \mathcal{A} , i.e., $\sigma(t, \mathbf{x}, \mathbf{c}) = 1$. Then, the following a-priori estimates hold true:*

$$\frac{d}{dt} u_k(t) = 0, \quad \frac{d}{dt} E_{kk}(t) \leq 0, \quad |E_{kj}(t)| \leq \frac{1}{2} (E_{kk} + E_{jj}) \quad (0)$$

for each $k, j = 1, \dots, d_1$.

Proof. For each fixed $k = 1, \dots, d_1$ we have

$$\begin{aligned} \frac{d}{dt} u_k(t) &= - \int_{\Omega} x_k \nabla_{\mathbf{x}} \cdot (\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}) f(t, \mathbf{x}, \mathbf{c})) \, d\mathbf{x} d\mathbf{c} \\ &= \int_{\Omega} \int_{\Omega} \chi_{\epsilon_1}(\|\mathbf{y} - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{z} - \mathbf{c}\|) (y_k - x_k) f(t, \mathbf{y}, \mathbf{z}) f(t, \mathbf{x}, \mathbf{c}) \, d\mathbf{y} d\mathbf{z} d\mathbf{x} d\mathbf{c} = 0. \end{aligned}$$

The last term vanishes due to anti-symmetry of the integrand by the interchange of variables $\mathbf{x} \leftrightarrow \mathbf{y}$. The second statement follows by observing that for each $k = 1, \dots, d_1$ we have

$$\begin{aligned} \frac{d}{dt} E_k(t) &= - \int_{\Omega} x_k^2 \nabla_{\mathbf{x}} \cdot (\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}) f(t, \mathbf{x}, \mathbf{c})) \, d\mathbf{x} d\mathbf{c} \\ &= - \int_{\Omega} \int_{\Omega} \chi_{\epsilon_1}(\|\mathbf{y} - \mathbf{x}\|) \chi_{\epsilon_2}(\|\mathbf{z} - \mathbf{c}\|) (y_k - x_k)^2 f(t, \mathbf{y}, \mathbf{z}) f(t, \mathbf{x}, \mathbf{c}) \, d\mathbf{y} d\mathbf{z} d\mathbf{x} d\mathbf{c} \leq 0. \end{aligned}$$

Hence, we obtain

$$0 \leq |E_{kj}(t)| \leq \int_{\Omega} |x_k x_j| f(t, \mathbf{x}, \mathbf{c}) \, d\mathbf{x} d\mathbf{c} \leq \frac{1}{2} (E_{kk} + E_{jj})(t) \leq \frac{1}{2} (E_{kk} + E_{jj})(0). \quad \square$$

Instead, in the case of global interactions and for a general kernel \mathcal{A} and we have

Corollary 2. *Assume that ϵ_1 and ϵ_2 are sufficiently large so that $\mathcal{A}(t, \mathbf{x}, \mathbf{c}, \mathbf{y}, \mathbf{z}; \epsilon_1, \epsilon_2) = 1$, for all $\mathbf{x}, \mathbf{y} \in \Omega_1, \mathbf{c}, \mathbf{z} \in \Omega_2$ and $t \geq 0$. Then, for all $k, j = 1, \dots, d_1$, the following relations hold true:*

$$u_k(t) = u_k(0), \quad \forall t \geq 0 \quad \text{and} \quad E_{kj}(t) \xrightarrow{t \rightarrow \infty} u_k(0) u_j(0).$$

Proof. Multiplying (12) by \mathbf{x}^α with $|\alpha| = p \in \mathbb{N}$, and integrating over Ω , we compute

$$\frac{d}{dt} \langle \mathbf{x}^\alpha \rangle(t) = -p \langle \mathbf{x}^\alpha \rangle(t) + \sum_{\ell=1}^{d_1} \alpha_\ell u_\ell(t) \langle \mathbf{x}^{\alpha^{(\ell)}} \rangle(t)$$

where $\alpha^{(\ell)} = \alpha - \mathbf{e}_\ell$ with \mathbf{e}_ℓ being the ℓ -th element of the standard basis vector of \mathbb{R}^{d_1} and $|\alpha^{(\ell)}| = p - 1$. Then, for $p = 1$ we still have conservation of the first moment since we obtain

$$\frac{d}{dt} u_k(t) = -u_k(t) + u_k(t) = 0, \quad k = 1, \dots, d_1.$$

For $p = 2$ we have

$$\frac{d}{dt}E_{kj}(t) = -2E_{kj}(t) + 2u_k(t)u_j(t) = -2E_{kj}(t) + 2u_k(0)u_j(0), \quad k, j = 1, \dots, d_1$$

and therefore

$$E_{kj}(t) = E_{kj}(0)e^{-2t} + u_k(0)u_j(0)(1 - e^{-2t}) \xrightarrow{t \rightarrow \infty} u_k(0)u_j(0), \quad k, j = 1, \dots, d_1. \quad \square$$

Observe that, compared to the local interaction case analyzed in Theorem 3.1 where in general the energy decay property of all second order moments is not guaranteed, in global interactions also the mixed second order moments decay in time. In other words, while in the local interaction model only variances go to zero, in the global interaction model both variances and covariances go to zero in the large time behavior.

Lemma 3.1 and Corollary 2 suggest that for $t \rightarrow \infty$ any initial distribution will tend to a stationary distribution that is concentrated on a finite number of points in Ω_1 and represents therefore clusters. Hence, in the following we investigate the asymptotic distribution of the model (12).

Lemma 3.2. *Let ϵ_1 and ϵ_2 be arbitrary positive bounded confidence levels. Then, the distribution*

$$f^\infty(\mathbf{x}, \mathbf{c}) = \sum_{k=1}^{n_1} f_k \delta(\mathbf{x} - \mathbf{x}_k) \sum_{\ell=1}^{n_2(k)} \delta(\mathbf{c} - \mathbf{c}_{k_\ell}) \quad (13)$$

with $f_k > 0$ and $\sum_{k=1}^{n_1} \sum_{\ell=1}^{n_2(k)} f_k = 1$ is a weak stationary solution of the model (12) if and only if either $\|\mathbf{x}_i - \mathbf{x}_k\|_{\mathbb{R}^{d_1}} > \epsilon_1$ for all $i \neq k$ or $\|\mathbf{c}_{i_j} - \mathbf{c}_{k_\ell}\|_{\mathbb{R}^{d_2}} > \epsilon_2$ for all $i \neq k$, for all j, ℓ , or both hold true.

Proof. Let us assume that at least one of $\|\mathbf{x}_i - \mathbf{x}_k\|_{\mathbb{R}^{d_1}} > \epsilon_1$ for all $i \neq k$ and $\|\mathbf{c}_{i_j} - \mathbf{c}_{k_\ell}\|_{\mathbb{R}^{d_2}} > \epsilon_2$ for all $i \neq k$, for all j, ℓ , is verified and prove that f^∞ is the asymptotic distribution of the equation (12), i.e. given a test function $\varphi \in C_0^1(\Omega)$

$$\int_{\Omega} \varphi(\mathbf{x}, \mathbf{c}) \nabla_{\mathbf{x}} \cdot (\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}) f^\infty(\mathbf{x}, \mathbf{c})) \, d\mathbf{x}d\mathbf{c} = 0.$$

We obtain

$$\begin{aligned} & \int_{\Omega} \varphi(\mathbf{x}, \mathbf{c}) \nabla_{\mathbf{x}} \cdot (\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}) f^\infty(\mathbf{x}, \mathbf{c})) \, d\mathbf{x}d\mathbf{c} \\ &= - \sum_{k=1}^{n_1} f_k \sum_{\ell=1}^{n_2(k)} \mathbf{V}(t, \mathbf{x}_k, \mathbf{c}_{k_\ell}; \epsilon_1, \epsilon_2) \cdot \nabla_{\mathbf{x}} \varphi(\mathbf{x}, \mathbf{c})|_{\mathbf{x}=\mathbf{x}_k} \\ &= - \sum_{k=1}^{n_1} f_k^2 \sum_{\ell=1}^{n_2(k)} \frac{(\mathbf{x}_k - \mathbf{x}_k)}{\sigma(\mathbf{x}_k, \mathbf{c}_{k_\ell})} \cdot \nabla_{\mathbf{x}} \varphi(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_k} = 0. \end{aligned}$$

Conversely, assume that f^∞ as in (13) is the steady state of (12). Assume by contradiction that there exist $k, i \in \{1 \dots, n_1\}$ such that $\|\mathbf{x}_k - \mathbf{x}_i\|_{\mathbb{R}^{d_1}} \leq \epsilon_1$ and $\ell \in \{1 \dots, n_2(k)\}, j \in \{1 \dots, n_2(i)\}$ such that $\|\mathbf{c}_{i_j} - \mathbf{c}_{k_\ell}\|_{\mathbb{R}^{d_2}} \leq \epsilon_2$. Then, using

similar computations as before we obtain

$$\begin{aligned} & \int_{\Omega} \varphi(\mathbf{x}, \mathbf{c}) \nabla_{\mathbf{x}} \cdot (\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}) f^{\infty}(\mathbf{x}, \mathbf{c})) \, d\mathbf{x} d\mathbf{c} \\ &= -f_k f_i \frac{(\mathbf{x}_k - \mathbf{x}_i)}{\sigma(\mathbf{x}_k, \mathbf{c}_{k\ell})} \cdot \nabla_{\mathbf{x}} \varphi(\mathbf{x}, \mathbf{c})|_{\mathbf{x}=\mathbf{x}_k} \\ & \quad - f_i f_k \frac{(\mathbf{x}_i - \mathbf{x}_k)}{\sigma(\mathbf{x}_i, \mathbf{c}_{ij})} \cdot \nabla_{\mathbf{x}} \varphi(\mathbf{x}, \mathbf{c})|_{\mathbf{x}=\mathbf{x}_i} \neq 0 \end{aligned}$$

which contradicts the hypothesis. □

Lemma 3.3. *Assume that ϵ_1 and ϵ_2 are sufficiently large so that $\mathcal{A}(t, \mathbf{x}, \mathbf{c}, \mathbf{y}, \mathbf{z}; \epsilon_1, \epsilon_2) = 1$, for all $\mathbf{x}, \mathbf{y} \in \Omega_1$, $\mathbf{c}, \mathbf{z} \in \Omega_2$ and $t \geq 0$. Then, the distribution*

$$f^{\infty}(\mathbf{x}, \mathbf{c}) = \delta(\mathbf{x} - \bar{\mathbf{x}}) \tilde{f}^c(0, \mathbf{c}) = \prod_{k=1}^{d_1} \delta(x_k - \bar{x}_k) \tilde{f}^c(0, \mathbf{c})$$

with $\tilde{f}^c(0, \mathbf{c}) = \int_{\Omega_1} f_0(\mathbf{x}, \mathbf{c}) \, d\mathbf{x}$ is a weak stationary solution of the model (12) if and only if $\bar{x}_k = u_k(0)$.

Proof. Substituting the expression of f^{∞} in the weak form of the kinetic equation, integrating by parts and using the fact that for ϵ_1 and ϵ_2 large enough so that

$$\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}) = \int_{\Omega} \mathbf{y} f(t, \mathbf{y}, \mathbf{z}) \, d\mathbf{z} d\mathbf{y} - \mathbf{x} = \int_{\Omega} \mathbf{y} f_0(\mathbf{y}, \mathbf{z}) \, d\mathbf{z} d\mathbf{y} - \mathbf{x}$$

we obtain

$$\begin{aligned} & \int_{\Omega} \varphi(\mathbf{x}, \mathbf{c}) \nabla_{\mathbf{x}} \cdot (\mathbf{V}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}) f^{\infty}(\mathbf{x}, \mathbf{c})) \, d\mathbf{x} d\mathbf{c} \\ &= \int_{\Omega_2} \tilde{f}^c(\mathbf{c}) \left(\bar{\mathbf{x}} - \int_{\Omega} \mathbf{y} f_0(\mathbf{y}, \mathbf{z}) \, d\mathbf{z} d\mathbf{y} \right) \cdot \nabla_{\mathbf{x}} \varphi(\mathbf{x}, \mathbf{c})|_{\mathbf{x}=\bar{\mathbf{x}}} d\mathbf{c} \end{aligned}$$

which is zero if $\bar{\mathbf{x}} = \int_{\Omega} \mathbf{y} f_0(\mathbf{y}, \mathbf{z}) \, d\mathbf{z} d\mathbf{y}$. □

Remark 2. As direct consequence of Theorem 3.2 and Theorem 3.3 we have that taking ϵ_2 sufficiently large or an initial distribution $f_0(\mathbf{x}, \mathbf{c})$ being atomic with respect the second variable, the same quantized steady-state of the kinetic formulation of the microscopic model (1) are preserved, cf. [13].

Remark 3. In Theorem 3.2 the number of clusters n_1 , as well the values f_j and the positions \mathbf{x}_j of the clusters, are functions of the initial distribution $f_0(\mathbf{x}, \mathbf{c})$ and of the bounded confidence levels ϵ_1 and ϵ_2 . As pointed-out also in [14], in general it is not possible to predict the number of Dirac deltas in the asymptotic configuration from a given initial distribution. However, for ϵ_2 sufficiently large and assuming $\Omega = [0, 1]^d$, the number of clusters is $1 \leq \tilde{n} \leq \lfloor \frac{1}{\epsilon_1^d} \rfloor$. This consideration is also observed at the Boltzmann level, see [9].

4. Numerical experiments and applications. The theoretical results on the asymptotic behavior of the mean-field equation (12) introduced in the previous section are here also numerically investigated. Moreover, we show the efficiency of the model as technique to solve realistic data-clustering problems and to this end we focus on applications to the field of shape detection and image segmentation.

In order to efficiently solve the kinetic model (12) we employ the Mean Field Interaction Algorithm introduced in [2] which is based on random subset evaluation

Algorithm 1 Mean Field Interaction Algorithm for the kinetic equation (12).

- 1: Given N sample pairs (x_i^0, c_i^0) , with $i = 1, \dots, N$ computed from the initial distribution $f^0(x, c)$ and $M \leq N$;
- 2: **for** $k = 0$ **to** $k_{tot} - 1$ **do**
- 3: **for** $i = 1$ **to** N **do**
- 4: sample M data j_1, \dots, j_M uniformly without repetition among all data;
- 5: compute

$$\bar{A}(x_i^k, c_i^0) = \sum_{\ell=1}^M \mathcal{A}_{\epsilon_1, \epsilon_2}(x_i^k, c_i^0, x_{j_\ell}^k, c_{j_\ell}^0), \quad \bar{x}_i^k = \sum_{\ell=1}^M \frac{\mathcal{A}_{\epsilon_1, \epsilon_2}(x_i^k, c_i^0, x_{j_\ell}^k, c_{j_\ell}^0)}{\bar{A}(x_i^k, c_i^0)} x_{j_\ell}$$

- 6: compute the data change

$$x_i^{k+1} = x_i^k (1 - \Delta t \bar{A}(x_i^k, c_i^0)) + \Delta t \bar{A}(x_i^k, c_i^0) \bar{x}_i^k$$

- 7: **end for**
 - 8: **end for**
-

of the kernel term $\mathcal{A}_{\epsilon_1, \epsilon_2}(t, \mathbf{x}, \mathbf{c}, \mathbf{y}, \mathbf{z})$ given in (10). The algorithm used in the numerical experiments is summarized by the steps in Algorithm 1 for a time interval $[0, T]$ discretized in k_{tot} subintervals of size Δt . The computational cost of the algorithm is $O = (MN)$, where M is the size of the subset of interacting particles, and for $M = N$ we obtain the explicit Euler scheme for the original N particle system (8) whose cost is $O(N^2)$. For further details on the Mean Field Interaction Algorithm we refer to [2].

4.1. Numerical steady–states and moment evolution. We use the Mean Field Interaction Algorithm to numerically investigate the properties of the kinetic model proved in the previous section. We analyze two typical situations which lead to the two applications we show later in this section.

4.1.1. Constant static feature. First, we consider an initial distribution being the uniform distribution with respect to \mathbf{x} and a constant distribution along the static feature variable \mathbf{c} , i.e. $f_0(\mathbf{x}, \mathbf{c}) = \chi_{[0,1]^{d_1}}(\mathbf{x})$. This choice is obviously also consistent with considering the bounded confidence level ϵ_2 very large. We show that, as discussed in Remark 2, equation (12) provides the same steady–states of the Hegselmann–Krause model.

One–dimension. The numerical steady–states of the mean–field equation (12) are first computed for the one–dimensional case and compared to the steady–states of the microscopic model (8). Observe that, under the assumption of an initial constant distribution along \mathbf{c} , the asymptotic behavior of (8) is the classical one prescribed by the one–dimensional Hegselmann–Krause model (1). The evolution in time of the first and second moment is also provided.

In Figure 1 we show the steady–state provided by the mean–field kinetic model (12) for the one–dimensional initial uniform distribution on $[0, 1]$. The final time is $T = 20$ and the time step is $\Delta t = 0.5$. We consider $N = 5 \times 10^5$ particles in the Monte Carlo method so that we reduce error due to the sampling. The number of interacting particles is taken as $M = 10$. We show the results for two values of the confidence bound, $\epsilon_1 = 0.5$ in the left panel and $\epsilon_1 = 0.15$ in the right panel. The evolution of the distributions up to final time is showed by normalizing with respect to the maximum value at each fixed time. For $\epsilon_1 = 0.5$ we observe the formation

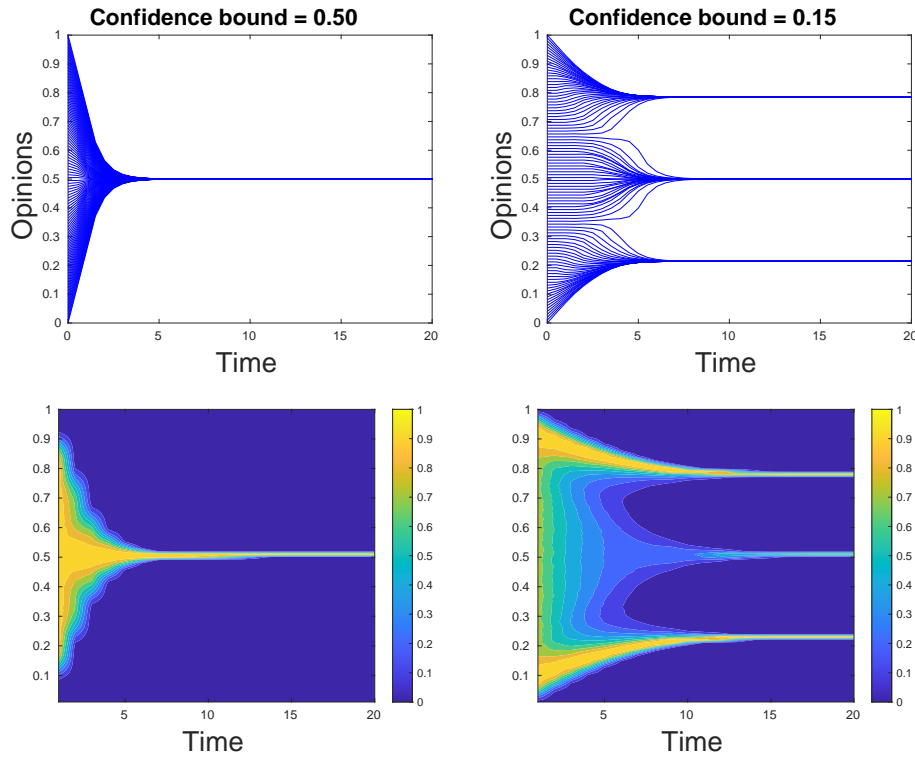


FIGURE 1. Trend to the steady-state of the one-dimensional Hegselmann-Krause model (1) with $n = 100$ agents equally spaced at initial time and non-symmetric interactions (top row) and of the mean-field model (12) computed with Algorithm 1 (bottom row) up to final time $T = 20$. Left panels show the case for $\epsilon_1 = 0.5$, the right panels show the case for $\epsilon_1 = 0.15$.

of a consensus state in large time behavior. In fact, the final distribution is a Dirac delta centered in the initial value of the first moment. For $\epsilon_1 = 0.15$, instead, three clusters arise at equilibrium, similarly the classical Hegselmann-Krause model for the same confidence bounds and equally distributed initial data.

As observed in Figure 2, for both values of the bounded confidence level, we have conservation of the first moment and energy dissipation in time. Although we are using a non-symmetric model, the first moment is preserved during the time evolution, with deviation of order 10^{-4} from the initial value, since the initial distribution is symmetric. The fluctuations in the first moment are due to the stochastic method.

In Figure 3 we show some results of a simple analysis of Algorithm 1 with respect to the values of M , i.e. the size of the random subset of interacting particles. Left panel shows the trend to the steady-state of the kinetic model computed with Algorithm 1 and same parameters as in the right plot of Figure 1. However, here we consider $N = 2 \times 10^4$ and $M = 2$. The same equilibrium state is reached but with a faster transient. The dependence of the velocity to the formation of cluster on the size M of the random subset is showed in the right panel of Figure 3. The velocity

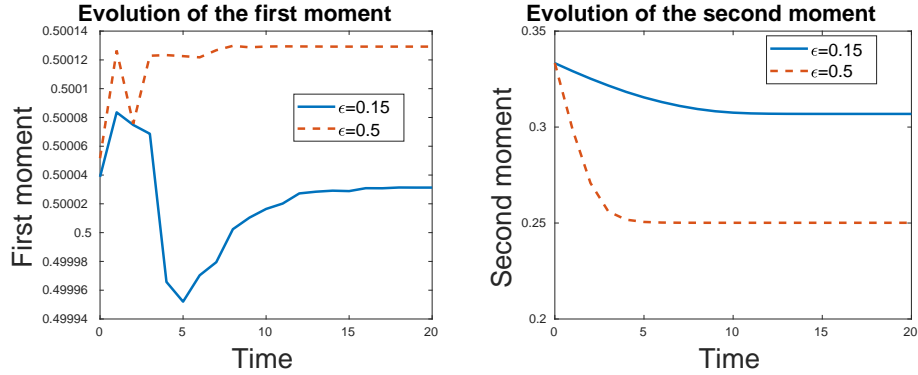


FIGURE 2. Evolution in time of the first moment (left) and of the second moment (right) for the two values of the bounded confidence level $\epsilon_1 = 0.5$ (dashed lines) and $\epsilon_1 = 0.15$ (solid lines).

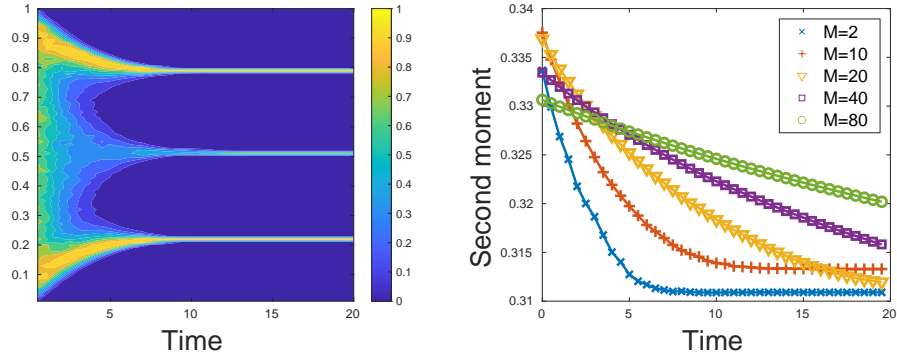


FIGURE 3. Left: trend to the steady-state of the mean-field model (12) computed with Algorithm 1 with $N = 2 \times 10^4$, $M = 2$, $\epsilon_1 = 0.15$ and up to final time $T = 20$. Right: energy decay of the mean-field model (12) for several values of interacting particles M .

to equilibrium decreases as M increases. In particular, the $M = 2$ case, where each particle is enforced to align to the velocity of the other particle instead of their “mean”, exhibits the fastest convergence. This analysis suggests that, although model (12) has a convolution structure and in principle can be efficiently solved with a Fast Fourier Transform at a $O(N \log N)$ cost, Algorithm 1 may be preferable since it has a comparable (or even lower) computational cost and it can be applied to more general kernels where the convolution structure is lost.

Two-dimensions. We also show one example of the numerical steady-states of the mean-field equation (12) in the two-dimensional case. Again we employ Algorithm 1. Figure 4 shows the steady-state for the particle density and the kinetic density at time $t = 4$ and final time $T = 50$. We use $N = 10000$ particles and the time step is $\Delta t = 0.5$. Using the Mean Field Interaction Algorithm 1, the number of interacting particles is taken as $M = 10$. We show the results for one value of the confidence bound, $\epsilon_1 = 0.15$ which leads to the formation of 8

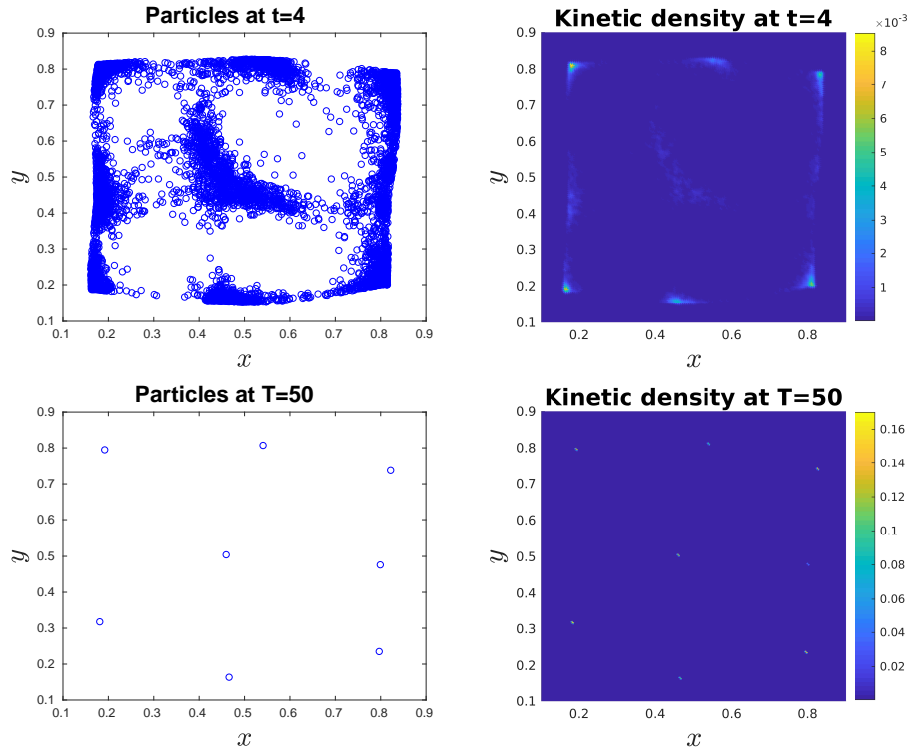


FIGURE 4. Particle solution (left plots) with $N = 10000$ and kinetic density (right plots). Results are provided at time $t = 4$ (top row) and final time $T = 50$ (bottom row). The bounded confidence level is $\epsilon_1 = 0.15$.

clusters at equilibrium. The evolution in time of the two-dimensional moments is also provided in Figure 5. We still observe conservation of the first moment, due to the initial symmetric distribution, and decay in time for the two non-mixed second moments.

4.1.2. *Non-constant static feature.* Next, we consider a non-constant initial distribution along the variable \mathbf{c} . We consider a one-dimensional setting both along \mathbf{x} and \mathbf{c} since the goal of this experiment is to provide evidence of the influence of the static feature on the clustering process. The initial kinetic distribution is given as a tensor product between a uniform and Gaussian distribution so that

$$f_0(x, c) = \chi_{[0,1]}(x) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(c-\mu)^2}{2\sigma^2}\right),$$

with mean $\mu = 0.5$, and variance $\sigma^2 = 0.3$. See the top left plot in Figure 6. All the simulations are performed by using $N = 5000$ samples up to equilibrium.

In this experiment, we expect that the behavior at equilibrium is influenced also by the interactions with respect to the static feature variable due to the corresponding characteristic function in the kernel (10). In Figure 6 we show the results provided by using $\epsilon_1 = 0.15$ and $\epsilon_2 = 0.025$. The top left plot shows particles and kinetic density at equilibrium. Observe that, compared to the case with constant

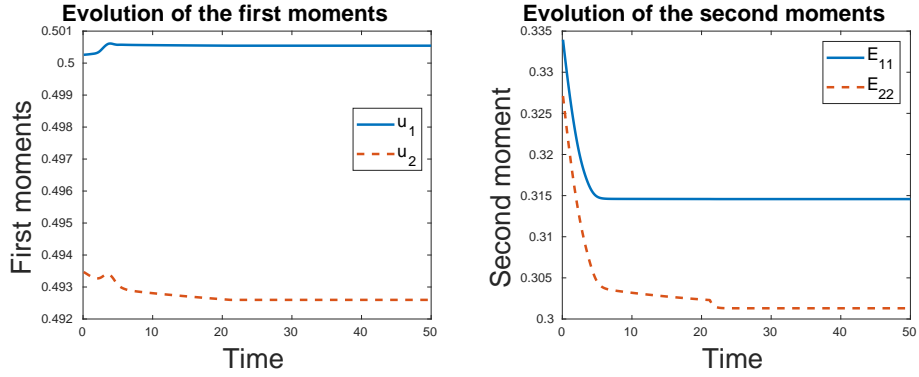


FIGURE 5. Evolution in time of the two-dimensional first moments (left) and second moments (right) for the bounded confidence level $\epsilon_1 = 0.15$.

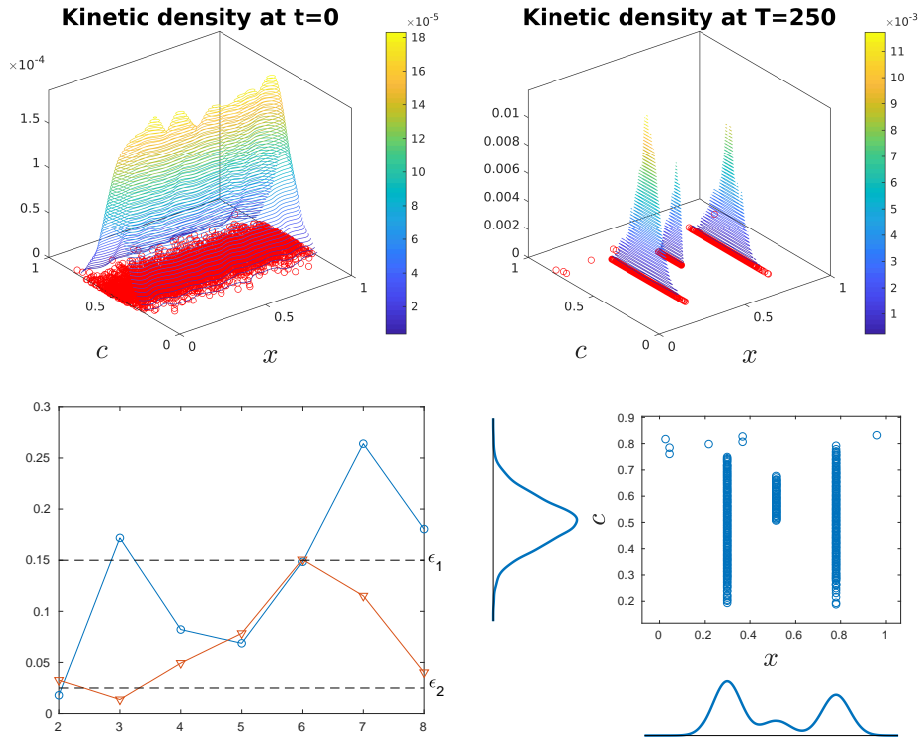


FIGURE 6. Top row: particles and kinetic density at initial time (left plot) and at equilibrium (right plot). Bottom row: at left, analysis of the distances between clusters in x (blue line with circle markers) and c direction (red line with triangle markers); at right, plot of the marginals. Confidence levels are $\epsilon_1 = 0.15$ and $\epsilon_2 = 0.025$.

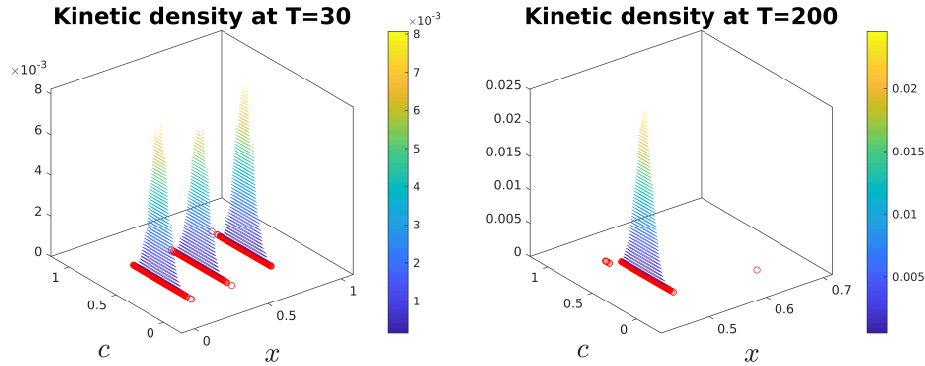


FIGURE 7. Particle and kinetic density at equilibrium with confidence levels $\epsilon_1 = 0.15$, $\epsilon_2 = 0.1$ (left) and $\epsilon_1 = 1$, $\epsilon_2 = 0.025$ (right).

distribution along the static feature, the low value of ϵ_2 allows for the formation of more clusters, precisely 8. Clusters arise because either their distance in x direction is larger than the confidence level ϵ_1 or the minimum distance in c direction is larger than ϵ_2 . This consideration is analyzed in the bottom left plot of Figure 6 where the blue line with circle markers shows the distance in x between two adjacent clusters and the red line with triangle markers the minimum distance between the static features of particles being in two adjacent clusters. Observe that when the distance in x is lower than the value ϵ_1 , the corresponding distance between the static features is larger than ϵ_2 , and therefore cluster is no longer possible. Moreover, we notice that time to reach equilibrium is larger due to two levels of clustering. Finally, the bottom right plot of Figure 6 shows the two marginal distributions.

In Figure 7 we show the behavior at equilibrium by considering two different pairs of confidence levels. In the left plot, $\epsilon_1 = 0.15$ and $\epsilon_2 = 0.1$. Here ϵ_2 is large enough to not influence the clustering process and in fact we recover 3 clusters, exactly the same provided in Figure 1 when the initial kinetic distribution was constant in c . In the right plot, $\epsilon_1 = 1$ and $\epsilon_2 = 0.025$. In this case, ϵ_1 is very large and we show the formation of clusters at equilibrium when only clustering due to the static feature is allowed.

4.2. Clustering and shape detection. The property of the equation (12), analytically studied in Section 3.1 and numerically investigated in Section 4.1, of having quantized steady-states makes the model suitable for solving data clustering problems. In particular, in this section we use the kinetic model as technique for shape detection.

Shape detection can be considered as a branch of pattern recognition problems that focuses on the discovering of previously unknown patterns in a big set of data. Machine learning and many clustering algorithms have already been applied to this class of problems, such as the k -means algorithm which however suffers from the necessity of defining a-priori the number of clusters. Instead, as already pointed out, opinion dynamics based models automatically find the number of clusters as functions of the confidence value.

TABLE 1. Number of clusters and errors depending on the bounded confidence value and the percentage of the noise when uniformly distributed.

$\alpha = 5\%$			$\alpha = 7.5\%$			$\alpha = 10\%$		
ϵ_1	\mathcal{E}	\tilde{n}	ϵ_1	\mathcal{E}	\tilde{n}	ϵ_1	\mathcal{E}	\tilde{n}
0.03	1.25e-02	30	0.03	3.47e-02	51	0.06	2.64e-02	11
0.04	4.10e-03	16	0.05	1.21e-02	14	0.07	1.48e-02	8
0.05	4.00e-03	12	0.07	7.70e-03	8	0.08	1.12e-02	8
0.06	4.60e-03	9	0.09	7.90e-03	8	0.09	1.63e-02	5
0.07	5.40e-03	8	0.11	1.66e-02	3	0.10	1.63e-02	5

More specifically, in this section we apply the models introduced in Section 2 to the example of a character recognition. We consider a letter “A” which is composed by n two-dimensional points $\mathcal{L} = \{\mathbf{x}_i\}_{i=1}^n$ defining three segments in $\Omega = [0, 1]^2$. To each of these points we apply an additive noise distributed according to a uniform or a Gaussian distribution. We obtain a new set of n points

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i + \alpha \boldsymbol{\theta}_i, \quad \alpha > 0, \quad \boldsymbol{\theta}_i \sim \mathcal{U}(-1, 1) \quad \text{or} \quad \boldsymbol{\theta}_i \sim \mathcal{N}(0, 1)$$

which define the noisy pattern. Here, the value α represents the percentage of noise and it is also chosen in such a way that $\tilde{\mathbf{x}}_i \in \Omega$, $i = 1, \dots, n$. The points $\{\tilde{\mathbf{x}}_i\}_{i=1}^n$ are used as initial condition of the model. They are seen as samples from a distribution defining the noisy pattern and the Mean Field Interaction Algorithm 1 is employed in order to solve the kinetic model.

The goal of this example is to cluster the noisy data into a set of points which give information on the shape of the exact pattern to be detected. This can be considered also as a dimensionality reduction problem which could help other algorithms to recognize the unknown pattern efficiently by using information on the position of the clusters. Here the efficiency of the results is studied by means of the following measure:

$$\mathcal{E}(\epsilon_1, \epsilon_2, \alpha) = \frac{1}{\tilde{n}} \sum_{k=1}^{\tilde{n}} \min_{\mathbf{x} \in \mathcal{L}} \|\mathbf{C}_k - \mathbf{x}\|_2$$

where $\{\mathbf{C}_k\}_{k=1}^{\tilde{n}}$ is the position of clusters in Ω , \tilde{n} the number of clusters at equilibrium. The quantity \mathcal{E} measures for each cluster the minimum 2-norm distance to the exact pattern and then computes the normalized 1-norm of these quantities. The definition of the measure \mathcal{E} is inspired by the idea of computing an average of the minimum distances between the clusters and the shape \mathcal{L} , when formation of multiple clusters distributed close the shape \mathcal{L} arises, as in the cases showed here. Certainly, \mathcal{E} cannot provide a general way to measure the quality of the results, since e.g. in case of one steady-state cluster positioned exactly on a point of \mathcal{L} , we would have $\mathcal{E} = 0$ providing absurdly the optimal choice.

In the following, all the results are given for an initial set of $n = 5000$ particles and a large enough final time to reach a steady-state $T = 50$. Moreover, we use the Euclidean norm and all particles are initialized with a constant static feature. For an extension of this application to clustering also with respect to the static feature see Remark 4 below.

In Table 1 and Table 2 we show a sensitivity study of the error \mathcal{E} as function of the percentage of noise α and of the bounded confidence value ϵ_1 , for initial noisy data uniformly and normally distributed, respectively. The number of clusters at

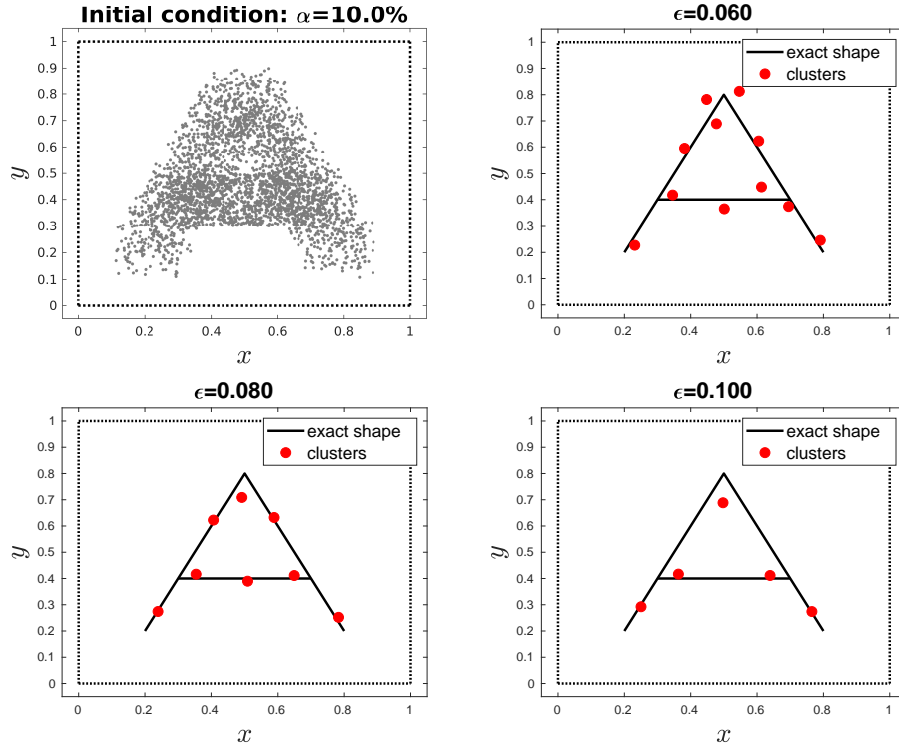


FIGURE 8. Shape detection of the letter “A” initialized with 10% of a uniform additive noise. Top left panel shows the initial condition. We show clusters obtained with bounded confidence values $\epsilon_1 = 0.06$ (top right), $\epsilon_1 = 0.08$ (bottom left) and $\epsilon_1 = 0.1$ (bottom right).

TABLE 2. Number of clusters and errors depending on the bounded confidence value and the percentage of the noise when normally distributed.

$\alpha = 5\%$			$\alpha = 5.5\%$			$\alpha = 6\%$		
ϵ_1	\mathcal{E}	\tilde{n}	ϵ_1	\mathcal{E}	\tilde{n}	ϵ_1	\mathcal{E}	\tilde{n}
0.05	4.44e-02	23	0.05	4.73e-02	24	0.05	6.37e-02	30
0.06	1.36e-02	11	0.06	2.62e-02	13	0.06	4.16e-02	16
0.065	6.40e-03	9	0.065	1.63e-02	11	0.07	2.12e-02	9
0.07	6.70e-03	7	0.0675	7.40e-03	10	0.075	9.70e-03	7
0.08	8.50e-03	7	0.07	8.00e-03	9	0.08	9.20e-03	7
0.09	1.00e-02	6	0.08	9.80e-03	7	0.085	1.10e-02	5

equilibrium is also provided. Results with minimal error value are highlighted with bold font. In particular, we point out as the increasing percentage of noise highly affects the results.

In Figure 8 and Figure 9 some of this cases are shown. Precisely, for the uniformly distributed initial noisy data, Figure 8, we consider the largest value of the percentage of noise $\alpha = 10\%$ and three values of the bounded confidence level ϵ_1 :

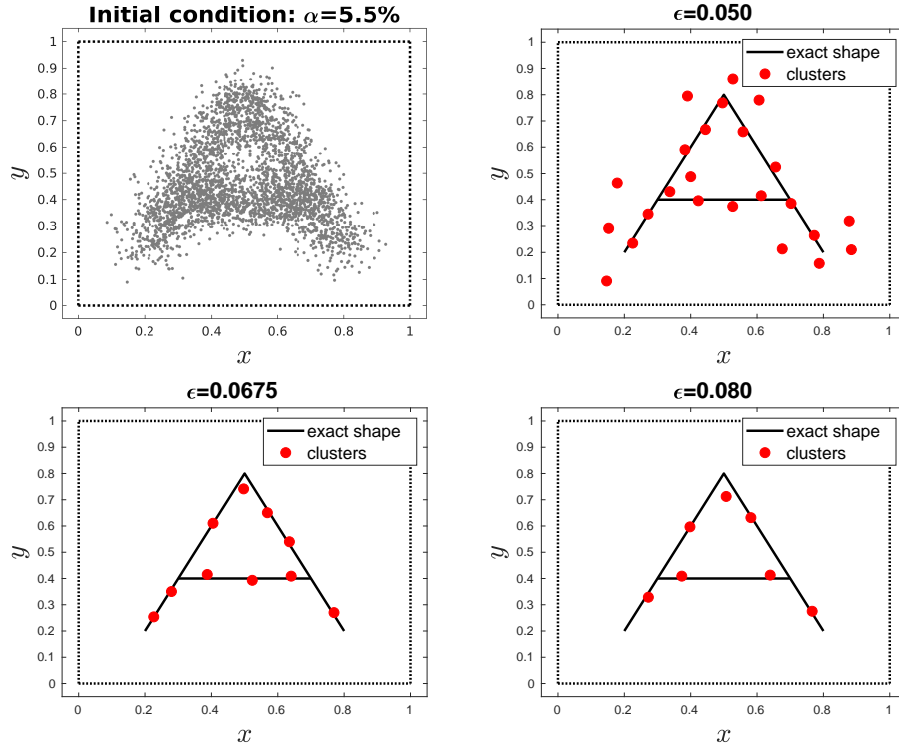


FIGURE 9. Shape detection of the letter “A” initialized with 5.5% of a Gaussian additive noise. Top left panel shows the initial condition. We show clusters obtained with bounded confidence values $\epsilon_1 = 0.05$ (top right), $\epsilon_1 = 0.0675$ (bottom left) and $\epsilon_1 = 0.08$ (bottom right).

the smallest $\epsilon_1 = 0.06$ in Table 1, the optimal $\epsilon_1 = 0.08$ (giving the smallest value of the error \mathcal{E}) and the largest one $\epsilon_1 = 0.1$. We observe that the result is highly influenced by the choice of the confidence bound which determines the number of final clusters. Smaller and larger values of ϵ_1 result in larger values of the error measure \mathcal{E} . In fact, in one case the model provides too many clusters which are very spread out, with several outliers. In the other case too few clusters so that all the information on the exact shape is lost. For the optimal value of ϵ_1 the model provides a good agreement with the exact pattern.

Similar considerations hold for Figure 9 where we show the noisy initial data obtained with $\alpha = 10\%$ percentage of noise and three values of the bounded confidence level $\epsilon_1 = 0.05$, $\epsilon_1 = 0.0675$ (giving the smallest value of the error \mathcal{E}) and $\epsilon_1 = 0.08$.

We point out that the goal of this application is to provide a preliminary step for shape feature extraction. This technique should be coupled with a further algorithm which is able to select the correct letter from a given alphabet. The dimensionality reduction provided by the data clustering approach allows to efficiently employ the subsequent recognition analysis.

Remark 4 (Shape detection with non-constant static feature). In the previous examples the static features is taken constant for all the data. In the non-constant case, one could think of this feature as an additional initial given input which measures the quality of the information of each single point and that can be modeled by the distance to its exact value in \mathcal{L} .

4.3. Color image segmentation. We now turn to the second application of the data clustering model 12 which is the segmentation of gray scale images. Image segmentation is widely used in medical and astronomical image processing, face recognition, etc. In fact, this technique allows to partition an image in sets of significant regions of pixels sharing same characteristics such as closeness and similar intensity of color. As in shape detection, the aim of such approach is to modify the description of an image into a structure which easier to be analyzed for subsequent computer vision algorithms.

Several mathematical techniques have been applied to image segmentation. For instance we mention level set methods [39, 43] methods based on the Kuramoto model [29, 37] and supervised convolution neural networks [17, 46]. A more complicated procedure to segmentation of images is the clustering approach and we refer e.g. to the k -means method [44, 47], c -means method [33] and hierarchical clustering method [45].

Here, we study the efficiency of the kinetic model (12) to solve image segmentation problems and therefore we still relies on the clustering technique. Thanks to the model introduced in this paper, segmentation can be performed by using two levels of clustering in order to determine regions of pixels with similar characteristics. More precisely, we cluster based on the Euclidean distance and the distance of the gray intensity between pixels. Taking into account spatial coordinates is needed in order to avoid the problem of selecting homogeneous regions of pixels which are however distinct in the original image. The clustering with respect the intensity of the gray colors is performed by using the static feature variable c .

The application of (12) works as described in the following. Number of pixels is the number of particles which are assumed to be equally spaced samples. The intensity of the gray color of each pixel is computed at initial time and defines the one-dimensional static feature c of each particle. The Mean Field Interaction Algorithm 1 is then applied to find clusters representing local regions of the original image with homogeneous intensity of color. At equilibrium, we compute the mean of the color intensity of pixels belonging to the same cluster and then they are mapped back to the original positions in the initial image to get the segmentation. Another possible approach, which is more suitable for images with very sharp color intensities, such as in astronomical images, is to apply thresholding technique. This method is based on replacing the color intensity of each cluster by black if the mean is below a certain threshold or by white otherwise. This technique results therefore in binary images.

In Figure 10 we first apply the segmentation to a benchmark test in order to show how the model works on a simple test. We initialize an gray scale image with 4096 pixels with four regions of different intensity colors. The top and the bottom left corners have intensity 1 and 0.75, respectively. The top and the bottom right corners have intensity 0 and 0.25, respectively. See the left panel of Figure 10. The positions of the pixels are rescaled on $\Omega = [0, 1]^2$. The clustering algorithm is applied with the two bounded confidence values $\epsilon_1 = 0.5$ and $\epsilon_2 = 0.3$. At equilibrium the two red clusters showed in the middle panel of Figure 10 are computed by the kinetic

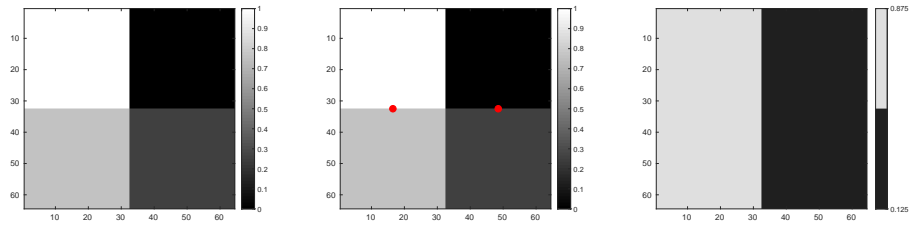


FIGURE 10. Left panel: initial image of 4096 pixels with four regions with different gray intensity. Middle panel: red dots show the positions of the clusters at equilibrium. Right panel: segmentation of the initial image in two regions.

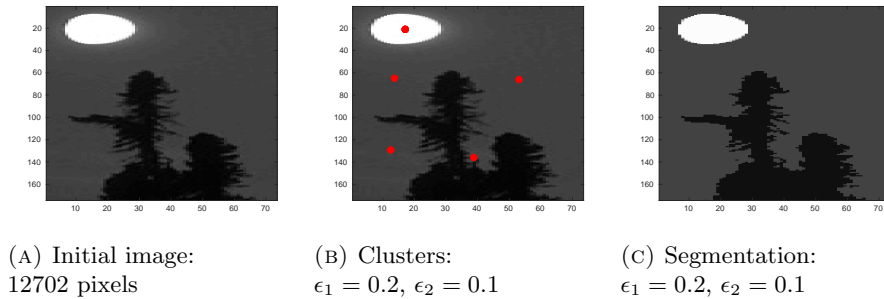


FIGURE 11. Image segmentation of 174×73 gray scale image taken by the data-set [6].

model. The mean of the intensity colors of the pixels in each cluster is computed and mapped back to the initial position obtaining the segmentation in the right panel of Figure 10. The two values of the gray intensity are 0.125 and 0.875.

We now apply the segmentation process to real images taken from different data-sets. In Figure 11 we consider an image of 174×73 pixels showing a night sky, trees and the moon. On this example, the method is able to provide good results with a wide range of confidence values since the variation of the gray scale is sharp between the objects which are also characterized by regions having homogeneous colors. In Figure 11 we report the result obtained with $\epsilon_1 = 0.2$ and $\epsilon_2 = 0.1$ which results in the formation of 5 clusters.

The second image in Figure 12 is also characterized by two backgrounds: the sky which is very homogeneous and the field which is less homogeneous. The goal of the segmentation process would be to identify the two cows, separating them from the background. We report the results obtained with $\epsilon_1 = 0.2$ and $\epsilon_2 = 0.05$, resulting in 6 clusters. Observe that ϵ_2 is taken small in order to avoid the formation of many clusters due to the low homogeneity of the field. The two cows are well identified by the method.

In the case of Figure 13 the confidence level ϵ_2 is taken larger in order to force clustering of regions with snow but different shadows. The clustering method identifies 8 clusters and the skier is well separated by the background.

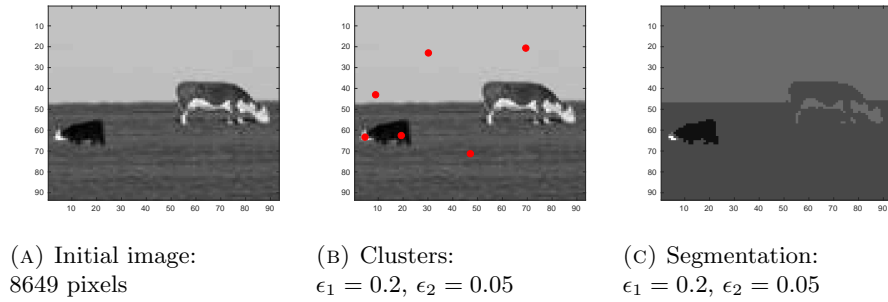


FIGURE 12. Image segmentation of 93×93 gray scale image taken by the data-set [24].

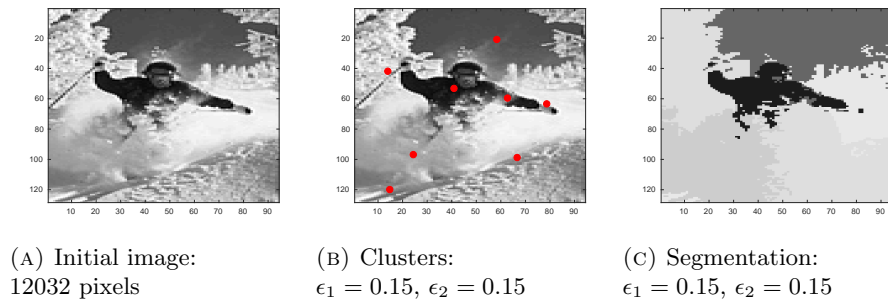


FIGURE 13. Image segmentation of 128×94 gray scale image taken by the data-set [32].

In Figure 14 we employ the kinetic model for clustering to segmentation of a real image taken by [32]. The true image is 67×67 pixels and several values of the bounded confidence levels ϵ_1 and ϵ_2 are applied in order to cluster with respect positions and intensity color of pixels. The best result is obtained with $\epsilon_1 = 0.3$ and $\epsilon_2 = 0.1$ which results in the formation of 8 clusters. Lower values of ϵ_1 allows to obtain more clusters and thus more gray intensity colors at equilibrium. Increasing ϵ_1 requires to not take ϵ_2 also large, otherwise many information are lost at equilibrium.

Finally, in Figure 15 we present an additional experiment aimed to provide a comparison between our segmentation process and the one published in [29] based on the Kuramoto model. The image is selected from the data-set [32]. Right panel of Figure 15 has to be compared with Figure (7c) in [29] and, in particular, we observe that our result is able to detect better some structures of the image that are lost by using the Kuramoto model.

5. Conclusions. In this paper we have proposed a new method for clustering of large data which is based on a suitable generalization of the Hegselmann-Krause opinion dynamics model. The extension accounts for clustering with respect to two different sets of features of each single datum: one set describes time dependent characteristics, the other one represents static characteristics.

The mean-field limit of the particle model has been formally derived and the resulting kinetic equation allows for the study of mathematical properties. Time

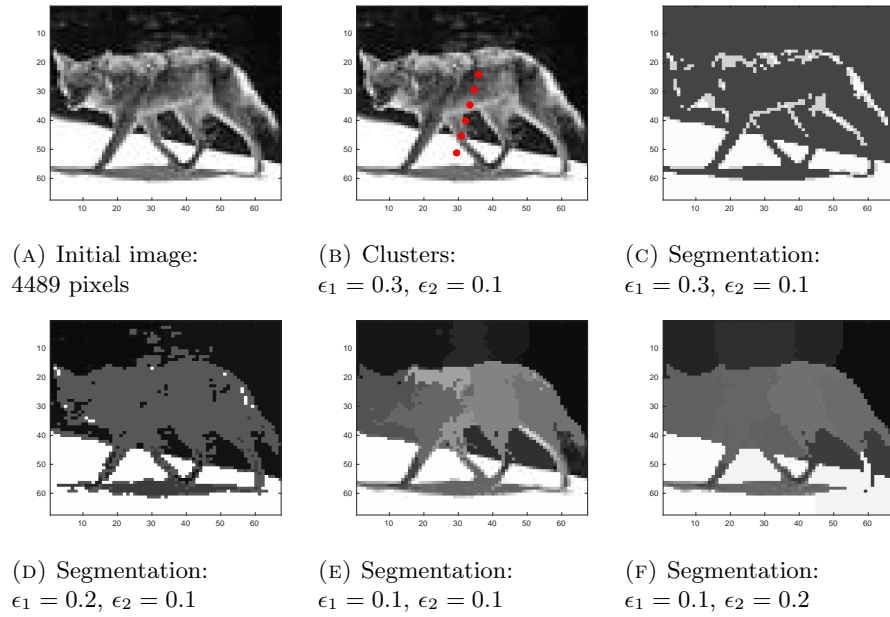


FIGURE 14. Image segmentation of 67×67 gray scale image taken by the data-set [32].

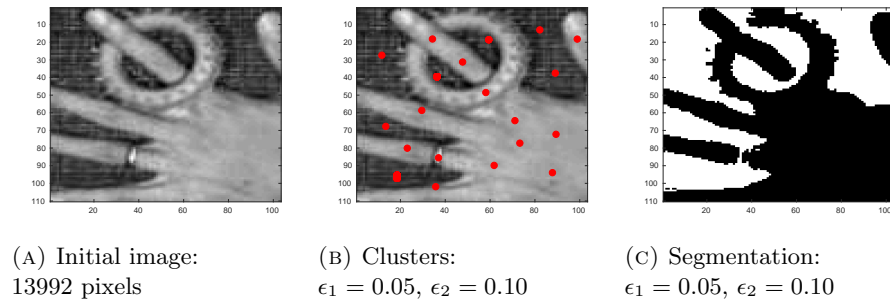


FIGURE 15. Image segmentation of 132×106 gray scale image taken by the data-set [32].

evolution of moments and asymptotic behavior of kinetic equation have been analytically characterized. Moreover, the derivation of the mean-field model allows for reduction of computational complexity thanks to a suitable random subset Monte Carlo algorithm.

Several applications to digital imaging have been proposed. In particular, we have focused on application to shape detection and image segmentation, showing the efficiency of the model to provide satisfying results. Finally, we emphasize that the present model has to be intended as a starting point towards more realistic applications, for example based on the use of non static features combined with suitable machine learning approaches.

Acknowledgments. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC-2023 Internet of Production – 390621612.

Lorenzo Pareschi and Giuseppe Visconti acknowledge the support of the “National Group for Scientific Computation (GNCS-INDAM)”, project “Numerical approximation of hyperbolic problems and applications”.

REFERENCES

- [1] G. Albi, N. Bellomo, L. Fermo, S.-Y. Ha and J. Kim, et al., [Vehicular traffic, crowds, and swarms.: From kinetic theory and multiscale methods to applications and research perspectives](#), *Math. Models Methods Appl. Sci.*, **29** (2019), 1901–2005.
- [2] G. Albi and L. Pareschi, [Binary interaction algorithms for the simulation of flocking and swarming dynamics](#), *Multiscale Model. Simul.*, **11** (2013), 1–29.
- [3] G. Albi and L. Pareschi, [Modeling of self-organized systems interacting with a few individuals: From microscopic to macroscopic dynamics](#), *Appl. Math. Lett.*, **26** (2013), 397–401.
- [4] G. Albi, L. Pareschi, G. Toscani and M. Zanella, [Recent advances in opinion modeling: Control and social influence](#), in *Active Particles, Vol. 1*, Model. Simul. Sci. Eng. Technol., Birkhäuser/Springer, Cham, 2017, 49–98.
- [5] G. Albi, L. Pareschi and M. Zanella, [Opinion dynamics over complex networks: Kinetic modeling and numerical methods](#), *Kinet. Relat. Models*, **10** (2017), 1–32.
- [6] P. Arbeláez, M. Maire, C. Fowlkes and J. Malik, [Contour detection and hierarchical image segmentation](#), *IEEE Trans. Pattern Anal. Mach. Intell.*, **33** (2011), 898–916.
- [7] V. D. Blondel, J. M. Hendrickx and J. N. Tsitsiklis, [On Krauses’s multi-agent consensus model with state-dependent connectivity](#), *IEEE Trans. Automat. Control*, **54** (2009), 2586–2597.
- [8] V. D. Blondel, J. M. Hendrickx and J. N. Tsitsiklis, [Continuous time average-preserving opinion dynamics with opinion-dependent communications](#), *SIAM J. Control Optim.*, **48** (2010), 5214–5240.
- [9] D. Borra and T. Lorenzi, [Asymptotic analysis of continuous opinion dynamics models under bounded confidence](#), *Commun. Pure Appl. Anal.*, **12** (2013), 1487–1499.
- [10] L. Boudin, R. Monaco and F. Salvarani, [Kinetic model for multidimensional opinion formation](#), *Phys. Rev. E*, **81** (2010), 9pp.
- [11] L. Boudin and F. Salvarani, [Opinion dynamics: Kinetic modelling with mass media, application to the Scottish independence referendum](#), *Phys. A*, **444** (2016), 448–457.
- [12] J. A. Cañizo, J. A. Carrillo and J. Rosado, [A well-posedness theory in measures for some kinetic models of collective motion](#), *Math. Models Methods Appl. Sci.*, **21** (2011), 519–539.
- [13] C. Canuto, F. Fagnani and P. Tilli, [An Eulerian approach to the analysis of rendez-vous algorithms](#), *IFAC Proceedings Volumes*, **41** (2008), 9039–9044.
- [14] C. Canuto, F. Fagnani and P. Tilli, [An Eulerian approach to the analysis of Krause’s consensus models](#), *SIAM J. Control Optim.*, **50** (2012), 243–265.
- [15] J. A. Carrillo, Y.-P. Choi, C. Totzeck and O. Tse, [An analytical framework for consensus-based global optimization method](#), *Math. Models Methods Appl. Sci.*, **28** (2018), 1037–1066.
- [16] J. A. Carrillo, M. Fornasier, G. Toscani and F. Vecil, [Particle, kinetic, and hydrodynamic models of swarming](#), in *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, Model. Simul. Sci. Eng. Technol., Birkhäuser Boston, Boston, MA, 2010, 297–336.
- [17] L.-C. Chen, P. G., I. Kokkinos, K. Murphy and A. Loddon Yuille, [DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs](#), *IEEE Trans. Pattern Anal. Machine Intelligence*, **40** (2018), 834–848.
- [18] I. D. Couzin, J. Krause, N. R. Franks and S. A. Levin, [Effective leadership and decision-making in animal groups on the move](#), *Nature*, **433** (2005), 513–516.
- [19] F. Cucker and S. Smale, [Emergent behavior in flocks](#), *IEEE Trans. Automat. Control*, **52** (2007), 852–862.
- [20] P. Degond and S. Motsch, [A macroscopic model for a system of swarming agents using curvature control](#), *J. Stat. Phys.*, **143** (2011), 685–714.

- [21] F. Dietrich, S. Martin and M. Jungers, [Transient cluster formation in generalized Hegselmann-Krause opinion dynamics](#), 2016 European Control Conference (ECC), Aalborg, Denmark, 2016.
- [22] J. C. Dittmer, [Consensus formation under bounded confidence](#), *Nonlinear Anal.*, **47** (2001), 4615–4622.
- [23] B. Düring, P. Markowich, J.-F. Pietschmann and M.-T. Wolfram, [Boltzmann and Fokker-Planck equations modelling opinion formation in the presence of strong leaders](#), *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, **465** (2009), 3687–3708.
- [24] S. Gould, R. Fulton and D. Koller, [Decomposing a scene into geometric and semantically consistent regions](#), 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009.
- [25] R. Hegselmann and U. Krause, [Opinion dynamics and bounded confidence: Models, analysis and simulation](#), *J. Artificial Societies Social Simulation*, **5** (2002).
- [26] J. M. Hendrickx, *Graphs and Networks for the Analysis of Autonomous Agent Systems*, Ph.D thesis, Ecole Polytechnique de Louvain, 2008.
- [27] P.-E. Jabin and S. Motsch, [Clustering and asymptotic behavior in opinion formation](#), *J. Differential Equations*, **257** (2014), 4165–4187.
- [28] J. Kennedy and R. Eberhart, [Particle swarm optimization](#), *Proc. Internat. Conference Neural Networks*, Perth, Australia, 1995.
- [29] X. Liu, Y. Qiao, X. Chen, J. Miao and L. Duan, [Color image segmentation based on modified Kuramoto model](#), *Procedia Comp. Sci.*, **88** (2016), 245–258.
- [30] J. Lorenz, [A stabilization theorem for dynamics of continuous opinions](#), *Phys. A*, **355** (2005), 217–223.
- [31] J. MacQueen, [Some methods for classification and analysis of multivariate observations](#), *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability*, Vol. I: Statistics, Univ. California Press, Berkeley, CA, 1967, 281–297.
- [32] D. Martin, C. Fowlkes, D. Tal and J. Malik, [A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics](#), *Proc. Eighth IEEE Internat. Conference Comp. Vision*, Vancouver, Canada, 2001.
- [33] K. H. Memon, S. Memon, M. A. Qureshi, M. B. Alvi, D. Kumar and R. A. Shah, [Kernel possibilistic fuzzy c-means clustering with local information for image segmentation](#), *Internat. J. Fuzzy Syst.*, **21** (2019), 321–332.
- [34] S. Motsch and E. Tadmor, [A new model for self-organized dynamics and its flocking behavior](#), *J. Stat. Phys.*, **144** (2011), 923–947.
- [35] S. Motsch and E. Tadmor, [Heterophilious dynamics enhances consensus](#), *SIAM Review*, **56** (2014), 577–621.
- [36] A. Nedić and B. Touri, [Multi-dimensional Hegselmann-Krause dynamics](#), 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), Maui, HI, 2012.
- [37] A. V. Novikov and E. N. Benderskaya, [Oscillatory neural networks based on the Kuramoto model for cluster analysis](#), *Pattern Recognition Image Anal.*, **24** (2014), 365–371.
- [38] G. Oliva, D. La Manna, A. Fagiolini and R. Setola, [Distributed data clustering via opinion dynamics](#), *Internat. J. Distributed Sensor Networks*, **11** (2015).
- [39] S. Osher and J. A. Sethian, [Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations](#), *J. Comput. Phys.*, **79** (1988), 12–49.
- [40] L. Pareschi and G. Toscani, *Interacting Multiagent Systems. Kinetic Equations and Monte Carlo Methods*, Oxford University Press, 2013.
- [41] R. Pinnau, C. Totzeck, O. Tse and S. Martin, [A consensus-based model for global optimization and its mean-field limit](#), *Math. Models Methods Appl. Sci.*, **27** (2017), 183–204.
- [42] G. Puppo, M. Semplice, A. Tosin and G. Visconti, [Kinetic models for traffic flow resulting in a reduced space of microscopic velocities](#), *Kinet. Relat. Models*, **10** (2017), 823–854.
- [43] J. A. Sethian, *Level Set Methods and Fast Marching Methods. Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*, Cambridge Monographs on Applied and Computational Mathematics, 3, Cambridge University Press, Cambridge, 1999.
- [44] P. Shan, [Image segmentation method based on K-mean algorithm](#), *EURASIP J. Image Video Processing*, **81** (2018).
- [45] L. Tian, L. Han and J. Yue, [Research on image segmentation based on clustering algorithm](#), *Internat. J. Signal Process. Image Process. Pattern Recognition*, **9** (2016), 1–12.

- [46] L. Zhang, Y. Gao, Y. Xia, K. Lu, J. Shen and R. Ji, [Representative discovery of structure cues for weakly-supervised image segmentation](#), *IEEE Transac. Multimedia*, **16** (2014), 470–479.
- [47] X. Zheng, Q. Lei, R. Yao, Y. Gong and Q. Yin, [Image segmentation based on adaptive K-means algorithm](#), *EURASIP J. Image Video Processing*, **68** (2018).

Received July 2019; revised March 2020.

E-mail address: herty@igpm.rwth-aachen.de

E-mail address: lorenzo.pareschi@unife.it

E-mail address: visconti@igpm.rwth-aachen.de