Learning Interaction Dynamics from Particle Trajectories and Density Evolution

Christos N. Mavridis, Amoolya Tirumalai and John S. Baras

***Abstract*— We propose a family of parametric interaction functions in the general Cucker-Smale model such that the mean-field macroscopic system of equations can be iteratively solved in an optimization scheme aiming to learn the interaction dynamics of the microscopic model from observations of macroscopic quantities. We treat the interaction functions as Green's functions for a semi-linear Poisson differential operator, which allows the transformation of the non-local interaction terms of the macroscopic model into a system of PDEs. The resulting system of hydrodynamic equations is efficiently solved as part of an iterative learning algorithm that estimates the interaction function from particle density evolution data. Finally, we utilize the proposed interaction function model to formulate an efficient learning algorithm based on observations from particle trajectories, and discuss the trade-offs associated with each approach.**

I. INTRODUCTION

Extracting the laws of interaction between agents of networked systems, ranging from power systems and chemical reaction networks, to animal flocks and UAV swarms, is a fundamental challenge in many areas of science and engineering [1], [2], [3], [4], [5], [6].

Statistical [7], [8], and, mainly, model-based [2], [3], [9] learning approaches have been used to infer interaction rules between particles. In [10] symbolic equations are generated from the numerically calculated derivatives of the system variables, in [11] the constitutive equations of physical components composing the system are learned, while in [12] the order of a fractional differential system of equations, which models the system, is estimated. Recently, Matei et al. in [9] have modeled the network as a port-Hamiltonian system [13] consisting of interconnected generalized mass-spring-damper systems, and reconstructed the laws of interaction and its dynamical properties.

There are generally two broad approaches in modeling the underlying dynamics of ensembles of self-organizing agents: the microscopic particle models described by ordinary or stochastic differential equations, and the macroscopic continuum models, described by partial differential equations. Agent-based models assume behavioral rules at the individual level, such as velocity alignment, attraction, and repulsion [2], [3], [4], [5], while macroscopic models, consider large number of interacting agents, approaching the mean-field limit, and consist of a system of compressible hydrodynamic PDEs [14], [15], defined on macroscopic quantities.

While particle models are widely used in numerical simulations and learning procedures, [9], [12], useful real-life data of particle trajectories is difficult to extract [5], [12], and may require substantial memory and computation resources. The experimental measurements usually involve video recordings and are subject to noise mainly because particle trajectories seem to overlap. Additional distortions are introduced from the computer vision methods used to reconstruct three-dimensional trajectories, and the estimation of velocities from position observations [5], [12], [16], [17]. On the other hand, useful approximations of the ensemble's density evolution can be easier to extract, often by applying simple morphological operators on video recordings. For this reason, developing learning algorithms based on the macroscopic quantities can play a crucial role in the analysis of collective motion, but remains inhibited due to computational expense, since the flocking dynamics can be non-local as well as nonlinear [15], which results in a costly computation of the solution of the corresponding hydrodynamic equations [12].

In this work, we modify the classical Cucker-Smale model of nonlocal particle interaction for velocity consensus [3], [18] so that we can efficiently solve the macroscopic hydrodynamic equations in an iterative optimization scheme to reconstruct the interaction function from density evolution data. We propose a family of parametric interaction functions that are treated as Green's functions associated with an appropriately defined semi-linear Poisson partial differential operator, which allows for the transformation of the macroscopic hydrodynamic integro-differential equations into an augmented system of PDEs and speeds-up the computation of the non-local interaction terms. We investigate the conditions under which time-asymptotic flocking is achieved, and utilize the computational advantages of the proposed methodology to construct a learning algorithm to estimate the interaction function based on observations of the particle density evolution. Finally, an efficient learning algorithm based on the particle trajectories is also studied. Making use of the proposed interaction function model significantly reduces the number of parameters to be estimated, and provides an approximation of the interaction function that is meaningful over its entire domain.

The rest of the manuscript is organized as follows: Section II defines the Cucker-Smale flocking dynamics and its mean-field limit. Section III introduces the semi-linear Poisson mediated interaction functions and the conversion of the macroscopic equations to a system of PDEs. In Sections IV and IV the learning problems based on observations of density evolution and particle trajectories, respectively, are formulated. Finally, Section VI presents the numerical results and Section VII concludes the paper.

Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742, USA. Emails: {mavridis,ast256,baras}@umd.edu

II. Mathematical Models

In this section we introduce the Cucker-Smale particle dynamics, define time-asymptotic flocking, and derive the mean-field macroscopic equations.

A. The Cucker-Smale Model

Consider an interacting system $\mathcal{G}$ of N identical particles (representing autonomous agents) with unit mass in $\mathbb{R}^d$, $d \in \{1,2,3\}$. Let $x_i(t),\ v_i(t) \in \mathbb{R}^d$ represent the position and velocity of the i^{th}-particle at each time $t \geq 0$, respectively, for $1 \leq i \leq N$. Then the general Cucker-Smale dynamical system [3] of $(2N)$ ODEs read as:

$$\begin{cases} \frac{dx_i}{dt} &= v_i \\ \frac{dv_i}{dt} &= \frac{1}{N}\sum_{j=1}^{N} \psi(x_j,x_i)(v_j - v_i) \end{cases} \tag{1}$$

where $x_i(0)$, are $v_i(0)$ are given for all $i = 1,\ldots,N$, and $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ represents the interaction function between each pair of particles.

The center of mass system (x_c, v_c) of $\mathcal{G} = \{(x_i,v_i)\}_{i=1}^N$ is defined as

$$x_c = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad v_c = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{2}$$

We are interested in ψ being symmetric, i.e., $\psi(x,s) = \psi(s,x)$, in which case system (1) implies

$$\frac{dx_c}{dt} = v_c, \quad \frac{dv_c}{dt} = 0 \tag{3}$$

which yields the unique solution

$$x_c(t) = x_c(0) + t v_c(0),\ t \geq 0 \tag{4}$$

Under additional assumptions on ψ (Section **??**), system (1) converges to a velocity consensus under spatial coherence, known as time-asymptotic flocking, and defined as:

Definition 1 (Asymptotic Flocking)**.** *An $N-$body interacting system $\mathcal{G} = \{(x_i,v_i)\}_{i=1}^N$ exhibits time-asymptotic flocking with bounded fluctuation if and only if the following two relations hold:*

- *(Velocity alignment): The velocity fluctuations approach zero asymptotically, i.e.*

$$\lim_{t\to\infty} \sum_{i=1}^{N} \|v_i(t) - v_c(t)\|^2 = 0$$

- *(Spatial coherence): The position fluctuations are uniformly bounded, i.e. for some $0 < \Lambda < \infty$,*

$$\sup_{0\leq t\leq\infty} \|x_i(t) - x_c(t)\| < \Lambda,\ \forall i \in \{1,\ldots,N\}$$

Throughout this paper, we will be working with the fluctuation variables around the center of mass system

$$(\hat{x}_i, \hat{v}_i) := (x_i - x_c, v_i - v_c) \tag{5}$$

which can be shown to satisfy the same dynamics (1). Following the spatial coherence of the flocking behavior, the position variables $\hat{x}_i$ will be defined in a compact support $D := \left\{x \in \mathbb{R}^d : \|x\| < L/2\right\}$ for some finite $L > 0$ and for all $i \in \{1,\ldots,N\}$, with $\|\cdot\|$ representing the standard l_2-norm in $\mathbb{R}^d$ and $d \in \{1,2,3\}$. We note that transformation (5) only requires the knowledge of the initial conditions $x_i(0)$ and $v_i(0)$, $i = 1,\ldots,N$.

B. The Mean-Field Limit

Consider the joint probability triple of the entire particle system $\{\Omega := \mathbb{R}^{2Nd}, \mathcal{B}(\Omega), P_{xv}\}$, the state space for each particle $\{\mathbb{R}^{2d}, \mathcal{B}(\mathbb{R}^{2d})\}$ and define the empirical (random) probability measure $F_{xv}^N : \Omega \times [t_0,t_f] \times \mathcal{B}(\mathbb{R}^{2d}) \to [0,1]$ such that

$$F_{xv}^N(t,A) := \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}_A((x_i(t), v_i(t))) \tag{6}$$

where $\mathbb{I}_A(\cdot)$ is the indicator function, $A \in \mathcal{B}(\mathbb{R}^{2d})$. Other authors use Dirac measures (not the Dirac delta function) in this definition. F_{xv}^N is a random measure which is purely atomic. Using arguments originally due to McKean and Vlasov [19], [20], it can be shown that there exists a deterministic and continuous F_{xv}^* such that $F_{xv}^N \overset{a.e.}{\to} F_{xv}^*$ in the weak-* sense, and that the joint probability density $f_{xv}^* : [t_0,t_f] \times \mathbb{R}^{2d} \to \mathbb{R}_+$ associated with this measure, evolves according to the forward Kolmogorov equation on $[t_0,t_f] \times \mathbb{R}^{2d}$:

$$\begin{aligned} &\partial_t f_{xv}^* + \nabla_x \cdot (v f_{xv}^*) + \nabla_v \cdot (\mathcal{F} f_{xv}^*) = 0 \\ &\mathcal{F}(t,x,v) := \int_{\mathbb{R}^{2d}} \psi(x,s)(w-v) f_{xv}^*(t,s,w) ds dw. \end{aligned} \tag{7}$$

We define the marginal probability density $\rho : [t_0,t_f] \times D \to \mathbb{R}_+$ (henceforth referred to only as density)

$$\rho(t,x) := \int_{\mathbb{R}^d} f_{xv}^*(t,x,v) dv \tag{8}$$

and the momentum density $m : [t_0,t_f] \times D \to \mathbb{R}^d$ and bulk velocity $u : [t_0,t_f] \times D \to \mathbb{R}^d$

$$m(t,x) := \int_{\mathbb{R}^d} v f_{xv}^*(t,x,v) dv := \rho(t,x) u(t,x) \tag{9}$$

where $D \subseteq \mathbb{R}^d$. It is additionally assumed that ρ, m, u are compactly supported. Substituting in (7), we obtain the $(d+1)$ compressible Euler equations (see also [21]) on $[t_0,t_f] \times D$:

$$\begin{cases} \partial_t \rho + \nabla_x \cdot m = 0 \\ \partial_t m + \nabla_x \cdot (m \otimes u) = \rho \mathcal{L}_\psi m - m \mathcal{L}_\psi \rho \end{cases} \tag{10}$$

where

$$\mathcal{L}_\psi \phi(t,x) = \int_D \psi(x,s)\phi(t,s) ds. \tag{11}$$

is an integral transform with kernel $\psi : D \times D \to \mathbb{R}$.

III. Semi-linear Poisson Mediated Flocking

Suppose that the kernel function ψ is a Green's function associated with some linear partial differential operator $\mathcal{L}_x : C^\infty_{\mathcal{C},\mathbb{R}}(D) \to C^\infty_{\mathcal{C},\mathbb{R}}(D)$, such that

$$\mathcal{L}_x y(t,x) = \phi(t,x) \tag{12}$$

implies

$$y(t,x) = \int_D \psi(x,s)\phi(t,s) ds \tag{13}$$

A classical example is the operator associated with the Poisson equation that arises in self-gravitational hydrodynamics [22]. Then, system (10) is equivalent to the augmented system of $(2d+2)$ partial differential equations:

$$\begin{cases} \partial_t \rho + \nabla_x \cdot m = 0 \\ \mathcal{L}_x z = \rho \\ \mathcal{L}_x y = m \\ \partial_t m + \nabla_x \cdot (m \otimes m\rho^{-1}) = \rho y - zm \end{cases} \tag{14}$$

Moreover system (14) defined on $[t_0, t_f] \times D$, with homogeneous Dirichlet boundary conditions, and initial conditions

$$\rho(0,\hat{x}) = \rho_0(\hat{x}),\ v(0,\hat{x}) = v_0(\hat{x}) \tag{15}$$

is a well-posed Boundary Value Problem (BVP).

We adopt the semi-linear Poisson-mediated flocking model of [23], and make use of the parametrized partial differential operator

$$\mathcal{L}_x := -\frac{1}{2k}(\partial_x^2 - \lambda^2) \tag{16}$$

in the BVP (14), (15) with $D := \left\{x \in \mathbb{R}^d : \|x\| < L/2\right\}$.

In the following, we will illustrate our analysis in the one-dimensional case ($d = 1$) which can be generalized to higher dimensions (Section VI-C). System (14) can now be compactly written as

$$\begin{aligned} &\partial_t U + \partial_x F(U) = S(Y,U) \\ &\mathcal{L}_x Y = U \end{aligned} \tag{17}$$

where $U := [\rho, m]^T$, $F := [m, m^2\rho^{-1}]^T$, $S := [0, \rho y - zm]^T$, and $Y := [z, y]^T$. The Green's function ψ associated with the BVP in (17) takes the form

$$\psi(x,s) = \begin{cases} K\alpha(s)\beta(x) & s \le x \\ K\alpha(x)\beta(s) & s > x \end{cases} \tag{18}$$

where

$$\begin{aligned} K &= -\frac{k}{\lambda}\frac{1}{e^{\lambda L} - e^{-\lambda L}} \\ \alpha(z) &= 2\cosh(\lambda(z + L/2)) \\ \beta(z) &= 2\cosh(\lambda(z - L/2)) \end{aligned} \tag{19}$$

We note that the interaction function ψ in (18), and, as a result, the flocking behavior of the system $\mathcal{G}$, depends on the bounded domain D in which it is defined as illustrated in [23]. In addition, it can be shown (see [23], Section 4.1) that under mild conditions on the initial conditions $v_i(0)$, $i = 1, \ldots, N$, and the size of the domain L, the solution $\{(x_i(t), v_i(t))\}_{i=1}^N$, $t \ge 0$, of system (1) satisfies the flocking conditions of Definition 1.

IV. Learning the Interaction Function from Density Evolution

We utilize the proposed interaction function form (18), (19), and the resulting system of PDEs (17), described in Section III, to efficiently compute the macroscopic quantities, i.e. the momentum and density, in a density-based learning scheme, aiming to estimate the interaction function ψ.

A. Computational Methods

We start with the following proposition on the conservation of mas and momentum which is proved in [23]:

Proposition 1. *If $Y \in C^\infty_{\mathbb{R},\mathcal{C}}(D)$, then mass and momentum are conserved, i.e.*

$$\frac{d}{dt}\int_D U dx = \int_D S dx = 0. \tag{20}$$

In order to solve the hyperbolic system and the elliptic equations of (17), we use the following solvers.

1) Hyperbolic Solver: To solve the hyperbolic system of (17), we apply the finite volume method [24]. We define the sequence of points $x_s = \{x_0, ..., x_i, ..., x_N\}$ which are the centers of the cells $I_i := [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$, and average the PDE over these cells, which gives

$$\frac{1}{\lambda(I_i)}\frac{d}{dt}\int_{I_i} U dx = -\frac{1}{\lambda(I_i)}\int_{I_i}\partial_x F dx + \frac{1}{\lambda(I_i)}\int_{I_i} S dx \tag{21}$$

where $\lambda(\cdot)$ is the Lebesgue measure. Assuming these are identical, such that $\Delta x := \lambda(I_i)\forall i$, we can make use of the divergence theorem, and replace the integrals of U, F, S with their cell-averages, i.e. their midpoint values $\bar{U}, \bar{F}, \bar{S}$, in order to obtain

$$\frac{d}{dt}\bar{U}_i = -\frac{1}{\Delta x}(\bar{F}_{i+\frac{1}{2}} - \bar{F}_{i-\frac{1}{2}}) + \bar{S}_i \tag{22}$$

where $\bar{U}_i := \bar{U}(x_i), \bar{F}_i := \bar{F}(x_i), \bar{S} := \bar{S}(x_i)$. We employ the second-order strong stability preserving Runge-Kutta scheme [25] for time integration. For the fluxes, we assume piecewise linearity and use the Kurganov-Tadmor flux [25] given by

$$\begin{aligned} \bar{F}_{i+\frac{1}{2}} &:= \frac{1}{2}[F_i^* + F_{i+1}^* - \max\{|u_i^*|, |u_{i+1}^*|\}(U_{i+1}^* - U_i^*)] \\ U_{i+1}^* &:= U_{i+1} - \frac{\Delta x}{2} minmod(\frac{U_{i+2} - U_{i+1}}{\Delta x}, \frac{U_{i+1} - U_i}{\Delta x}) \\ U_i^* &:= U_i + \frac{\Delta x}{2} minmod(\frac{U_{i+1} - U_i}{\Delta x}, \frac{U_i - U_{i-1}}{\Delta x}) \end{aligned} \tag{23}$$

where $minmod(a,b) := \frac{1}{2}(sign(a) + sign(b))\min(|a|, |b|)$.

2) Elliptic Solver: To solve the elliptic equations of (17), we apply the classical second-order finite difference method, which reads as

$$\frac{y_{i+1}^j - 2y_i^j + y_{i-1}^j}{\Delta x^2} - \lambda^2 y_i^j = -2kU_i^j \tag{24}$$

Over the interior points, this yields linear equations

$$(\frac{1}{\Delta x^2}\mathbf{A} - \lambda^2\mathbf{I})y_{int}^j = -2kU_{int}^j - \frac{1}{\Delta x^2}\begin{bmatrix} y_0^j & 0 & \ldots & 0 & y_N^j \end{bmatrix}^T \tag{25}$$

where $\mathbf{A}$ is the $(1, -2, 1)$ tridiagonal matrix.

Therefore banded matrix algorithms [26] can be used to solve the corresponding system of equations quickly, as shown in Fig. 1.

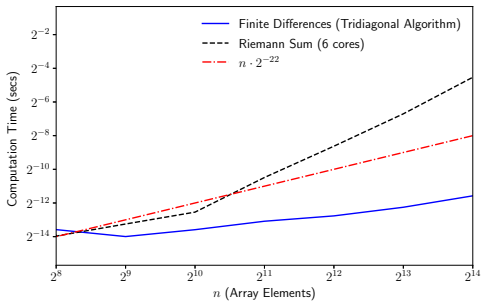

Fig. 1. Computation Times for Nonlocal Terms. Using finite differences is much faster than computing a convolution (Riemann) sum, even when parallelism of the sum is exploited.

B. The learning algorithm

We formulate the process of learning the interaction function ψ from density data as an optimization problem:

$$\min_{k,\lambda} \sum_{\tau=t_0}^{t_f} D_{KL}(F_x^N[\tau,x_s]||F_x^*[\tau,x_s]) \tag{26}$$

where $\{F_x^*[\tau,x_s]\}_{\tau=t_0}^{t_f}$ and $\{F_x^N[\tau,x_s]\}_{\tau=t_0}^{t_f}$ are the discretized time sequences of the empirical and mean-field probability measures, respectively, evaluated at the sequence of points x_s defined in Section IV-A.1. The density associated with F_x^*, the mean-field density ρ, is subject to (17) (and therefore dependent on $\theta := (k,\lambda)$) and the Kullback-Leibler (KL) divergence D_{KL} between two distinct probability measures F_i, F_j, both defined in $\{\Omega, \mathcal{B}(\Omega)\}$, with associated densities ρ_i, ρ_j is given by

$$D_{KL}(F_i||F_j) := \int_\Omega \log_2 \frac{dF_i}{dF_j} dF_i = \int_\Omega \rho_i \log_2 \frac{\rho_i}{\rho_j} dx \tag{27}$$

We approach the solution $\theta^* := (k^*,\lambda^*)$ of (26) with respect to $V_d(\theta) := \sum_{\tau=t_0}^{t_f} D_{KL}(F_x^N[\tau,x_s]||F_x^*[\tau,x_s])$, with the iterative relation

$$\theta^{n+1} = -\hat{\mathbf{H}}^{-1}(\theta^n)\nabla_\theta V_d(\theta^n) \tag{28}$$

where $\hat{\mathbf{H}}$ is a positive-definite approximation of the Hessian computed via the Lanczos iteration [27].

We note that in each iteration of the learning algorithm, the solution of the BVP associated with the system of PDEs (17) must be numerically computed, which has become feasible due to the computational advantages originating from the use of the proposed linear operator $\mathcal{L}_x$ (16), and presented in Section IV-A and Fig.1.

V. Learning the Interaction Function from Particle Trajectories

The proposed interaction function form (18), (19), described in Section III, can also be used to efficiently estimate the interaction function ψ in a particle-based learning scheme given observations of the particle trajectories.

The advantage of using the interaction function model (18), (19), in learning the interaction dynamics of a swarm from particle trajectories is twofold. First, the number of parameters to be estimated becomes really small, especially compared to a general regression function such as a neural network [9], which reduces the amount of data required for convergence. Secondly, every update in the optimization algorithm will improve the estimate of the interaction function over the entire domain D, and not only over a small subset $D_o \subset D$ where the distances between each pair of interacting particles happen to be observed, as opposed to using general regression models such as polynomial functions and neural networks.

As shown in [9], the Cucker-Smale model (1) is equivalent to a fully connected N-dimensional network of generalized mass-spring-dampers with appropriately defined Hamiltonian functions, that can be written in an input-state-output port-Hamiltonian form [13]:

$$\dot{z} = [J(z) - R(z)]\frac{\partial H(z)}{\partial z} \tag{29}$$

where $z = (q,p)$, with $q,p \in \mathbb{R}^{\frac{N(N-1)}{2}}$ being the vectors of relative distances and momenta between each pair of particles, and the quantities $J = -J^T$, H and R are appropriately defined. The dependence of (29) on the interaction function ψ is introduced by the resistive term $R = R(\psi)$ [9].

We formulate the learning process as a least-squares optimization problem

$$\min_{k,\lambda} \sum_{\tau=t_0}^{t_f} \|\dot{z}^*(\tau) - \dot{z}(\tau)\|^2 \tag{30}$$

where z^* represent the observed trajectories, and z are subject to (29).

We approach the solution $\theta^* := (k^*,\lambda^*)$ of (30) with respect to $V_p(\theta) := \sum_{\tau=t_0}^{t_f} \|\dot{z}^*(\tau) - \dot{z}(\tau)\|^2$, with an iterative gradient descent method

$$\theta^{n+1} = \theta^n - \alpha_n(\nabla_\theta V_p(\theta^n)),\ n = 0,1,2,\dots \tag{31}$$

where the iteration maps $\alpha_n : \mathbb{R}^2 \to \mathbb{R}^2$, $n \geq 0$ are defined in accordance with the Adam method of moments for stochastic optimization [28], and the computation of the gradient vectors is implemented using automatic differentiation [29].

VI. Numerical Results and Higher Dimensions

We illustrate our results in the domain $D = [-\pi,\pi]$ ($L = 2\pi$), with initial density and bulk velocity given by

$$\rho_0(\hat{x}) = \frac{\pi}{2L}\cos(\frac{\pi\hat{x}}{L}), \tag{32}$$

$$u_0(\hat{x}) = -c\sin(\frac{\pi\hat{x}}{L}),\ \hat{x} \in D,\ c > 0 \tag{33}$$

i.e. assuming that $\rho_0(\hat{x}) = u_0(\hat{x}) = 0,\ \forall \hat{x} \notin D$, where $\hat{x}$ is as defined in (5).

The system of particle equations (1) is numerically solved using the velocity Verlet algorithm [12]. In all simulations, we take $\lambda = 1$, $k = 4$. The agreement between the solutions of the particle model (1) and the macro-scale model (14) for $N = 10^4$, $\Delta\hat{x} = \frac{2\pi}{600}$, and $\Delta t = .001$ is shown in Fig. 2.

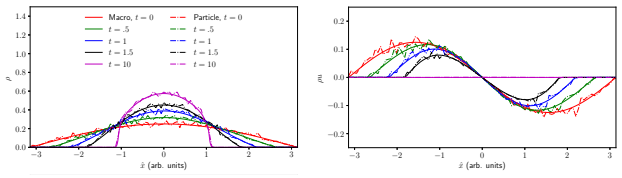

Fig. 2. Evolution of the densities $\rho(t,\hat{x})$ and momentum profiles $m(t,\hat{x})$ as computed by solving the macro-scale model and the particle model (dashed-line).

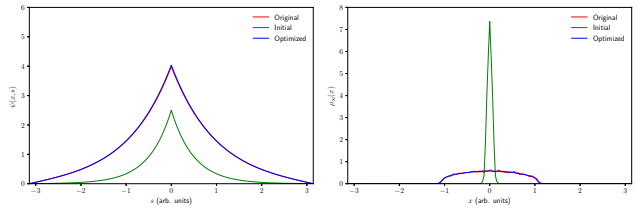

Fig. 3. Reconstruction of the interaction function ψ using observations of density evolution data. (*up*) Estimated interaction functions $\psi(\hat{x}=0,\hat{s})$. (*down*) Estimated equilibrium densities $\rho_\infty(\hat{x})$

A. Learning from density data

In order to accurately evaluate the learning scheme defined in Section IV, we obtain the empirical density evolution data ρ^N by first simulating the particle equations (1) with initial conditions randomly generated from the initial density and bulk velocity (33), and then taking the piecewise-constant density discretization

$$\rho^N[t_i,x_s] := \frac{1}{N\lambda(I_j)}\mu(\{x_k[t_i]\in I_j\}) \tag{34}$$

where $\lambda(\cdot)$ is the Lebesgue measure, $\mu(\cdot)$ is the counting measure, and I_i, x_j are defined as in the formulation of the finite volume method (Section IV-A).

The results are illustrated in Fig. 3. The parameters $(\hat{k}^*,\hat{\lambda}^*) = (4.0782, 1.0069) \sim (4,1)$ of the interaction function ψ were recovered and the Newton's iteration converged in 23 iterations.

We note that problem (26) is generally a non-convex optimization problem, and may be sensitive to initial estimates of the parameters (k,λ) leading to sub-optimal solutions $(\hat{k}^*,\hat{\lambda}^*) \neq (k^*,\lambda^*)$. In addition, the objective function $V_d := \sum_{\tau=t_0}^{t_f} D_{KL}(F_x^N[\tau,x_s]||F_x^*[\tau,x_s])$ may practically converge to zero event though $(\hat{k}^*,\hat{\lambda}^*) \neq (k^*,\lambda^*)$, suggesting non-unique solutions of the learning problem (26) for some observation data. This is a standard problem in non-convex iterative optimization algorithms and is usually addressed with observation data of better quality (usually of larger quantity as well), or repetitive application of the optimization algorithm for randomly chosen initial conditions [30]. However, the reconstructed $\hat{\psi}^* := \psi(\hat{k}^*,\hat{\lambda}^*)$ may be used in the microscopic model (1) to accurately reconstruct the actual observed trajectories.

B. Learning from Particle Trajectories

In order to be able to assess the learning performance we obtain the observed data of particle trajectories z^* by simulating the particle equations (1). The results are illustrated in Fig. 4 where trajectory data of 100 particles have been used in a time window of 0.2sec, corresponding to 20 time steps. In the upper figure, the trajectories were generated using the proposed interaction function (18), (19) for $(k^*,\lambda^*) = (4.0, 1.0)$ and the reconstructed parameters were $(k^*,\lambda^*) = (4.0781, 1.0004)$. In the lower figure, the trajectories were generated using the Cucker-Smale interaction function

$$\psi_{cs}(x,s) = \frac{K}{(1+\|x-s\|^2)^\gamma} \tag{35}$$

for $(K,\gamma) = (1.0, 3.0)$, and the algorithm converged with a training loss $V_p := \sum_{\tau=t_0}^{t_f} \|\dot{z}^*(\tau) - \dot{z}(\tau)\|^2 \leq 10^{-5}$.

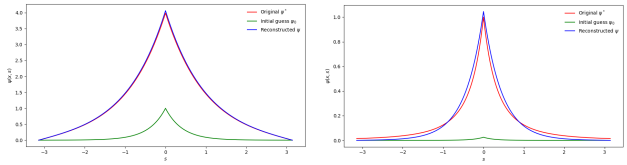

Fig. 4. Reconstruction of the interaction function ψ using observed trajectories of interacting particles for 0.2 seconds. Illustration for $x=0$. (*up*) Particle trajectories simulated using the proposed interaction function. (*down*) Particle trajectories simulated using the original Cucker-Smale interaction function.

C. Higher Dimensions

The proposed methodology can be generalized in higher dimensions, with the radial symmetry of the interaction function ψ suggesting the use of singular kernels, which have been extensively studied in the literature and, under mild assumptions in the initial conditions, have been shown to result in flocking behavior while, at the same time, avoiding collisions [31].

In the BVP of the augmented system of PDEs (14), with $D := \{x\in\mathbb{R}^d : \|x\| < L/2\}$, we select the linear differential operator

$$\mathcal{L}_x = -k^{-d/2}(\nabla_x^2 - \lambda^2) \tag{36}$$

which is associated with a Green's function of the form

$$\psi(x,s) = \hat{\psi}(x-s) + \phi(x,s) \tag{37}$$

where $\hat{\psi}$ is given by

$$\begin{aligned}\hat{\psi}(x,s) &= \tilde{\psi}(\|x-s\|)\\ &= \left(\frac{k}{2\pi}\right)^{d/2}\left(\frac{\lambda}{\|x-s\|}\right)^{d/2-1} K_{d/2-1}(\lambda\|x-s\|)\end{aligned} \tag{38}$$

with $K_\alpha(\cdot)$ being the modified Bessel function of the second kind of order α, and, it can be shown that ([23]),

$$\phi(x,s) = -\tilde{\psi}(\frac{2}{L}\|x\|\|s - \frac{L^2}{4}\frac{x}{\|x\|^2}\|). \tag{39}$$

The interaction function ψ depends on the parameter values k and λ as illustrated, for the 2-dimensional case, in Fig.5.

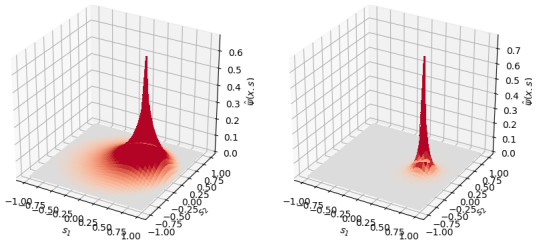

Fig. 5. The effect of the parameters k, λ on the profile of the interaction function $\psi((0,0.5),s)$, $s \in B_2(0,1)$. Left: $(k,\lambda) = (1,0.5)$. Right: $(k,\lambda) = (2,10)$.

VII. Conclusion

Particle-based learning requires the solution of a large system of ODEs, which can be computed using fast existing numerical algorithms when the number of particles in not too large. At the same time, actual data of particle trajectories with high signal-to-noise ratio can hardly be found. On the other hand, density-based learning assumes large number of interacting particles and is based on density data, which can be easier to extract from experimental recordings. Yet, it requires the solution of a system of PDEs, which, although much smaller, can be significantly difficult to compute.

In this work, we studied both a particle-based and a density-based learning algorithm. First we proposed an iterative optimization algorithm for learning the interaction dynamics of a general Cucker-Smale particle interaction model from observations of the particle density evolution. In doing so, we introduced a family of compactly supported parametric interaction functions that allow for the mean-field macroscopic system of hydrodynamic equations to be efficiently solved as part of the learning scheme. Finally, we made use of the proposed interaction function model to formulate an efficient particle-based learning algorithm that uses observations from particle trajectories.

VIII. Acknowledgements

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Agreement No. HR00111990027 and partially by ONR grant N00014-17-1-2622.

References

[1] A. Okubo, "Dynamical aspects of animal grouping: swarms, schools, flocks, and herds," *Advances in biophysics*, vol. 22, pp. 1–94, 1986.

[2] C. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in *ACM SIGGRAPH computer graphics*, vol. 21, no. 4. ACM, 1987, pp. 25–34.

[3] F. Cucker and S. Smale, "Emergent behavior in flocks," *IEEE Transactions on automatic control*, vol. 52, no. 5, pp. 852–862, 2007.

[4] I. Giardina, "Collective behavior in animal groups: theoretical models and empirical studies," *HFSP journal*, vol. 2, no. 4, pp. 205–219, 2008.

[5] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, *et al.*, "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study," *Proceedings of the national academy of sciences*, vol. 105, no. 4, pp. 1232–1237, 2008.

[6] I. L. Bajec and F. H. Heppner, "Organized flight in birds," *Animal Behaviour*, vol. 78, no. 4, pp. 777–789, 2009.

[7] F. Lu, M. Zhong, S. Tang, and M. Maggioni, "Nonparametric inference of interaction laws in systems of agents from trajectory data," *arXiv preprint arXiv:1812.06003*, 2018.

[8] S. Brunton, J. Proctor, and J. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, 2016.

[9] I. Matei, C. Mavridis, J. S. Baras, and M. Zhenirovskyy, "Inferring particle interaction physical models and their dynamical properties," in *2019 IEEE Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 4615–4621.

[10] J. Bongard and H. Lipson, "Automated reverse engineering of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 104, no. 24, pp. 9943–9948, 2007.

[11] I. Matei, J. de Kleer, and R. Minhas, "Learning constitutive equations of physical components with constraints discovery," in *2018 Annual American Control Conference (ACC)*, June 2018, pp. 4819–4824.

[12] Z. Mao, Z. Li, and G. Karniadakis, "Nonlocal flocking dynamics: Learning the fractional order of pdes from particle simulations," *arXiv preprint arXiv:1810.11596*, 2018.

[13] A. van der Schaft and D. Jeltsema, "Port-hamiltonian systems theory: An introductory overview," *Foundations and Trends® in Systems and Control*, vol. 1, no. 2-3, pp. 173–378, 2014. [Online]. Available: http://dx.doi.org/10.1561/2600000002

[14] 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*. Springer, 2010, pp. 297–336.

[15] R. Shvydkoy and E. Tadmor, "Eulerian dynamics with a commutator forcing ii: Flocking," *arXiv preprint arXiv:1701.07710*, 2017.

[16] D. Chen, T. Vicsek, X. Liu, T. Zhou, and H.-T. Zhang, "Switching hierarchical leadership mechanism in homing flight of pigeon flocks," *EPL (Europhysics Letters)*, vol. 114, no. 6, p. 60008, 2016.

[17] M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, "Hierarchical group dynamics in pigeon flocks," *Nature*, vol. 464, pp. 890–893, 2010.

[18] S.-Y. Ha, J.-G. Liu, *et al.*, "A simple proof of the cucker-smale flocking dynamics and mean-field limit," *Communications in Mathematical Sciences*, vol. 7, no. 2, pp. 297–325, 2009.

[19] C. Lancellotti, "On the vlasov limit for systems of nonlinearly coupled oscillators without noise," *Transport theory and statistical physics*, vol. 34, no. 7, pp. 523–535, 2005.

[20] F. Golse, "The mean-field limit for the dynamics of large particle systems," *Journées équations aux dérivées partielles*, pp. 1–47, 2003.

[21] J. Carrillo, M. Fornasier, G. Toscani, and F. Vecil, *Particle, kinetic, and hydrodynamic models of swarming*, G. Naldi, L. Pareschi, and G. Toscani, Eds. Boston: Birkhäuser Boston, 2010.

[22] J. K. Truelove, R. I. Klein, C. F. McKee, J. H. H. II, L. H. Howell, J. A. Greenough, and D. T. Woods, "Self-gravitational hydrodynamics with three-dimensional adaptive mesh refinement: Methodology and applications to molecular cloud collapse and fragmentation," *The Astrophysical Journal*, vol. 495, no. 2, pp. 821–852, mar 1998.

[23] C. N. Mavridis, A. Tirumalai, J. S. Baras, and I. Matei, "Semilinear poisson-mediated flocking in a cucker-smale model," in *24th International Symposium on Mathematical Theory of Networks and Systems (MTNS)*. IFAC, 2020.

[24] R. J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, ser. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2002.

[25] A. Kurganov and E. Tadmor, "New high-resolution central schemes for nonlinear conservation laws and convection–diffusion equations," *Journal of Computational Physics*, vol. 160, no. 1, pp. 241 – 282, 2000.

[26] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. The Johns Hopkins University Press, 2013.

[27] S. G. Nash, "Newton-type minimization via the lanczos method," *SIAM Journal on Numerical Analysis*, vol. 21, no. 4, pp. 770–788, 1984. [Online]. Available: https://doi.org/10.1137/0721052

[28] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[29] D. Maclaurin, D. Duvenaud, M. Johnson, and J. Townsend, "Autograd," https://github.com/HIPS/autograd, 2018.

[30] C. T. Kelley, *Iterative methods for optimization*. SIAM, 1999.

[31] S. Ahn, H. Choi, S.-Y. Ha, and H. Lee, "On collision-avoiding initial configurations to cucker-smale type flocking models," *Communications in Mathematical Sciences*, vol. 10, 06 2012.