Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-30 (2022) 103–108

Supervised learning for kinetic consensus control

Giacomo Albi* Sara Bicego** Dante Kalise**

* Department of Computer Science, University of Verona, Strada le Grazie 15 - 37134 Verona, Italy (email: giacomo.albi@univr.it).
** Department of Mathematics, Imperial College London, South Kensington Campus - SW72AZ London, UK (email: s.bicego21,dkaliseb@imperial.ac.uk)

Abstract: In this paper, how to successfully and efficiently condition a target population of agents towards consensus is discussed. To overcome the curse of dimensionality, the mean field formulation of the consensus control problem is considered. Although such formulation is designed to be independent of the number of agents, it is feasible to solve only for moderate intrinsic dimensions of the agents space. For this reason, the solution is approached by means of a Boltzmann procedure, i.e. quasi-invariant limit of controlled binary interactions as approximation of the mean field PDE. The need for an efficient solver for the binary interaction control problem motivates the use of a supervised learning approach to encode a binary feedback map to be sampled at a very high rate. A gradient augmented feedforward neural network for the Value function of the binary control problem is considered and compared with direct approximation of the feedback law.

Keywords: Multi-agent systems, optimal feedback control, mean field models, supervised learning, opinion dynamics.

1. INTRODUCTION

Social behaviours can be seen as the result of a suitable combination of endogenous population interactions and external influences. How to successfully condition a population of agents towards a designed purpose is a fascinating question, whose answer is being widely researched (Li and Tan, 2019).

The formulation of such a problem in a dynamic optimization framework ensures the availability of control synthesis methods, which nonetheless come with the huge drawback of the curse of dimensionality. The problem reads as the minimization of a cost functional subject to individual-based interaction dynamics, thus its solution easily becomes unfeasible to compute as the number of agents in the population grows. The natural way of circumventing this is using a multiscale approach working with the population density instead of its microscopic state. For a number of $N \to \infty$ interacting agents, this leads to a mean field formulation of the control problem. Although mean field optimal control problems are designed to be independent of the number of agents, they are computationally feasible only for moderate intrinsic dimensions d of the agents' state space. For this reason, we rely on the approximation of the mean field PDE governing the evolution of probability distribution characterizing the agents' population. An approximate solution is obtained as a result of a Boltzmann type dynamics (Albi et al., 2017b; Albi and Pareschi, 2013; Albi et al., 2017a). This procedure provides the approximated mean field solution as limit of a reduced problem, modeling the interactions taking place only within controlled couples of agents. We refer the reader to (Albi and Kalise, 2018) for a generalization of this procedure when allowing only a subset of agents in the system to be influenced by an external control signal, and to (Albi et al., 2015) for a system-generalized control action.

The efficiency of this Boltzmann approach is linked to the availability of a sufficiently fast solver for a binary interaction control problem, that is, an optimal control problem for a reduced system of two agents, which is sampled at a very high-frequency rate. We address this computational requirement by enconding a binary feedback map by means of a supervised learning procedure (Kang et al., 2021; Darbon and Osher, 2016; Azmi et al., 2021; Albi et al., 2022), which is trained upon synthetic data from sampling a feedback law generated by a state-dependent Riccati equation approach (SDRE) (Cloutier, 1997; Banks et al., 2007; Jones and Astolfi, 2020). This feedback law corresponds to an approximation of the associated optimal feedback law characterized by the solution of a Hamilton-Jacobi-Bellman PDE. Despite being suboptimal, the SDRE law locally asymptotically stabilizes the dynamics, and can be easily computed by the sequential solution of algebraic Riccati equations, providing a reasonable alternative in high-dimensional settings where the numerical approximation of optimal feedback laws is prohibitively expensive.

The rest of the paper is organized as follows. In Section 2 we introduce the mean field formulation of the addressed consensus problem, and in Section 3 we present a consistent alternative description of Boltzmann type. In Section 4 the state-dependent Riccati equation approach is presented, and its numerical approximation through

10.1016/j.ifacol.2022.11.036

supervised learning is discussed in Section 5. A computational assessment dealing with control of first order opinion dynamics can be found in Section 6.

2. MEAN FIELD CONSENSUS PROBLEM

We consider a population of N_a agents evolving according to interaction dynamics of form:

$$\dot{x}_i = \frac{1}{N_a}\sum_{j=1}^{N_a} P(x_i,x_j)(x_j - x_i) + u_i \quad x_i(0) = x_i^0\,, \quad (1)$$

where the kernel $P(x_i,x_j)$ models the communication between agents with states $x_i \in \mathbb{R}^d$, and the control variable $\mathbf{u} = (u_1,...,u_{N_a})$ aims at steering the system towards a consensus state $\bar{x} = \frac{1}{N_a}\sum_{i=1}^{N_a} x_i$. We express this goal as an infinite horizon nonlinear stabilization problem

$$\min_{\mathbf{u}(\cdot)\in\mathcal{L}^2(\mathbb{R}_+;\mathbb{R}^{d\times N_a})} \int_0^\infty \frac{1}{N_a}\sum_{i=1}^{N_a} \|x_i - \bar{x}\|^2 + \beta||u_i||_2^2 dt\,, \quad (2)$$

subject to (1). A natural feature of agent-based models is that the number of interacting agents can become prohibitively large. Hence, as the number of agents N_a grows, instead studying the microscopic, individual-based optimal control problem (1)-(2), one can conveniently model the population by means of the density distribution of agents

$$f = f(t;x), \qquad t \geq 0, \qquad x \in \mathbb{R}^d\,, \quad (3)$$

which evolves in time according with dynamics of the form

$$\partial_t f = -\nabla_x \cdot \left[(\mathcal{P}[f] + u)f\right], \quad (4)$$

where the mean field interaction force $\mathcal{P}$ relative to the distribution f reads

$$\mathcal{P}[f(x)] = \int_\Omega P(x,s)(s-x)f(s)ds\,. \quad (5)$$

The optimal solution of the mean field optimal control problem – obtained as combination of (4) with a suitable cost functional – is, by construction, independent of the number of agents, since it models the macroscopic behaviour of the population as a whole. However, the mean field optimal control solutions are meant to be computed via first-order optimality conditions, whose complexity is linked to the dimensionality d of the state space: even for moderate values of d, the computational cost can be formidably high (Bensoussan et al., 2013)(Fornasier and Solombrino, 2014).

3. BOLTZMANN-TYPE FORMULATION

To circumvent the difficulties related to the solution of the mean field control problem, here we aim at modeling the evolution in time of the population density function $f(t,x)$ from a kinetic viewpoint instead. To this end, we assume two agents with states $x_i, x_j \in \mathbb{R}^d$ interacting according to the binary rule

$$\begin{aligned} x_i^* &= x_i + \eta\left(P(x_i,x_j)(x_j - x_i) + u(x_i,x_j)\right) \\ x_j^* &= x_j + \eta\left(P(x_j,x_i)(x_i - x_j) + u(x_j,x_i)\right), \end{aligned} \quad (6)$$

where η measures the strength of the interaction, and (x_i^*, x_j^*) are the post-interaction states. Hence, the evolution of $f(t,x)$ is driven by a Boltzmann-type dynamics:

$$\partial_t f(t,x) = \lambda\mathcal{Q}_{\eta,u}(f,f)(t,x)\,, \quad (7)$$

where λ is a parameter describing the interaction frequency, and the operator $Q_{\eta,u}(f,f)$ accounts for the gain and loss of particles located a certain position x at time t, as follows

$$\mathcal{Q}_{\eta,u}(f,f) = \mathcal{Q}^+_{\eta,u}(f,f) - \mathcal{Q}^-_{\eta,u}(f,f) \quad (8)$$

with

$$\mathcal{Q}^+_{\eta,u}(f,f)(t,x) = \int_\Omega \frac{1}{\mathcal{J}_\eta} f(t,{}^*x)f(t,{}^*s)d\mathbf{s}\,,$$

$$\mathcal{Q}^-_{\eta,u}(f,f)(t,x) = f(t,x)\int_\Omega f(t,s)ds\,,$$

and where $({}^*x_i, {}^*x_j) \longmapsto (x_i,x_j)$ are the pre-interaction states associated to (6), and $\mathcal{J}_\eta$ represents the Jacobian of the binary interaction (6). The interest in solving (7), arises when considering under a quasi-invariant scaling (i.e. $\eta = \varepsilon, \lambda = \varepsilon^{-1}$), as this provides us with the following consistency theorem between the mean field evolution of the dynamics and their Boltzmann formulation. We refer the reader to (Albi et al., 2017a) for detailed derivation and proof of the result.

Theorem 1 Let $\eta \geq 0$, $\varepsilon > 0$, $P(\cdot,\cdot) \in \mathcal{L}^2_{loc}$ at all times $t \in [0,+\infty)$, and we consider a weak solution f of (7) from initial condition $f_0(x)$. Furthermore, we introduce the scaling $\eta = \varepsilon$, $\lambda = \varepsilon^{-1}$ for the binary interaction rule, and we define $f^\varepsilon(t;x)$ to be a solution for the associated scaled version of (7). Then, as $\varepsilon \to 0$, we have pointwise convergence (up to subsequences) of the scaled solution $f^\varepsilon(t;x)$ to the solution $f(t;x)$ of (4).

Different numerical schemes can be derived to simulate the kinetic dynamics, (Albi and Pareschi, 2013). In particular, the evolution of $f = f(t,x)$ can be approximated by means of Direct Simulation Monte Carlo Methods, introducing a forward Euler discretization as follows

$$\begin{aligned} f_{n+1} &= f_n + \Delta t\lambda\left(\mathcal{Q}^+_{\eta,u}(f_n,f_n) - \mathcal{Q}^-_{\eta,u}(f_n,f_n)\right) \quad (9) \\ &= \big(1 - \Delta t\lambda\big)f_n + \Delta t\lambda\, Q^+_{\eta,u}(f_n,f_n), \quad (10) \end{aligned}$$

with $\Delta t \leq \varepsilon$ to preserve positivity of the solution f_{n+1}. Thus, sampling N_{sample} particles from the initial distribution $f_0(x) = f(0,x)$ we can approximate the solution of (9) via stochastic simulation of the binary interaction (6).

The convenience of this Boltzmann-type description relies on the possibility of approximating the behaviour of the population as the quasi-invariant limit of binary interactions, meaning that at each time step the agents are influenced only within couples. This heavily tackles down the computational complexity involved, since we are now considering the combination of many 2−agents subproblems. The number of interacting couples depends on the frequency parameter $\lambda = 1/\varepsilon$: a choice $\varepsilon \ll 1$ leads to weak, but frequent interactions, which is the typical case of mean-field models. Nonetheless, this requires an efficient solver for the reduced consensus problem.

4. STATE DEPENDENT RICCATI EQUATION

In this section, we aim at solving the reduced 2-agents problem, for which the states –encoding the position of both the coupled agents i-j – are denoted as a single variable $\mathbf{x}(t) = (x_i(t), x_j(t))^\top \in \mathbb{R}^{2d}$. Similarly, we use bold notation when referring to the interaction force and the control variable associated to the dynamical system for $\mathbf{x}$.

The binary consensus problem resulting from the microscopic formulation(1)-(2) can be written as a nonlinear quadratic regulator problem (NLQR)

$$\min_{\mathbf{u}(\cdot)\in\mathbf{U}} \mathcal{J}_{\mathbf{x}_0}(\mathbf{u}(\cdot)) := \int_0^\infty \mathbf{x}^\top(s)\mathbf{Q}\mathbf{x}(s) + \mathbf{u}^\top(s)\mathbf{R}\mathbf{u}(s)\, ds\,, \quad (11)$$

subject to nonlinear, control-affine dynamics

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{B}\mathbf{u}(t)\,, \quad \mathbf{x}(0) = \mathbf{x}_0\,, \quad (12)$$

where $\mathbf{u}(\cdot) \in \mathbf{U} = \{\mathbf{u}(t) : \mathbb{R}^+ \to \mathbb{R}^{2d}, \text{measurable}\}$ is an unbounded control variable, $\mathbf{Q} \in \mathbb{R}^{2d\times 2d}$ is a symmetric positive semidefinite matrix, and $\mathbf{R} \in \mathbb{R}^{2d\times 2d}$ is symmetric positive definite. The control operator $\mathbf{B} : \mathbb{R}^{2d\times 2d}$, and the system dynamics $\mathbf{f}(\mathbf{x}) : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ are $\mathcal{C}^1(\mathbb{R}^{2d})$ and such that $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ and $\mathbf{B}(\mathbf{0}) = \mathbf{0}$. Using dynamic programming, the optimal feedback law $\mathbf{u}(\cdot)$ solving (11) can be recovered in terms of the value function of the control problem

$$V(\mathbf{x}) = \inf_{\mathbf{u}(\cdot)\in\mathbf{U}} \mathcal{J}_{\mathbf{x}}(\mathbf{u}(\cdot))\,, \quad (13)$$

solving the following first-order, static, nonlinear Hamilton-Jacobi-Bellman PDE

$$\nabla V(\mathbf{x})^\top \mathbf{f}(\mathbf{x}) - \frac{1}{4}\nabla V(\mathbf{x})^\top \mathbf{W}(\mathbf{x})\nabla V(\mathbf{x}) + \mathbf{x}^\top\mathbf{Q}\mathbf{x} = 0\,, \quad (14)$$

where $\mathbf{W} = \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^\top$. Once the function $V(\mathbf{x})$ is computed, the associated optimal feedback is given by

$$\mathbf{u}(\mathbf{x}) = -\frac{1}{2}\mathbf{R}^{-1}\mathbf{B}^\top\nabla V(\mathbf{x})\,. \quad (15)$$

Solving (14) can be in general difficult and expensive from a computational point of view. The value function $V(\cdot)$ maps variables living in $\mathbb{R}^{2d}$, where the dimension d can be arbitrarily high. Equation (14) is a nonlinear PDE, thus it can be unfeasible to solve via standard methods even for moderate dimensional optimal control problems ($d > 3$).

4.1 Algebraic Riccati Equation and state-dependence

In a simplified setting, under further assumptions of linearity for the free dynamics $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ and making the ansatz $V(\mathbf{x}) = \mathbf{x}^\top\Pi\mathbf{x}$ with $\Pi \in \mathbb{R}^{2d\times 2d}$, the optimality condition (15) can be written as

$$\mathbf{u}(\mathbf{x}) = -\mathbf{R}^{-1}\mathbf{B}^\top\Pi\mathbf{x}\,, \quad (16)$$

where Π is a positive definite solution of the Algebraic Riccati Equation (ARE)

$$\mathbf{A}^\top\Pi + \Pi\mathbf{A} - \Pi\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^\top\Pi + \mathbf{Q} = 0\,. \quad (17)$$

Even though the class of problems being addressed in this paper is characterized by non-linearity in the free dynamics, a similar approach arises when casting (12) in semilinear form:

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}(\mathbf{x})\mathbf{x}\,, \qquad \dot{\mathbf{x}} = \mathbf{A}(\mathbf{x})\mathbf{x} + \mathbf{B}\mathbf{u}\,. \quad (18)$$

In this setting, the solution of the HJB PDE (14) can be approximated by a state-dependent Riccati equation (SDRE)

$$\mathbf{A}^\top(\mathbf{x})\Pi(\mathbf{x}) + \Pi(\mathbf{x})\mathbf{A}(\mathbf{x}) - \Pi(\mathbf{x})\mathbf{W}\Pi(\mathbf{x}) + \mathbf{Q} = 0\,. \quad (19)$$

In the linear case, the ARE (17) directly comes from the HJB PDE (14) by considering the ansatz $V(\mathbf{x}) = \mathbf{x}^\top\Pi\mathbf{x}$ for the associated value function. Thus, the feedback (16) resulting from the ARE solution Π coincides with the optimal control variable resulting from (15). The consistency between the ARE and the Dynamic Programming solutions is not readily available when dealing with nonlinear dynamics of the form (18). This is due to the state-dependence in the SDRE solution $\Pi(\mathbf{x})$, which leads to

$$\begin{aligned} V(\mathbf{x}) &= \mathbf{x}^\top\Pi(\mathbf{x})\mathbf{x}\,, \\ \nabla V(\mathbf{x}) &= 2\Pi(\mathbf{x})\mathbf{x} + \varphi(\mathbf{x})\,, \end{aligned} \quad (20)$$

where $\varphi(\mathbf{x})$ is a $2d$-dimensional vector-valued function such that

$$\varphi_k(\mathbf{x}) = \sum_{i,j=1}^{2d} x_i x_j \frac{\partial\Pi(\mathbf{x})_{i,j}}{\partial x_k}\,. \quad (21)$$

Thus, when substituting (20) in the HJB PDE (14), we do not recover the SDRE (19), due to the presence of an additional term associated to the $\varphi(\mathbf{x})$ component in $\nabla V(\mathbf{x})$. For this reason, the feedback law

$$\mathbf{u}(\mathbf{x}) = -\mathbf{R}^{-1}\mathbf{B}^\top\Pi(\mathbf{x})\mathbf{x}\,, \quad (22)$$

is a suboptimal approximation to the HJB feedback. Nevertheless, under stabilizability assumptions, the SDRE feedback law is locally asymptotically stabilizing (Banks et al., 2007).

4.2 Freezing coefficients in the Riccati Equation

The main computational bottleneck associated to the synthesis of the SDRE feedback law is that eq. (19) cannot be solved analytically for for $\Pi(\mathbf{x})$, and needs to be realized in a model predictive control fashion along a trajectory, as proposed in (Banks et al., 2007). Given the current state $\mathbf{x}$ of the system, we assume the operator $\Pi(\mathbf{x})$ to be a positive definite matrix in $\Pi \in \mathbb{R}^{d\times d}$, meaning that (19) reduces to its algebraic form (17).

This procedure can be useful to generate suboptimal approximations of the controlled trajectories associated to infinite horizon control problems of the form (11)-(12). While evolving along a trajectory, we assume the system to be in a configuration $\bar{\mathbf{x}}$. By freezing every state-dependent operator accordingly with the current state $\bar{\mathbf{x}}$, we obtain an ARE to be solved for the frozen SDRE operator $\Pi(\bar{\mathbf{x}})$, and the associated feedback law $\mathbf{u}(\bar{\mathbf{x}})$ can be recovered via (22). Then, we let the system evolve according with the $\mathbf{u}(\bar{\mathbf{x}})$-controlled dynamics for a short time frame, after which the procedure is repeated.

Even if we assume that this SDRE approach generates asymptotically stable closed-loop solutions (Banks et al., 2007), a main limitation persists, residing in the implementation of a sufficiently efficient ARE solver to enable a high-frequency sampling of controlled binary interactions (6). For this efficiency purpose, we rely on supervised learning approximation models (Kang et al., 2021), (Darbon and Osher, 2016) to encode the control action in a neural network.

5. SUPERVISED LEARNING APPROXIMATION

We populate a training set for the control law by solving in an offline phase the frozen SDREs for a collection of states associated to N_s sampled couples in $\mathbb{R}^d \times \mathbb{R}^d$ of interacting agents $\mathcal{X}_t = \{(\mathbf{x} = x_i, x_j)^{(k)}\}_{k=1}^{N_s}$. Aiming at approximating the solution of the binary infinite horizon optimal control problem (11), we consider models within the family of feed-forward neural networks (FNNs), for which the choice of $\mathbf{u}(\cdot) \in \mathbb{R}^{2d}$ as learning target variable can be suboptimal in terms of goodness of fit of the model. A variety of alternatives has been proposed and compared in literature (Wang and Wu, 1998), (Kang et al., 2021), (Darbon and Osher, 2016), and a widely popular choice can be to target the associated scalar *value function* $V_\theta(\cdot) \approx V(\cdot) \in \mathbb{R}$, and then recover the feedback as a function of the gradient of $V_\theta(\cdot)$:

$$\mathbf{u}_V(\mathbf{x}) = -\frac{R^{-1}B^T\nabla V_\theta(\mathbf{x})}{2}\,, \tag{23}$$

where $\nabla V_\theta(\cdot)$ can be efficiently retrieved via automatic differentiation.

5.1 *Gradient-augmented supervised learning*

Since the learning final goal is to approximate the feedback law $\mathbf{u}(\cdot)$, the accuracy of the gradient approximation $\nabla V_\theta(\cdot)$ is fundamental. In this direction, our training is strengthened thanks to a *gradient-augmented* loss function, accounting for not only the approximation error in the target variable, but also for the discrepancy in terms of its gradient. This requires an enriched dataset for the training phase of the supervised learning procedure, including both the value function associated to the binary infinite horizon problem and its gradient $\mathcal{T} = \{\mathbf{x}^{(k)}, V(\mathbf{x}^{(k)}), \nabla V(\mathbf{x}^{(k)})\}_{k=1}^{N_s}$. In particular, at every sampled state $\mathbf{x}$, once the frozen SDRE has been solved for Π, we consider the ansatz $V(\mathbf{x}) = \mathbf{x}^T\Pi\mathbf{x}$ and we approximate the space derivative $\nabla V(\mathbf{x}) \approx 2\Pi\mathbf{x}$, obtained by neglecting the state dependency of the SDRE solution.

Even if this ansatz for $V(\cdot)$ approximates the HJB PDE only around a neighborhood of the origin when applied to nonlinear problems, and the target gradient is not exact, this choice allows to conveniently collect the enriched data-set without any computational cost additional to the solution of the ARE associated to the current state.

5.2 *Network Architecture*

Feed-forward neural networks approximate a function via a chain of composition of layers $l_1, ..., l_M$, consisting of an *activation function* $\sigma(\cdot)$ applied component-wise to a linear combination of the layer input variable:

$$l_m(\mathbf{y}) = \sigma_m(\mathbf{A}_m\mathbf{y} + \mathbf{b}_m)\,. \tag{24}$$

The weight matrices $\{\mathbf{A}_m, \mathbf{b}_m\}_m = \theta$ are parameters to be optimized during the training phase, so that the associated ANN minimizes a suitable *loss function*. In the gradient-augmented settings, we consider a compromise between a fitting functional and a gradient regulation:

$$\mathcal{L}oss(V, V_\theta) = \mathcal{L}_2(V, V_\theta) + \mu_{dV}\mathcal{L}_2(\nabla V, \nabla V_\theta)\,, \tag{25}$$

where $\mathcal{L}_2(f, f_\theta)$ denotes the *mean squared error* (MSE):

$$\mathcal{L}_2(f, f_\theta) := \frac{1}{N_s}\sum_{k=1}^{N_s} \|f(\mathbf{x}^{(k)}) - f_\theta(\mathbf{x}^{(k)})\|^2\,. \tag{26}$$

The number M of layers, their width (i.e. the number of neurons per layer), the activation functions $\sigma_m(\cdot)$, and the loss weight μ_{dV} are *hyper-parameters* of the model to be optimally tuned so that the trained model not only reaches a good approximation for the training set $\mathcal{X}_t$, but also generalizes outside the training data.

6. CONTROLLING OPINION DYNAMICS

The aforementioned methodology has been assessed with a numerical test from (Albi et al., 2017a) dealing with a high-dimensional consensus problem for first-order opinion dynamics governed by the Sznajd model (Sznajd-Weron and Sznajd, 2000). Here, the evolution of the state variables is described through the asymmetric interaction kernel $P(\cdot,\cdot)$ defined as follows:

$$P(x_i, x_j) = \beta(1 - x_i^2)\,, \qquad \beta \in \mathbb{R}. \tag{27}$$

We limit our state space to samples in $\Omega = [-1, 1]$ describing the opinions of a large population of voters between two extremal positions $\{-1, 1\}$. The interaction kernel models the propensity of agents to change their opinions when interacting with others: the more the agent's opinion is close to the boundary of the domain Ω, the less they are going to influence their peers. A choice of a parameter $\beta < 0$ leads to separation of opinions, meaning that without any external action, the population's opinion is going to concentrate around $x = 1$ and $x = -1$ (Aletti et al., 2007). Here, we fix $\beta = -1$.

As previously discussed, for a sufficiently high number of agents (here we consider $N_a = 10^5$), the individual-based model (1) can be cast in its mean field formulation (4), which is consistent with a kinetic-like equation (7) for the evolution in time of the population probability density of having an agent with opinion $x \in \Omega$. This Boltzmann description allows us to approximate the evolution of the mean field dynamics as a limit of binary interactions of sampled couples of agents within the population.

In order to cast the problem under consideration in semilinear form, we consider the following change of variables

$$(x_i, x_j)^\top \mapsto (x_i, \bar{x}) \tag{28}$$

This allows us to write the cost functional (2) in quadratic form (11) with weights $\mathbf{R} = \gamma/2$ and $\mathbf{Q} = 2\mathbb{I}_2 - \mathbb{J}_2$, where respectively $\mathbb{I}_2$ is the identity, and $\mathbb{J}_2$ denotes the matrix full of ones in $\mathbb{R}^{2\times 2}$. Similarly, the binary dynamics (1) with Sznajd kernel (27)

$$\dot{\mathbf{x}} = \begin{cases} \dot{x}_i = \dfrac{\beta}{2}(1 - x_i^2)(x_j - x_i) + u_i \\ \dot{x}_j = \dfrac{\beta}{2}(1 - x_j^2)(x_i - x_j) + u_j \end{cases} \tag{29}$$

can be written, after the change of variable (28), in semilinear form as

$$\begin{bmatrix} x_i \\ \bar{x} \end{bmatrix} = \begin{bmatrix} -P(x_i, \bar{x}) & P(x_i, \bar{x}) \\ -\bar{P}(x_i, \bar{x}) & \bar{P}(x_i, \bar{x}) \end{bmatrix} \begin{bmatrix} x_i \\ \bar{x} \end{bmatrix} + \mathbb{I}_2\mathbf{u} \tag{30}$$

where

$$P(x_i, \bar{x}) = \beta(1 - x_i^2) \qquad \bar{P}(x_i, \bar{x}) = \beta\big((2\bar{x} - x_i)^2 - x_i^2\big).$$

For populating the dataset $\mathcal{X}_t$ we uniformly sample $N_s = 10^3$ states $\mathbf{x}^{(k)} = (x_i, \bar{x})^{(k)}$ and we apply (28). For every sampled current state of the system, we compute the state-dependent SDRE coefficients and we rely on the `lqr` routine in MATLAB for solving the associated ARE for $\Pi(\mathbf{x}^{(k)})$. With this suboptimal SDRE solution, the training set $\mathcal{T}_V = \{\mathbf{x}^{(k)}, V^{(k)}, \nabla V^{(k)}\}_{k=1}^{Ns}$ is computed with $V^{(k)} = \mathbf{x}^{(k)T}\Pi(\mathbf{x}^{(k)})\mathbf{x}^{(k)}$, and $\nabla V^{(k)} = 2\Pi(\mathbf{x}^{(k)})\mathbf{x}^{(k)}$. An additional dataset is generated, with the purpose of comparing the gradient-augmented approximation $\mathbf{u}_V$ with the direct approximation of the feedback law $\mathbf{u}_\theta$: $\mathcal{T}_u = \{\mathbf{x}^{(k)}, u^{(k)}\}_{k=1}^{Ns}$, where $\mathbf{u}^{(k)} = -\mathbf{R}^{-1}\mathbf{B}^T\Pi(\mathbf{x}^{(k)})\mathbf{x}^{(k)}$. For the training of the model $\mathbf{u}_\theta$ the loss function was the MSE (26).

Once the datasets $\mathcal{T}_V$, $\mathcal{T}_u$ have been computed, they have been split into *training sets* and *validation sets*, with a ratio of 80/20. The ANN architectures for both the ANN $\mathbf{u}_\theta$ and V_θ (together with the loss weight $\mu_{dV} \in [0, 2]$) were chosen accordingly with the goodness of fit of the model evaluated within the validation samples: this hyperparameter tuning phase has been dealt with via a grid search in the parameter space by maximizing the precision of the trained model, by means of minimization of the *mean relative error* (MRE).

The desired architecture was identified in both cases to be a FFN with $M = 4$, having identity activation function for the input and output layers $\sigma_{1,4}(\mathbf{y}) = \mathbf{y}$, and a sigmoid function for the remaining ones $\sigma_{2,3}(\mathbf{y}) = (1 + e^{-\mathbf{y}})^{-1}$. The hidden layers were populated by 100 neurons per each, while the dimension of the state space Ω, and the scalar nature of the target of V_θ defined the depths of the remaining layers. For V_θ, the best configuration of hyperparameters set the loss weight to $\mu_{dV} = 0.05$.

The goodness of fit of the trained models is finally evaluated in a *test set*, a uniform grid of $N_v = 10^5$ points within the state space, where the approximated control is compared with the pointwise computation through the SDRE solution. Goodness of fit of trained models in both tests are presented in Table 1.

Table 1. Goodness of fit in terms of: MSE, *coefficient of determination* r^2, and MRE.

target		*MSE*	r^2	*MRE*
V_θ	$\mu_{dv} = 0.05$	9.24×10^{-6}	0.96480	0.0195
	$\mu_{dv} = 0$	3.50×10^{-5}	0.86674	0.2779
dV_θ	$\mu_{dv} = 0.05$	3.71×10^{-7}	0.99992	0.0079
	$\mu_{dv} = 0$	4.48×10^{-5}	0.99017	0.0987
u_V	$\mu_{dv} = 0.05$	5.94×10^{-4}	0.99992	0.0079
	$\mu_{dv} = 0$	0.071668	0.98417	0.0987
u_θ		0.002778	0.99962	0.0195

With the trained models V_θ and $\mathbf{u}_\theta$ for the binary interactions, we compare the evolution of a sampled couple under the action of the different controls: the suboptimal feedback law $\mathbf{u}$ obtained with the SDRE approach and its approximations $\mathbf{u}_V$ and $\mathbf{u}_\theta$. A further comparison is done w.r.t. the open-loop control variable obtained by solving Pontryagin's optimality conditions (PMP) holding in finite horizon settings. Aiming at approaching the feedback behaviour in PMP settings, we consider a time horizon T large enough for the system to reach consensus. A plot of the dynamics of a sampled couple of agents $x_i, x_j \in \Omega = [-1, 1]$ can be seen in Figure 1.

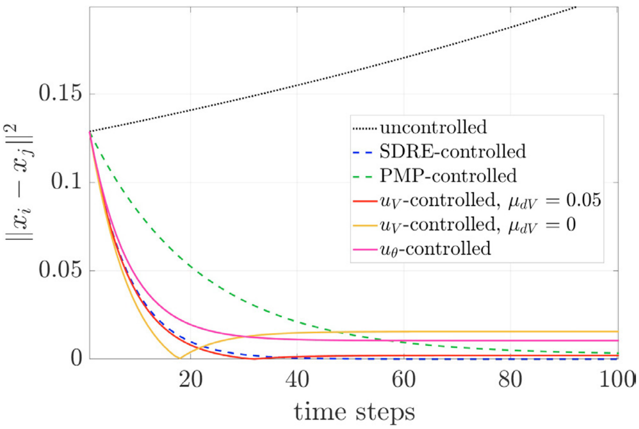

Fig. 1. Evolution of the Euclidean distance within a sampled couple of agents in both uncontrolled ($u(x) \equiv 0 \, \forall x \in \Omega$) and controlled settings. The SDRE feedback law $\mathbf{u}$ succeeds at steering the couple system towards consensus much faster than the open-loop control variable (for which we need $T > 100$ to reach consensus). Among the different approximations, the feedback $\mathbf{u}_V$, resulting from the gradient-augmented model, leads to the best performance.

Finally, we can rely on the binary interactions controlled via $\mathbf{u}_V$, $\mathbf{u}_\theta$ in order to approximate the behaviour of the whole population. In particular, with the choice of time-step $\Delta t = \varepsilon$ in (9), we allow each one of the agents to interact with someone else at every update. This means that at each time step, we can sample $N_a/2$ couples within the population and then act on their interactions by means of a feedback variable. Every couple evolves according to a forward Euler scheme with time-step Δt, after which the population density function is updated to be the sampled density of all the agents. In this way we approximate the behaviour of the controlled population from a mean field viewpoint, by only solving many 2-dimensional infinite horizon optimal control problems. The time evolution of the population probability density function influenced through $\mathbf{u}_V$ can be seen in Figure 3. In Figure 2, we compare the given initial distribution $f_0(x)$ with its time evolution according to controlled mean field dynamics by means of a variety of feedback laws.

7. CONCLUSIONS

In this paper a mixture of approximation techniques for solving optimal control of multi-agent systems has been discussed and numerically tested. The first approximation step coincides with considering a mean field formulation of the controlled dynamics, so that the number of agents populating the system no longer contributes to the dimensionality of the problem. Then, the complexity of the solution of such a mean field optimal control problem has been further reduced thanks to a description of the population dynamics from a kinetic viewpoint, by means of a Bolzmann equation for the time evolution of the population density. This formulation has the advantage that the complexity of the associated solution is dramatically reduced with respect to the mean field optimal control, still retaining the ability to influence the population as a whole. Finally, a gradient-augmented supervised learning model has been trained for approximating the suboptimal

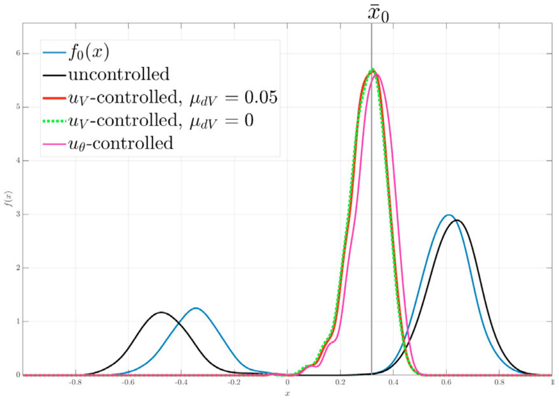

Fig. 2. Comparison between the initial density function $f_0(x)$ with the sampled probability density obtained as limit of binary interactions controlled by approximated control variables $\mathbf{u}_V$ and $\mathbf{u}_\theta$. The controlled system has only been evolved for 10 time steps and yet it is already concentrating around the target opinion (consensus) for both the standard and the gradient-augmented $\mathbf{u}_V$ feedback laws. The action of $\mathbf{u}_\theta$ can be seen to steer the agents towards a slightly different configuration. The uncontrolled dynamics are leading to opinion separation, consistently with the parameter choice $\beta = -1 < 0$.

SDRE solution of the reduced Bolzmann binary interactions. A comparison between the proposed model and the direct approximation of the feedback law in a numerical example has highlighted an outstanding performance of the former. In the future we will assess the proposed methodology in higher dimensional dynamics, where the supervised learning of the feedback map is essential to enable computational feasibility of the kinetic approach.

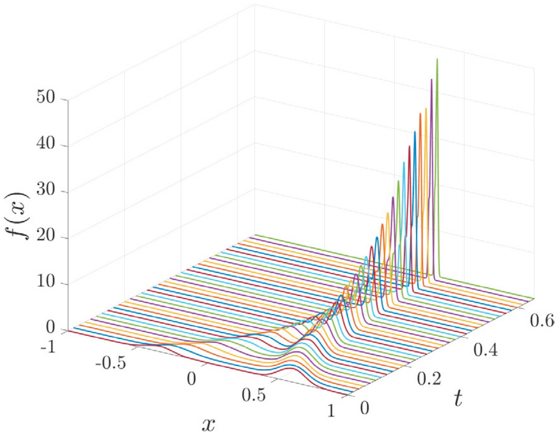

Fig. 3. Evolution of the $\mathbf{u}_V$-controlled probability density of finding an agent with opinion x versus time (seconds). From a double-peaked density obtained as a mixture of normal densities, the opinions of the agents rapidly converge towards consensus.

REFERENCES

Albi, G., Bicego, S., and Kalise, D. (2022). Gradient-augmented Supervised Learning of Optimal Feedback Laws Using State-Dependent Riccati Equations. *Systems Control Lett.*, 6, 836–841.

Albi, G., Choi, Y.P., Fornasier, M., and Kalise, D. (2017a). Mean field control hierarchy. *Appl. Math. Optim.*, 76(1), 93–135.

Albi, G., Fornasier, M., and Kalise, D. (2017b). A boltzmann approach to mean-field sparse feedback control. *IFAC-PapersOnLine*, 50(1), 2898–2903. 20th IFAC World Congress.

Albi, G., Herty, M., and Pareschi, L. (2015). Kinetic description of optimal control problems and applications to opinion consensus. *Commun. Math. Sci.*, 13(6), 1407–1429.

Albi, G. and Kalise, D. (2018). (sub)optimal feedback control of mean field multi-population dynamics. *IFAC-PapersOnLine*, 51(3), 86–91. 6th IFAC Workshop on Lagrangian and Hamiltonian Methods for Nonlinear Control LHMNC 2018.

Albi, G. and Pareschi, L. (2013). Binary interaction algorithms for the simulation of flocking and swarming dynamics. *Multiscale Model. Simul*, 11, 1–29.

Aletti, G., Naldi, G., and Toscani, G. (2007). First order continuous models of opinion formation. *SIAM Journal on Applied Mathematics*, 67(3), 837–853.

Azmi, B., Kalise, D., and Kunisch, K. (2021). Optimal feedback law recovery by gradient-augmented sparse polynomial regression. *J. Mach. Learn. Res.*, 22, Paper No. 48, 32.

Banks, H.T., Lewis, B.M., and Tran, H.T. (2007). Nonlinear feedback controllers and compensators: a state-dependent riccati equation approach. *Computational Optimization and Applications*, 37(2), 177–218.

Bensoussan, A., Frehse, J., and Yam, P. (2013). *Mean Field Games and Mean Field Type Control Theory*. Springer, New York.

Cloutier, J.R. (1997). State-dependent riccati equation techniques: an overview. In *Proceedings of the 1997 American Control Conference (Cat. No.97CH36041)*, volume 2, 932–936 vol.2.

Darbon, J. and Osher, S. (2016). Algorithms for overcoming the curse of dimensionality for certain hamilton-jacobi equations arising in control theory and elsewhere. *Res. Math. Sci.*, 3, 26, Paper No. 19.

Fornasier, M. and Solombrino, F. (2014). Mean-Field Optimal Control. *ESAIM: COCV*, 20(4), 1123–1152.

Jones, A. and Astolfi, A. (2020). On the solution of optimal control problems using parameterized state-dependent riccati equations. In *2020 59th IEEE Conference on Decision and Control (CDC)*, 1098–1103.

Kang, W., Gong, Q., Nakamura-Zimmerer, T., and Fahroo, F. (2021). Algorithms of data generation for deep learning and feedback design: a survey. *Phys. D*, 425, Paper No. 132955, 10.

Li, Y. and Tan, C. (2019). A survey of the consensus for multi-agent systems. *Systems Science & Control Engineering*, 7(1), 468–482.

Sznajd-Weron, K. and Sznajd, J. (2000). Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(06), 1157–1165.

Wang, J. and Wu, G. (1998). A multilayer recurrent neural network for solving continuous-time algebraic riccati equations. *Neural Networks*, 11(5), 939–950.