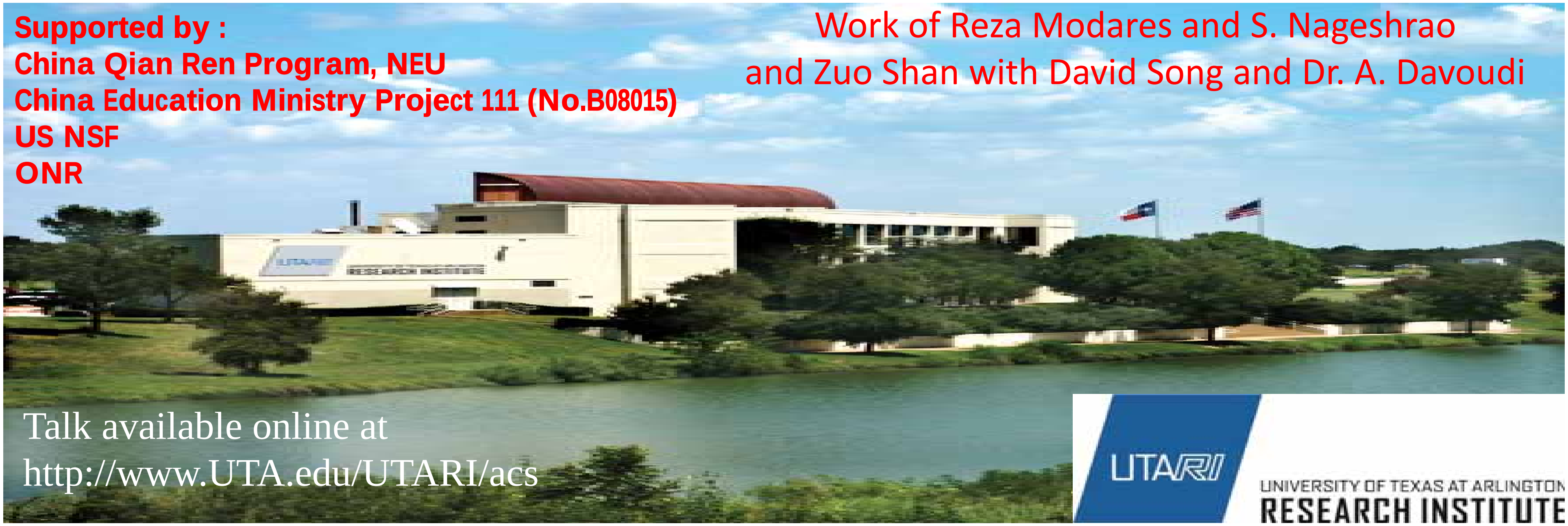

Zhongping Jiang
Tengfei Liu

F.L. Lewis, NAI

Moncrief-O'Donnell Chair, UTA Research Institute (UTARI)
The University of Texas at Arlington, USA
and
Qian Ren Consulting Professor, State Key Laboratory of Synthetical Automation for Process Industries,
Northeastern University, Shenyang, China

Output Regulation of Heterogeneous MAS- Reinforcement Learning Solution for Unknown Systems



Supported by :
China Qian Ren Program, NEU
China Education Ministry Project 111 (No.B08015)
US NSF
ONR

Work of Reza Modares and S. Nageshrao and Zuo Shan with David Song and Dr. A. Davoudi

Talk available online at
http://www.UTA.edu/UTARI/acs

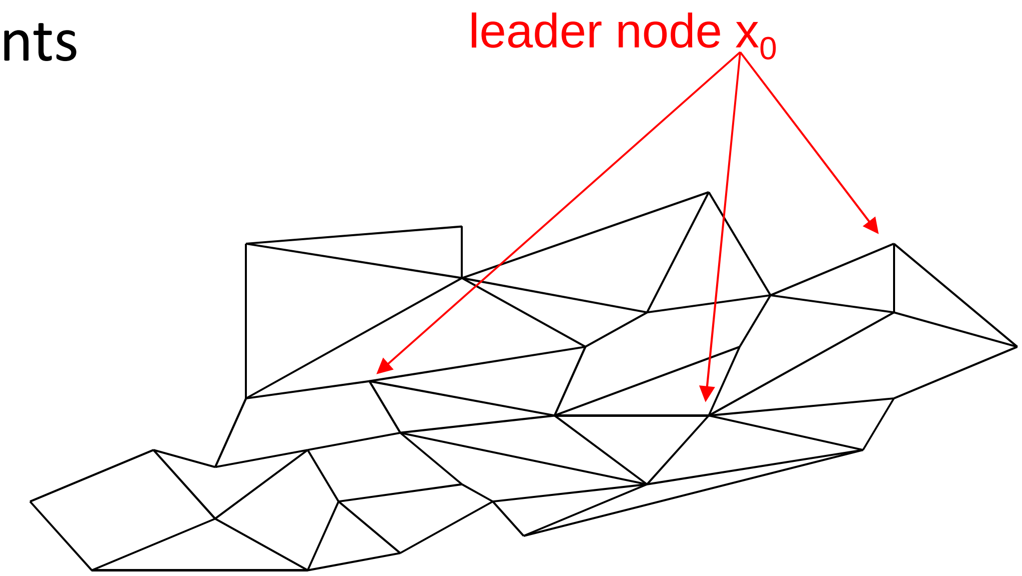
Output Synchronization of Heterogeneous MAS

Heterogeneous Multi-Agents

$$\dot{x}_i = A_i x_i + B_i u_i$$
$$y_i = C_i x_i$$

Leader

$$\dot{\zeta}_0 = S \zeta_0$$
$$y_0 = R \zeta_0$$



Output regulation error $\eta_i(t) = y_i(t) - y_0(t) \to 0$

Output regulator equations

$$A_i \Pi_i + B_i \Gamma_i = \Pi_i S$$
$$C_i \Pi_i = R$$

Dynamics are different, state dimensions can be different
o/p reg eqs capture the common core of all the agents dynamics
And define a synchronization manifold

Heterogeneous Multi-Agents

$$\dot{x}_i = A_i x_i + B_i u_i$$
$$y_i = C_i x_i$$

Leader

$$\dot{\zeta}_0 = S \zeta_0$$
$$y_0 = R \zeta_0$$

Output regulator equations
$$A_i \Pi_i + B_i \Gamma_i = \Pi_i S$$
$$C_i \Pi_i = R$$

Tracking error
$$\bar{\varepsilon}_i = \bar{x}_i - \Pi_i \zeta_0$$

Output regulation error
$$\eta_i(t) = y_i(t) - y_0(t) \to 0$$

Π_i is the insertion map of *S* in *A*

Dynamics are different, state dimensions can be different
o/p reg eqs capture the common core of all the agents dynamics
And define a synchronization manifold

Two Control Methods

Output regulator equations

$$A_i \Pi_i + B_i \Gamma_i = \Pi_i S$$

$$C_i \Pi_i = R$$

Control Method #1

$$\dot{\zeta}_i = S\zeta_i + c\left[\sum_{j=1}^{N} a_{ij}(\zeta_j - \zeta_i) + g_i(\zeta_0 - \zeta_i)\right]$$

$$u_i = K_{1i}(x_i - \Pi_i \zeta_i) + \Gamma_i \zeta_i = K_{1i} x_i + (\Gamma_i - K_{1i}\Pi_i)\zeta_i \equiv K_{1i} x_i + K_{2i}\zeta_i$$

Control Method #2- more intriguing

Local neighborhood output tracking error

$$e_{y_i} \equiv \sum_{j \in N_i} a_{ij}\left(y_j - y_i\right) + g_i\left(y_0 - y_i\right).$$

compensator

$$\begin{cases} \dot{z}_i = F_i z_i + G_i e_{y_i} \\ u_i = K_i x_i + H_i z_i \end{cases}$$

either

$$u_i = K_i x_i + H_i z_i = K_i x_i + (\Gamma_i - K_i \Pi_i) z_i$$

Or, assume p-copy in compensator
Then K_i, H_i are independent

Must know agent and leader's dynamics S,R

Data-driven Adaptive Solution of o/p Reg Eqs in Real Time

Output regulator equations

$$A_i \Pi_i + B_i \Gamma_i = \Pi_i S$$

$$C_i \Pi_i = R$$

Solve o/p regulator eqs online using measured data by Reinforcement Learning

Automatica 71 (2016) 334–341

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

ELSEVIER

Brief paper

Optimal model-free output synchronization of heterogeneous systems using off-policy reinforcement learning☆

Hamidreza Modares[a], Subramanya P. Nageshrao[b,1], Gabriel A. Delgado Lopes[b], Robert Babuška[b], Frank L. Lewis[a,c]

Optimal Output Synchronization of Heterogeneous MAS Using Off-policy IRL

Nageshrao, Modares, Lopes, Babuska, Lewis

MAS $\dot{x}_i = A_i x_i + B_i u_i$

$y_i = C_i x_i$

Leader $\dot{\zeta}_0 = S \zeta_0$

$y_0 = R \zeta_0$

Optimal Tracker Problem

Augmented Systems

$$X(t) = \left[x_i(t)^T \;\; \zeta_0^{\;T} \right]^T \in \mathbb{R}^{n_i + p}$$

$$\dot{X}_i = T_i X_i + B_{1i} u_i$$

$$T_i = \begin{bmatrix} A_i & 0 \\ 0 & S \end{bmatrix}, B_{1i} = \begin{bmatrix} B_i \\ 0 \end{bmatrix}$$

Performance index

$$V(X_i(t)) = \int_t^{\infty} e^{-\gamma_i(\tau - t)} X_i^{\;T} (C_{1i}^{\;T} Q_i C_{1i} + K_i^{\;T} W_i K_i) X_i \, d\tau$$
$$= X_i(t)^T P_i X_i(t)$$

Control

$$u_i = K_{1i} x_i + K_{2i} \zeta_0 = K_i X_i$$

Optimal Tracker Solution by Reinforcement Learning

$$K_i = [K_{1i}, K_{2i}] = -W_i^{-1} B_{1i}^T P_i$$

Tracker ARE

$$T_i^T P_i + T_i P_i - \gamma_i P_i + C_{1i}^T Q_i C_{1i} - P_i B_{1i} W_i^{-1} B_{1i}^T P_i = 0$$

Algorithm 1. *On-policy IRL State-feedback algorithm*

Policy Evaluation- Solve IRL Bellman equation

$$e^{-\gamma_i \delta t} X_i(t+\delta t)^T P_i^\kappa X_i(t+\delta t) - X_i(t)^T P_i^\kappa X_i(t) = -\int_t^{t+\delta t} e^{-\gamma_i(\tau - t)} (y_i - y_0)^T Q_i (y_i - y_0) d\tau$$

Policy Update-

$$K_i^{\kappa+1} = [K_{1i}^{\kappa+1}, K_{2i}^{\kappa+1}] = -W_i^{-1} B_{1i}^T P_i^\kappa$$

Theorem- Algorithm 1 converges to the solution to the ARE

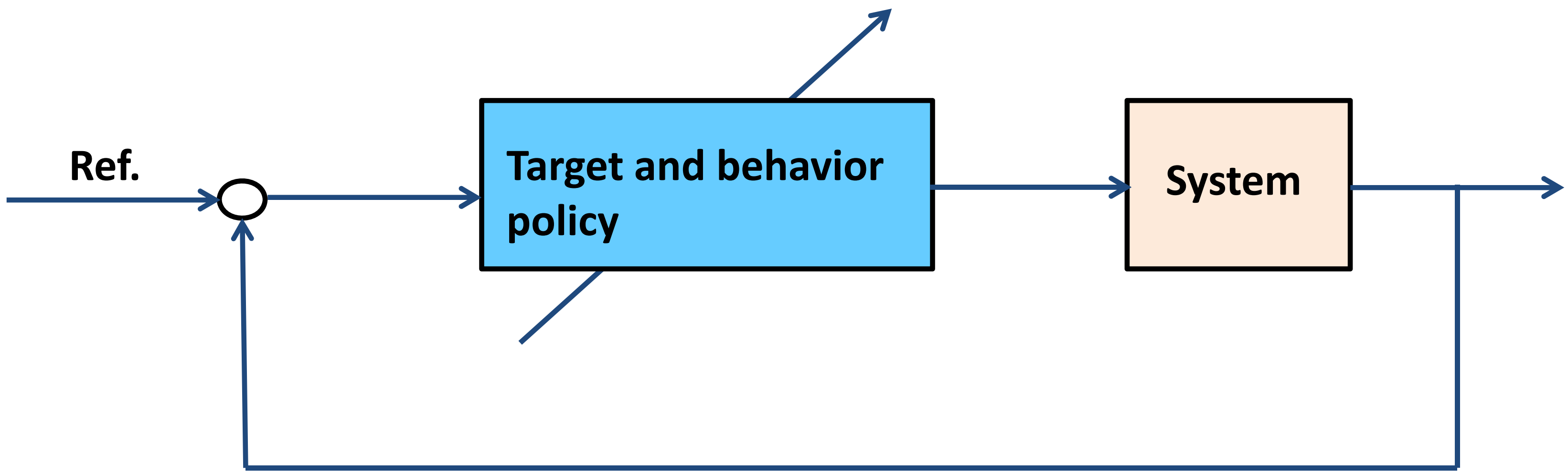
Bellman equation is solved using RLS or batch LS
It requires a Persistence of Excitation (PE) condition that may be hard to satisfy

Must know B_{1i}

On-policy RL

Target policy: The policy that we are learning about.

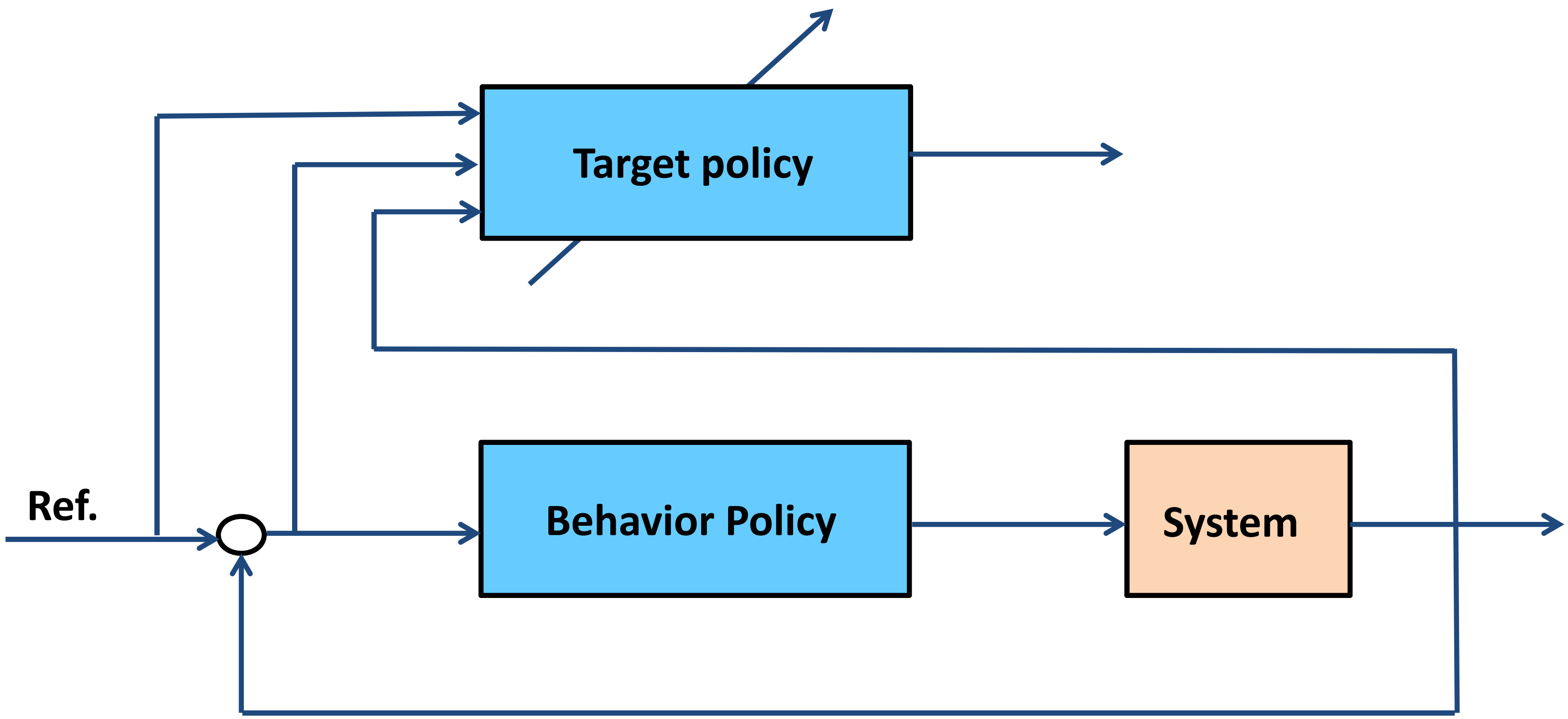
Behavior policy: The policy that generates actions and behavior



Target policy and behavior policy are the same

Off-policy RL

Humans can learn optimal policies while actually applying suboptimal policies



Target policy and behavior policy are different

Off-Policy RL

Tracker dynamics

$$\dot{X}_i = T_i X_i + B_{1i} u_i$$

Rewrite as

$$\dot{X}_i = (T_i + B_{1i} K_i^{\kappa}) X_i + B_{1i}(u_i - K_i^{\kappa} X_i) \equiv \bar{T}_i X_i + B_{1i}(u_i - K_i^{\kappa} X_i)$$

Now the Bellman equation becomes

$$e^{-\gamma_i \delta t} X_i(t+\delta t)^T P_i^{\kappa} X_i(t+\delta t) - X_i(t)^T P_i^{\kappa} X_i(t) = -\int_t^{t+\delta t} e^{-\gamma_i(\tau - t)} (y_i - y_0)^T Q_i (y_i - y_0) d\tau$$
$$+ \quad 2\int_t^{t+\delta t} e^{-\gamma_i(\tau - t)} (u_i - K_i^{\kappa} X_i)^T W_i K_i^{\kappa+1} X_i \, d\tau$$

Extra term containing $K_i^{\kappa+1}$

Algorithm 2. *Off-policy IRL Data-based algorithm*

Iterate on this equation and solve for $P_i^{\kappa}, K_i^{\kappa+1}$ simultaneously at each step

Note about probing noise If $u_i = K_i^{\kappa} X_i + e$ then $(u_i - K_i^{\kappa} X_i) = e$

Do not have to know any dynamics

agent $\dot{x}_i = A_i x_i + B_i u_i$, $y_i = C_i x_i$ Or leader $\dot{\zeta}_0 = S \zeta_0$, $y_0 = R \zeta_0$

Theorem- Off-policy Algorithm 2 converges to the solution to the ARE

$$T_i^T P_i + T_i P_i - \gamma_i P_i + C_{1i}^T Q_i C_{1i} - P_i B_{1i} W_i^{-1} B_{1i}^T P_i = 0$$

Theorem- o/p reg eq solution

Let $P_i = \begin{bmatrix} P^i_{11} & P^i_{12} \\ P^i_{21} & P^i_{22} \end{bmatrix}$

Then the solution to the output regulator equations

$$A_i \Pi_i + B_i \Gamma_i = \Pi_i S$$
$$C_i \Pi_i = R$$

Is given by

$$\Pi_i = -(P^i_{11})^{-1} P^i_{12}$$
$$\Gamma_i = K_{2i} - K_{1i} (P^i_{11})^{-1} P^i_{12}$$

Do not have to know the
Agent dynamics or the leader's dynamics (S,R)

Observer for Leader's State and Dynamics

To avoid knowledge of leader's state in

$$u_i = K_{1i} x_i + K_{2i} \zeta_0 = K_i X_i$$

Use adaptive observer for leader's state

$$\dot{\zeta}_i = \hat{S}_i \zeta_i + c \left[\sum_{j=1}^{N} a_{ij} (\zeta_j - \zeta_i) + g_i (\zeta_0 - \zeta_i) \right]$$

$$\dot{\hat{S}}_{\text{vec}i} = -\Gamma_{Si} (I_q \otimes \zeta_i) \left[\sum_{j=1}^{N} a_{ij} (\zeta_j - \zeta_i) + g_i (\zeta_0 - \zeta_i) \right]$$

Then use control

$$u_i = K_{1i} x_i + K_{2i} \zeta_i \equiv K_i \hat{X}_i \equiv K_i \begin{bmatrix} x_i \\ \zeta_i \end{bmatrix}$$

So this is the Control Method #1

Do not have to know the leader's dynamics (S,R)

Two Control Methods

Output regulator equations

$$A_i\Pi_i + B_i\Gamma_i = \Pi_i S$$

$$C_i\Pi_i = R$$

We just found a Controller #1 using data-based control

$$\dot{\zeta}_i = S\zeta_i + c\left[\sum_{j=1}^{N} a_{ij}(\zeta_j - \zeta_i) + g_i(\zeta_0 - \zeta_i)\right]$$

$$u_i = K_{1i}(x_i - \Pi_i\zeta_i) + \Gamma_i\zeta_i = K_{1i}x_i + (\Gamma_i - K_{1i}\Pi_i)\zeta_i \equiv K_{1i}x_i + K_{2i}\zeta_i$$

Now we seek a Controller using Method #2- more intriguing

Local neighborhood output tracking error

$$e_{y_i} \equiv \sum_{j\in N_i} a_{ij}\left(y_j - y_i\right) + g_i\left(y_0 - y_i\right).$$

compensator

$$\begin{cases} \dot{z}_i = F_i z_i + G_i e_{y_i} \\ u_i = K_i x_i + H_i z_i \end{cases}$$

either

$$u_i = K_i x_i + H_i z_i = K_i x_i + (\Gamma_i - K_i\Pi_i)z_i$$

Or, assume p-copy in compensator
Then K_i, H_i are independent

Do not want to solve o/p reg eqs or know any dynamics

Overall Dynamics Structure

Agents

$$\dot{x}_i = A_i x_i + B_i u_i + R_i x_0,$$
$$y_i = C_i x_i,$$

Leader

$$\dot{x}_0 = A_0 x_0,$$
$$y_0 = F x_0,$$

Compensator

$$\dot{z}_i = G_{1i} z_i + G_{2i} e_{iv},$$

Control Input

$$u_i = K_{y_i} y_i + K_{z_i} z_i$$

Local neighborhood output error

$$e_{iv} = \frac{1}{d_i + g_i}\left[\sum_{j \in N_i} e_{ij}(y_i - y_j) + g_i(y_i - y_0)\right],$$

Global form

$$e = \left[\left(diag\{\tfrac{1}{d_i+g_i}\}(L+G)\right) \otimes I_p\right] diag\{C_i\} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix}$$

Global Dynamics

$$\varrho_g = [x_1^T, \ldots, x_N^T, z_1^T, \ldots, z_N^T]^T$$

$$r_G = [\underbrace{x_0^T, x_0^T, \ldots, x_0^T}_{N}]^T$$

$$\dot{\varrho}_g = A_{c_1}\varrho_g + B_c r_G,$$

$$\dot{r}_G = \tilde{A}_0 r_G,$$

$$e = [\tilde{C} \quad 0]\,\varrho_g - \tilde{F} r_G,$$

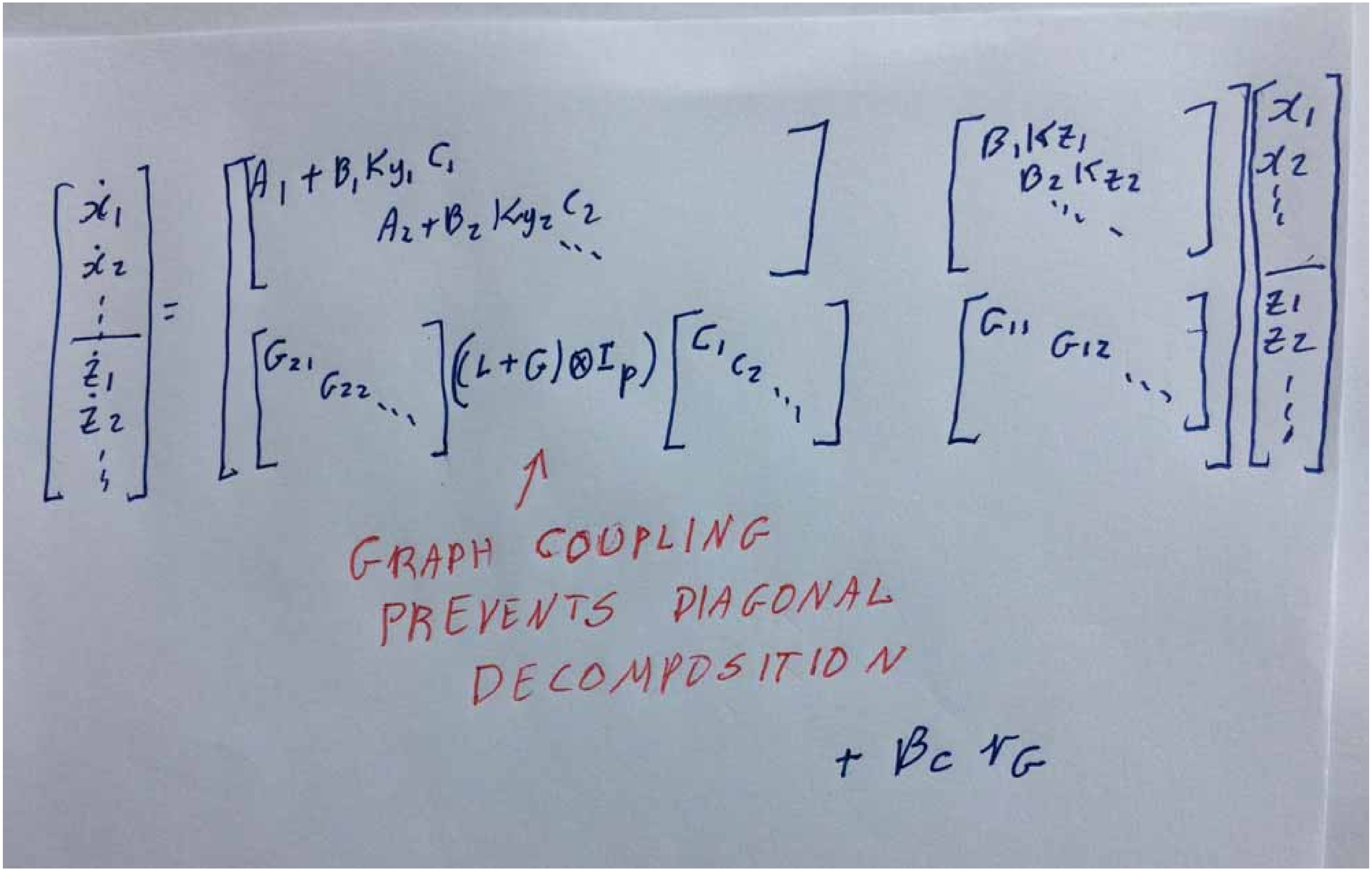
$$\tilde{A} = diag\{A_i\}$$

$$\begin{bmatrix} \dot{x} \\ \dot{z} \end{bmatrix}$$

$$A_{c_1} = \begin{bmatrix} \tilde{A} + \tilde{B}\tilde{K}_y\tilde{C} & \tilde{B}\tilde{K}_z \\ \tilde{G}_2\tilde{H}\tilde{C} & \tilde{G}_1 \end{bmatrix}, \qquad B_c = \begin{bmatrix} \tilde{R} \\ -\tilde{G}_2\tilde{H}\tilde{F} \end{bmatrix}.$$

$$\tilde{H} = \left(diag\{\tfrac{1}{d_i + g_i}\}(L+G) \right) \otimes I_p$$

Problems to Get Local Design Procedure




Joel Santos

A unified approach to output synchronization of heterogeneous multi-agent systems via $\mathcal{L}_2$-gain design ⋆

Shan Zuo [a,b], Yongduan Song [a], Hamidreza Modares [b], Frank L. Lewis [b], Ali Davoudi [b]

[a] *School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, P.R. China*

[b] *The University of Texas at Arlington Research Institute, Fort Worth, TX 76118, USA*

Heterogeneous Multi-agent Dynamics

$$\begin{cases} \dot{x}_i = A_i x_i + B_i u_i, \\ y_i = C_i x_i, \end{cases}$$

Leader Dynamics

$$\begin{cases} \dot{\zeta}_0 = S\zeta_0, \\ y_0 = R\zeta_0, \end{cases}$$

Output Regulation Problem

$$\eta_i \equiv y_i - y_0 = C_i x_i - R\zeta_0.$$

$$\lim_{t \to \infty} \eta_i(t) = 0, \forall i \in N.$$

Local Neighborhood Output Error

$$e_{y_i} \equiv \sum_{j \in N_i} a_{ij} (y_j - y_i) + g_i (y_0 - y_i)$$

Compensator 1 – local state feedback

$$\begin{cases} \dot{z}_i = F_i z_i + G_i e_{y_i}, \\ u_i = K_i x_i + H_i z_i, \end{cases}$$

Compensator 2 – local output feedback

$$\begin{cases} \dot{z}_i = F_i z_i + G_i e_{y_i}, \\ u_i = K_i y_i + H_i z_i, \end{cases}$$

Compensator 3 – no local system feedback

$$\begin{cases} \dot{z}_i = F_i z_i + G_i e_{y_i}, \\ u_i = H_i z_i, \end{cases}$$

Local Systems with Interactions from Neighbors

Rewrite local o/p error

$$\begin{aligned} e_{y_i} &= \sum_{j\in N_i} a_{ij}\left(y_j - y_i\right) + g_i\left(y_0 - y_i\right) \\ &= -\left(d_i + g_i\right) y_i + \sum_{j\in N_i} a_{ij} y_j + g_i y_0 \\ &= -y_i + \sum_{j\in N_i} a_{ij} y_j + g_i y_0. \end{aligned}$$

Local Dynamics

$$\begin{bmatrix} \dot{x}_i \\ \dot{z}_i \end{bmatrix} = \begin{bmatrix} A_i + B_i K_i & B_i H_i \\ -G_i C_i & F_i \end{bmatrix} \begin{bmatrix} x_i \\ z_i \end{bmatrix}$$

Interaction terms

$$+\begin{bmatrix} 0 \\ G_i \end{bmatrix} \sum_{j\in N_i} a_{ij} \begin{bmatrix} C_j & 0 \end{bmatrix} \begin{bmatrix} x_j \\ z_j \end{bmatrix} + \begin{bmatrix} 0 \\ G_i \end{bmatrix} g_i R \zeta_0,$$

$$\eta_i = \begin{bmatrix} C_i & 0 \end{bmatrix} \begin{bmatrix} x_i \\ z_i \end{bmatrix} - R\zeta_0.$$

State Regulation Errors

$$\bar{\varepsilon}_i = \bar{x}_i - X_i \zeta_0$$

o/p reg eqs

$$\begin{cases} \bar{A}_i X_i + \bar{G}_i R = X_i S \\ \bar{C}_i X_i = R \end{cases}.$$

Error dynamics

$$\dot{\bar{\varepsilon}}_i = \bar{A}_i \bar{\varepsilon}_i + \bar{G}_i \sum_{j \in N_i} a_{ij} \bar{C}_j \bar{\varepsilon}_j$$

$$= \begin{bmatrix} A_i + B_i K_i & B_i H_i \\ -G_i C_i & F_i \end{bmatrix} \bar{\varepsilon}_i + \bar{G}_i \sum_{j \in N_i} a_{ij} \bar{C}_j \bar{\varepsilon}_j$$

$$= \begin{bmatrix} A_i + B_i K_i & B_i H_i \\ -G_i C_i & F_i \end{bmatrix} \bar{\varepsilon}_i + \bar{G}_i \bar{z}_i$$

$$\eta_i = \bar{C}_i \bar{\varepsilon}_i$$

Local transfer functions

$$T_i(s) \equiv \bar{C}_i \left(s - \bar{A}_i \right)^{-1} \bar{G}_i$$

State Regulation Errors

$$\bar{\varepsilon}_i = \bar{x}_i - X_i \zeta_0.$$

Error Dynamics

$$\begin{cases} \dot{\bar{\varepsilon}}_i = \bar{A}_i \bar{\varepsilon}_i + \bar{G}_i \sum_{j \in N_i} a_{ij} \bar{C}_j \bar{\varepsilon}_j, \\ \eta_i = \bar{C}_i \bar{\varepsilon}_i. \end{cases}$$

Global Form

$$\begin{cases} \dot{\zeta} = \operatorname{diag}(\bar{A}_i)\, \zeta + \operatorname{diag}(\bar{G}_i)\, \bar{\mathcal{A}} \operatorname{diag}(\bar{C}_i)\, \zeta, \\ \xi = \operatorname{diag}(\bar{C}_i)\, \zeta. \end{cases}$$

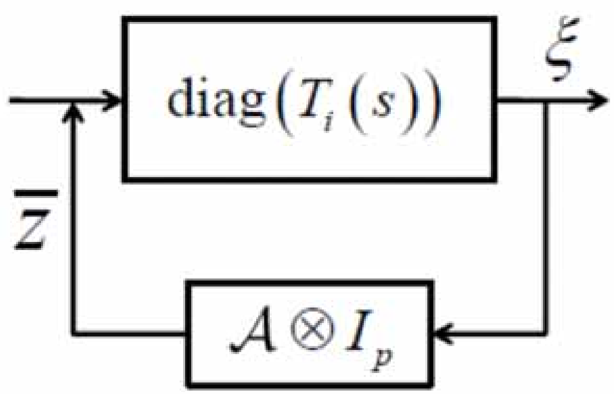


Figure 1. Closed-loop system of (10).

Main Theorem

Theorem 1. *For systems (16), suppose the following three conditions hold for each $i \in N$:*

(i) Matrix $\bar{A}_i$ is Hurwitz. Local Design

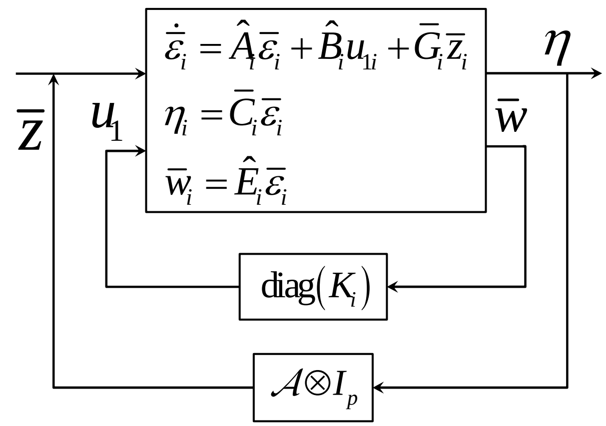
(ii) $\max_{i \in N} \|T_i\|_\infty < 1/\rho(\mathcal{A})$. Interaction Small Gain Condition

(iii) There exists a unique solution X_i to the output regulator equations

$$\begin{cases} \bar{A}_i X_i + \bar{G}_i R = X_i S, \\ \bar{C}_i X_i = R. \end{cases} \quad (17)$$

Then, Problem 1 is solved.

Local Design Procedure



$$\dot{\bar{\varepsilon}}_i = \bar{A}_i\bar{\varepsilon}_i + \bar{G}_i\sum_{j\in N_i} a_{ij}\bar{C}_j\bar{\varepsilon}_j$$

Closed-loop Systems

$$= \begin{bmatrix} A_i + B_iK_i & B_iH_i \\ -G_iC_i & F_i \end{bmatrix}\bar{\varepsilon}_i + \bar{G}_i\sum_{j\in N_i} a_{ij}\bar{C}_j\bar{\varepsilon}_j$$

$$= \begin{bmatrix} A_i + B_iK_i & B_iH_i \\ -G_iC_i & F_i \end{bmatrix}\bar{\varepsilon}_i + \bar{G}_i\bar{z}_i$$

Open-loop Systems

$$= \begin{bmatrix} A_i & 0 \\ -G_iC_i & F_i \end{bmatrix}\bar{\varepsilon}_i + \begin{bmatrix} B_i \\ 0 \end{bmatrix}\begin{bmatrix} K_i & H_i \end{bmatrix}\bar{\varepsilon}_i + \bar{G}_i\bar{z}_i$$

$$= \begin{bmatrix} A_i & 0 \\ -G_iC_i & F_i \end{bmatrix}\bar{\varepsilon}_i + \begin{bmatrix} B_i \\ 0 \end{bmatrix}\bar{K}_i\bar{\varepsilon}_i + \bar{G}_i\bar{z}_i.$$

Then problem is formulated as

$$\begin{cases} \dot{\bar{\varepsilon}}_i = \hat{A}_i\bar{\varepsilon}_i + \hat{B}_i u_{1i} + \bar{G}_i\bar{z}_i, \\ \eta_i = \bar{C}_i\bar{\varepsilon}_i \end{cases}$$

Static State feedback

$$u_{1i} = \bar{K}_i\bar{\varepsilon}_i$$

Optimal Local Design Algorithm

H-infinity design

Theorem 3. *Under Assumptions 1, 2, 3 and 4, select* $\gamma_i < 1/\rho(A)$. *Then, Problem 1 is solved if, for some positive definite matrices* R_i *and scalar* $\alpha_i > 0$, *one takes*

$$\bar{K}_i = -\frac{1}{\alpha_i} R_i^{-1} \hat{B}_i^T P_i$$

where $P_i^T = P_i > 0$ *is a solution of*

$$P_i \hat{A}_i + \hat{A}_i^T P_i + \bar{C}_i^T \bar{C}_i + \frac{1}{\gamma_i^2} P_i \bar{G}_i \bar{G}_i^T P_i - \frac{1}{\alpha_i} P_i \hat{B}_i R_i^{-1} \hat{B}_i^T P_i = 0.$$

[Gadewadikar et al., 2007] Gadewadikar, J., Lewis, F. L., Xie, L., Kucera, V., and Abu-Khalaf, M. (2007). Parameterization of all stabilizing H_∞ static state-feedback gains: application to output-feedback design. *Automatica*, 43(9):1597–1604.

Off-policy RL Algorithm to Solve Output Regulation for Heterogeneous MAS

$$
\begin{aligned}
& e^{-\delta_i \theta t}\bar{\varepsilon}_i(t+\theta t)^T P_i^\kappa \bar{\varepsilon}_i(t+\theta t) - \bar{\varepsilon}_i(t)^T P_i^\kappa \bar{\varepsilon}_i(t) \\
&= \int_t^{t+\theta t} e^{-\delta_i(\tau-t)}\left(-\eta_i^T Q_i \eta_i - \alpha_i \bar{\varepsilon}_i^T \left(\bar{K}_i^\kappa\right)^T R_i \bar{K}_i^\kappa \bar{\varepsilon}_i\right) d\tau \quad + \int_t^{t+\theta t} e^{-\delta_i(\tau-t)} \gamma_i^2 \left(\bar{z}_i^{\ \kappa}\right)^T \bar{z}_i^{\ \kappa} d\tau \\
&\quad - \int_t^{t+\theta t} e^{-\delta_i(\tau-t)} 2\alpha_i \bar{\varepsilon}_i^T \underline{\left(\bar{K}_i^{\kappa+1}\right)^T R_i \left(\bar{K}_i - \bar{K}_i^\kappa\right)} \bar{\varepsilon}_i d\tau \quad + \int_t^{t+\theta t} e^{-\delta_i(\tau-t)} 2\gamma_i^2 \left(\bar{z}_i^{\ \kappa+1}\right)^T \left(\bar{z}_i - \bar{z}_i^{\ \kappa}\right) d\tau.
\end{aligned}
$$

This is a standard parameter ID equation from Adaptive Control
It is linear in the unknown parameters
Solve using Batch LS or RLS to get the updated values

$$P_i^\kappa, \bar{K}_i^{\kappa+1}, \bar{z}_i^{\ \kappa+1}$$

This solves the o/p regulation problem for heterogeneous MAS
Without solving o/p reg equations and
without knowing any agent or leader's dynamics