Team Optimal Control of Coupled Subsystems with Mean-Field Sharing

Jalal Arabneydi and Aditya Mahajan

Electrical and Computer Engineering Department, McGill University

Email: jalal.arabneydi@mail.mcgill.ca

Date: December 15th, 2014

Outline

Motivation

- **What do we mean by team control problem?** Any setup in which agents (decision makers) need to collaborate with each other to achieve a common task.

- Team optimal control of decentralized stochastic systems arises in applications in:
  - Networked control systems
  - Robotics
  - Communication networks
  - Transportation networks
  - Sensor networks
  - Smart grids
  - Economics
  - Etc.

- No solution approach exists for general **infinite-horizon** decentralized control systems.

- In general, these problems belong to **NEXP complexity** class.

Brief Literature Review

- **Classical** information structure: All agents have identical information.
- **Non-classical** information structure: Agents have different information sets.

Examples of **non-classical** information structure:

- Static team (Radner 1962, Marschack and Radner 1972)
- Dynamic team (Witsenhausen 1971, Witsenhausen 1973)
- Specific information structure
  - Partially nested (Ho and Chu 1972)
  - One-step delayed sharing (Witsenhausen 1971, Yoshikawa 1978)
  - n-step delayed sharing (Witsenhausen 1971, Varaiya 1978, Nayyar 2011)
  - Common past sharing (Aicardi 1978)
  - Periodic sharing (Ooi 1997)
  - Belief sharing (Yuksel 2009)
  - Partial history sharing (Nayyar 2013)
  - This work introduces a new information structure : **Mean-field sharing**

Notation:

- N : Number of homogeneous subsystems (not necessarily large).
- $X_t^i \in \mathcal{X}$: State of subsystem $i \in \{1, \ldots, N\}$ at time t.
- $U_t^i \in \mathcal{U}$: Action of subsystem $i \in \{1, \ldots, N\}$ at time t.
- Mean-Field:

$$Z_t(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(X_t^i = x),\ x \in \mathcal{X} \qquad or \qquad Z_t = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_t^i}.$$

- All system variables are finite-valued.

Problem statement:

- Dynamics of subsystem i: $X^i_{t+1} = f_t(X^i_t, U^i_t, W^i_t, Z_t), \quad i \in \{1, \dots, N\}$.
- **Mean-field sharing** Information structure: $U^i_t = g^i_t(Z_{1:t}, X^i_t)$, where g^i_t is called control law of subsystem i at time t.
- Control strategy: The collection $\mathbf{g}^i = (g^i_1, \dots, g^i_T)$ of control laws of subsystem i over time is control strategy of subsystem i. The collection $\mathbf{g} = (\mathbf{g}^1, \dots, \mathbf{g}^N)$ of control strategies is control strategy of the system.
- Optimization problem: Let $X_t = (X^i_t)_{i=1}^N$ and $U_t = (U^i_t)_{i=1}^N$. We are interested in finding a strategy $\mathbf{g}$ that minimizes

$$J(\mathbf{g}) = \mathbb{E}^{\mathbf{g}} \left[\sum_{t=1}^{T} \ell_t(X_t, U_t) \right].$$

Problem Formulation

Assumptions:

(A1) Initial states $(X_1^i)_{i=1}^N$ are i.i.d. random variables.

(A2) Disturbances at time t, $(W_t^i)_{i=1}^N$, are i.i.d. random variables.

(A3) Let $X_t := (X_t^i)_{i=1}^N$ and $W_t := (W_t^i)_{i=1}^N$; then, $\{X_1, \{W_t\}_{t=1}^T\}$ are mutually independent.

(A4) All controllers use **identical control laws**.

Note that:

- (A1), (A2), and (A3) are standard assumptions in Markov decision problems.
- In general,(A4) leads to a loss in performance. However, it is a standard assumption in the literature on large scale systems for reasons of **simplicity**, **fairness**, and **robustness**.

Main Results

We identify a dynamic program to compute an optimal strategy. In particular,

Theorem 2:

Let ψ_t^* be a solution to the following dynamic program: at time t for every z_t

$$V_t(z_t) = \min_{\gamma_t}(\mathbb{E}[\ell_t(X_t, U_t) + V_{t+1}(Z_{t+1})|Z_t = z_t, \Gamma_t = \gamma_t])$$

where $\gamma_t : \mathcal{X} \to \mathcal{U}$ and $\gamma_t = \psi_t(z_t)$. Define $g_t^*(z, x) := \psi_t^*(z)(x), \forall x \in \mathcal{X}, \forall z$. Then, $\mathbf{g}^* = (g_1^*, \ldots, g_T^*)$ is an optimal strategy.

Salient feature of the model:

- Very few assumptions on the model.
- Allow for mean-field coupled dynamics.
- Allow for arbitrary coupled cost. (We do not assume cost to be weakly coupled.)

Main Results

Salient feature of the results:

- Computing **globally optimal** solution.
- Solution approach works for **arbitrary number of controllers**.
- State space of dynamic program increases **polynomially** (rather than exponentially) w.r.t. the number of controllers.
- Action space of dynamic program does not depend on the number of controllers.
- The size of information state does not increase with time; hence, the results naturally extend to **infinite horizon** under standard assumptions.
- The results extend naturally to randomized strategies by considering $\Delta(\mathcal{U})$ as the action space.
- Since the dynamic program is based on common information, each agent can independently solve the dynamic program and **compute the optimal strategy in a decentralized manner**.

Proof Approach

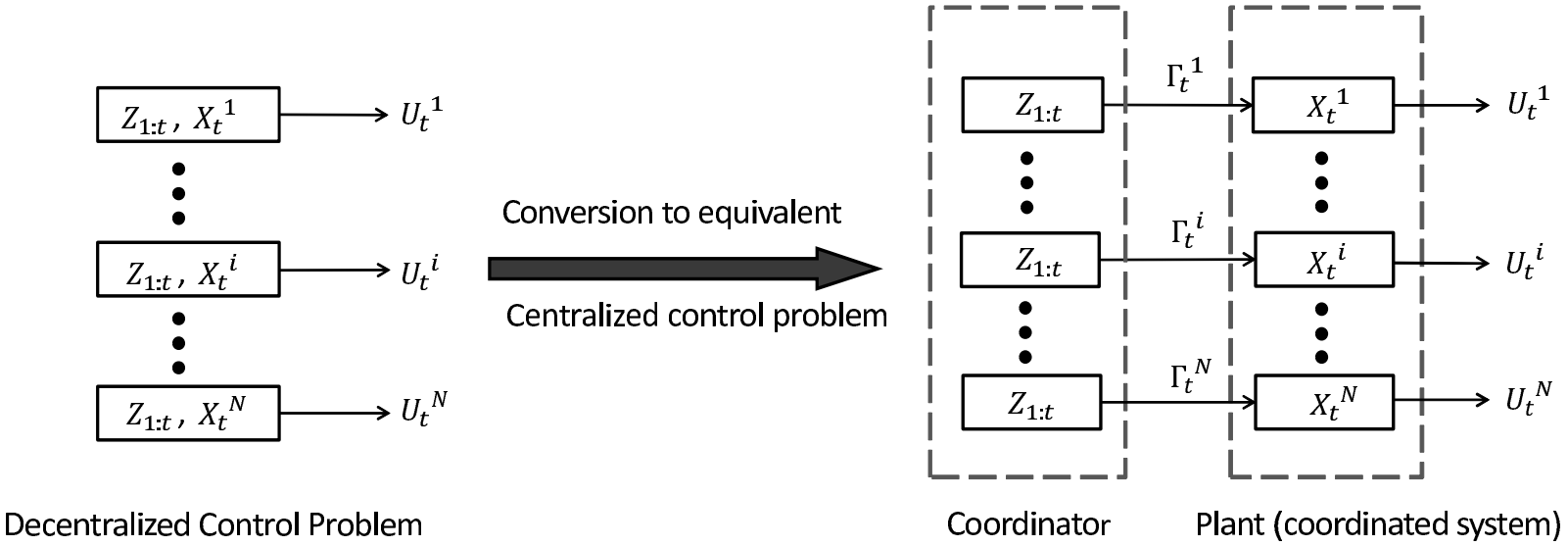
- **Step 1**: We follow common information approach [Nayyar, Mahajan, and Teneketzis 2013], and convert the decentralized control problem into a centralized control problem.

- **Step 2**: We exploit the symmetry of the problem (with respect to the controllers) to show that the **mean-field Z_t is an information state** for the centralized problem identified in Step 1. We then use this information state Z_t to obtain a dynamic programming decomposition.

Step 1: An Equivalent Centralized System

We define Γ_t and ψ_t as follows:

$$\Gamma_t(\cdot) := g_t(Z_{1:t}, \cdot), \Gamma_t : \mathcal{X} \mapsto \mathcal{U} \quad , \quad \Gamma_t = \psi_t(Z_{1:t}) := g_t(Z_{1:t}, \cdot).$$



Symmetric control laws assumption $g_t^i =: g_t, \forall i$, implies that $\Gamma_t^i =: \Gamma_t, \forall i$.

Equivalent Centralized Control Problem

The objective is to minimize

$$\hat{J}(\psi) = \mathbb{E}^{\psi}\left[\sum_{t=1}^{T} \ell_t(X_t, \Gamma_t(X_t^1), \ldots, \Gamma_t(X_t^N))\right].$$

Lemma 2:

For any choice $\gamma_{1:t}$ of $\Gamma_{1:t}$, any realization $z_{1:t}$ of $Z_{1:t}$, and any $x \in \mathcal{X}^N$,

$$\mathbb{P}(X_t = x | Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}) = \mathbb{P}(X_t = x | Z_t = z_t) = \frac{\mathbb{1}(x \in H(z_t))}{|H(z_t)|}$$

where $H(z) := \{x \in \mathcal{X}^N : \frac{1}{N} \sum_{i=1}^{N} \delta_{x^i} = z\}$.

Proof Outline:

- By induction, it is shown above conditional probability is **indifferent to permutation of** x; hence, mean-field is sufficient to characterize it.
- The latter property is proved using the symmetry of the model and the control laws.

Lemma 3:

The expected per-step cost may be written as a function of Z_t and Γ_t. In particular, there exists a function $\hat{\ell}_t$ (that does not depend on strategy $\boldsymbol{\psi}$) s.t.

$$\mathbb{E}[\ell_t(X_t, \Gamma_t(X_t^1), \dots, \Gamma_t(X_t^N)) | Z_{1:t}, \Gamma_{1:t}] =: \hat{\ell}_t(Z_t, \Gamma_t).$$

Proof Outline: Consider

$$\begin{aligned}
&\mathbb{E}[\ell_t(X_t, \Gamma_t(X_t^1), \dots, \Gamma_t(X_t^N)) | Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}] \\
&= \sum_{x} \ell_t(x, \gamma_t(x^1), \dots, \gamma_t(x^N)) \mathbb{P}(X_t = x | Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}) =: \hat{\ell}_t(Z_t, \Gamma_t).
\end{aligned}$$

Substituting the result of Lemma 2, and simplifying gives the result.

Step 2: Identifying an Information State

Lemma 4:

For any choice $\gamma_{1:t}$ of $\Gamma_{1:t}$, any realization $z_{1:t}$ of $Z_{1:t}$, and any z,

$$\mathbb{P}(Z_{t+1} = z | Z_{1:t} = z_{1:t}, \Gamma_{1:t} = \gamma_{1:t}) = \mathbb{P}(Z_{t+1} = z | Z_t = z_t, \Gamma_t = \gamma_t).$$

Also, the above conditional probability does not depend on strategy ψ.

Proof Outline: The result relies on the independence of the noise processes across subsystems and Lemma 2.

Theorem 1:

In the equivalent centralized problem, there is no loss of optimality in restricting attention to Markov strategy i.e. $\Gamma_t = \psi_t(Z_t)$. Furthermore, optimal policy $\boldsymbol{\psi}^*$ is obtained by solving the following dynamic program

$$V_t(z_t) = \min_{\gamma_t}(\hat{\ell}_t(z_t, \gamma_t) + \mathbb{E}[V_{t+1}(Z_{t+1})|Z_t = z_t, \Gamma_t = \gamma_t])$$

where $\gamma_t : \mathcal{X} \to \mathcal{U}$.

Proof Outline: Z_t is an information state for the equivalent centralized problem because:

- As shown in Lemma 3, the per-step cost can be written as a function of Z_t and Γ_t.
- As shown in Lemma 4, $\{Z_t\}_{t=1}^T$ a controlled Markov process with control action Γ_t.

Thus, the result follows from standard results in Markov decision theory.

Theorem 2:

Let ψ_t^* be a solution to the following dynamic program: at time t for every z_t

$$V_t(z_t) = \min_{\gamma_t}(\mathbb{E}[\ell_t(X_t, U_t) + V_{t+1}(Z_{t+1})|Z_t = z_t, \Gamma_t = \gamma_t])$$

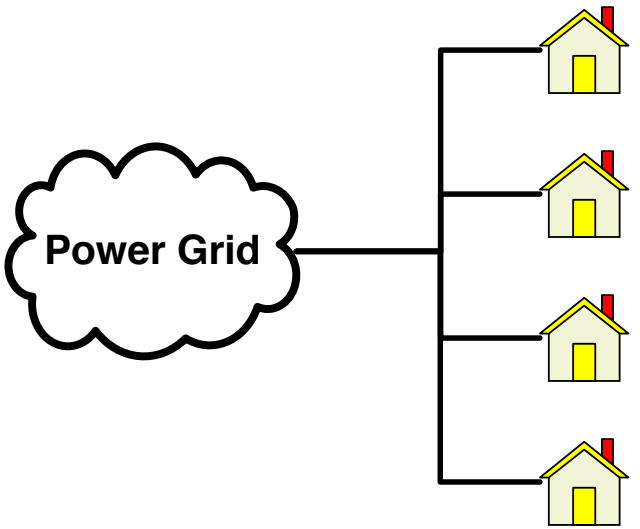
where $\gamma_t : \mathcal{X} \to \mathcal{U}$ and $\gamma_t = \psi_t(z_t)$. Define $g_t^*(z, x) := \psi_t^*(z)(x), \forall x \in \mathcal{X}, \forall z$. Then, $\mathbf{g}^* = (g_1^*, \ldots, g_T^*)$ is an optimal strategy.

Proof Outline:

In step 1, we converted the decentralized control problem to an equivalent centralized control problem. Now, we translate the answer of the equivalent centralized control problem back to that of the original decentralized control problem and obtain Theorem 2 from Theorem 1.

Example: Demand Response in Smart Grids

- $X_t^i \in \mathcal{X} = \{OFF, ON\}$
- $Z_t = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}(X_t^i = OFF)$
- Dynamics: $\mathbb{P}(X_{t+1}^i | X_t^i, U_t^i) =: [P(u_t^i)]_{x_t^i x_{t+1}^i}$
- Actions: $U_t^i \in \mathcal{U} = \{DoNothing, TurnOFF, TurnON\}$
- Cost of action: $C(U_t^i)$



- Objective: Keep the demand distribution Z_t close to a desired distribution ζ_t with minimum intervention such that following cost is minimized.

$$\mathbb{E}^{\mathbf{g}} \left[\sum_{t=1}^{\infty} \beta^{t-1} \left(\frac{1}{N} \sum_{i=1}^{N} C(U_t^i) + D(Z_t \parallel \zeta_t) \right) \right]$$

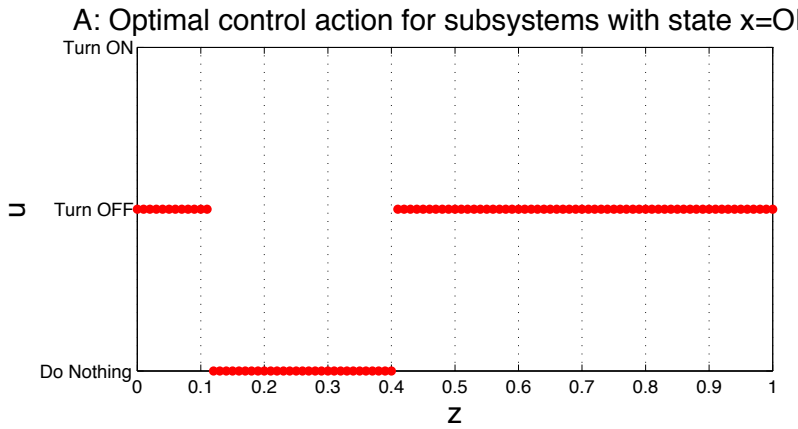
Numerical Result of the Example

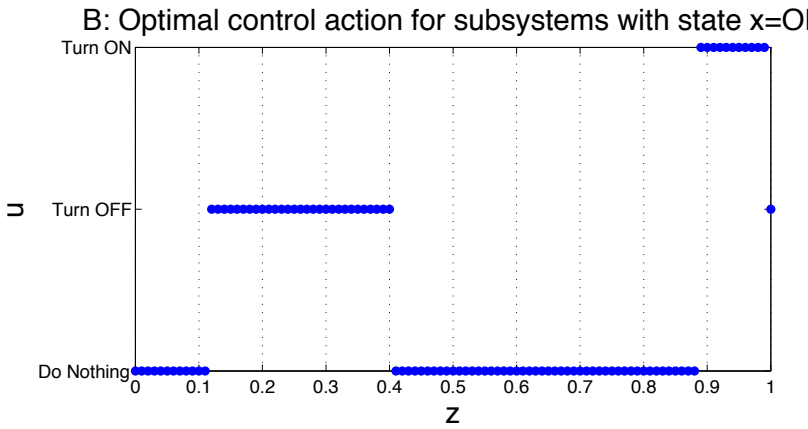
- Parameters:

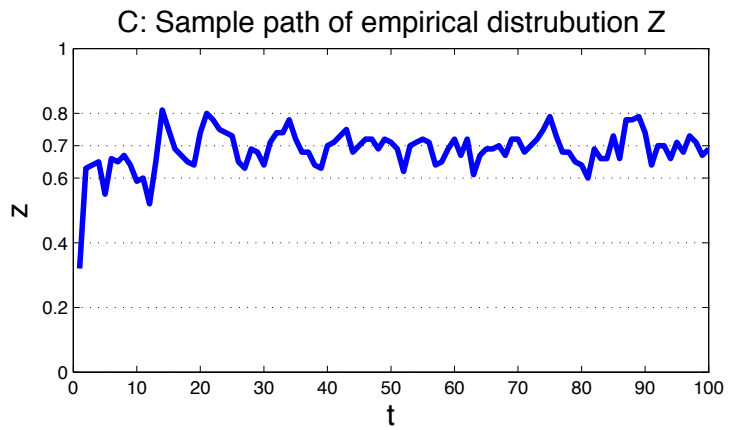
$$N = 100 \quad \beta = 0.9 \quad , \zeta_t = \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix},$$

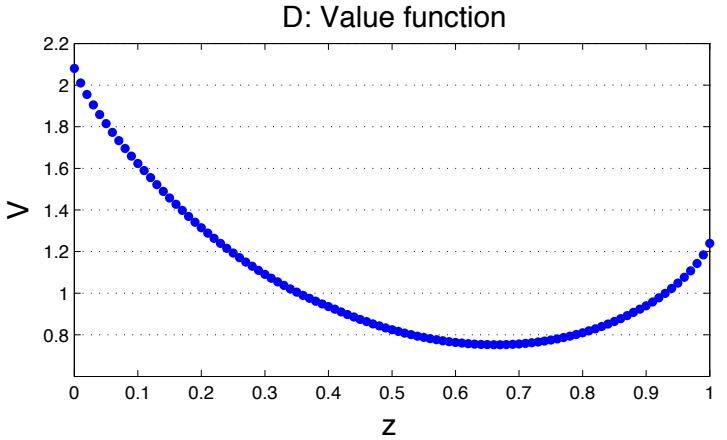
u	Do Nothing	Turn OFF	Turn ON
$c(u)$	0	0.1	0.2
$P(u)$	$\begin{bmatrix} 0.25 & 0.75 \\ 0.375 & 0.625 \end{bmatrix}$	$\begin{bmatrix} 0.85 & 0.15 \\ 0.875 & 0.125 \end{bmatrix}$	$\begin{bmatrix} 0.05 & 0.95 \\ 0.075 & 0.925 \end{bmatrix}$

- Optimal solution:









Generalization 1: Noisy Observation of Mean-Field

The solution methodology and dynamic programming decomposition extend to the scenario where all controllers observe a noisy version of the mean-field.

- $Y_t = h_t(Z_t, V_t)$: Noisy observation of the mean-field.
- $U_t^i = g_t(Y_{1:t}, X_t^i)$: Agents observe a noisy version of the mean-field.
- $\Pi_t = \mathbb{P}(Z_t | Y_{1:t}, \Gamma_{1:t})$: Information state for the coordinated system.
- A dynamic program is derived to obtain an optimal strategy. In particular,

Theorem 3:

Let ψ_t^* be a solution to the following dynamic program: at time t for every π_t

$$V_t(\pi_t) = \min_{\gamma_t}(\mathbb{E}[\ell_t(X_t, U_t) + V_{t+1}(\Pi_{t+1}) | \Pi_t = \pi_t, \Gamma_t = \gamma_t])$$

where $\gamma_t : \mathcal{X} \to \mathcal{U}$ and $\gamma_t = \psi_t(\pi_t)$. Define $g_t^*(\pi, x) := \psi_t^*(\pi)(x), \forall x \in \mathcal{X}, \forall \pi$. Then, $\mathbf{g}^* = (g_1^*, \ldots, g_T^*)$ is an optimal strategy.

Generalization 2: Multiple type of Subsystems

So far, we have assumed homogeneous subsystems. Our results generalize to multiple types where subsystem i has a type $k \in \{1, \dots, K\}$.

- **Dynamics** of subsystem i depends on its type k: $X^i_{t+1} = f^k_t(X^i_t, U^i_t, W^i_t, Z_t)$.
- **Mean-field**: $Z_t = (Z^1_t, \dots, Z^K_t)$, where Z^k_t is the mean-field of subsystems with type k.
- **Control law** of subsystem i depends on its type k: $U^i_t = g^k_t(Z_{1:t}, X^i_t)$.
- Empirical distribution of number of types is common knowledge between subsystems.
- Subsystems are **arbitrarily** coupled in the cost.

Theorem 4:

Let ψ^*_t be a solution to the following dynamic program: at time t for every z_t

$$V_t(z_t) = \min_{\gamma_t}(\mathbb{E}[\ell_t(X_t, U_t) + V_{t+1}(Z_{t+1})|Z_t = z_t, \Gamma_t = \gamma_t])$$

where $\gamma_t = (\gamma^1_t, \dots, \gamma^K_t)$ and $\gamma^k_t : \mathcal{X}^k \to \mathcal{U}^k$. Define

$$g^{*,k}_t(z, x) := \psi^{*,k}_t(z)(x), \forall x \in \mathcal{X}^k, \forall z.$$

Then, $\mathbf{g}^* = (g^{*,1}, \dots, g^{*,K})$ is an optimal strategy, where $\mathbf{g}^{*,k} = (g^{*,k}_1, \dots, g^{*,k}_T)$.

Generalization 3: Major Minor Setup

Consider **one major** subsystem distinguished by index 0 and N **minor** subsystems.

- **Dynamics** : $X^0_{t+1} = f^0_t(X^0_t, U^0_t, W^0_t, Z_t)$ and $X^i_{t+1} = f_t(X^i_t, U^i_t, W^i_t, Z_t, X^0_t)$.
- **Mean-field**:Z_t is the mean-field of **minor subsystems.**
- **Control laws**: $U^0_t = g^0_t(Z_{1:t}, X^0_{1:t})$ and $U^i_t = g_t(Z_{1:t}, X^0_{1:t}, X^i_t)$.
- Subsystems are **arbitrarily** coupled in the cost.

Theorem 5:

Let ψ^*_t be a solution to the following dynamic program: at time t for every z_t

$$V_t(z_t, x^0_t) = \min_{u^0_t, \gamma_t} (\mathbb{E}[\ell_t(X^0_t, X_t, U^0_t, U_t) + V_{t+1}(Z_{t+1}, X^0_{t+1}) | , Z_t = z_t, \Gamma_t = \gamma_t, X^0_t = x^0_t, U^0_t = u^0_t])$$

where $\gamma_t : \mathcal{X} \to \mathcal{U}$. Define

- $g^{*,0}_t(z, x^0) := \psi^{*,1}_t(z, x^0), \forall x \in \mathcal{X}^0, \forall z.$
- $g^*_t(z, x^0, x) := \psi^{*,2}_t(z, x^0)(x), \forall x \in \mathcal{X}, \forall x \in \mathcal{X}^0, \forall z.$

Then, $(g^{*,0}, g^*)$ is an optimal strategy, where $\mathbf{g}^{*,0} = (g^{*,0}_1, \ldots, g^{*,0}_T)$ and $\mathbf{g}^* = (g^*_1, \ldots, g^*_T)$.

Summary

- We identified a dynamic program that obtains a **global optimal** strategy for **arbitrary number** of controllers.
- The state space of dynamic program increases **polynomially** (rather than exponentially) w.r.t. the number of controllers.
- We illustrated our approach by an example in smart grids with $N = 100$ subsystems.
- The results naturally extend to **infinite horizon** and **randomized strategies**.
- We showed that the results generalize to **noisy mean-field**, **multiple types**, and **major-minor setup**.

- The proposed setup is practical, because it is:
  - **Realistic:** There are very few assumptions imposed on the model. In many real-world applications such as smart grids, social networks, etc., the assumed symmetry is reasonable (even desirable) for reasons of fairness, robustness, and simplicity.
  - **Implementable:** Mean-field sharing information structure is **physically** and **economically** efficient.
  - **Solvable:** The solution approach is **computationally** efficient.

Thank You