2024 American Control Conference (ACC)
July 8-12, 2024. Toronto, Canada

Large-Scale Multi-Agent System Optimization with Fixed Final Density Constraints: An Imbalanced Mean-Field Game Theory

Shawon Dey and Hao Xu

Abstract— **This paper presents a novel distributed optimization algorithm for large-scale multi-agent systems (LS-MAS), particularly with a given fixed final density constraint. Although the Mean field game (MFG) theory provides a distribution solution to overcome the "Curse of dimensionality" in LS-MAS, it significantly sacrifices LS-MAS optimality and also not be capable of achieving arbitrary fixed final probability density function (PDF) constraint. To overcome these challenges, a novel Imbalanced Mean-Field Game (Imb-MFG) theory is developed along with an adaptive PDF decomposition algorithm and distributed reinforcement learning. Specifically, an induction-based PDF parameter estimation is developed to decompose the final density constraints into multiple imbalanced norm distributions. Then, the Imb-MFG theory is designed by integrating multi-group MFG with a constrained K-means clustering algorithm. To solve the developed Imb-MFG and further obtain the distributed optimal solution, a multi-actor-critic-mass (Multi-ACM) algorithm is designed to learn the solution of multi-group coupled Hamilton-Jacobi-Bellman (HJB) and Fokker-Planck-Kolmogorov (FPK) equations simultaneously. Finally, the convergence of the developed Multi-ACM algorithm is guaranteed through Lyapunov analysis.**

I. Introduction

In recent years, there has been widespread adoption and a notable surge of interest in multi-agent systems (MAS) [1], especially with emphasis on applications such as traffic management [2], autonomous UAV [3] in military applications, and so on. With the rapid advancements in game theory [4] and distributed control [5], there has been an effective exploration of decision-making and control policies for MAS, supported by robust mathematical foundations. However, while extending MAS to LS-MAS, two significant challenges emerge. First, the data exchange in LS-MAS is needed but exceedingly difficult to maintain in practice due to communication intricacies. Second, the issue of the "*Curse of Dimensionality*" arises with the exponential expansion of agent interactions while solving the partial differential equation (PDE)-based optimal control. Addressing these two challenges in LS-MAS, the previous studies [6] employed the Mean-Field Game (MFG) theory [7]. In MFG theory, by using a locally computed PDF to represent the states of the massive agents in LS-MAS without engaging in interaction with other agents, each agent can effectively obtain group information without introducing extra communication demands and computational complexities. Although MFG addresses the challenge from notorious "*Curse of Dimensionality*" in LS-MAS, it is unrealistic and significantly limits the capability by assuming all the agents in LS-MAS are homogeneous and follow a single PDF. In addition, due to the homogeneous assumption, existing MFG theoretical control [7] has always led the overall PDF of LS-MAS to Gaussian distribution. It is very difficult to force the final PDF of LS-MAS to follow a given distribution different from the Gaussian distribution which limits the feasibility of existing MFG in the real world. To tackle these challenges, a novel Imbalanced Mean-Field Game (Imb-MFG) theory has been developed by integrating an adaptive PDF decomposition algorithm along with multi-group MFG [8]. Furthermore, adapting the distributed reinforcement learning (RL) [9], the optimal control strategy for LS-MAS with a fixed final density constraint can be obtained by learning the solution of Imb-MFG. Specifically, a massive number of agents are deployed, and none of the distributed agents possesses knowledge of the fixed final density constraint. Then, an induction-based PDF decomposition parameter estimation is developed to estimate the appropriate parameters that can decompose the fixed final density function into a combination of multiple imbalanced norm distributions. Next, a constrained K-means clustering [10] approach is utilized along with estimated PDF decomposition parameters to divide the LS-MAS into multiple groups. After that, the initial PDF distributions for individual groups in LS-MAS are obtained. Then, using multi-group MFG [8] along with obtained initial and desired final imbalanced norm distributions for individual groups, an Imb-MFG is formulated. Within each group, the MFG will force its PDF to converge to the desired imbalanced norm distribution. Eventually, the multiple imbalanced norm distributions of multi-group LS-MAS can be mixed jointly to satisfy the fixed final PDF constraint. Similar to other MFG theory [11], finding the optimal solution of Imb-MFG needs solving multiple coupled forward and backward PDEs, called the FPK and HJB equation, for decomposed multi-group LS-MAS. However, solving these PDEs directly is quite challenging [8]. To tackle this issue, a multi-actor-critic-mass (Multi-ACM) learning algorithm is developed by adopting adaptive dynamic programming [12] and reinforcement learning [9] techniques.

The major contributions are: 1) An Imbalanced Mean-Field Game (Imb-MFG) theory is developed along with an induction-based PDF decomposition method to formulate the distributed optimal control problem for LS-MAS with a fixed final PDF constraint. 2) A multi-actor-critic-mass (Multi-ACM) learning algorithm is developed to solve the Imb-MFG theory and attain a distributed optimal solution for LS-MAS

The authors are with the Department of Electrical and Biomedical Engineering, University of Nevada, Reno, NV, 89557 USA. E-mail: sdey@unr.edu; haoxu@unr.edu. This work was supported by the National Science Foundation under Grant 2144646.

979-8-3503-8265-5/$31.00 ©2024 AACC

with a fixed final density constraint.

II. Problem Formulation

Consider an LS-MAS of M agents. Then, the stochastic homogeneous dynamic system of any agent $\mathcal{A}$ is defined as:

$$dx(t) = [f(x(t)) + g(x(t))u(t)]dt + \sigma d\omega \tag{1}$$

with state $x(t) \in \mathbb{R}^n$ and control $u(t) \in \mathbb{R}^m$. The dynamics function $f(x) \in \mathbb{R}^n$ and $g(x) \in \mathbb{R}^{n\times m}$ are smooth and known. The term $\omega \in \mathbb{R}^n$ is the wiener process represents environmental noise, and $\sigma \in \mathbb{R}^{n\times n}$ is the coefficient matrix. According to classical statistical theory [13], the final PDF constraint, $m_d(x;\theta_d)$, in LS-MAS optimization is assumed to be represented as a linear combination of multiple norm distributions with different means and variances, i.e.

$$m_d(x;\theta_d) = \sum_{j=1}^{N} w_{d,j} m_{d,j}(x;\theta_{d,j}) \quad j = 1, ..., N. \tag{2}$$

where $m_{d,j}(x,\theta_{d,j})$ is the norm distribution for j^{th} group, with a total N groups in LS-MAS and $\theta_{d,j} = \{\mu_{d,j}, \Sigma_{d,j}\}$ is a parameter set with mean $\mu_{d,j}$ and covariance matrix $\Sigma_{d,j}$ of the group j. The collection of all the parameters including weight for group j is denoted as $\theta_{c,j} = \{w_{d,j}, \mu_{d,j}, \Sigma_{d,j}\}$. Then, the mixture-PDF is parameterized by the weight $\boldsymbol{w}$, mean $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$. The parameters collection in mixture-PDF is denoted as $\theta = \{\boldsymbol{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$. Next, the cost function for agent $\mathcal{A}$ in LS-MAS can be formulated as

$$J(x, m(x;\theta)) = \mathbb{E}\Big\{ \int_0^\infty \begin{bmatrix} r(x(t),u(t))+ \\ \Phi(m(x;\theta)) \end{bmatrix} dt \Big\} \tag{3}$$

where the first term is defined as $r(x(t),u(t)) = \|x - \mathbb{E}\{m_{d,j}(x,\theta_{d,j})\}\|_Q^2 + \|u\|_R^2$, captures the state error and control input's quadratic norms weighted by Q and R, which are positive definite matrix. Here, $\mathbb{E}\{m_{d,j}(x,\theta_{d,j})\}$ represents the expected mean value of the desired PDF for group j. Then any agent $\mathcal{A}$ error term from group j can be defined as $e = x - \mathbb{E}\{m_{d,j}(x,\theta_{d,j})\}$ with error dynamic

$$de = [f_a(e) + g_a(e)u]dt + \sigma d\omega \tag{4}$$

where $f_a(e) = f(e + \mathbb{E}\{m_{d,j}(x,\theta_{d,j})\})$ and $g_a(e) = g(e + \mathbb{E}\{m_{d,j}(x,\theta_{d,j})\})$. The second term in the cost function is a coupling function that is used to achieve the target mixture-PDF of LS-MAS. The coupling function can be written as:

$$\Phi(m(x;\theta)) = \|m_j(x;\theta) - m_{d,j}(x;\theta_{d,j})\|_2^2 \tag{5}$$

This function quantifies the discrepancy between the real-time PDF of individual groups, denoted by $m_j(x;\theta)$, and the desired final PDF constraint, denoted by $m_{d,j}(x;\theta_{d,j})$.

The optimal control formulation: Considering continuous dynamics of agent $\mathcal{A}$ in (1), an admissible control policy need to be evaluated to minimize the cost function in (3). Then, according to the optimal control [14] and Bellman's optimality principle [8], the Hamiltonian is defined as follows:

$$H[x, \partial_x J(x, m_j(x;\theta))] = \mathbb{E}\Big\{ \Phi(m_j(x;\theta)) + \partial_x J^T(x, m_j(x;\theta))[f(x) + g(x)u] \Big\} \tag{6}$$

Next, the optimal control for each agent can be derived as

$$u(x) = -1/2\ \mathbb{E}\Big\{ R^{-1} g^T(x) \partial_x J(x, m_j(x;\theta)) \Big\} \tag{7}$$

Then the corresponding HJB equation is obtained by substituting the optimal evaluation function into the Hamiltonian which is shown in Eq. (16). To obtain the HJB equation, the PDF m_j is required. The PDF function can be obtained by solving the FPK equation shown in Eq. (17).

III. Imbalanced-MFG Theory with Induction-based PDF Decomposition and Multi-ACM Learning-based Neural Network Estimators

An Imb-MFG theory-based optimal control framework is designed to collectively achieve the final mixture-PDF through the efforts of individual agents. Particularly, an induction-based adaptive PDF decomposition method is designed to decompose the final PDF constraint into a combination of multi-imbalanced norm distributions. Then, the overall final PDF constraint is achieved by dividing LS-MAS into multi-groups and ensuring individual groups converge to imbalanced norm distribution. Specifically, a constrained k-means clustering algorithm is adopted to decompose the LS-MAS into multiple groups. Then, an Imb-MFG theory is developed to formulate the distributed optimal control problem for individual agents, aiming to achieve the desired final mixture-PDF. Due to the property of MFG [7], Imb-MFG can ensure agents within the same group achieve to desired imbalanced norm distribution.

A. Induction-based adaptive PDF Decomposition

In this section, an induction-based method is developed to estimate the parameters of the final mixture-PDF function. The ideal final PDF constraint is as follows

$$m_d(x;\theta_d) = \sum_{j=1}^{N} w_{d,j} m_{d,j}(x;\mu_{d,j},\Sigma_{d,j}) \tag{8}$$

where $w_{d,j} \in \mathbb{R}$ are the ideal weights, $\mu_{d,j} \in \mathbb{R}^n$ are the mean, and $\Sigma_{d,j} \in \mathbb{R}^{n\times n}$ are the covariance of individual norm distribution. Next, the equation (8) is rewritten as:

$$m_d(x;\theta_d) = \boldsymbol{w}_d^T m_d(x : \boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d) \tag{9}$$

Here, $\boldsymbol{w}_d \in \mathbb{R}^N$ is the weight of the ideal mixture-PDF. Now the final mixture-PDF function can be estimated as follows:

$$\hat{m}_d(x;\hat{\theta}_d) = \hat{\boldsymbol{w}}_d^T m_d(x;\hat{\boldsymbol{\mu}}_d, \hat{\boldsymbol{\Sigma}}_d) \tag{10}$$

The estimation error of the PDF function can be defined as:

$$e_m = \boldsymbol{w}_d^T m_d(x;\boldsymbol{\mu}_d,\boldsymbol{\Sigma}_d) - \hat{\boldsymbol{w}}_d^T m_d(x;\hat{\boldsymbol{\mu}}_d,\hat{\boldsymbol{\Sigma}}_d) \tag{11}$$

The weights estimation error is $\tilde{\boldsymbol{w}}_d = \boldsymbol{w}_d - \hat{\boldsymbol{w}}_d$. The PDF function estimation error is $\tilde{m}_d(x;\tilde{\boldsymbol{\mu}}_d,\tilde{\boldsymbol{\Sigma}}_d) = m_d(x;\boldsymbol{\mu}_d,\boldsymbol{\Sigma}_d) - m_d(x;\hat{\boldsymbol{\mu}}_d,\hat{\boldsymbol{\Sigma}}_d)$. The equation (11) can be represented as:

$$e_m = \boldsymbol{w}_d^T m_d(x;\boldsymbol{\mu}_d,\boldsymbol{\Sigma}_d) - \boldsymbol{w}_d^T m_d(x;\hat{\boldsymbol{\mu}}_d,\hat{\boldsymbol{\Sigma}}_d) + \boldsymbol{w}_d^T m_d(x;\hat{\boldsymbol{\mu}}_d,\hat{\boldsymbol{\Sigma}}_d) - \hat{\boldsymbol{w}}_d^T m_d(x;\hat{\boldsymbol{\mu}}_d,\hat{\boldsymbol{\Sigma}}_d)$$

$$= \boldsymbol{w}_d^T \tilde{m}_d(x; \tilde{\boldsymbol{\mu}}_d, \tilde{\boldsymbol{\Sigma}}_d) + \tilde{\boldsymbol{w}}_d^T m_d(x; \hat{\boldsymbol{\mu}}_d, \hat{\boldsymbol{\Sigma}}_d) \quad (12)$$

Assumption 1: The PDF function is Lipschitz continuous, implies the existence of Lipschitz functions L_μ and L_Σ, satisfy the inequality $\|\tilde{m}_d(x; \tilde{\boldsymbol{\mu}}_d, \tilde{\boldsymbol{\Sigma}}_d)\| \leq L_\mu \|\tilde{\boldsymbol{\mu}}_d\| + L_\Sigma \|\tilde{\boldsymbol{\Sigma}}_d\|$.

Now the squared residual error is defined as $E_m = 1/2\ e_m^T e_m$. If $\hat{\boldsymbol{w}}_d \to \boldsymbol{w}_d$, $\hat{\boldsymbol{\mu}}_d \to \boldsymbol{\mu}_d$ and $\hat{\boldsymbol{\Sigma}}_d \to \boldsymbol{\Sigma}_d$, then $e_m \to 0$. Then, an induction-based gradient descent method is designed to update the parameters of the mixture PDF. The iteration index is denoted as l. In the remaining sections, the bold notation is omitted to simplify the presentation. The gradient descent-based parameters update law is defined as

$$\hat{w}_d^{[l+1]} = \hat{w}_d^{[l]} + \alpha_w m_d(x; \hat{\mu}_d, \hat{\Sigma}_d) e_m^T \quad (13)$$

$$\hat{\mu}_d^{[l+1]} = \hat{\mu}_d^{[l]} + \alpha_\mu \hat{w}_d^{[l]} m_\mu(x; \hat{\mu}_d, \hat{\Sigma}_d) e_m^T \quad (14)$$

$$\hat{\Sigma}_d^{[l+1]} = \hat{\Sigma}_d^{[l]} + \alpha_\Sigma \hat{w}_d^{[l]} m_\Sigma(x; \hat{\mu}_d, \hat{\Sigma}_d) e_m^T \quad (15)$$

where α_w, α_μ and α_Σ are the mixture PDF weights learning gains. Also, $m_\mu(x; \hat{\mu}_d, \hat{\Sigma}_d)$ and $m_\Sigma(x; \hat{\mu}_d, \hat{\Sigma}_d)$ are the first derivative of activation function $m_d(x; \hat{\mu}_d, \hat{\Sigma}_d)$ with respect to imbalanced means and covariances, respectively.

The convergence of the mixture PDF's parameters estimation is described by the induction-based Lyapunov analysis:

Theorem 1*:* The update law of the mixture PDF's parameters is given by (13)-(15), with positive tuning gains. According to mathematical induction theory [15], in the base case, the parameters approximation error $\tilde{w}_d^{[l]}$, $\tilde{\mu}_d^{[l]}$, and $\tilde{\Sigma}_d^{[l]}$ at the lth iteration, are uniformly and ultimately bounded (UUB). The bounds for these errors are denoted as B_{w_d}, B_{μ_d}, and B_{Σ_d}. Moving to the induction step at the $(l+1)$th iteration, if the UUB statement holds for the base case at the lth iteration, it must also hold for the next case at the $(l+1)$th iteration.

Proof: Consider the Lyapunov candidate function as follows:

$$\begin{aligned}
&\Delta L_{w_d} = \tilde{w}_d^{T^{[l+1]}} \tilde{w}_d^{[l+1]} - \tilde{w}_d^{T^{[l]}} \tilde{w}_d^{[l]} \\
&= [\tilde{w}_d^{[l]} - \alpha_w m_d(x; \hat{\mu}_d, \hat{\Sigma}_d) e_m^T]^T [\tilde{w}_d^{[l]} - \alpha_w m_d(x; \hat{\mu}_d, \hat{\Sigma}_d) e_m^T] \\
&- \tilde{w}_d^{T^{[l]}} \tilde{w}_d^{[l]} \quad (18)
\end{aligned}$$

Now substituting equation (12) and considering assumption 1 of Lipschitz function, the equation (18) is rewritten as:

$$\begin{aligned}
&\Delta L_{w_d} \leq -2\alpha_w \tilde{w}_d^{T^{[l]}} m_{w_d} \{w_d [L_\mu \|\tilde{\mu}_d^{[l-1]}\| + L_\Sigma \|\tilde{\Sigma}_d^{[l-1]}\|] \\
&+ \tilde{w}_d^{T^{[l]}} m_{w_d}\} + \alpha_w^2 \|m_{w_d}\|^2 \{w_d [L_\mu \|\tilde{\mu}_d^{[l-1]}\| + L_\Sigma \|\tilde{\Sigma}_d^{[l-1]} \\
&\|] + \tilde{w}_d^{T^{[l]}} m_{w_d}\}^2 \quad (19)
\end{aligned}$$

where $m_{w_d} = m_d(x; \hat{\mu}_d, \hat{\Sigma}_d)$. While iterative updating the parameters, certain dependencies are employed. Specifically, during the weight updates, the mean and covariance approximation errors from the previous iteration are utilized. In the subsequent mean update, both the current iteration's weight update and the previous iteration's covariance are taken into account. Lastly, during the covariance update, the current iteration's weight and mean approximations are utilized, given that these parameters have already been updated in the current iteration. Now, the Eq. (19) is rewritten as follows

$$\begin{aligned}
&\Delta L_{w_d} \leq -2\alpha_w \tilde{w}_d^{T^{[l]}} w_d m_{w_d} L_\mu \|\tilde{\mu}_d^{[l-1]}\| - 2\alpha_w \tilde{w}_d^{T^{[l]}} w_d \\
&m_{w_d} L_\Sigma \|\tilde{\Sigma}_d^{[l-1]}\| - 2\alpha_w \|m_{w_d}\|^2 \|\tilde{w}_d^{[l]}\|^2 + 4\alpha_w^2 \|m_{w_d}\|^2 w_d \\
&L_\mu^2 \|\tilde{\mu}_d^{[l-1]}\|^2 + 4\alpha_w^2 \|m_{w_d}\|^2 \|w_d\|^2 L_\Sigma^2 \|\tilde{\Sigma}_d^{[l-1]}\|^2 + 2\alpha_w^2 \\
&\|m_{w_d}\|^4 \|\tilde{w}_d^{[l]}\|^2 \\
&\leq -[2\alpha_w \|m_{w_d}\|^2 - 2\alpha_w^2 \|m_{w_d}\|^2 \|w_d\|^2 - 2\alpha_w^2 \|m_{w_d}\|^4] \\
&\|\tilde{w}_d^{[l]}\|^2 + \Phi_{\text{con}}^w(\tilde{\mu}_d^{[l-1]}, \tilde{\Sigma}_d^{[l-1]}) \quad (20)
\end{aligned}$$

where,

$$\begin{aligned}
&\Phi_{\text{con}}^w(\tilde{\mu}_d^{[l-1]}, \tilde{\Sigma}_d^{[l-1]}) = [L_\mu^2 + 4\alpha_w^2 \|m_{w_d}\|^2 \|w_d\|^2 L_\mu^2] \\
&\|\tilde{\mu}_d^{[l-1]}\|^2 + [L_\Sigma^2 + 4\alpha_w^2 \|m_{w_d}\|^2 \|w_d\|^2 L_\Sigma^2] \|\tilde{\Sigma}_d^{[l-1]}\|^2 \quad (21)
\end{aligned}$$

Then ΔL_{w_d} is less than zero outside a compact set if:

$$\|\tilde{w}_d^{[l]}\| > \sqrt{\frac{\Phi_{\text{con}}^w(\tilde{\mu}_d^{[l-1]}, \tilde{\Sigma}_d^{[l-1]})}{\alpha_w \left[\begin{matrix} 2\|m_{w_d}\|^2 - 2\alpha_w \\ \|m_{w_d}\|^2 \|w_d\|^2 - 2\alpha_w \|m_{w_d}\|^4 \end{matrix}\right]}} \equiv B_{w_d} \quad (22)$$

Next, Consider a Lyapunov candidate function as follows:

$$\Delta L_{\mu_d} = \tilde{\mu}_d^{T^{[l+1]}} \tilde{\mu}_d^{[l+1]} - \tilde{\mu}_d^{T^{[l]}} \tilde{\mu}_d^{[l]} \quad (23)$$

Substituting parameters estimation error dynamics and subsequently utilizing equation (12) to replace e_m, the expression represented by equation (23) can be reformulated as:

$$\begin{aligned}
&\Delta L_{\mu_d} \leq -2\alpha_\mu \tilde{\mu}_d^{T^{[l]}} \|\hat{w}_d^{[l]}\| L_{m_{\hat{\mu}}} \{w_d [L_\mu \|\tilde{\mu}_d^{[l]}\| + L_\Sigma \|\tilde{\Sigma}_d^{[l-1]}\|] \\
&+ \tilde{w}_d^{T^{[l]}} m_{w_d}\} + \alpha_\mu^2 \|\hat{w}_d^{[l]}\|^2 L_{m_{\hat{\mu}}}^2 \{w_d [L_\mu \|\tilde{\mu}_d^{[l]}\| + L_\Sigma \|\tilde{\Sigma}_d^{[l-1]}\|] \\
&+ \tilde{w}_d^{T^{[l]}} m_{w_d}\}^2 \\
&\leq -[2\alpha_\mu \|\hat{w}_d^{[l]}\| L_{m_{\hat{\mu}}} w_d L_\mu - \alpha_\mu^2 \|\hat{w}_d^{[l]}\|^2 L_{m_{\hat{\mu}}}^2 - 4\alpha_\mu^2 \|\hat{w}_d^{[l]}\|^2 \\
&L_{m_{\hat{\mu}}}^2 \|w_d\|^2 L_\mu^2] \|\tilde{\mu}_d^{[l]}\|^2 + \Phi_{\text{con}}^\mu(\tilde{w}_d^{[l]}, \Sigma_d^{[l-1]}) \quad (24)
\end{aligned}$$

where,

$$\begin{aligned}
&\Phi_{\text{con}}^\mu(\tilde{w}_d^{[l]}, \Sigma_d^{[l-1]}) = [L_\Sigma^2 \|w_d\|^2 + 4\alpha_\mu^2 \|\hat{w}_d^{[l]}\|^2 L_{m_{\hat{\mu}}}^2 \|w_d\|^2 L_\Sigma^2] \\
&\|\tilde{\Sigma}^{[l-1]}\|^2 + [m_{w_d}^2 + 2\alpha_\mu^2 \|\hat{w}_d^{[l]}\|^2 L_{m_{\hat{\mu}}}^2 m_{w_d}^2] \|\tilde{w}_d^{[l]}\|^2 \quad (25)
\end{aligned}$$

Then ΔL_{μ_d} is less than zero outside a compact set if:

$$\|\tilde{\mu}_d^{[l]}\| > \sqrt{\frac{\Phi_{\text{con}}^\mu(\tilde{w}_d^{[l]}, \Sigma_d^{[l-1]})}{\alpha_\mu \left[\begin{matrix} 2\|\hat{w}_d^{[l]}\| L_{m_{\hat{\mu}}} w_d L_\mu - \alpha_\mu \|\hat{w}_d^{[l]}\|^2 \\ L_{m_{\hat{\mu}}}^2 - 4\alpha_\mu \|\hat{w}_d^{[l]}\|^2 L_{m_{\hat{\mu}}}^2 \|w_d\|^2 L_\mu^2 \end{matrix}\right]}} \equiv B_{w_\mu} \quad (26)$$

Finally, Consider a Lyapunov candidate function as follows:

$$\Delta L_{\Sigma_d} = \tilde{\Sigma}_d^{T^{[l+1]}} \tilde{\Sigma}_d^{[l+1]} - \tilde{\Sigma}_d^{T^{[l]}} \tilde{\Sigma}_d^{[l]} \quad (27)$$

Now the equation (27) is reformulated as:

$$\begin{aligned}
&\Delta L_{\Sigma_d} \leq -[2\alpha_\Sigma \hat{w}_d^{T^{[l]}} w_d L_{m_{\hat{\Sigma}}} L_\Sigma - 2\alpha_\Sigma^2 \|\hat{w}_d^{[l]}\|^2 L_{m_{\hat{\Sigma}}}^2 - 4\alpha_\Sigma^2 \\
&\|\tilde{w}_d^{[l]}\|^2 L_{m_{\hat{\Sigma}}}^2 \|w_d\|^2 L_\Sigma^2] \|\tilde{\Sigma}_d^{[l]}\|^2 + \Phi_{\text{con}}^\Sigma(\tilde{w}_d^{[l]}, \tilde{\mu}_d^{[l]}) \quad (28)
\end{aligned}$$

with,

$$\begin{aligned}
&\Phi_{\text{con}}^\Sigma(\tilde{w}_d^{[l]}, \tilde{\mu}_d^{[l]}) = [L_\mu^2 \|w_d\|^2 + 4\alpha_\Sigma^2 \|\tilde{w}_d^{[l]}\|^2 L_{m_{\hat{\Sigma}}}^2 \|w_d\|^2 L_\mu^2] \\
&\|\tilde{\mu}_d^{[l]}\|^2 + [m_{w_d}^2 + 2\alpha_\Sigma^2 \|\tilde{w}_d^{[l]}\|^2 L_{m_{\hat{\Sigma}}}^2 m_{w_d}^2] \|\tilde{w}_d^{[l]}\|^2 \quad (29)
\end{aligned}$$

$$HJB: \mathbb{E}\{\Phi(x, m_j(x;\theta))\} = \mathbb{E}\{-\partial_t J(x, m_j(x;\theta)) - 0.5\sigma^2 \Delta J(x, m_j(x;\theta)) + H[x, \partial_x J(x, m_j(x;\theta))]\} \tag{16}$$

$$FPK: \mathbb{E}\{\partial_t m_j(x;\theta) - 0.5\sigma^2 \Delta m_j(x;\theta) - div(m_j D_p H[x, \partial_x J(x, m_j(x;\theta))])\} = 0 \tag{17}$$

Then ΔL_{Σ_d} is less than zero outside a compact set if:

$$\|\tilde{\Sigma}_d^{[l]}\| > \sqrt{\frac{\Phi_{\text{con}}^{\Sigma}(\tilde{w}_d^{[l]}, \mu_d^{[l]})}{\alpha_\Sigma \begin{bmatrix} 2\|\hat{w}_d^{[l]}\| w_d L_{m_{\hat{\Sigma}}} L_\Sigma - 2\alpha_\Sigma \|\hat{w}_d^{[l]}\|^2 \\ L_{m_{\hat{\Sigma}}}^2 - 4\alpha_\Sigma \|\tilde{w}_d^{[l]}\|^2 L_{m_{\hat{\Sigma}}}^2 \|w_d\|^2 L_\Sigma^2 \end{bmatrix}}} \equiv B_{w_\Sigma} \tag{30}$$

Note that, $L_{m_{\hat{\mu}}}$ and $L_{m_{\hat{\Sigma}}}$ are the Lipschitz constants of the functions $m_\mu(x; \hat{\mu}_d, \hat{\Sigma}_d)$ and $m_\Sigma(x; \hat{\mu}_d, \hat{\Sigma}_d)$, respectively. Following a similar approach as the previous method, we can deduce the subsequent condition for the next iteration:

$$\|\tilde{w}_d^{[l+1]}\| > \sqrt{\frac{\Phi_{\text{con}}^{w}(\tilde{\mu}_d^{[l]}, \tilde{\Sigma}_d^{[l]})}{\alpha_w \begin{bmatrix} 2\|m_{w_d}\|^2 - 2\alpha_w \\ \|m_{w_d}\|^2 \|w_d\|^2 - 2\alpha_w \|m_{w_d}\|^4 \end{bmatrix}}} \tag{31}$$

with, $\Phi_{\text{con}}^{w}(\tilde{\mu}_d^{[l]}, \tilde{\Sigma}_d^{[l]}) < \Phi_{\text{con}}^{w}(\tilde{\mu}_d^{[l-1]}, \tilde{\Sigma}_d^{[l-1]})$, $\|\tilde{\mu}_d^{[l]}\| < \|\tilde{\mu}_d^{[l-1]}\|$ and $\|\tilde{\Sigma}_d^{[l]}\| < \|\tilde{\Sigma}_d^{[l-1]}\|$. Similarly,

$$\|\tilde{\mu}_d^{[l+1]}\| > \sqrt{\frac{\Phi_{\text{con}}^{\mu}(\tilde{w}_d^{[l+1]}, \Sigma_d^{[l]})}{\alpha_\mu \begin{bmatrix} 2\|\hat{w}_d^{[l+1]}\| L_{m_{\hat{\mu}}} w_d L_\mu - \alpha_\mu \|\hat{w}_d^{[l+1]}\|^2 \\ L_{m_{\hat{\mu}}}^2 - 4\alpha_\mu \|\hat{w}_d^{[l+1]}\|^2 L_{m_{\hat{\mu}}}^2 \|w_d\|^2 L_\mu^2 \end{bmatrix}}} \tag{32}$$

with, $\Phi_{\text{con}}^{\mu}(\tilde{w}_d^{[l+1]}, \tilde{\Sigma}_d^{[l]}) < \Phi_{\text{con}}^{\mu}(\tilde{w}_d^{[l]}, \tilde{\Sigma}_d^{[l-1]})$, $\|\tilde{\Sigma}_d^{[l]}\| < \|\tilde{\Sigma}_d^{[l-1]}\|$ and $\|\tilde{w}_d^{[l+1]} < \|\tilde{w}_d^{[l]}\|\|$. And,

$$\|\tilde{\Sigma}_d^{[l+1]}\| > \sqrt{\frac{\Phi_{\text{con}}^{\Sigma}(\tilde{w}_d^{[l+1]}, \mu_d^{[l+1]})}{\alpha_\Sigma \begin{bmatrix} 2\|\hat{w}_d^{[l+1]}\| w_d L_{m_{\hat{\Sigma}}} L_\Sigma - 2\alpha_\Sigma \|\hat{w}_d^{[l+1]}\|^2 \\ L_{m_{\hat{\Sigma}}}^2 - 4\alpha_\Sigma \|\tilde{w}_d^{[l+1]}\|^2 L_{m_{\hat{\Sigma}}}^2 \|w_d\|^2 L_\Sigma^2 \end{bmatrix}}} \tag{33}$$

with, $\Phi_{\text{con}}^{\Sigma}(\tilde{w}_d^{[l+1]}, \mu_d^{[l+1]}) < \Phi_{\text{con}}^{\Sigma}(\tilde{w}_d^{[l]}, \mu_d^{[l]})$, $\|\tilde{\mu}_d^{[l+1]}\| < \|\tilde{\mu}_d^{[l]}\|$ and $\|\tilde{\Sigma}_d^{[l+1]}\| < \|\tilde{\Sigma}_d^{[l]}\|$. Given the mathematical induction-based method, the base case establishes that the parameter approximation errors are bounded. The induction step then states that if the base case holds, the boundedness of the errors carries over to subsequent iterations. Now, the weight set of the final desired PDF functions is defined as $\hat{w}_d = \{\hat{w}_{d,1}, \hat{w}_{d,2}, ...\hat{w}_{d,N}\}$. Then, an iterative constrained K-means clustering algorithm [10] is employed to break down the LS-MAS system into N groups. The constraints are determined by the estimated weights, ensuring that each group contains at least the minimum required number of agents to achieve the desired final mixture-PDF. Here, the cluster number is defined as $K = N$. Also, j is the cluster index with $j = 1, 2, ..., K$. The minimum number of agents in cluster j can be defined as $p_j = (\frac{\hat{w}_{d,j}}{\sum_{j=1}^{K} \hat{w}_{d,j}})M$, with $\sum_{j=1}^{K} p_j \leq M$. The iterative constrained K-means algorithm [10] with the redefined constraint on agent number p_j is provided as follows.

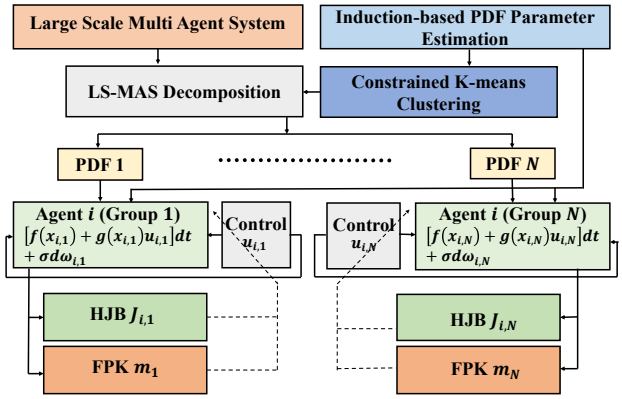


Fig. 1: Imb-MFG theory based LS-MAS adaptive PDF decomposition and Multi-ACM learning

Iterative Constrained K-Means Algorithm: Given the initial cluster center at iteration t that is $C_{1,t}, C_{2,t}, ..., C_{K,t}$, then the cluster assignment and update step are as follows:
(1) *Cluster Assignment:* For i^{th} agent with state x_i, assign the point x_i to any cluster j by minimizing the following function while ensuring center $C_{j,t}$ is nearest to the position x_i and the selection variable $q_{i,j} \geq p_j$.

$$\min_{C,T} \sum_{i=1}^{M} \sum_{j=1}^{K} q_{i,j} (\frac{1}{2} \|x_i - C_{j,t}\|_2^2) \tag{34}$$
$$\text{s.t.} \sum_{i=1}^{M} q_{i,j} \geq p_j, \sum_{j=1}^{K} q_{i,j} = 1, q_{i,j} \geq 0$$

(2) *Cluster Update:* Update $C_{j,t+1}$ as

$$C_{j,t+1} = \begin{cases} \frac{\sum_{i=1}^{M} q_{i,j}^t x_i}{\sum_{i=1}^{M} q_{i,j}^t} & \text{if } \sum_{i=1}^{M} q_{i,j}^t > 0 \\ C_{j,t} & \text{otherwise} \end{cases} \tag{35}$$

The algorithm will be terminated if the condition $C_{j,t+1} = C_{j,t}$ is satisfied.

B. Multi-ACM Based Neural Network Estimator

In this section, the Multi-ACM algorithm is developed. To achieve the final goal, each agent maintains three neural networks (NN), i.e. the *actor NN* approximates the optimal control policy, the *critic NN* approximates the optimal evaluation function and the *mass NN* estimates the density of the entire population. The optimal cost, control, and mass function can be represented as:

$$\begin{aligned} \text{Critic: } J(x, m_j) &= \mathbb{E}\{W_J^T \phi_J(x, m_j) + \varepsilon_{\text{HJB}}\} \\ \text{Actor: } u(x, m_j) &= \mathbb{E}\{W_u^T \phi_u(x, m_j) + \varepsilon_{\text{u}}\} \\ \text{Mass: } m_j(x, t) &= \mathbb{E}\{W_{m_j}^T \phi_{m_j}(x, J, t) + \varepsilon_{\text{FPK}}\} \end{aligned} \tag{36}$$

where, W_J, W_u, and W_{m_j} are the critic, actor, and mass neural network weights of agent $\mathcal{A}$ in group j, respectively. The activation functions are ϕ_J, ϕ_u, and ϕ_{m_j}. The reconstruction errors of the critic, actor, and mass NN are represented as ε_{HJB}, ε_u and ε_{FPK}, respectively. Next, the approximation of the optimal cost, control, and mass distribution function are:

$$\begin{aligned} \text{Critic: } \hat{J}(x,\hat{m}_i) &= \mathbb{E}\{\hat{W}_J^T \hat{\phi}_J(x,\hat{m}_j)\} \\ \text{Actor: } \hat{u}(x,\hat{m}_j) &= \mathbb{E}\{\hat{W}_u^T \hat{\phi}_u(x,\hat{m}_j)\} \\ \text{Mass: } \hat{m}_j(x,t) &= \mathbb{E}\{\hat{W}_{m_j}^T \hat{\phi}_{m_j}(x,\hat{J},t)\} \end{aligned} \tag{37}$$

By substituting equations (37) into the HJB, FPK, and optimal control equations (16), (17) and (7), the residuals errors can be used to tune the critic, actor, and mass NNs:

$$\mathbb{E}\{e_{\text{HJB}}\} = \mathbb{E}\left\{ \begin{bmatrix} \Phi(x,\tilde{m}_j) - \tilde{W}_J^T \hat{\Psi}_J(x,\hat{m}_j) \\ -W_J^T \tilde{\Psi}_J(x,\tilde{m}_j) - \varepsilon_{\text{HJB}} \end{bmatrix} \right\} \tag{38}$$

$$\mathbb{E}\{e_{\text{FPK}}\} = \mathbb{E}\left\{ \begin{bmatrix} -\tilde{W}_{m_j}^T \hat{\Psi}_{m_j}(x,\hat{J},t) - W_{m_j}^T \\ \tilde{\Psi}_{m_j}(x,\tilde{J},t) - \varepsilon_{\text{FPK}} \end{bmatrix} \right\} \tag{39}$$

Similarly, actor residual error is obtained as follows

$$\mathbb{E}\{e_{\text{u}}\} = \mathbb{E}\left\{ \begin{bmatrix} -\tilde{W}_u^T \hat{\phi}_u(x,\hat{m}_j) - W_u^T \tilde{\phi}_u(x,\tilde{m}_j) \\ -\frac{1}{2} R^{-1} g^T(x) \partial_x \tilde{J}(x,\tilde{m}_j) - \varepsilon_{\text{u}} \end{bmatrix} \right\} \tag{40}$$

where, $\tilde{W}_J = W_J - \hat{W}_J$, $\tilde{W}_u = W_u - \hat{W}_u$ and $\tilde{W}_{m_j} = W_{m_j} - \hat{W}_{m_j}$. Next, applying the gradient descent algorithm, the critic, mass, and actor update law is as follows:

$$\mathbb{E}\{\dot{\hat{W}}_J\} = \mathbb{E}\{-\alpha_J \frac{\Psi_J(x,\hat{m}_j) e_{\text{HJB}}^T}{1 + \|\Psi_J(x,\hat{m}_j)\|^2}\} \tag{41}$$

$$\mathbb{E}\{\dot{\hat{W}}_{m_j}\} = \mathbb{E}\{-\alpha_{m_j} \frac{\Psi_{m_j}(x,\hat{J},t) e_{\text{FPK}}^T}{1 + \|\Psi_{m_j}(x,\hat{J},t)\|^2}\} \tag{42}$$

$$\mathbb{E}\{\dot{\hat{W}}_u\} = \mathbb{E}\{-\alpha_u \frac{\phi_u(x,\hat{m}_j) e_{\text{u}}^T}{1 + \|\phi_u(x,\hat{m}_j)\|^2}\} \tag{43}$$

where α_J, α_{m_j} and α_u are the learning rates.

Lemma 1: There exists optimal control policy u for the stochastic system dynamic equation given in (4)

$$\mathbb{E}\{e^T[f_a(e(t)) + g_a(e(t))u(t) + \frac{\sigma d\omega}{dt}]\} \leq -\gamma \mathbb{E}\{\|e\|^2\} \tag{44}$$

Theorem 2: The critic, mass, and actor NNs' weights are updated by (41)-(43), with the learning rates α_J, α_{m_j} and α_u are positive. Then, $\mathbb{E}\{\tilde{W}_J\}$, $\mathbb{E}\{\tilde{W}_{m_j}\}$, $\mathbb{E}\{\tilde{W}_u\}$ and $\mathbb{E}\{e\}$ are all UUB. Moreover, $\mathbb{E}\{\tilde{W}_J\}$, $\mathbb{E}\{\tilde{W}_{m_j}\}$, $\mathbb{E}\{\tilde{W}_u\}$ and $\mathbb{E}\{e\}$ are asymptotically stable with zero reconstruction error [8].

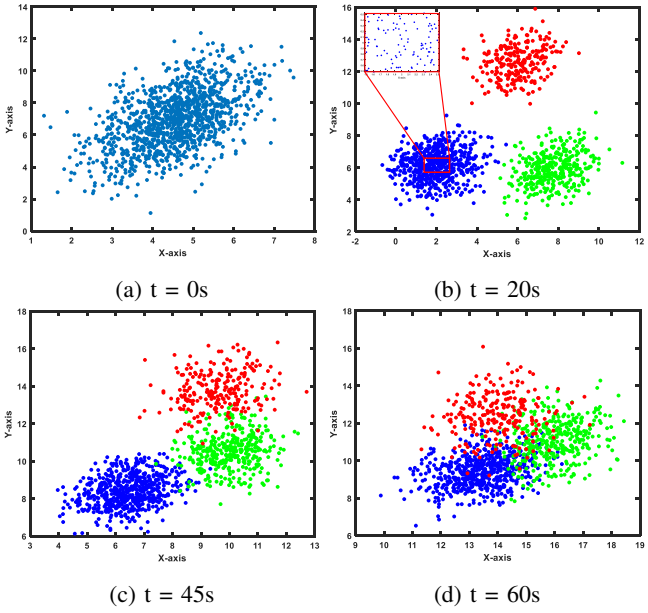
Proof: Omitted due to page limitation.

IV. SIMULATION RESULTS

In this section, the LS-UAV system is employed to show the efficiency of the Imb-MFG theory and multi-ACM algorithm. Initially, the system has been populated with a total of 1200 UAVs. The primary goal for each agent within this system is to collaboratively attain a final mixture distribution constraint. This kind of shape formation in the context of LS-MAS can prove to be crucial, especially in battlefield scenarios, as it enables efficient capture of evaded UAVs. Let the initial states of the agents be generated using the normal distribution: $\mathcal{N}(\mu = [4.5 \quad 7], \Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 3 \end{bmatrix})$. Also, the dynamic of each agent:

$$f(x) = \begin{bmatrix} -x_1 + \frac{1}{2}x_2^2 \\ -0.4x_2^2 \end{bmatrix}, \quad g(x) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

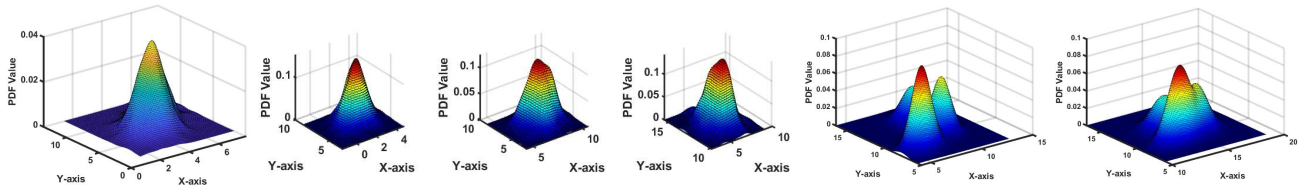
with $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^T$ is the state. Here, each agent doesn't possess knowledge of the desired mixture distribution before embarking on the mission. Here, the final distribution is a Gaussian mixture-PDF defined in (2) as $m_d(x;\theta_d) = \sum_{j=1}^{N} w_{d,j} m_{d,j}(x;\theta_{d,j})$. Here N is the number of Gaussian components and $m_{d,j}(x;\theta_{d,j}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_{d,j}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(x - \mu_{d,j})^T \Sigma_{d,j}^{-1}(x - \mu_{d,j}))$. The tuning gain is defined as $\alpha_w = 1 \times 10^{-3}$, $\alpha_\mu = 1.7 \times \times 10^{-4}$ and $\alpha_\Sigma = 1 \times 10^{-4}$. Then, the PDF function estimation error threshold is given as $\delta_{e_m} = 1 \times 10^{-5}$. A total of 305 iterations have been performed. Then, the estimated parameters of the desired mixture-PDF are obtained. The weights of the mixture are $\hat{w}_{d,1} = 0.495$, $\hat{w}_{d,2} = 0.3025$, $\hat{w}_{d,3} = 0.2025$, with the cluster number $N = 3$. Also, the mean and covariance of the mixture PDF are estimated as $\hat{\mu}_1 = \begin{bmatrix} 13.4256 & 9.5846 \end{bmatrix}^T$, $\hat{\mu}_2 = \begin{bmatrix} 15.9451 & 10.8812 \end{bmatrix}^T$, $\hat{\mu}_3 = \begin{bmatrix} 14.0557 & 12.4397 \end{bmatrix}^T$, $\hat{\Sigma}_1 = \begin{bmatrix} 1.1956 & 0.334 \\ 0.334 & 0.7661 \end{bmatrix}$, $\hat{\Sigma}_2 = \begin{bmatrix} 0.7845 & 0.1826 \\ 0.1826 & 1.3426 \end{bmatrix}$ and $\hat{\Sigma}_3 = \begin{bmatrix} 0.9882 & 0.2105 \\ 0.2105 & 1.2259 \end{bmatrix}$. Then a constrained K-means clustering algorithm is employed to decompose the LS-MAS. To determine the minimum number of agents for 3 clusters, we utilize the estimated weight parameters, resulting in mini-



(a) t = 0s (b) t = 20s

(c) t = 45s (d) t = 60s

Fig. 2: Large-scale UAVs positions over a period of time (a) $t = 0s$ (b) $t = 20s$ (C) $t = 45s$ (d) $t = 60s$ Different colors represent different groups.



(a) PDF $t = 0s$ (b) Groups PDF $t = 20s$ (c) PDF, $t = 45s$ (d) PDF, $t = 60s$

Fig. 3: PDF decomposition of large-scale UAVs in 3-D view. (a) The initial PDF distribution at $t = 0s$. (b) The decomposed PDFs (c) The mixture PDF at time $t = 45s$. (d) The final mixture-PDF of UAVs at time $t = 60s$.

mum numbers of agents as follows: $p_1 = 596$, $p_2 = 363$, and $p_3 = 241$. Next, we assign agents to clusters using equations (34) and (35). The initial distributions for the decomposed groups, which are provided to their respective group of agents, are as follows: Group 1: $\mathcal{N}(\mu = [2 \quad 6.2], \Sigma = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix})$, Group 2: $\mathcal{N}(\mu = [7.8 \quad 6], \Sigma = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix})$, Group 3: $\mathcal{N}(\mu = [6 \quad 12.5], \Sigma = \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix})$. The learning rates of the Multi-ACM NNs are selected as $\alpha_J = 2 \times 10^{-5}$, $\alpha_u = 2 \times 10^{-4}$, $\alpha_{m_j} = 2 \times 10^{-3}$. Also, the thresholds of the HJB, FPK and actor residual error are selected as $\delta_{\text{HJB}} = 1 \times 10^{-5}$, $\delta_{\text{FPK}} = 1 \times 10^{-3}$ and $\delta_u = 1 \times 10^{-2}$. Figure 2 shows the positions of UAVs evolving over time. The initial position of the 1200 UAVs is depicted in figure 2(a). Then, the position of UAVs is demonstrated after the decomposition in figure 2(b). Figure 2(c) shows the position of the UAVs

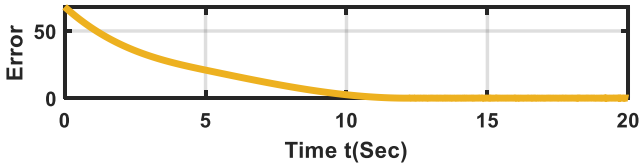


Fig. 4: Final PDF estimation error

at time $t = 45s$. Finally, at the end of the simulation at $t = 60s$, the UAVs in each group successfully reach a position that satisfies the ε-Nash equilibrium by solving the HJB and FPK equations. This allows them to achieve the intended final arbitrary PDF. Also, we have plotted the PDF of agents at various time intervals in Figure 3. In Figure 3(a), the initial PDF is presented. The resulting PDFs of the decomposed groups are displayed in Figure 3(b). Then figure 3(c) shows the PDF of all UAVs at time $t = 45s$. The LS-UAVs achieve the desired final mixture-PDF constraint with time progress. The final PDF of the UAVs is shown in 3(d). Now, the final mixture PDF percentage estimation error is shown in figure 4. The error is calculated using equation, $\text{Error}_1 = \frac{m_d(x;\theta_d) - \hat{m}_d(x;\theta_d)}{m_d(x;\theta_d)} \times 100$. From this figure, it is clear that the error of the PDF function approximation converges to zero after a certain time period.

V. CONCLUSION

This study presented a novel distributed optimization algorithm for LS-MAS with a fixed final PDF constraint in uncertain environments. This algorithm incorporated an MFG theory to address the computational and communication complexities associated with LS-MAS. It also tackles the limitations of MFG theory, which sacrifices optimality and struggles to achieve an arbitrary final PDF constraint. The developed algorithm includes a novel Imb-MFG theory along with PDF decomposition and distributed reinforcement learning. Particularly, an induction-based PDF parameter estimation is designed and a constrained K-means clustering algorithm is applied to decompose the LS-MAS into multiple groups, aiming to achieve the desired final arbitrary PDF constraint. Moreover, a Multi-ACM learning algorithm is designed to solve the Imb-MFG and find the optimal solution.

REFERENCES

[1] A. Dorri, S. S. Kanhere, and R. Jurdak, "Multi-agent systems: A survey," *Ieee Access*, vol. 6, pp. 28 573–28 593, 2018.
[2] M. Balmer, K. Nagel, and B. Raney, "Large-scale multi-agent simulations for transportation applications," in *Intelligent Transportation Systems*, vol. 8, no. 4. Taylor & Francis, 2004, pp. 205–221.
[3] W. Wang, L. Wang, J. Wu, X. Tao, and H. Wu, "Oracle-guided deep reinforcement learning for large-scale multi-uavs flocking and navigation," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 10 280–10 292, 2022.
[4] S. Parsons and M. Wooldridge, "Game theory and decision theory in multi-agent systems," *Autonomous Agents and Multi-Agent Systems*, vol. 5, pp. 243–254, 2002.
[5] F.-L. Lian, J. Moyne, and D. Tilbury, "Network design consideration for distributed control systems," *IEEE transactions on control systems technology*, vol. 10, no. 2, pp. 297–307, 2002.
[6] S. Dey and H. Xu, "Distributed adaptive flocking control for large-scale multiagent systems," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2024.
[7] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese journal of mathematics*, vol. 2, no. 1, pp. 229–260, 2007.
[8] S. Dey and H. Xu, "Hierarchical game theoretical distributed adaptive control for large scale multi-group multi-agent system," *IET Control Theory & Applications*, 2023.
[9] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996.
[10] P. S. Bradley, K. P. Bennett, and A. Demiriz, "Constrained k-means clustering," *Microsoft Research, Redmond*, vol. 20, no. 0, p. 0, 2000.
[11] Z. Zhou and H. Xu, "Large-scale multiagent system tracking control using mean field games," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5602–5610, 2021.
[12] F.-Y. Wang, H. Zhang, and D. Liu, "Adaptive dynamic programming: An introduction," *IEEE computational intelligence magazine*, vol. 4, no. 2, pp. 39–47, 2009.
[13] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical distributions*. John Wiley & Sons, 2011.
[14] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.
[15] A. Baker, "Mathematical induction and explanation," *Analysis*, vol. 70, no. 4, pp. 681–689, 2010.