U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND – ARMY RESEARCH LABORATORY

Large-Scale, Multi-Agent, Reinforcement Learning Control

Jemin George, Ph.D.

Computational & Information Sciences Directorate (CISD)

Army Research Laboratory (ARL)

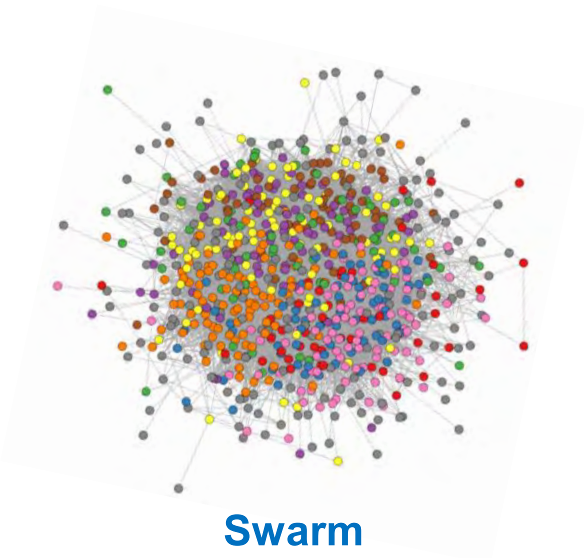
DISTRIBUTION STATEMENT A

09 June 2021

Outline

- Overview
- Hierarchical Reinforcement Learning (HRL) Control
  - RL Control (backgound)
  - Problem Formulation
  - Proposed HRL Solution
    - HRL for approximate control of heterogeneous swarm
    - HRL for optimal control of homogeneous swarm
  - Example: Formation Control
- Swarm Decomposition
  - Decomposition Objectives
  - Example: Formation Maneuver
- AirSim Experiments
- Conclusions

Reinforcement Learning (RL) Control of Swarms: Overview



Swarm

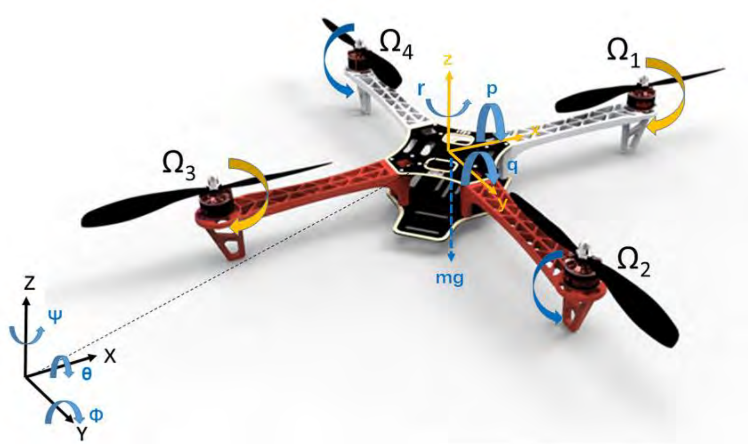
$$J = \int_0^{\infty} \Phi(\boldsymbol{x}(t), \boldsymbol{u}(t))\, dt$$

State — Input — Time

Objective

$$\boldsymbol{u}^{\star}(t) = \min_{\boldsymbol{u}(t)} \; J$$

Problem

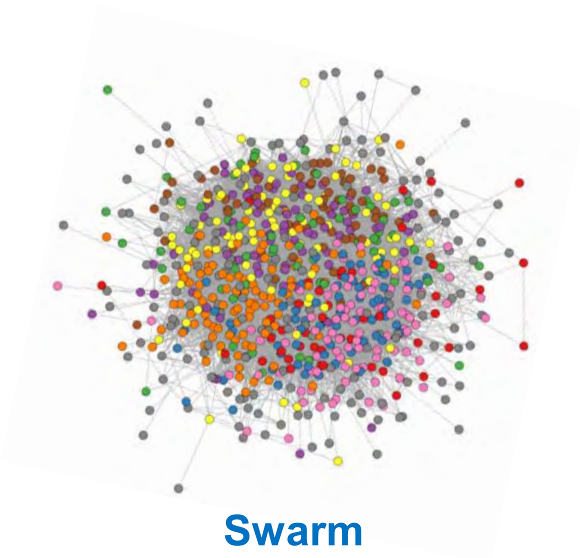


$$x = \begin{bmatrix} X \\ Y \\ Z \\ \theta \\ \phi \\ \psi \\ \dot{X} \\ \dot{Y} \\ \dot{Z} \\ p \\ q \\ r \end{bmatrix} \qquad u = \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \\ \Omega_4 \end{bmatrix}$$

System Dynamics

$$\frac{d}{dt} x(t) = f(x(t), u(t))$$

Reinforcement Learning (RL) Control of Swarms: Overview



Swarm

State

Input

$$J = \int_0^\infty \Phi(\boldsymbol{x}(t), \boldsymbol{u}(t))\, dt$$

Time

Objective

$$\boldsymbol{u}^\star(t) = \min_{\boldsymbol{u}(t)} J$$

Problem

Linear System Dynamics

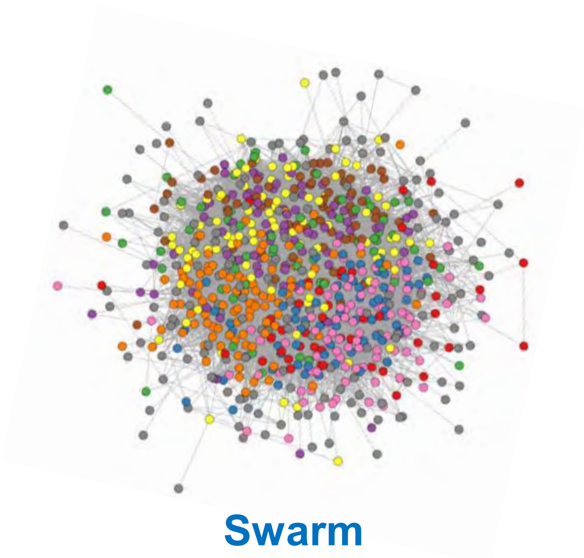
$$\dot{x}(t) = Ax(t) + Bu(t)$$

Quadratic Objective

$$\Phi(x, u) = x^\top Q x + u^\top R u$$

Linear Quadratic Regulator

Reinforcement Learning (RL) Control of Swarms: Overview



Swarm

$$J = \int_0^\infty \Phi(\boldsymbol{x}(t), \boldsymbol{u}(t))\, dt$$

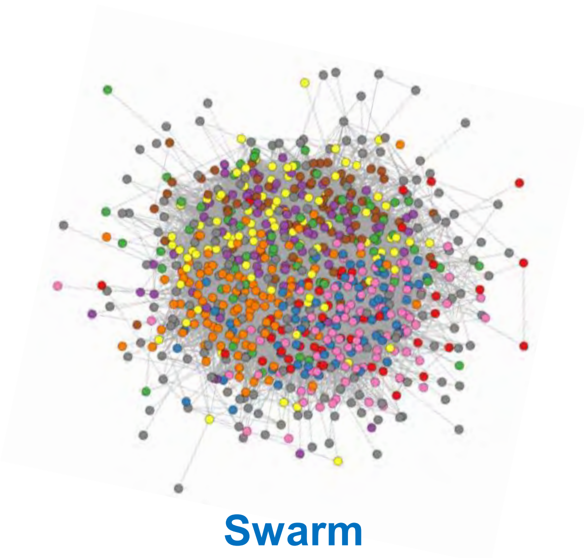
Objective

$$\boldsymbol{u}^\star(t) = \min_{\boldsymbol{u}(t)} \; J$$

Problem

- ☐ Why is it difficult?
  - Uncertainty
  - Size
- ☐ Reinforcement Learning Control $\Longleftrightarrow$ Adaptive Optimal Control
  - Adaptive: unknown/uncertain dynamics and environment
  - Optimal: $\min_{\boldsymbol{u}(t)} \; J$

Reinforcement Learning (RL) Control of Swarms: Overview



Swarm

$$J = \int_0^\infty \Phi(\boldsymbol{x}(t), \boldsymbol{u}(t))\, dt$$

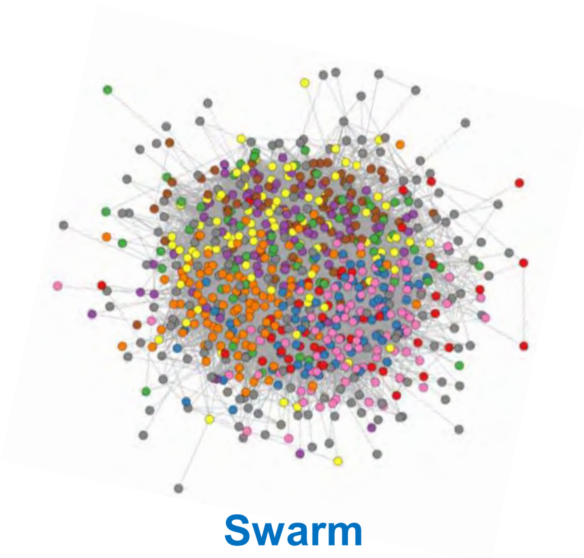
Objective

$$\boldsymbol{u}^\star(t) = \min_{\boldsymbol{u}(t)} \; J$$

Problem

- Why is it difficult?
  - Uncertainty
  - Size
- Reinforcement Learning Control $\Longleftrightarrow$ Adaptive Optimal Control
  - Adaptive: unknown/uncertain dynamics and environment
  - Optimal: $\min_{\boldsymbol{u}(t)} \; J$

Reinforcement Learning (RL) Control of Swarms: Overview



Swarm

$$J = \int_0^\infty \Phi(\boldsymbol{x}(t), \boldsymbol{u}(t))\, dt$$

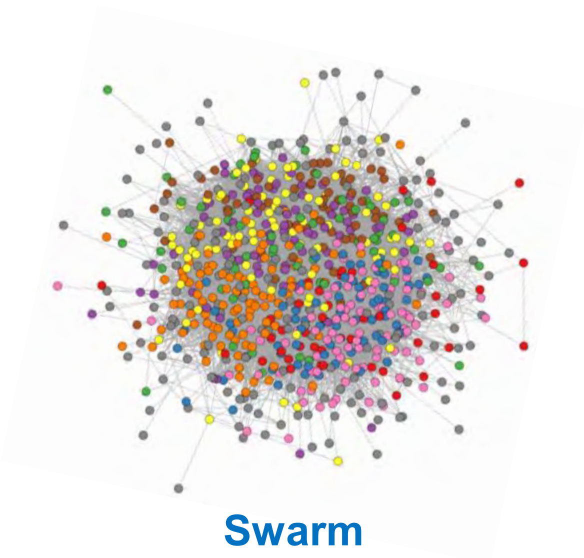
Objective

$$\boldsymbol{u}^\star(t) = \min_{\boldsymbol{u}(t)} \; J$$

Problem

- ☐ Why is it difficult?
  - Uncertainty
  - Size ← **This Talk**
- ☐ Reinforcement Learning Control $\Longleftrightarrow$ Adaptive Optimal Control
  - Adaptive: unknown/uncertain dynamics and environment
  - Optimal: $\min_{\boldsymbol{u}(t)} \; J$

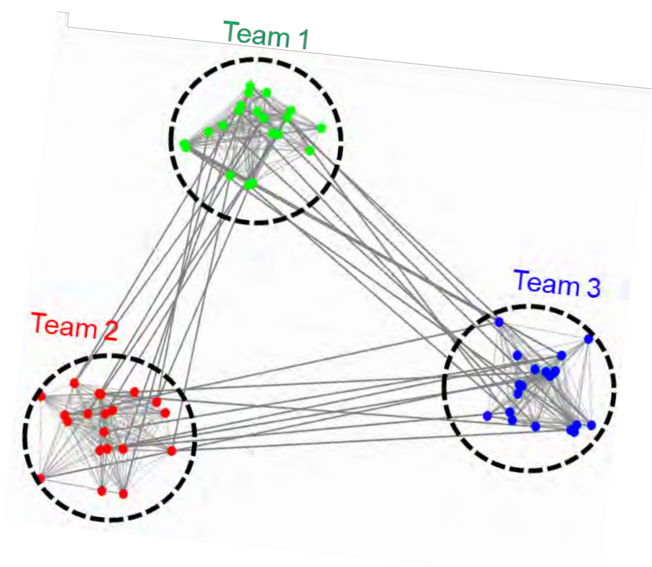
Reinforcement Learning (RL) Control of Swarms: Overview



$$J = \int_0^\infty \Phi(\boldsymbol{x}(t), \boldsymbol{u}(t))\, dt$$

$$\boldsymbol{u}^\star(t) = \min_{\boldsymbol{u}(t)} J$$

Swarm **Objective** **Problem**



$$J = \sum_{j=1}^{N} J_j + J_g$$

$$\boldsymbol{u}_1^\star(t) + \boldsymbol{u}_{g_1}(t)$$
$$\vdots$$
$$\boldsymbol{u}_N^\star(t) + \boldsymbol{u}_{g_N}(t)$$

Outline

- ☐ Overview
- ☐ Hierarchical Reinforcement Learning (HRL) Control
  - RL Control (backgound)
  - Problem Formulation
  - Proposed HRL Solution
    - HRL for approximate control of heterogeneous swarm
    - HRL for optimal control of homogeneous swarm
  - Example: Formation Control
- ☐ Swarm Decomposition
  - Decomposition Objectives
  - Example: Formation Maneuver
- ☐ AirSim Experiments
- ☐ Conclusions

Reinforcement Learning (RL) based Optimal Control

Optimal Control Problem (LQR)

- System: $\dot{x} = Ax + Bu$
- Cost functional: $J = \int_0^\infty (x^T Q x + u^T R u)\, dt$
- Control law that minimizes the value of the cost: $u = -Kx$
  - ⋆ $K = R^{-1} B^T P$
  - ⋆ $A^T P + PA - PBR^{-1} B^T P + Q = 0$

Linear state feedback controller

Algebraic matrix Riccati equation

Reinforcement Learning (RL) based Optimal Control

Optimal Control Problem

- System: $\dot{x} = Ax + Bu$
- Cost functional: $J = \int_0^\infty (x^T Q x + u^T R u)\, dt$
- Control law that minimizes the value of the cost: $u = -Kx$
  - ⋆ $K = R^{-1} B^T P$
  - ⋆ $A^T P + PA - PBR^{-1}B^T P + Q = 0$

RL learns *K* by solving the Riccati equation using only *x*(*t*) and *u*(*t*), no model

Unknown system dynamics
A & B are unknown

→

System response to control input is unknown

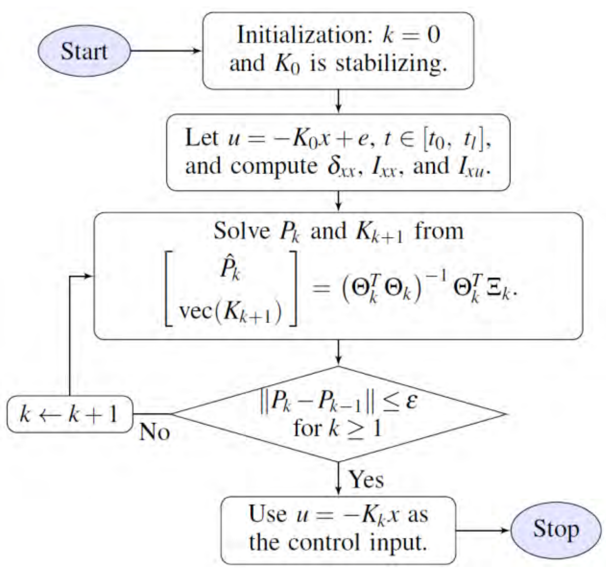
Reinforcement Learning (RL) based Optimal Control

Optimal Control Problem

- System: $\dot{x} = Ax + Bu$
- Cost functional: $J = \int_0^\infty (x^T Q x + u^T R u)\, dt$
- Control law that minimizes the value of the cost: $u = -Kx$
  - ⋆ $K = R^{-1} B^T P$
  - ⋆ $A^T P + PA - PBR^{-1} B^T P + Q = 0$

RL learns K by solving the Riccati equation using only $x(t)$ and $u(t)$, no model

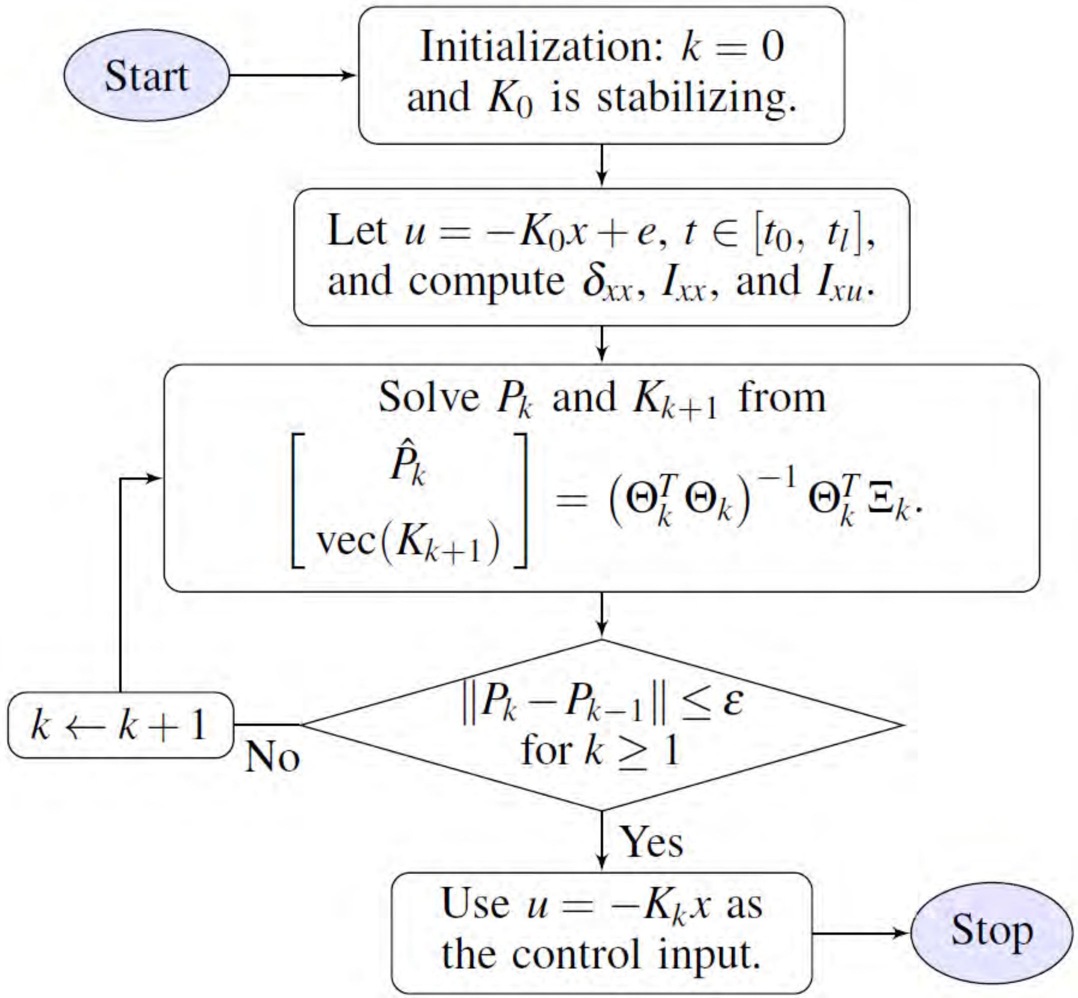
Adaptive Dynamic Programing (A and B are unknown)



Jiang, Y. and Jiang, Z.P., 2012. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. Automatica, 48(10), pp.2699-2704.

Reinforcement Learning (RL) based Optimal Control

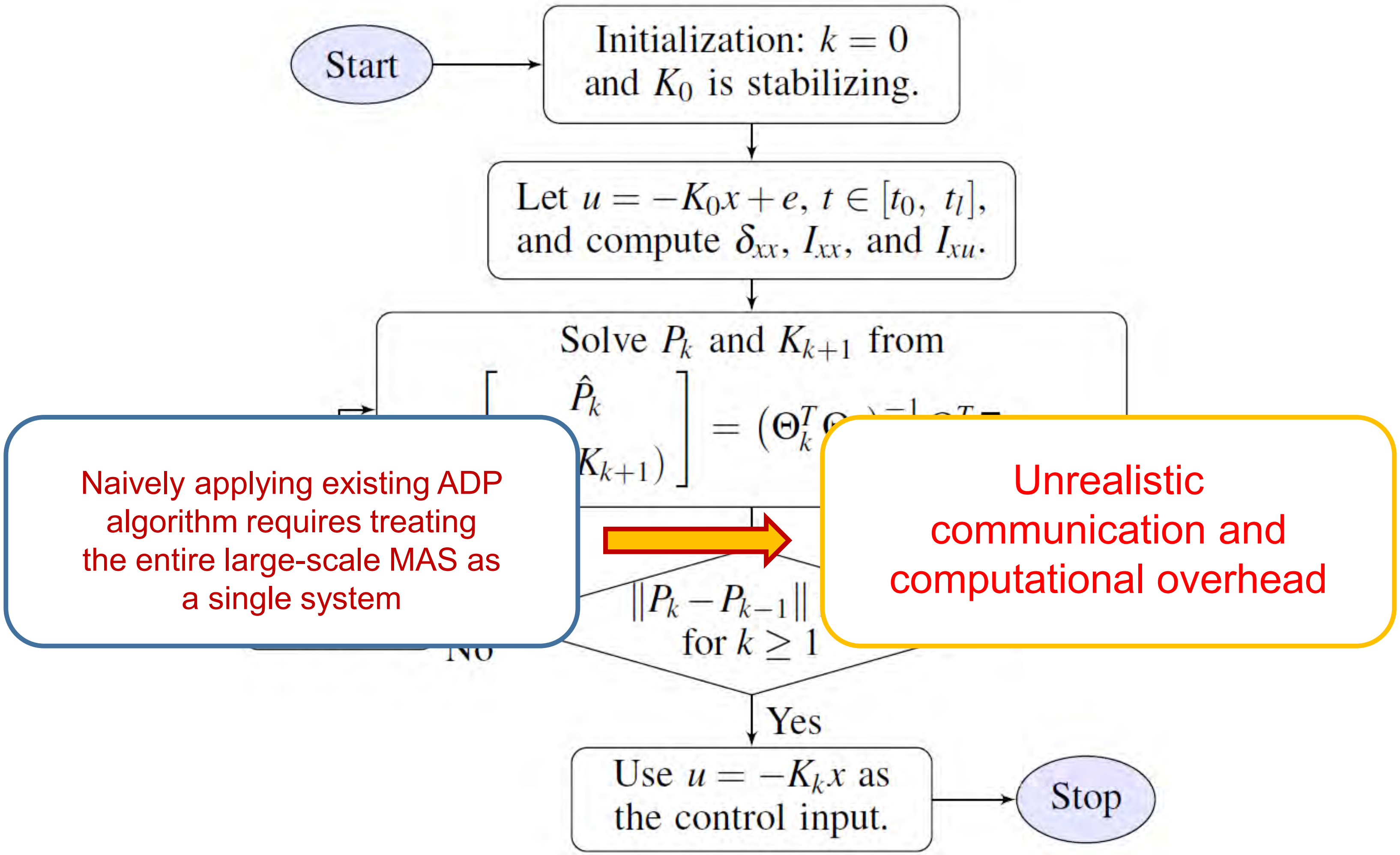
Adaptive Dynamic Programing (A and B are unknown)



Jiang, Y. and Jiang, Z.P., 2012. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. Automatica, 48(10), pp.2699-2704.

Reinforcement Learning (RL) based Optimal Control

Adaptive Dynamic Programing (A and B are unknown)



Jiang, Y. and Jiang, Z.P., 2012. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. Automatica, 48(10), pp.2699-2704.

Outline

- ☐ Overview
- ☐ Hierarchical Reinforcement Learning (HRL) Control
  - RL Control (backgound)
  - Problem Formulation
  - Proposed HRL Solution
    - HRL for approximate control of heterogeneous swarm
    - HRL for optimal control of homogeneous swarm
  - Example: Formation Control
- ☐ Swarm Decomposition
  - Decomposition Objectives
  - Example: Formation Maneuver
- ☐ AirSim Experiments
- ☐ Conclusions

Hierarchical RL for Multi-Agent Systems: Formulation

- Swarm of p agents consisting of N groups: $p = \sum_{j=1}^{N} p_j$
- Group-level dynamics:

$$\dot{\mathbf{x}}_j = A_j \mathbf{x}_j + B_j \mathbf{u}_j$$

- Swarm model:

$$\dot{\mathbf{x}} = \mathcal{A}\mathbf{x} + \mathcal{B}\mathbf{u}$$

- Control objective:

$$J = \int_0^\infty \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \, dt$$

- Optimal controller:

$$\mathbf{u} = -K^* \mathbf{x} = -R^{-1} \mathcal{B}^\top P^* \mathbf{x}$$

- Riccati equation

$$P^* \mathcal{A} + \mathcal{A}^\top P^* + Q - P^* \mathcal{B} R^{-1} \mathcal{B}^\top P^* = 0$$

Can’t solve since model is unknown!

Naively applying existing ADP algorithm requires treating the entire large-scale MAS as a single system

Hierarchical RL for Multi-Agent Systems: Formulation

- Swarm of p agents consisting of N groups: $p = \sum_{j=1}^{N} p_j$

- Group-level dynamics:

$$\dot{\mathbf{x}}_j = A_j \mathbf{x}_j + B_j \mathbf{u}_j$$

- Swarm model:

$$\dot{\mathbf{x}} = \mathcal{A}\mathbf{x} + \mathcal{B}\mathbf{u}$$

No physical (dynamical) coupling

$$\mathcal{A} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & A_N \end{bmatrix}$$

- Control objective:

$$J = \int_0^\infty \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \, dt$$

- Optimal controller:

$$\mathbf{u} = -K^* \mathbf{x} = -R^{-1} \mathcal{B}^\top P^* \mathbf{x}$$

- Riccati equation

$$P^* \mathcal{A} + \mathcal{A}^\top P^* + Q - P^* \mathcal{B} R^{-1} \mathcal{B}^\top P^* = 0$$

Hierarchical RL for Multi-Agent Systems: Formulation

- Swarm of p agents consisting of N groups: $p = \sum_{j=1}^{N} p_j$
- Group-level dynamics:

$$\dot{\mathbf{x}}_j = A_j \mathbf{x}_j + B_j \mathbf{u}_j$$

- Swarm model:

$$\dot{\mathbf{x}} = \mathcal{A}\mathbf{x} + \mathcal{B}\mathbf{u}$$

- Control objective:

$$J = \int_0^{\infty} \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \, dt$$

- Optimal controller:

$$\mathbf{u} = -K^* \mathbf{x} = -R^{-1} \mathcal{B}^\top P^* \mathbf{x}$$

- Riccati equation

$$P^* \mathcal{A} + \mathcal{A}^\top P^* + Q - P^* \mathcal{B} R^{-1} \mathcal{B}^\top P$$

Coupled control objective

$R = \mathrm{diag}\{R_1, ..., R_N\} \succ 0$

$Q = \bar{Q} + L_w \odot \tilde{Q}$

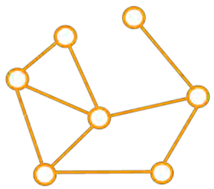
$$\bar{Q} = \begin{pmatrix} Q_{11} & & \\ & \ddots & \\ & & Q_{NN} \end{pmatrix}$$

$Q_{ii} \in \mathbb{R}^{n \times n}$ and $Q_{ii} \succeq 0$

$$\tilde{Q} = \begin{pmatrix} \tilde{Q}_{11} & \cdots & \tilde{Q}_{1N} \\ \vdots & \ddots & \vdots \\ \tilde{Q}_{N1} & \cdots & \tilde{Q}_{NN} \end{pmatrix}$$

$\tilde{Q}_{ij} \in \mathbb{R}^{n \times n}$ and $\tilde{Q}_{ij} \succeq 0$

L_w is the Laplacian matrix between the groups
Within a group, assume a fully connected network



A coupling graph is involved in the cost function

Hierarchical RL for Multi-Agent Systems: Formulation

$$J = \int_0^\infty \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \, dt$$

$$R = \mathrm{diag}\{R_1, ..., R_N\} \succ 0$$

$$Q = \bar{Q} + L_w \odot \tilde{Q}$$

$$\bar{Q} = \begin{pmatrix} Q_{11} & & \\ & \ddots & \\ & & Q_{NN} \end{pmatrix}$$

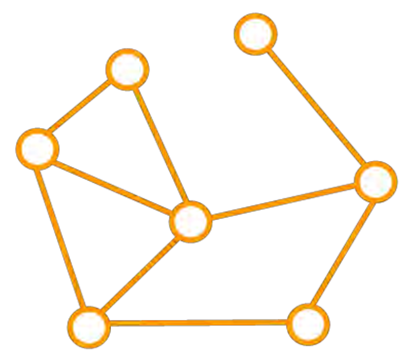
$Q_{ii} \in \mathbb{R}^{n \times n}$ and $Q_{ii} \succeq 0$

$$\tilde{Q} = \begin{pmatrix} \tilde{Q}_{11} & \cdots & \tilde{Q}_{1N} \\ \vdots & \ddots & \vdots \\ \tilde{Q}_{N1} & \cdots & \tilde{Q}_{NN} \end{pmatrix}$$

$\tilde{Q}_{ij} \in \mathbb{R}^{n \times n}$ and $\tilde{Q}_{ij} \succeq 0$

L_w is the Laplacian matrix between the groups

Within a group, assume a fully connected network



A coupling graph is involved in the cost function

Outline

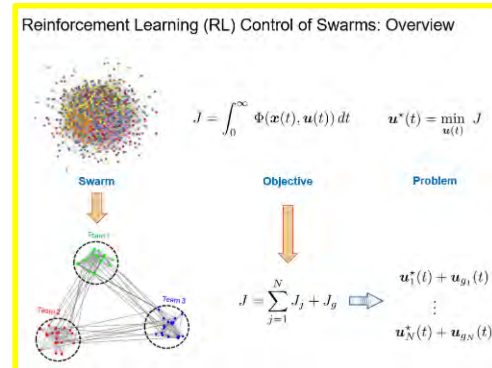
- ☐ Overview
- ☐ Hierarchical Reinforcement Learning (HRL) Control
  - ○ RL Control (backgound)
  - ○ Problem Formulation
  - ○ Proposed HRL Solution
    - HRL for approximate control of heterogeneous swarm
    - HRL for optimal control of homogeneous swarm
  - ○ Example: Formation Control
- ☐ Swarm Decomposition
  - ○ Decomposition Objectives
  - ○ Example: Formation Maneuver
- ☐ AirSim Experiments
- ☐ Conclusions

Approximate Control for Multi-Agent Systems

- Control objective:

$$J = \int_0^\infty \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \; dt = \underbrace{\int_0^\infty \mathbf{x}^\top \bar{Q} \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \; dt}_{\sum_{j=1}^N \; J_j} + \underbrace{\int_0^\infty \mathbf{x}^\top \left(L_w \odot \tilde{Q} \right) \mathbf{x} \; dt}_{J_g}$$

$$J_j = \int_0^\infty \mathbf{x}_j^\top \bar{Q}_j \mathbf{x}_j + \mathbf{u}_j^\top R_j \mathbf{u}_j \; dt$$



- $\mathbf{u}_j = -\underbrace{R_j^{-1} B_j^\top P_j}_{K_j} \, \mathbf{x}_j$, where $P_j \in \mathbb{R}^{np_j \times np_j}$ are from

$$\underbrace{P_j A_j + A_j^\top P_j + \bar{Q}_j - P_j B_j R_j^{-1} B_j^\top P_j = 0}_{J_j = \int_0^\infty \mathbf{x}_j^\top \bar{Q}_j \mathbf{x}_j + \mathbf{u}_j^\top R_j \mathbf{u}_j \; dt}$$

Individual agents or teams can solve for local optimal controllers in parallel using existing ADP algorithms

Approximate Control for Multi-Agent Systems

- Control objective:

$$J = \int_0^\infty \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \; dt = \underbrace{\int_0^\infty \mathbf{x}^\top \bar{Q} \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \; dt}_{\sum_{j=1}^N \; J_j} + \underbrace{\int_0^\infty \mathbf{x}^\top \left(L_w \odot \tilde{Q} \right) \mathbf{x} \; dt}_{J_g}$$

$$J_j = \int_0^\infty \mathbf{x}_j^\top \bar{Q}_j \mathbf{x}_j + \mathbf{u}_j^\top R_j \mathbf{u}_j \; dt$$

- $\mathbf{u}_j = -\underbrace{R_j^{-1} B_j^\top P_j}_{K_j} \mathbf{x}_j$, where $P_j \in \mathbb{R}^{np_j \times np_j}$ are from

$$P_j A_j + A_j^\top P_j + \bar{Q}_j - P_j B_j R_j^{-1} B_j^\top P_j = 0$$

- Define $\mathcal{R}^{-1} = R^{-1} + \tilde{R}$
- Consider a new Ricatti equation

$$\mathcal{P}\mathcal{A} + \mathcal{A}^\top \mathcal{P} + Q - \mathcal{P}\mathcal{B}\mathcal{R}^{-1}\mathcal{B}^\top \mathcal{P} = \mathcal{P}\mathcal{A} + \mathcal{A}^\top \mathcal{P} + \bar{Q} - \mathcal{P}\mathcal{B}R^{-1}\mathcal{B}^\top \mathcal{P} + L_w \odot \tilde{Q} - \mathcal{P}\mathcal{B}\tilde{R}\mathcal{B}^\top \mathcal{P}$$

Approximate Control for Multi-Agent Systems

- Control objective:

$$J = \int_0^\infty \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u}\ dt = \underbrace{\int_0^\infty \mathbf{x}^\top \bar{Q} \mathbf{x} + \mathbf{u}^\top R \mathbf{u}\ dt}_{\sum_{j=1}^N J_j} + \underbrace{\int_0^\infty \mathbf{x}^\top \left(L_w \odot \tilde{Q} \right) \mathbf{x}\ dt}_{J_g}$$

$$J_j = \int_0^\infty \mathbf{x}_j^\top \bar{Q}_j \mathbf{x}_j + \mathbf{u}_j^\top R_j \mathbf{u}_j\ dt$$

$$P_j A_j + A_j^\top P_j + \bar{Q}_j - P_j B_j R_j^{-1} B_j^\top P_j = 0$$

- Define $\mathcal{R}^{-1} = R^{-1} + \tilde{R}$

$\tilde{R}$ is selected so that $\mathcal{P}\mathcal{B}\tilde{R}\mathcal{B}^\top \mathcal{P} = L_w \odot \tilde{Q}$

- Consider a new Ricatti equation

$$\mathcal{P}\mathcal{A} + \mathcal{A}^\top \mathcal{P} + Q - \mathcal{P}\mathcal{B}\mathcal{R}^{-1}\mathcal{B}^\top \mathcal{P} = \mathcal{P}\mathcal{A} + \mathcal{A}^\top \mathcal{P} + \bar{Q} - \mathcal{P}\mathcal{B}R^{-1}\mathcal{B}^\top \mathcal{P} + L_w \odot \tilde{Q} - \mathcal{P}\mathcal{B}\tilde{R}\mathcal{B}^\top \mathcal{P}$$

  - Decoupled Ricatti equation with $\mathcal{P} = \mathrm{diag}\{P_1, \ldots, P_N\}$

$$\mathcal{P}\mathcal{A} + \mathcal{A}^\top \mathcal{P} + \bar{Q} - \mathcal{P}\mathcal{B}R^{-1}\mathcal{B}^\top \mathcal{P} = \mathrm{diag}\{P_j A_j + A_j^\top P_j + \bar{Q}_j - P_j B_j R_j^{-1} B_j^\top P_j\} = 0$$

- Then the control gain follows as: $K = \mathcal{R}^{-1}\mathcal{B}^\top \mathcal{P} = \underbrace{R^{-1}\mathcal{B}^\top \mathcal{P}}_{local} + \underbrace{\tilde{R}\mathcal{B}^\top \mathcal{P}}_{global}$.

Approximate Control for Multi-Agent Systems

- Control objective:

$$J = \int_0^\infty \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \; dt = \underbrace{\int_0^\infty \mathbf{x}^\top \bar{Q} \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \; dt}_{\sum_{j=1}^N J_j} + \underbrace{\int_0^\infty \mathbf{x}^\top \left(L_w \odot \tilde{Q} \right) \mathbf{x} \; dt}_{J_g}$$

$$J_j = \int_0^\infty \mathbf{x}_j^\top \bar{Q}_j \mathbf{x}_j + \mathbf{u}_j^\top R_j \mathbf{u}_j \; dt$$

$$P_j A_j + A_j^\top P_j + \bar{Q}_j - P_j B_j R_j^{-1} B_j^\top P_j = 0$$

- Let $\mathcal{P} = \operatorname{diag}\{P_1, \ldots, P_N\}$
- Define $\mathcal{R}^{-1} = R^{-1} + \tilde{R}$, where $\tilde{R}$ is selected so that $\mathcal{P}\mathcal{B}\tilde{R}\mathcal{B}^\top\mathcal{P} = L_w \odot \tilde{Q}$
- Then the control gain follows as: $K = \mathcal{R}^{-1}\mathcal{B}^\top\mathcal{P} = \underbrace{R^{-1}\mathcal{B}^\top\mathcal{P}}_{local} + \underbrace{\tilde{R}\mathcal{B}^\top\mathcal{P}}_{global}$.
  - What we are effectively minimizing:

$$\mathcal{J} = \int_0^\infty \mathbf{x}^\top Q' \mathbf{x} + \mathbf{u}^\top \mathcal{R} \mathbf{u} \; dt,$$

$$Q' = \bar{Q} + \mathcal{P}\mathcal{B}\tilde{R}\mathcal{B}^\top\mathcal{P}$$
$$\mathcal{R}^{-1} = R^{-1} + \tilde{R}$$

We are relaxing control penalty term to account for coupled state penalty term

Approximate Control for Multi-Agent Systems: Algorithm

- Control objective:

$$J = \int_0^\infty \mathbf{x}^\top Q \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \; dt = \underbrace{\int_0^\infty \mathbf{x}^\top \bar{Q} \mathbf{x} + \mathbf{u}^\top R \mathbf{u} \; dt}_{\sum_{j=1}^{N} J_j} + \underbrace{\int_0^\infty \mathbf{x}^\top \left(L_w \odot \tilde{Q} \right) \mathbf{x} \; dt}_{J_g}$$

Step 1: Solve in parallel using ADP

$$J_j = \int_0^\infty \mathbf{x}_j^\top \bar{Q}_j \mathbf{x}_j + \mathbf{u}_j^\top R_j \mathbf{u}_j \; dt$$

$$P_j A_j + A_j^\top P_j + \bar{Q}_j - P_j B_j R_j^{-1} B_j^\top P_j = 0$$

- Let $\mathcal{P} = \mathrm{diag}\{P_1, \ldots, P_N\}$

Step 2: Construct $\tilde{R}$

- Define $\mathcal{R}^{-1} = R^{-1} + \tilde{R}$, where $\tilde{R}$ is selected so that $\mathcal{P}\mathcal{B}\tilde{R}\mathcal{B}^\top\mathcal{P} = L_w \odot \tilde{Q}$

- Then the control gain follows as: $K = \mathcal{R}^{-1}\mathcal{B}^\top\mathcal{P} = \underbrace{R^{-1}\mathcal{B}^\top\mathcal{P}}_{local} + \underbrace{\tilde{R}\mathcal{B}^\top\mathcal{P}}_{global}$.
  - What we are effectively minimizing:

Step 3: Compute K

$$\mathcal{J} = \int_0^\infty \mathbf{x}^\top Q' \mathbf{x} + \mathbf{u}^\top \mathcal{R} \mathbf{u} \; dt,$$

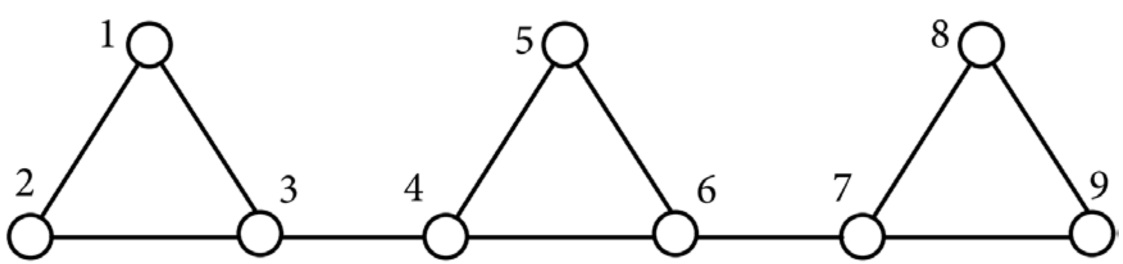
Comparison between Centralized RL and HRL

Heterogeneous Agents

Dimension				Time(sec)		Performance		
N	c	n	m	RL	HRL	OPT	HRL	SOP $(J - J^*)/J^*$
3	2	4	2	0.57	**0.07**	28.76	**40.57**	41.08%
3	3	4	2	7.04	**0.08**	52.75	**62.24**	18.00%
3	4	4	2	29.69	**0.24**	81.03	**87.57**	8.08%
4	4	8	4	> 60	**9.83**	198.89	**209.29**	5.23%

$N = 3$ & $c = 3$

N: # of teams (cliques)
c: # of agents per team
n: size of state-vector
m: size of input-vector



Graph $\mathcal{G}$ with 3 cliques, each clique contains 3 agents.

Bai, H., George, J. and Chakrabortty, A., "*Hierarchical Control of Multi-Agent Systems using Online Reinforcement Learning*," **American Control Conference (ACC)**, Denver, CO, July 2020

Reducing Learning Time via Hierarchical Approximation

Homogeneous Agents

- *Identical* agent dynamics: $\dot{x}_i = Ax_i + Bu_i,\ i = 1, \cdots, N$
- *Identical* performance metrics: $Q = (I_N + G) \otimes Q_0, \quad R = I_N \otimes R_0$
- **Decompose** into solving N smaller-sized LQR problems

$$\min_{v_i} J_i(\xi_i, v_i) = \int_0^\infty (g_i \xi_i^\top Q_0 \xi_i + v_i^\top R_0 v_i) dt$$
$$s.t. \quad \dot{\xi}_i = A\xi_i + Bv_i.$$

  - ⋆ g_i: eigenvalues of $I_N + G$
  - ⋆ v_i^* learned using ADP with a smaller dimension (n)

- Combined optimal control: $u^* = \sum_{i=1}^N (S_i \otimes I_m) v_i^*$.
- "Learn in parallel, implement centrally"

Final controller is optimal !

Comparison between Centralized RL and HRL

Homogeneous Agents

Similarity transformation allows to decouple the problem.

Final controller is optimal !

Comparisons Between Different Algorithms

Dimension			Computational Time (s)		
p	n	m	RL	HRL	C-HRL
5	6	4	0.8829	0.0863	**0.0770**
3	12	8	5.5812	0.1639	**0.1218**
3	18	12	43.9159	1.2854	**0.8772**
5	18	12	**	2.1959	**1.5911**
50	18	12	**	22.7517	**16.5972**

C-HRL: Apply a customized HRL algorithm to decomposed problems

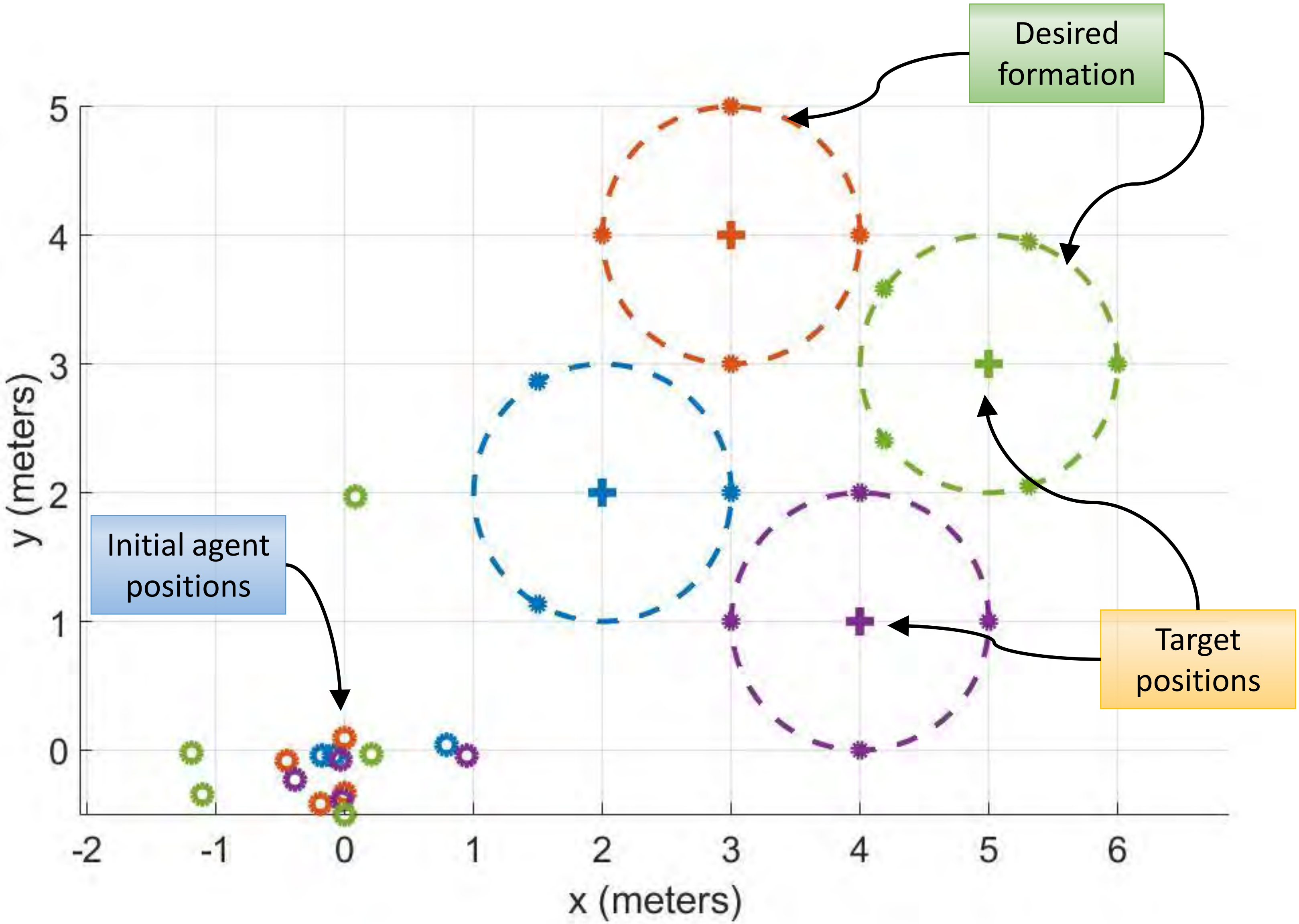
** : Computational time is longer than 60s

G. Jing, H. Bai, J. George and A. Chakrabortty, "Decomposability and Parallel Computation of Multi-Agent LQR", ***in FrA12 Regular Session (11:15-11:30) American Control Conference***, to appear, 2021.

Outline

- ☐ Overview
- ☐ Hierarchical Reinforcement Learning (HRL) Control
  - ○ RL Control (backgound)
  - ○ Problem Formulation
  - ○ Proposed HRL Solution
    - HRL for approximate control of heterogeneous swarm
    - HRL for optimal control of homogeneous swarm
  - ○ Example: Formation Control
- ☐ Swarm Decomposition
  - ○ Decomposition Objectives
  - ○ Example: Formation Maneuver
- ☐ AirSim Experiments
- ☐ Conclusions

HRL Example: Formation Control of Multiple Groups



HRL Example: Formation Control of Multiple Groups

Heterogeneous Agents

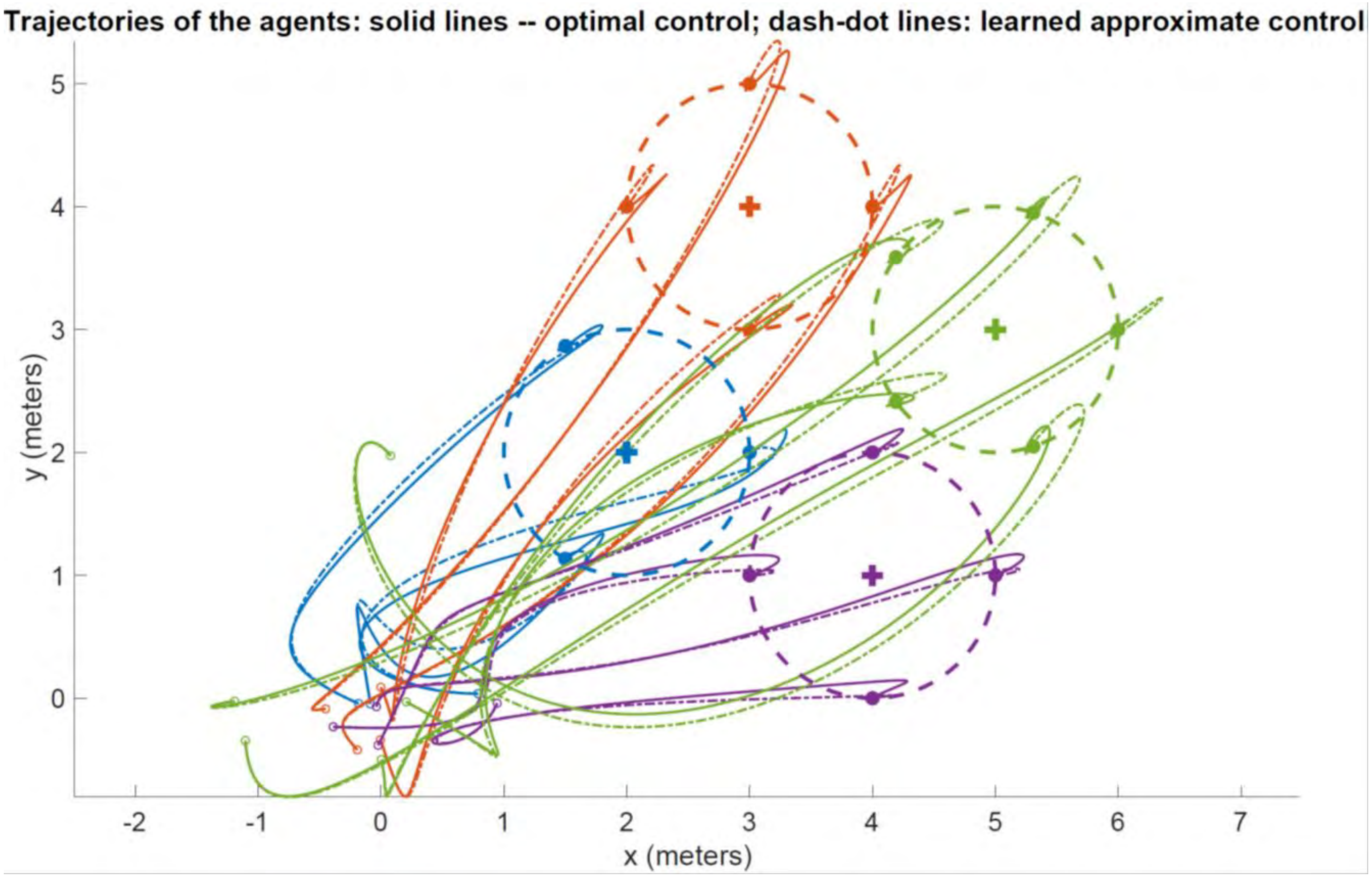
- 2D robots: $M_i \ddot{q}_i + C_i \dot{q}_i = u_i, \quad i = 1, \cdots, N$
- Unknown M_i and C_i
- The robots are divided into 4 groups to track 4 different targets of known locations.
  - Linear Quadratic Integral (LQI) approach
- Control objectives:
  - ⋆ each group converges to a desired formation
  - ⋆ its assigned target is at the center of the formation
  - ⋆ keep the group centroid as close as possible

$$J_j = \int_0^\infty X_j^\top \bar{Q}_j X_j + \mathbf{u}_j^\top R_j \mathbf{u}_j \, dt \quad J_g = \int_0^\infty X^\top (L_w \otimes S^\top S) X \, dt$$

$$J = \sum_{j=1}^{s} J_j + J_g = \int_0^\infty X^\top (\bar{Q} + \tilde{Q}) X + \mathbf{u}^\top R \mathbf{u} \, dt$$

Simulation results:

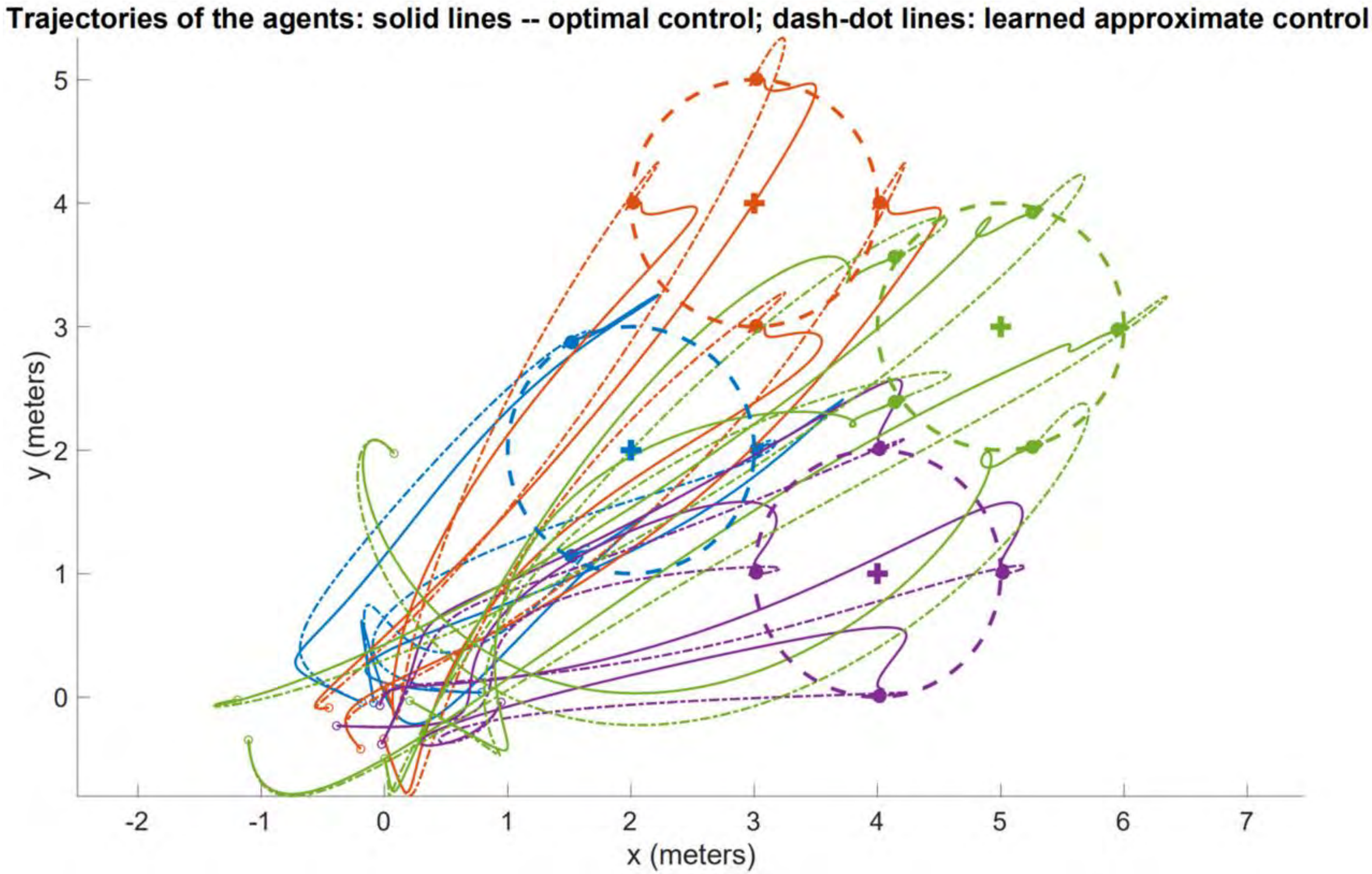
$$Q = 0.1 \times I + 0.5 \times \left(L_w \otimes \tilde{Q}\right)$$



Trajectories of agents under optimal and approximated optimal controllers. Targets are denoted by +'s. Different colors indicate different groups.

Simulation results:

$$Q = 0.1 \times I + 5 \times \left(L_w \otimes \tilde{Q}\right)$$



Trajectories of agents under optimal and approximated optimal controllers. Targets are denoted by +'s. Different colors indicate different groups.

Outline

- ☐ Overview
- ☐ Hierarchical Reinforcement Learning (HRL) Control
  - ◦ RL Control (backgound)
  - ◦ Problem Formulation
  - ◦ Proposed HRL Solution
    - HRL for approximate control of heterogeneous swarm
    - HRL for optimal control of homogeneous swarm
  - ◦ Example: Formation Control
- ☐ Swarm Decomposition
  - ◦ Decomposition Objectives
  - ◦ Example: Formation Maneuver
- ☐ AirSim Experiments
- ☐ Conclusions

Decomposition

- ☐ Recall $Q = \bar{Q} + L_w \otimes \tilde{Q}$, where $\bar{Q} = \mathrm{diag}\{\bar{Q}_{11}, \cdots, \bar{Q}_{NN}\}$.
- ☐ Decompose L_w into $L_w = G_1 + G_2$
  - ○ G_1: block diagonal Laplacian matrix with $s \leq N$ blocks
  - ○ G_2: describes couplings between the groups
- ☐ Now $Q = \underbrace{\bar{Q} + G_1 \otimes \tilde{Q}}_{\widehat{Q}:\ s \text{ groups}} + \underbrace{G_2 \otimes \tilde{Q}}_{\text{group coupling}}$

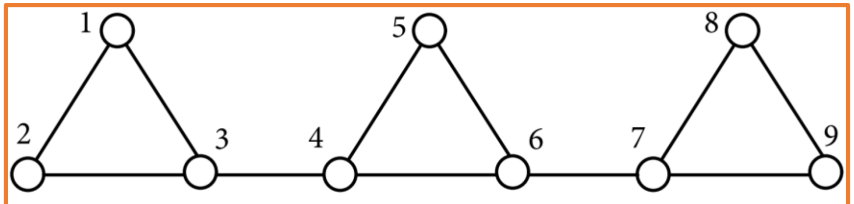
$$L_w = \begin{pmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 3 & -1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{pmatrix} \longrightarrow G_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \quad G_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 2 & -1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$L_w = \begin{pmatrix} 2 & -1 & 0 & 0 & -1 \\ -1 & 3 & -1 & -1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 2 & -1 \\ -1 & 0 & 0 & -1 & 2 \end{pmatrix} \longrightarrow G_1 = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \quad G_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Decomposition Strategies

- Given # of groups and L_w, find an optimal decomposition: challenging
- Explored two approaches
  - Reduce the optimality gap $J(x(0), u_h) - J(x(0), u^*)$
    - An upper bound of the gap depends on $\mathrm{tr}(G_2)$ & $\mathrm{cond}(\mathcal{P})$
    - Minimizing $\mathrm{tr}(G_2)$: k-cut graph partitioning problem
  - Limit required inter-agent communication links
    - maximize $\kappa = \sum_{i \nsim j} N_i N_j$
    - mixed-integer quadratic program (MIQP)

number of pairs of agents that do not need to communicate with each other.



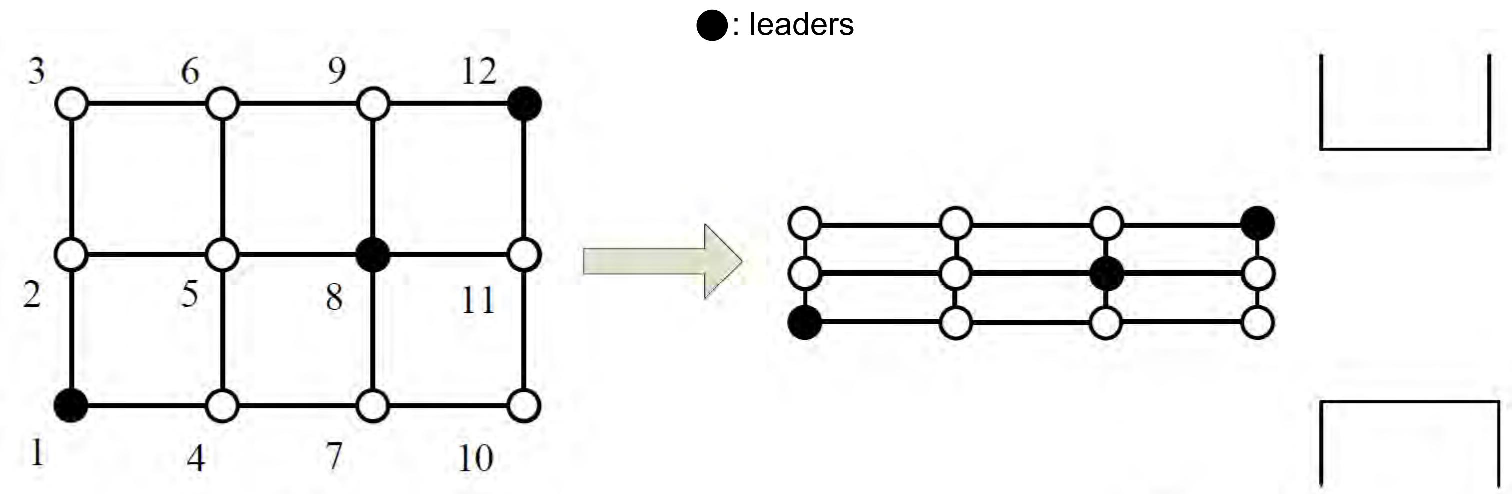
Graph $\mathcal{G}$ with 3 cliques, each clique contains 3 agents.

COMPARISONS BETWEEN DIFFERENT DECOMPOSITIONS.

Decomposition	κ	$\mathrm{tr}(G_2)$	$\mathrm{cond}(\mathcal{P})$	J	n_c	SOP $(J - J^*)/J^*$
{1,2},{3,...,7},{8,9}	4	8	16.4	15.9	32	23.57%
{1,2,3},{4,5,6},{7,8,9}	9	4	17.1	14.3	27	10.20%
{1,...,3},{4},{5,...,9}	15	6	17.0	15.8	21	22.15%
Undecomposed	n/a	n/a	n/a	12.9	36	0

Jing, et al. Model-Free Optimal Control of Linear Multi-Agent Systems via Decomposition and Hierarchical Approximation IEEE TCNS, 2021.

Multi-Agent Formation Maneuver Control



(a) (b)

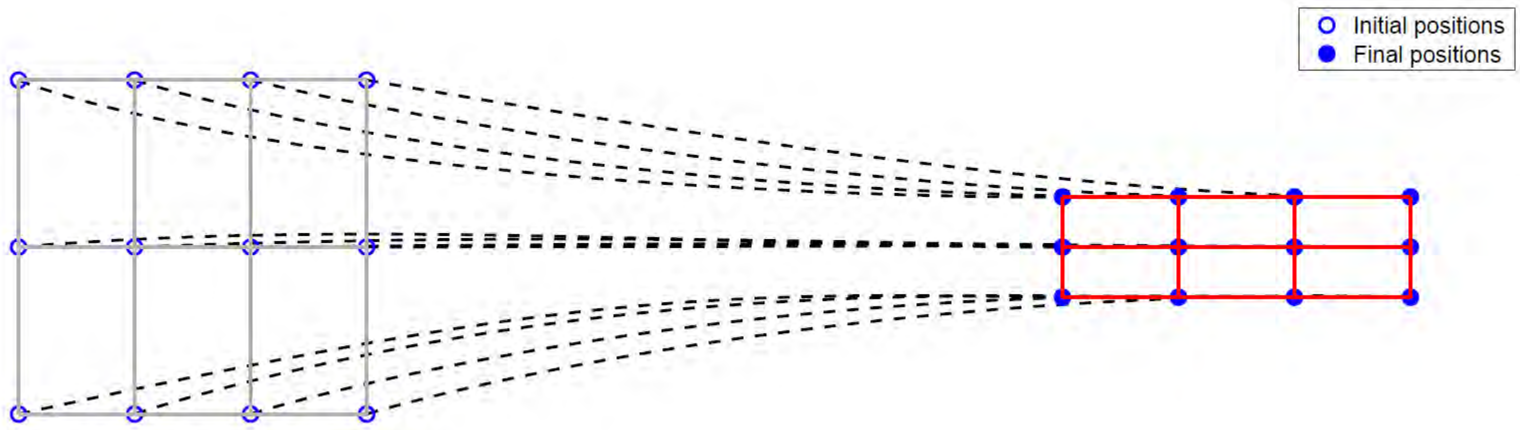
- ☐ Agent dynamics: $M_i\ddot{q}_i + C_i\dot{q}_i = u_i, \quad i = 1, \ldots, N$, (unknown M_i, C_i)
- ☐ Objective function

$$J_1 = \int_0^\infty \sum_{(i,j)\in\mathcal{E}_f} ||q_i - q_j - (h_i - h_j)||^2 + \sum_{i\in\mathcal{L}} ||q_i - h_i||^2 dt$$

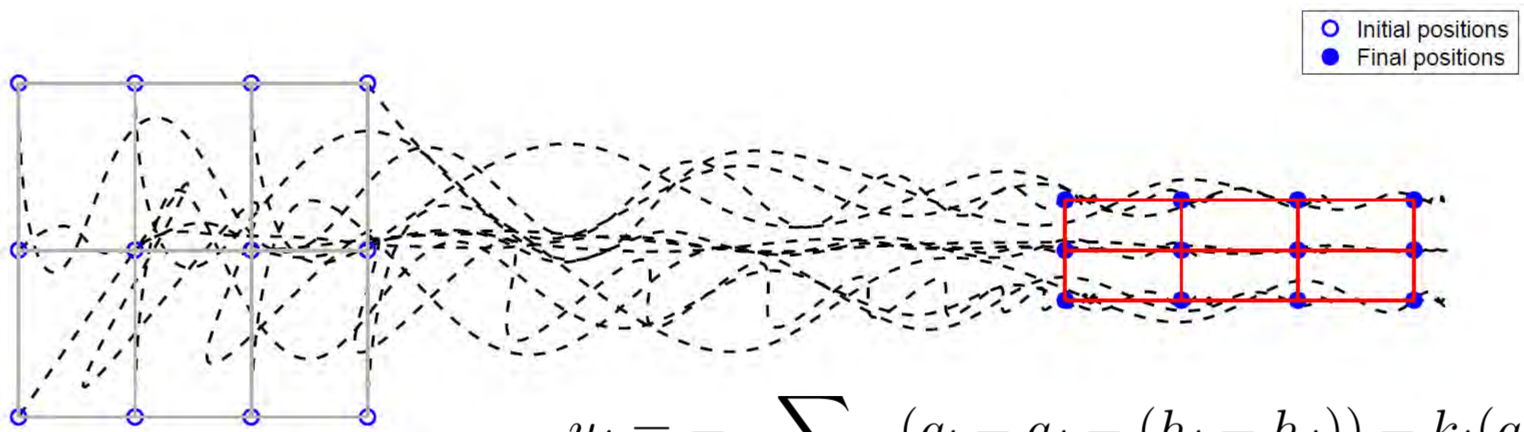
$$J_2 = \int_0^\infty \dot{q}^\top (L \otimes I_2)\dot{q} dt$$

$$\boxed{J = J_1 + J_2 = \int_0^\infty \left[x^\top((L + \Lambda) \otimes I_4)x + u^\top u\right] dt}$$

Multi-Agent Formation Maneuver Control



Optimal (centralized, complete communication graph) $J = 1112.64 \quad \& \quad J_u = 359.11$

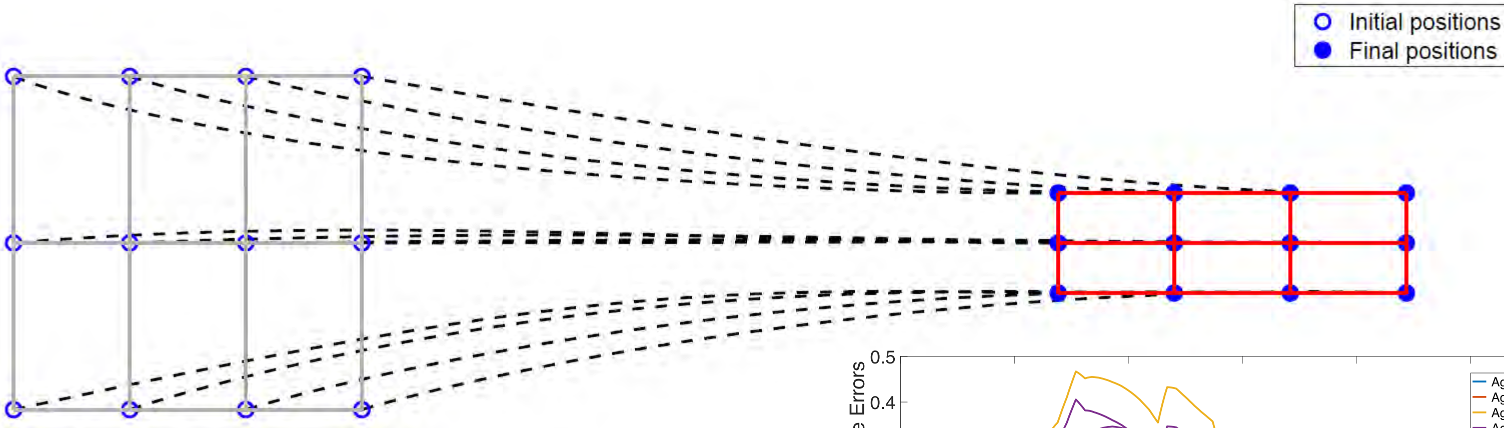


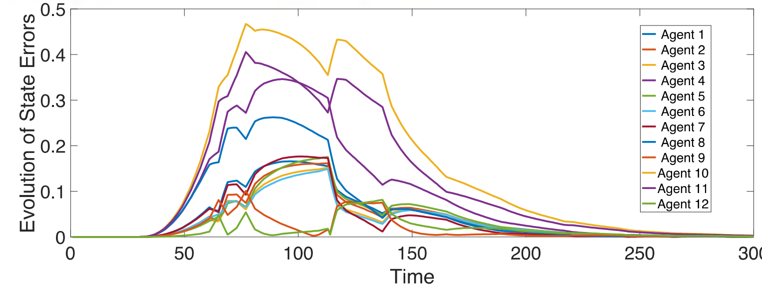
$$u_i = - \sum_{(i,j)\in\mathcal{E}_c} (q_i - q_j - (h_i - h_j)) - k_i(q_i - h_i)$$

Non-optimal (distributed, stable)

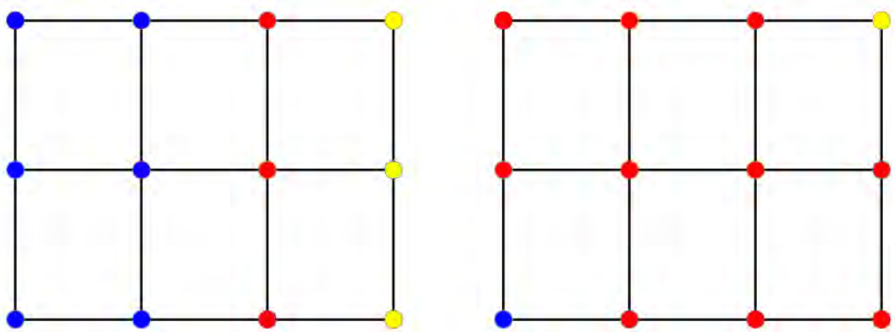
$$J = 2011.21 \quad \& \quad J_u = 945.56$$

Simulation results: HRL

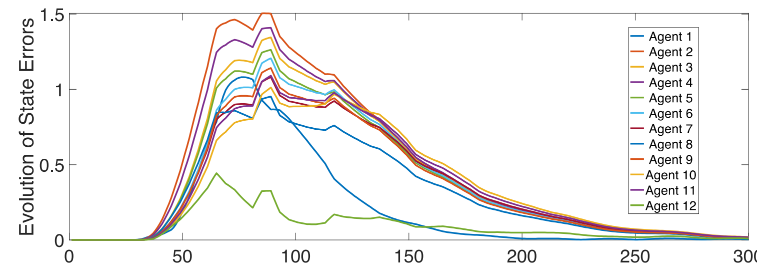




(a)



(a) (b)



(b)

COMPARISONS BETWEEN DIFFERENT DECOMPOSITIONS.

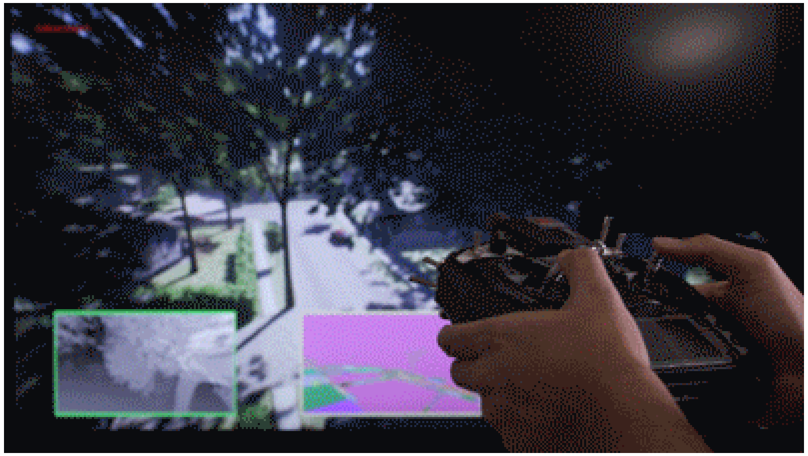
	Decomposition			Performance Indices								
	N_1	N_2	N_3	κ	$\mathrm{tr}(G_2)$	$\mathrm{cond}(\mathcal{P})$	$\mathrm{cond}(\hat{Q})$	J	J_u	n_c	Time(sec)	SOP
(a)	6	3	3	18	12	248.7647	46.1346	1259.7985	426.9677	48	0.9248	0.82%
(b)	1	10	1	1	8	339.4430	83.7524	1347.0390	431.6374	65	13.9719	7.96%
	7	2	3	0	12	285.1161	55.3510	1267.7974	442.4165	66	2.1165	1.61%

Outline

- ☐ Overview
- ☐ Hierarchical Reinforcement Learning (HRL) Control
  - RL Control (backgound)
  - Problem Formulation
  - Proposed HRL Solution
    - HRL for approximate control of heterogeneous swarm
    - HRL for optimal control of homogeneous swarm
  - Example: Formation Control
- ☐ Swarm Decomposition
  - Decomposition Objectives
  - Example: Formation Maneuver
- ☐ AirSim Experiments
- ☐ Conclusions

Microsoft AirSim (Aerial Informatics and Robotics Simulation)

An open-source, cross platform simulator for drones, ground vehicles such as cars and various other objects, built on Epic Games' Unreal Engine 4 as a platform for AI research.



https://microsoft.github.io/AirSim/

AirSim Simulation – 2 Teams

- Trajectory: Minimum Snap (compute @ 10 Hz)
- Position Control: LQR (compute @ 20 Hz, update gains @ 10 Hz)
- Formation:
- Circle of radius 4
- 10 meters above target

AirSim Simulation – Tracking & Formation Control

Conclusions

- Decomposition and hierarchical approximation can speed up reinforcement learning control of large-scale multi-agent systems (MAS).
- For heterogeous MAS,
  - Agents decomposed into groups & Control decomposed into a local control and a global control
  - Local control is learned (in parallel) and global control is approximated
  - Optimizing decompositions of the agents can reduce optimality gap and inter-agent communication.
- For homogeneous MAS, decomposition into N smaller, parallel problems leads to optimal control.
- Several options to decompose the large-scale MAS

Publications

1. G. Jing, H. Bai, J. George and A. Chakrabortty, "Model-Free Optimal Control of Linear Multi-Agent Systems via Decomposition and Hierarchical Approximation," in IEEE Transactions on Control of Network Systems, doi: 10.1109/TCNS.2021.3074256.
2. G. Jing, H. Bai, J. George, A. Chakrabortty and P. K. Sharma, "Learning Distributed Stabilizing Controllers for Multi-Agent Systems," in IEEE Control Systems Letters, doi: 10.1109/LCSYS.2021.3072007.
3. G. Jing, H. Bai, J. George and A. Chakrabortty, “Model-Free Reinforcement Learning of Minimal-Cost Variance Control,” IEEE Control Systems Letters, vol. 4, no. 4, pp. 916-921, 2020.
4. Bai, H., George, J. and Chakrabortty, A., “Hierarchical Control of Multi-Agent Systems using Online Reinforcement Learning,” American Control Conference (ACC), Denver, CO, July 2020.
5. G. Jing, H. Bai, J. George and A. Chakrabortty, “Model-Free Optimal Control of Linear Multi-Agent Systems via Decomposition and Hierarchical Approximation,” arXiv preprint, arXiv:2008.06604, Aug. 2020.
6. G. Jing, H. Bai, J. George and A. Chakrabortty, “Hierarchical Reinforcement Learning for Optimal Control of Linear Multi-Agent Systems: the Homogeneous case,” submitted to American Control Conference (ACC), New Orleans, LA, July 2021.

Aranya Chakrabortty
Gangshan Jing
North Carolina State University

He Bai
Collin Thornton
Oklahoma State University