Team Optimal Control with Mean-Field Sharing

Jalal Arabneydi and Aditya Mahajan
ECE Department, McGill University

Motivation

- Team optimal control of decentralized stochastic systems arises in many applications ranging from **networked control systems**, **robotics**, **communication networks**, **transportation networks**, **sensor networks**, **economics**, and **smart grids**.

Figure 1: Smart Grids; taken from www.3m.com.

- In general, these problems belong to **NEXP complexity** class which are harder than NP problems.
- There is no solution methodology for general (infinite-horizon) decentralized control systems.

Model

- $n \in \mathbb{N}$ homogeneous subsystems, $t \in \mathbb{N}$ denotes time.
- State and action of subsystem i: $X_t^i \in \mathcal{X}$ and $U_t^i \in \mathcal{U}$, respectively.
- Mean-field of system: $Z_t = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_t^i}$.
- System dynamics (weakly coupled): $X_{t+1}^i = f_t(X_t^i, U_t^i, W_t^i, Z_t)$, where noise $\{W_t^i\}_{t=1}^{T}$ is an independent process.
- Information structure: $U_t^i = g_t^i(Z_{1:t}, X_t^i)$.
- **Assumption**: identical control laws i.e. $g_t^i = g_t^j = g_t, \forall i, j$.
  In large-scale systems, it is reasonable to treat subsystems (with minor differences) identically, for the purposes of **simplicity**, **robustness**, and **fairness**.
- Cost structure for horizon T given control strategy $\mathbf{g} := (g_1, \dots, g_T)$: $J(\mathbf{g}) = \mathbb{E}^{\mathbf{g}}\left[\sum_{t=1}^{T} \ell_t(X_t^1, \dots, X_t^n, U_t^1, \dots, U_t^n)\right]$.
- Optimization problem: $min_{\mathbf{g}} J(\mathbf{g})$.

Main Results

- Let $\Gamma_t : \mathcal{X} \mapsto \mathcal{U}$ denote the mapping from $\mathcal{X}$ to $\mathcal{U}$ and $U_t^i = \Gamma_t(X_t^i)$.
  - The expected per-step cost may be written as a function of Z_t and Γ_t:
  $$\mathbb{E}[\ell_t(X_t^1, \dots, X_t^n, U_t^1, \dots, U_t^n) | Z_{1:t}, \Gamma_{1:t}] =: \hat{\ell}_t(Z_t, \Gamma_t). \quad (1)$$
  - Mean-field evolves in Markovian manner:
  $$\mathbb{P}(Z_{t+1} | Z_{1:t}, \Gamma_{1:t}) = \mathbb{P}(Z_{t+1} | Z_t, \Gamma_t). \quad (2)$$
- **Main Theorem:**
  - Each controller can drop $Z_{1:t-1}$ without loss of optimality that is we may restrict attention to controllers of the form $U_t^i = g_t(Z_t, X_t^i)$.
  - Let $\psi_t^*(z_t)$ denote any argmin of the right-hand side of following dynamic program. For $t = T, \dots, 1$, and for z_t,
  $$V_t(z_t) := \min_{\gamma_t}(\hat{\ell}_t(z_t, \gamma_t) + \mathbb{E}[V_{t+1}(Z_{t+1}) | Z_t = z_t, \Gamma_t = \gamma_t]) \quad (3)$$
  where $V_{T+1}(z_{T+1}) = 0, \forall z_{T+1}$, and the minimization is over all functions $\gamma_t : \mathcal{X} \to \mathcal{U}$. Define $g_t^*(z, x) := \psi_t^*(z)(x)$. Then, $\mathbf{g}^* = (g_1^*, \dots, g_T^*)$ is an optimal strategy.

Generalizations

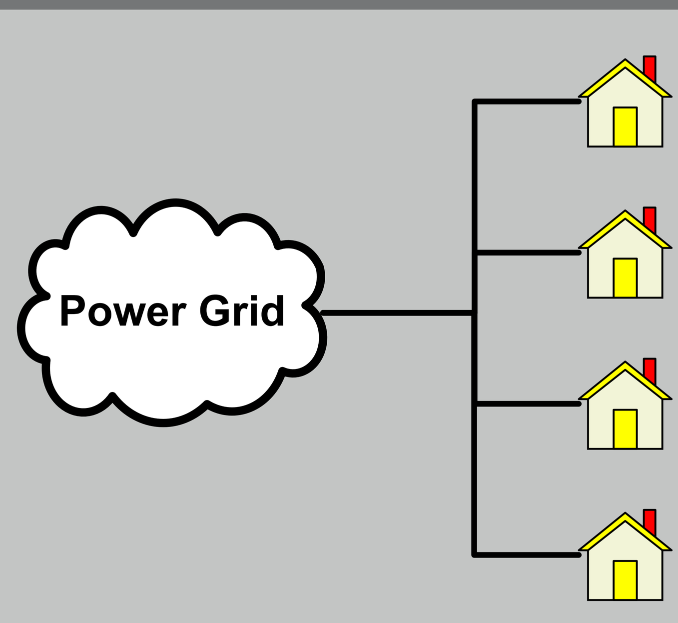
- Dynamic coupling through joint mean-field of state and action
- Correlated primitive random variables across space
- Randomized strategies
- Infinite horizon
- Multiple types
- Noisy observation of mean-field
- More general information structure
- Reinforcement learning algorithms such as Q-learning

Salient Features

- Above dynamic program obtains **globally optimum** control strategies.
- The size of the corresponding information state **does not increase with time**. Therefore, the results extend to infinite-horizon control setups.
- The size of the corresponding information state **increases polynomially with the number of controllers** rather than exponentially. This allows us to solve problems with moderate number of controllers.
- Since the dynamic program is based on common observed data, each agent can independently solve the dynamic program and **compute the optimal strategy in a decentralized manner**.
- Mean-field of the system can be **computed and communicated easily**.

Example 1: Demand Response

- $X_t^i \in \mathcal{X} = \{OFF, ON\}$
- $Z_t = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_t^i = OFF)$
- 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: $U_t^i \in \mathcal{U} = \{FREE, OFF, ON\}$
- Cost of action: $C(U_t^i)$



- Objective: Keep the demand distribution Z_t close to a desired distribution ζ_t with minimum force such that following cost is minimized.
$$\mathbb{E}^{\mathbf{g}}\left[\sum_{t=1}^{\infty} \beta^t \left(\frac{1}{n}\sum_{i=1}^{n} C(U_t^i) + D(Z_t \parallel \zeta_t)\right)\right]$$
- Numerical result:
  - Parameters

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

u	FREE	OFF	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

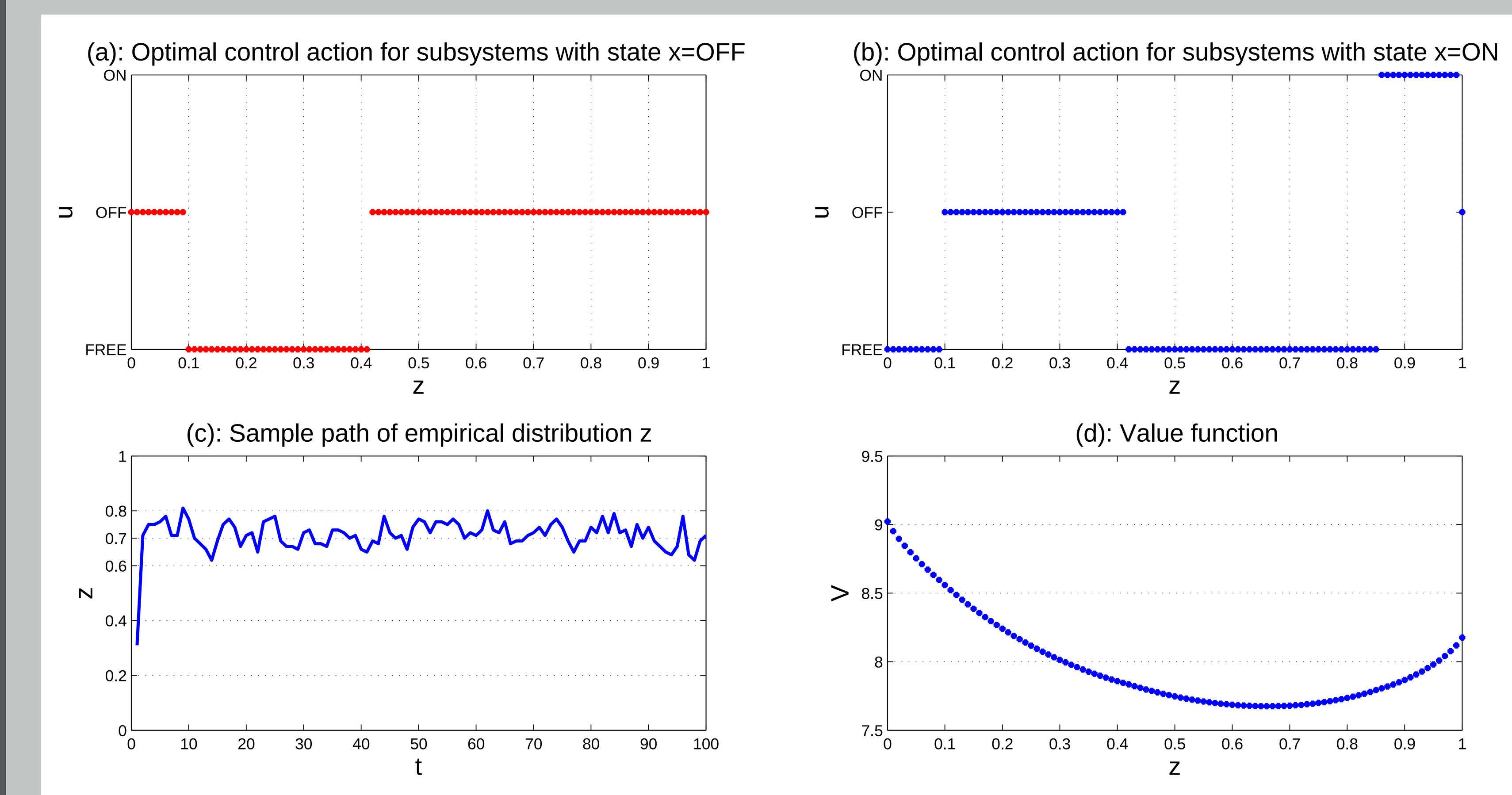
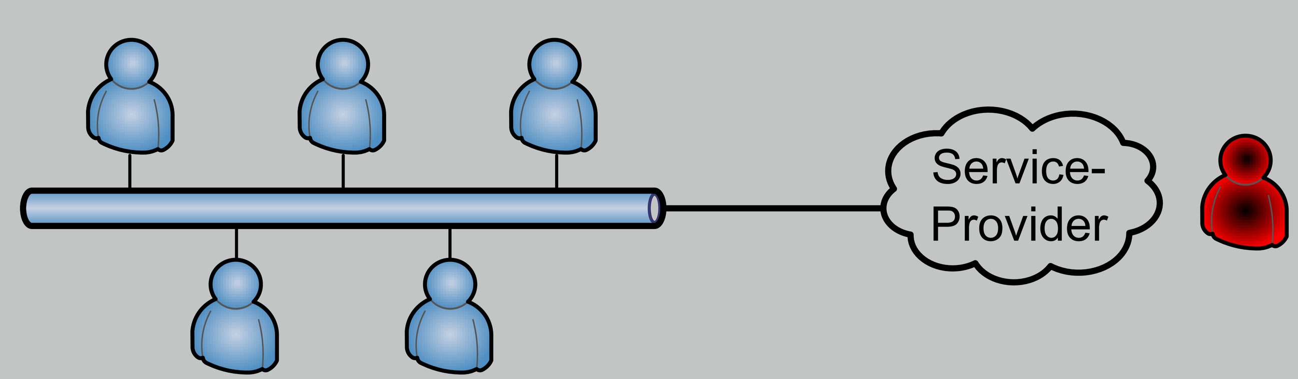


Figure 2: This figure displays numerical results associated with time-invariant reference ζ.

Example 2: Service-provider and customers

- Objective: Find a team optimal strategy such that the service is not only profitable but also customer-satisfactory.



- In this example, we use one of the generalizations known as team optimal control of coupled major-minor subsystems with mean-field sharing.

References

[1] Jalal Arabneydi and Aditya Mahajan.
Team optimal control of coupled subsystems with mean-field sharing.
Accepted in Conference on Decision and Control (CDC), 2014.

[2] Jalal Arabneydi and Aditya Mahajan.
Team optimal control of coupled major-minor subsystems with mean-field sharing.
Accepted to Indian Control Conference (ICC), 2014.

Email:jalal.arabneydi@mail.mcgill.ca
Email:aditya.mahajan@mcgill.ca