Mean-field teams

Aditya Mahajan
McGill University

Joint work with Jalal Arabneydi

CommNets Seminar
University of Southern California, 26 Aug 2015

What is team theory?

A brief overview of decision making

Decision making by a single agent

Static optimization

$$\min_{u \in \mathcal{C}} c(u)$$

- Linear programming
- Convex optimization
- Non-convex optimization

Bayesian optimization

$$\min_{g} \mathbb{E}[c(\omega, g(Y(\omega)))]$$

- Stochastic programming
- Stochastic approximation
- Markov Chain Monte Carlo

Dynamic optimization/ Stochastic control

$$\min_{(g_1, \dots, g_T)} \mathbb{E}\left[\sum_{t=1}^{T} c_t(x_t, u_t)\right]$$

where

$$\begin{aligned} x_{t+1} &= f_t(x_t, u_t, W_t), \\ y_t &= h_t(x_t, N_t), \\ u_t &= g_t(y_{1:t}, u_{1:t-1}) \end{aligned}$$

- Dynamic programming
- Pontryagin maximum principle
- Multi-stage stochastic programming

Decision making by multiple agents

Game theory

Each agent has an individual objective. Agents compete to minimize individual costs.

- Static games
- Bayesian games
- Dynamic games with imperfect information

Decision making by multiple agents

Game theory

Each agent has an individual objective. Agents compete to minimize individual costs.

- Static games
- Bayesian games
- Dynamic games with imperfect information

Team theory/ Decentralized stochastic control

All agents have a common objective. Agents cooperate to minimize team costs.

- Static (Bayesian) teams
- Dynamic teams or decentralized stochastic control

Research in team theory started in Economics in mid 50's in the context of organizational behaviour. It has been studied in Systems and Control since the late 60's and in Artificial Intelligence since late 90's.

Decision making by multiple agents

Game theory

Each agent has an individual objective. Agents compete to minimize individual costs

tion

The motivation of team theory/ decentralized control is not that it is more powerful than centralized control;

rather it is necessary in systems where centralized information is not available or is not practical.

ts cooperate

Dec

stic control

Research in team theory started in Economics in mid 50's in the context of organizational behaviour. It has been studied in Systems and Control since the late 60's and in Artificial Intelligence since late 90's.


Common theme: multi-stage multi-agent
decision making under uncertainty

Conceptual difficulties in dynamic teams

Witsenhausen Counterexample

- A two step dynamical system with two controllers
- Linear dynamics, quadratic cost, and Gaussian dist.
- Non-linear controllers outperform linear control strategies: cannot use Kalman filtering + Riccati eqn

Whittle and Rudge Example

- Infinite horizon system with two symmetric controllers
- Linear dynamics, quadratic cost, and Gaussian dist.
- A priori restrict attention to linear controllers
- Best linear controllers not representable by recursions of finite order

Complexity analysis

- All random variables are finite valued
- Finite horizon setup
- The problem of finding the best control strategy is in NEXP

- Witsenhausen, "A counterexample in stochastic optimum control," SICON 1969.
- Whittle and Rudge, "The optimal linear solution of a symmetric team control problem," App. Prob. 1974.
- Bernstein, et al, "The complexity of decentralized control of Markov decision processes," MOR 2002.

Brief overview of research in team theory

Research theme

- Identify specific information structures that capture key features of applications but, at the same time, are amenable to analysis.
- Develop analytic and computation approaches to optimally design controllers for these information structures.

Solution Approaches

- The person-by-person approach
  Radnar (1962), Ho (1970's)
- The designer's approach
  Witsenhausen (1970's), Mahajan (2008)
- The common-information approach
  Nayyar (2011), Nayyar Mahajan Tenenketzis (2013)

Specific Information Structures

Delayed state sharing, delayed observation sharing, control sharing, periodic sharing, belief sharing, . . .

Mean-field teams

Decentralized multi-agent systems with mean-field coupled dynamics

Motivating examples: Systems with homogeneous agents

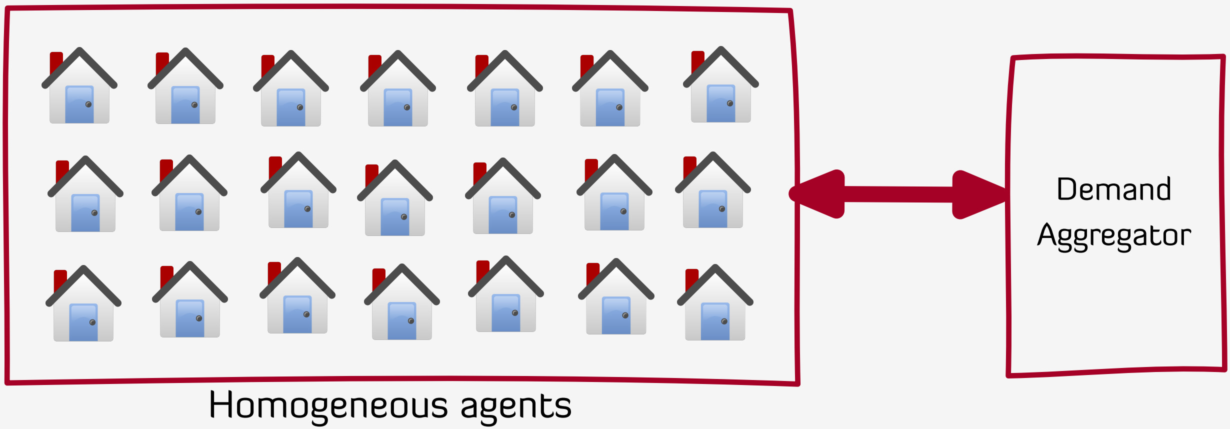
Homogeneous agents

- Population of homogeneous agents
- Influence the dynamics of each other through their mean-field (or empirical distribution)
- Equivalent to an interacting particle model
- Arbitrarily coupled cost

Motivating examples: Systems with homogeneous agents

Homogeneous agents

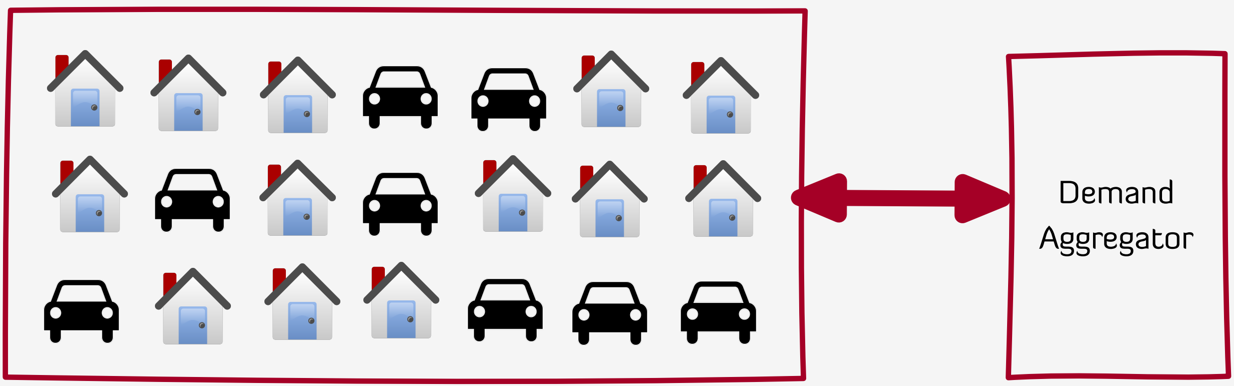
- Population of homogeneous agents
- Influence the dynamics of each other through their mean-field (or empirical distribution)
- Equivalent to an interacting particle model
- Arbitrarily coupled cost



Motivating examples: Systems with homogeneous agents

Homogeneous agents

- Population of homogeneous agents
- Influence the dynamics of each other through their mean-field (or empirical distribution)
- Equivalent to an interacting particle model
- Arbitrarily coupled cost

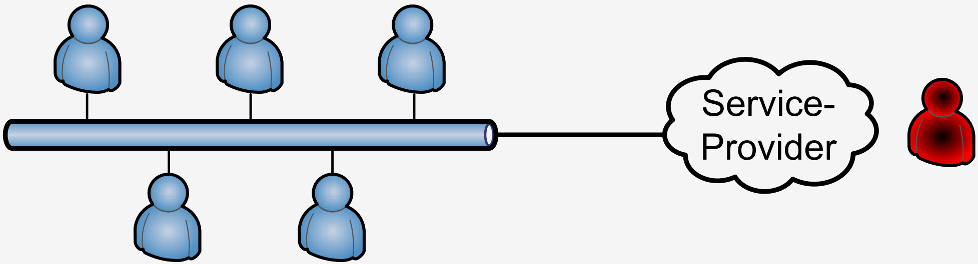


Homogeneous agents with multiple types

Motivating examples: Systems with homogeneous agents

Homogeneous agents

- Population of homogeneous agents
- Influence the dynamics of each other through their mean-field (or empirical distribution)
- Equivalent to an interacting particle model
- Arbitrarily coupled cost



Major and minor homogeneous agents

Decentralized multi-agent systems with mean-field coupled dynamics

Is it possible to obtain a scalable dynamic programming decomposition?

Mean-field approximation: Literature overview

Mean-field approximation in statistical physics (Weiss 1907; Landau 1937)

> It is a well-known phenomenon in many branches of the exact and physical sciences that very great numbers are often easier to handle than those of medium size. An almost exact theory of a gas, containing about 10^{25} freely moving particles, is incomparably easier than that of the solar system, made up of 9 major bodies... This is, of course, due to the excellent possibility of applying the laws of statistics and probabilities in the first case.
>
> — von Neumann and Morgenstern,
> Theory of Games and Economic Behavior §2.4.2

Mean-field approximation: Literature overview

Mean-field approximation in statistical physics (Weiss 1907; Landau 1937)

- . . .

Mean-field approximations in Game Theory

- Jovanovic Rosenthal 1988
- Bergin Bernhardt 1995
- Weintraub Benkard Van Roy 2008
- . . .

Mean-field approximation: Literature overview

Mean-field approximation in statistical physics (Weiss 1907; Landau 1937)

- . . .

Mean-field approximations in Game Theory

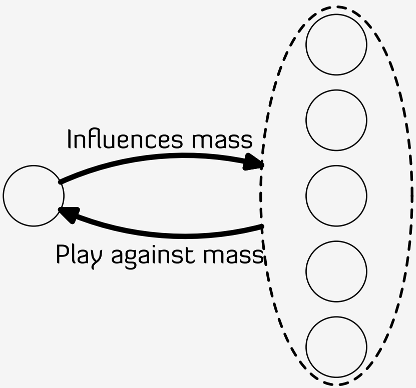
- Jovanovic Rosenthal 1988
- Bergin Bernhardt 1995
- Weintraub Benkard Van Roy 2008
- . . .

Mean-field approximations in Systems and Control

- Huang Caines Malhalmé 2003, . . .
- Larsy Lions 2006, . . .
- . . .

Mean-field approximation

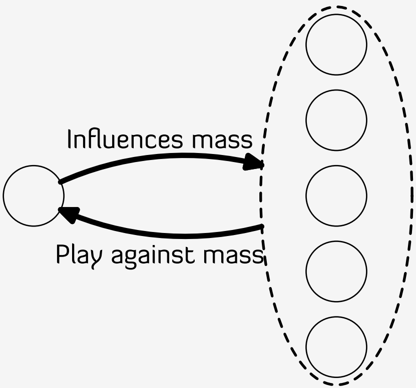
The main idea



- Assume asymptotically large population.
- Instead of keeping track of the individual behavior of agents, each agent keeps track of the mean behavior of all agents (or the mass) and chooses a strategy that is the best response to mean-field of the mass.
- Given the strategy of each agent, a collective mean-field behavior of the mass emerges.
- The mean-field equilibrium is a consistent set of strategies of the agents and mean-field of the mass.

Mean-field approximation

The main idea



- Assume asymptotically large population.
- Instead of keeping track of the individual behavior of agents, each agent keeps track of the mean behavior of all agents (or the mass) and chooses a strategy that is the best response to mean-field of the mass.
- Given the strategy of each agent, a collective mean-field behavior of the mass emerges.
- The mean-field equilibrium is a consistent set of strategies of the agents and mean-field of the mass.

Our motivation

- There are two phenomenon at play in mean-field approximations: symmetry (agents behave identically) and asymptotics (LLN). **Can we separate the two?**
- For dynamical systems, mean-field approximations give Nash equilibrium or PBPO solution. **Can we obtain team optimal solutions?**

Model and Problem Formulation

Mean-field of a sequence

Given a finite alphabet $\mathcal{X}$ and a positive integer n, define:

Let Δ_n denote the set of probability distribution on $\mathcal{X}$ with denominator n

- Note that $|\Delta_n| \leqslant (n+1)^{|\mathcal{X}|}$.

Example

Let $\mathcal{X} = \{0, 1\}$ and $n = 3$. Then,

$$\Delta_n = \left\{ \begin{bmatrix} 0 & 1 \end{bmatrix}, \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}, \begin{bmatrix} 1 & 0 \end{bmatrix} \right\}.$$

Mean-field of a sequence

Given a finite alphabet $\mathcal{X}$ and a positive integer n, define:

Let Δ_n denote the set of probability distribution on $\mathcal{X}$ with denominator n

- Note that $|\Delta_n| \leqslant (n+1)^{|\mathcal{X}|}$.

For a sequence $\boldsymbol{x} \in \mathcal{X}^n$, let $\xi(\boldsymbol{x}) \in \Delta_n$ denote the empirical distribution of $\boldsymbol{x}$

- We refer to $\xi(x)$ as the mean-field of a sequence.
- In information theory, $\xi(\boldsymbol{x})$ is called the type of a sequence.

Example

Let $\mathcal{X} = \{0, 1\}$, $n = 3$, and $\boldsymbol{x} = (0, 0, 1)$. Then,

$$\xi(x) = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

Mean-field of a sequence

Given a finite alphabet $\mathcal{X}$ and a positive integer n, define:

Let Δ_n denote the set of probability distribution on $\mathcal{X}$ with denominator n

- Note that $|\Delta_n| \leqslant (n+1)^{|\mathcal{X}|}$.

For a sequence $\mathbf{x} \in \mathcal{X}^n$, let $\xi(\mathbf{x}) \in \Delta_n$ denote the empirical distribution of $\mathbf{x}$

- We refer to $\xi(x)$ as the mean-field of a sequence.
- In information theory, $\xi(\mathbf{x})$ is called the type of a sequence.

For a distribution $z \in \Delta_n$, let $\Xi(z) \subset \mathcal{X}^n$ denote all sequences with mean-field z

- We refer to $\Xi(z)$ as the mean-field class of a distribution.
- In information theory, $\Xi(z)$ is called type class of a distribution.

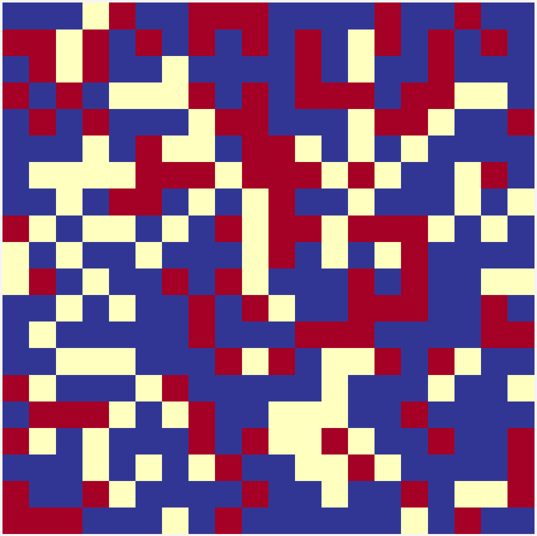
Example

Let $\mathcal{X} = \{0, 1\}$, $n = 3$, and $z = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \end{bmatrix}$. Then,

$$\Xi(z) = \{(0,0,1), (0,1,0), (1,0,0)\}.$$

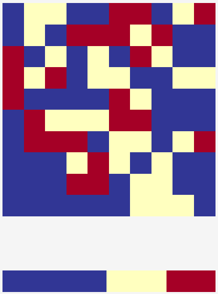
An example: color swatches

$\boldsymbol{x}$



$\xi(\boldsymbol{x})$

Mean-field coupled dynamical systems



Model

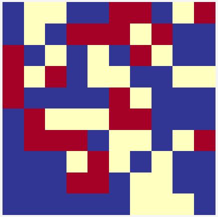
A finite population $\mathcal{N}$ of homogeneous agents.

X^n_t : State of agent n at time t

U^n_t : Control action of agent n at time t

$\xi(\mathbf{X}_t)$: Mean-field of the population at time t

Mean-field coupled dynamical systems



Model

A finite population $\mathcal{N}$ of homogeneous agents.

X^n_t : State of agent n at time t

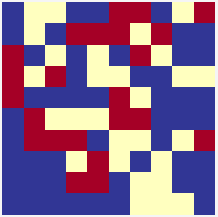
U^n_t : Control action of agent n at time t

$\xi(\mathbf{X}_t)$: Mean-field of the population at time t

Dynamics

$$X^n_{t+1} = f(X^n_t, \xi(\mathbf{X}_t), U^n_t, W^n_t)$$

Mean-field coupled dynamical systems



Model

A finite population $\mathcal{N}$ of homogeneous agents.

X^n_t : State of agent n at time t

U^n_t : Control action of agent n at time t

$\xi(\mathbf{X}_t)$: Mean-field of the population at time t

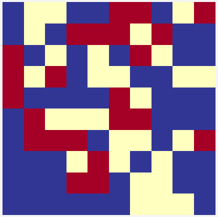
Dynamics

$$X^n_{t+1} = f(X^n_t, \xi(\mathbf{X}_t), U^n_t, W^n_t)$$

Per-stage Cost

Arbitrary coupled cost: $\ell(\mathbf{X}_t, \mathbf{U}_t)$

Mean-field coupled dynamical systems



Model

A finite population $\mathcal{N}$ of homogeneous agents.

X^n_t : State of agent n at time t

U^n_t : Control action of agent n at time t

$\xi(\mathbf{X}_t)$: Mean-field of the population at time t

Dynamics

$$X^n_{t+1} = f(X^n_t, \xi(\mathbf{X}_t), U^n_t, W^n_t)$$

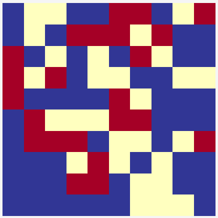
Per-stage Cost

Arbitrary coupled cost: $\ell(\mathbf{X}_t, \mathbf{U}_t)$

Information Structure

$$U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t})).$$

Mean-field coupled dynamical systems



Model

A finite population $\mathcal{N}$ of homogeneous agents.

X^n_t : State of agent n at time t

U^n_t : Control action of agent n at time t

$\xi(\mathbf{X}_t)$: Mean-field of the population at time t

Dynamics

$$X^n_{t+1} = f(X^n_t, \xi(\mathbf{X}_t), U^n_t, W^n_t)$$

Per-stage Cost

Arbitrary coupled cost: $\ell(\mathbf{X}_t, \mathbf{U}_t)$

Information Structure

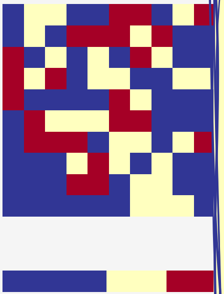
$$U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t})).$$

Objective

Identify strategies $\mathbf{g} = (\mathbf{g}^n)_{n \in \mathcal{N}}$,
where $\mathbf{g}^n = (g^n_1, \dots, g^n_T)$ to minimize

$$\mathbb{E}\Big[\sum_{t=1}^{T} \ell(\mathbf{X}_t, \mathbf{U}_t)\Big]$$

Mean-field coupled dynamical systems



Results extend naturally to
populations with multiple types.

Restrict to homogeneous population
for simplicity of presentation.

Per-stage Cost	Arbitrary coupled cost: $\ell(\mathbf{X}_t, \mathbf{U}_t)$
Information Structure	$U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t}))$.
Objective	Identify strategies $\mathbf{g} = (\mathbf{g}^n)_{n \in \mathcal{N}}$, where $\mathbf{g}^n = (g^n_1, \dots, g^n_T)$ to minimize $$\mathbb{E}\Big[\sum_{t=1}^{T} \ell(\mathbf{X}_t, \mathbf{U}_t)\Big]$$

Assumptions on the model

Assumption (A1)

The primitive random variables:

- initial states $(X_1^n)_{n \in \mathcal{N}}$ of all agents
- process noises $\{(W_t^n)_{n \in \mathcal{N}}\}_{t=1}^{T}$

are indepedent

Furthermore the initial states $(X_1^n)_{n \in \mathcal{N}}$ and the process noise $(W_t^n)_{n \in \mathcal{N}}$ of the minor subsystem are identically distributed.

Assumptions on the model

Assumption (A1)

The primitive random variables:

- initial states $(X_1^n)_{n \in \mathcal{N}}$ of all agents
- process noises $\{(W_t^n)_{n \in \mathcal{N}}\}_{t=1}^{T}$

are indepedent

Furthermore the initial states $(X_1^n)_{n \in \mathcal{N}}$ and the process noise $(W_t^n)_{n \in \mathcal{N}}$ of the minor subsystem are identically distributed.

Assumption (A2)

All agents use identical control strategies

Identical control strategies are not optimal in a system with homogeneous agents

Identical control strategies are not optimal in a system with homogeneous agents

A counterexample

- N homogeneous agents and horizon $T = 2$.
- $\mathcal{X} = \mathcal{U} = \{1, \dots, N\}$
- $X_1^n \sim \text{Unif}\{1, \dots, N\}$.
- Dynamics: $X_2^n = U_1^n$

$$\ell_1(\mathbf{x}_1, \mathbf{u}_1) = 0 \quad \text{and} \quad \ell_2(\mathbf{x}_2, \mathbf{u}_2) = k\,\mathbb{1}[\xi(\mathbf{x}_2) \neq \{\tfrac{1}{N}, \dots \tfrac{1}{N}\}]$$

Identical control strategies are not optimal in a system with homogeneous agents

A counterexample

- N homogeneous agents and horizon $T = 2$.
- $\mathcal{X} = \mathcal{U} = \{1, \dots, N\}$
- $X_1^n \sim \text{Unif}\{1, \dots, N\}$.
- Dynamics: $X_2^n = U_1^n$

$$\ell_1(\mathbf{x}_1, \mathbf{u}_1) = 0 \quad \text{and} \quad \ell_2(\mathbf{x}_2, \mathbf{u}_2) = k\, \mathbb{1}[\xi(\mathbf{x}_2) \neq \{\tfrac{1}{N}, \dots \tfrac{1}{N}\}]$$

If all agents use identical strategy

Total cost $= k\, \mathbb{E}[\xi(\mathbf{X}_2) \neq \{\tfrac{1}{N}, \dots \tfrac{1}{N}\}] > 0$

Identical control strategies are not optimal in a system with homogeneous agents

A counterexample

- N homogeneous agents and horizon $T = 2$.
- $\mathcal{X} = \mathcal{U} = \{1, \dots, N\}$
- $X_1^n \sim \text{Unif}\{1, \dots, N\}$.
- Dynamics: $X_2^n = U_1^n$

$$\ell_1(\mathbf{x}_1, \mathbf{u}_1) = 0 \quad \text{and} \quad \ell_2(\mathbf{x}_2, \mathbf{u}_2) = k\,\mathbb{1}[\xi(\mathbf{x}_2) \neq \{\tfrac{1}{N}, \dots \tfrac{1}{N}\}]$$

If all agents use identical strategy

$$\text{Total cost } = k\,\mathbb{E}[\xi(\mathbf{X}_2) \neq \{\tfrac{1}{N}, \dots \tfrac{1}{N}\}] > 0$$

Optimal strategy

$$g_1^n(x_1^n, \xi(\mathbf{x}_1)) = n. \qquad \text{Total cost } = 0.$$

Assuming that all agents use identical control strategies leads to loss of performance

The assumption is without loss for certain models (e.g., LQG) and for asymptotically large population

Ensures simplicity, fairness, and robustness

The main results

Structure of optimal strategies

Under (A1) and (A2), there is no loss of optimality is using control strategies of the form:

$$U_t^n = g_t(X_t^n, \xi(\mathbf{X}_t))$$

instead of

$$U_t^n = g_t^n(X_t^n, \xi(\mathbf{X}_{1:t})).$$

Salient Features

- By (A2), all agents use identical control strategies.
- Each agent only uses the the current mean-field instead of the history of mean-fields.

The main results

Structure of optimal strategies

Under (A1) and (A2), there is no loss of optimality is using control strategies of the form:

$$U^n_t = g_t(X^n_t, \xi(\mathbf{X}_t))$$

instead of

$$U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t})).$$

Dynamic programming decomposition

For a given $z \in \Delta_n$, define $\gamma_t \colon \mathcal{X} \to \mathcal{U}$ as $\gamma_t(\cdot) = g_t(\cdot, z)$.

Recursively define $V_t \colon \Delta_n \to \mathbb{R}$ as follows:
$V_{T+1}(z) = 0$, and for $t = T, \dots, 1$,

$$V_t(z) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\xi(\mathbf{X}_{t+1})) \mid \xi(\mathbf{X}_t) = z, \{U^n_t = \gamma_t(X^n_t)\}_{n \in \mathcal{N}}\right]$$

Let $\psi_t(z)$ denote the arg min of the right hand side. Then, the strategy $g_t(x^n_t, z) = \psi_t(z)(x^n_t)$ is optimal.

The main results

Structure of optimal strategies

Under (A1) and (A2), there is no loss of optimality is using control strategies of the form:

$$U^n_t = g_t(X^n_t, \xi(\mathbf{X}_t))$$

instead of

$$U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t})).$$

Dynamic programming decomposition

For a given $z \in \Delta_n$, define $\gamma_t \colon \mathcal{X} \to \mathcal{U}$ as $\gamma_t(\cdot) = g_t(\cdot, z)$.

Recursively define $V_t \colon \Delta_n \to \mathbb{R}$ as follows:
$V_{T+1}(z) = 0$, and for $t = T, \dots, 1$,

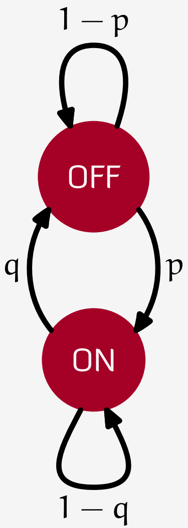
$$V_t(z) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\xi(\mathbf{X}_{t+1})) \mid \xi(\mathbf{X}_t) = z, \{U^n_t = \gamma_t(X^n_t)\}_{n \in \mathcal{N}}\right]$$

Let $\psi_t(z)$ denote the arg min of the right hand side. Then, the strategy $g_t(x^n_t, z) = \psi_t(z)(x^n_t)$ is optimal.

Salient Features

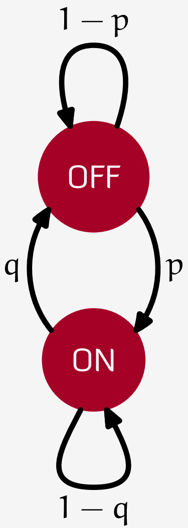
- At each step, we optimize over a function.
- A single DP gives optimal strategies at all agents.

An example: demand response in smart grids



$$Q = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

An example: demand response in smart grids

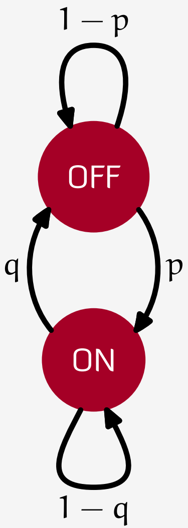


$$Q = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

Three actions

Action	Transition Matrix	Cost
DO NOTHING	Q	0
FORCE OFF	$(1-\varepsilon)\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \varepsilon Q$	c
FORCE ON	$(1-\varepsilon)\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} + \varepsilon Q$	c

An example: demand response in smart grids



$$Q = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

Three actions

Action	Transition Matrix	Cost
DO NOTHING	Q	0
FORCE OFF	$(1-\varepsilon)\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \varepsilon Q$	c
FORCE ON	$(1-\varepsilon)\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} + \varepsilon Q$	c

Reference Signal $\{\zeta_t\}_{t=1}^{T}$

Per stage cost $\ell(\mathbf{X}_t, \mathbf{U}_t) = D(\xi(\mathbf{X}_t) \| \zeta_t) + \dfrac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} c(U_t^i)$

Dumb strategy: Do nothing

examples/mean-field/demand-response-uncontrolled.mov

Naive strategy: Force all agents

examples/mean-field/demand-response-forced.mov

Each agent uses action FORCE ON With probability α and FORCE OFF with probability $1 - \alpha$ where α is chosen such that $\mathbb{E}[\xi(\mathbf{X}_t)] = \zeta_t$.

Optimal strategy: Based on solving the DP

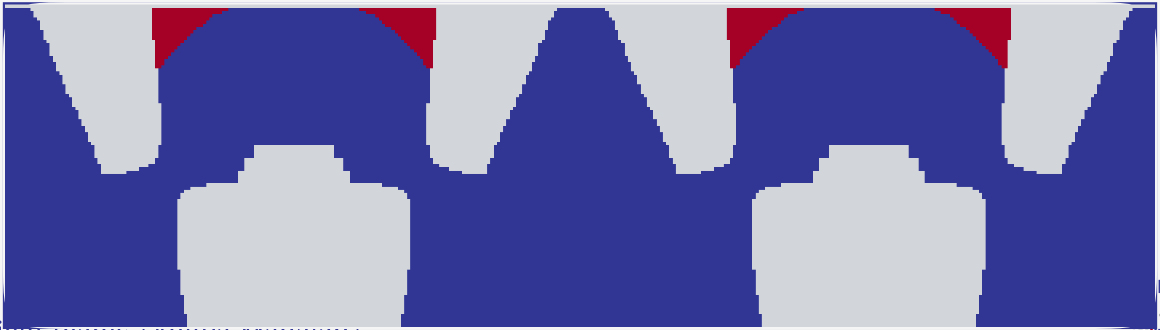
examples/mean-field/demand-response-optimal.mov

Optimal strategy: Based on solving the DP

x-axis: time, y-axis: mean-field, ■: FORCE ON, ■: FORCE OFF, ■: DO NOTHING.

Optimal strategy for $x = 0$

Optimal strategy for $x = 1$



Proof Outline

Step 1 Follow the common information approach to convert the decentralized control problem into a centralized one.

Preliminaries Exchangeable Markov processes

Step 2 Exploit exchangeability to identify a simpler information state

Proof Outline

Step 1 Follow the common information approach to convert the decentralized control problem into a centralized one.

Preliminaries Exchangeable Markov processes

Step 2 Exploit exchangeability to identify a simpler information state

From decentralized to centralized control: the common information approach

Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

From decentralized to centralized control: the common information approach

f_t $(X_t^1, \dots, X_t^N)$

g_t $X_t^1, Z_{1:t}$ U_t^1

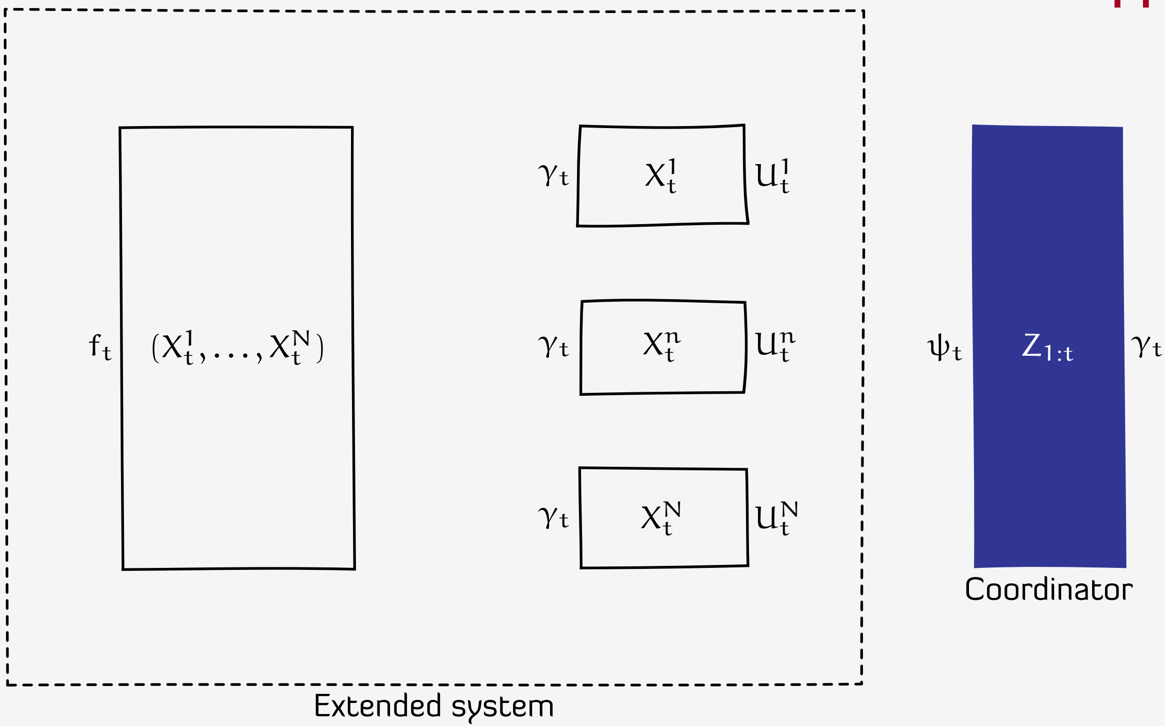
g_t $X_t^n, Z_{1:t}$ U_t^n

g_t $X_t^N, Z_{1:t}$ U_t^N

where $Z_t = \xi(\mathbf{X}_t)$

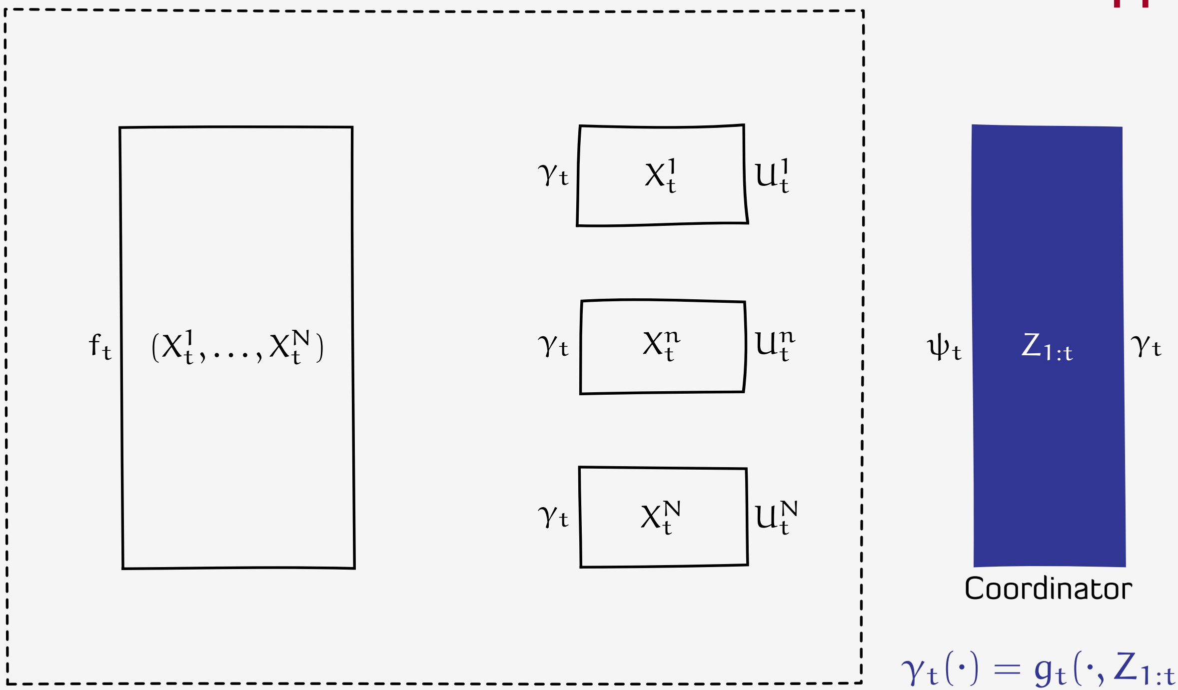
▷ Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

From decentralized to centralized control: the common information approach



▷ Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

From decentralized to centralized control: the common information approach



Extended system

$$\gamma_t(\cdot) = g_t(\cdot, Z_{1:t})$$

▶ Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Equivalent centralized problem

Dynamical system

State	:	$(X_t^1, \dots, X_t^N)$
Observations	:	Z_t
Control actions	:	γ_t where $\gamma_t : \mathcal{X} \mapsto \mathcal{U}$.
Control law	:	

$$\gamma_t = \psi_t(Z_{1:t})$$

Equivalent centralized problem

Dynamical system

"Standard"
centralized POMDP

State : $(X^1_t, \dots, X^N_t)$

Observations : Z_t

Control actions: γ_t where $\gamma_t : \mathcal{X} \mapsto \mathcal{U}$.

Control law :

$$\gamma_t = \psi_t(Z_{1:t})$$

Equivalent centralized problem

Dynamical system

"Standard" centralized POMDP

State : $(X_t^1, \dots, X_t^N)$

Observations : Z_t

Control actions: γ_t where $\gamma_t : \mathcal{X} \mapsto \mathcal{U}$.

Control law :

$$\gamma_t = \psi_t(Z_{1:t})$$

Information state

$$\Pi_t = \mathbb{P}(\text{state} \mid \text{observations}) = \mathbb{P}(X_t^1, \dots, X_t^N \mid Z_{1:t})$$

Dynamic Program

$$V_t(\pi) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\Pi_{t+1}) \mid \Pi_t = \pi, \{U_t^n = \gamma_t(X_t^n)\}_{n \in \mathcal{N}}\right]$$

Equivalent centralized problem

Dynamical system

"Standard" centralized POMDP

State : $(X_t^1, \dots, X_t^N)$

Observations : Z_t

Control actions: γ_t where $\gamma_t : \mathcal{X} \mapsto \mathcal{U}$.

Control law :

$$\gamma_t = \psi_t(Z_{1:t})$$

Information state $\Pi_t = \mathbb{P}(\text{state} \mid \text{observations}) = \mathbb{P}(X_t^1, \dots, X_t^N \mid Z_{1:t})$

Dynamic Program

$$V_t(\pi) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\Pi_{t+1}) \mid \Pi_t = \pi, \{U_t^n = \gamma_t(X_t^n)\}_{n \in \mathcal{N}}\right]$$

Complexity The size of the information state increases double exponentially with number of agents.

Equivalent centralized problem

Dynamical system — State : $(X_t^1, \dots, X_t^N)$

Exploit exchageability to identify
a simpler information state

Information state — $\Pi_t = \mathbb{P}(\text{state} \mid \text{observations}) = \mathbb{P}(X_t^1, \dots, X_t^N \mid Z_{1:t})$

Dynamic Program

$$V_t(\pi) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\Pi_{t+1}) \mid \Pi_t = \pi, \{U_t^n = \gamma_t(X_t^n)\}_{n \in \mathcal{N}}\right]$$

Complexity — The size of the information state increases double exponentially with number of agents.

Proof Outline

Step 1 Follow the common information approach to convert the decentralized control problem into a centralized one.

Preliminaries Exchangeable Markov processes

Step 2 Exploit exchangeability to identify a simpler information state

Exchangeable random vector

Permutation — For any sequence $\mathbf{x}$, let $\sigma\mathbf{x}$ denote a permutation of $\mathbf{x}$.

Note — Given any sequence $\mathbf{x}$ and permutation σ, $\xi(\sigma\mathbf{x}) = \xi(\mathbf{x})$.

Exchangeable random vector

Permutation For any sequence x, let σx denote a permutation of x.

Note Given any sequence x and permutation σ, $\xi(\sigma x) = \xi(x)$.

Definition A random vector $\mathbf{X} \in \mathcal{X}^n$ is called exchangeable if for any permutation σ, the distribution of $\sigma\mathbf{X}$ coincides with $\mathbf{X}$. Formally, for any $x \in \mathcal{X}^n$,

$$\mathbb{P}(\mathbf{X} = \sigma x) = \mathbb{P}(\mathbf{X} = x).$$

Equivalently, a random vector $\mathbf{X} \in \mathcal{X}^n$ is exchangeable if for any $x, y \in \mathcal{X}^n$ such that $\xi(x) = \xi(y)$, we have

$$\mathbb{P}(\mathbf{X} = x) = \mathbb{P}(\mathbf{X} = y).$$

Exchangeable random vector

Permutation For any sequence x, let σx denote a permutation of x.

Note Given any sequence x and permutation σ, $\xi(\sigma x) = \xi(x)$.

Definition A random vector $\mathbf{X} \in \mathcal{X}^n$ is called exchangeable if for any permutation σ, the distribution of $\sigma\mathbf{X}$ coincides with $\mathbf{X}$. Formally, for any $x \in \mathcal{X}^n$,

$$\mathbb{P}(\mathbf{X} = \sigma x) = \mathbb{P}(\mathbf{X} = x).$$

Equivalently, a random vector $\mathbf{X} \in \mathcal{X}^n$ is exchangeable if for any $x, y \in \mathcal{X}^n$ such that $\xi(x) = \xi(y)$, we have

$$\mathbb{P}(\mathbf{X} = x) = \mathbb{P}(\mathbf{X} = y).$$

Example Let $\mathcal{X} = \{0, 1\}$, $n = 3$, and p_{ijk} denotes $\mathbb{P}(\mathbf{X} = (i, j, k))$. Then $\mathbf{X}$ is exchangeable if

$$p_{001} = p_{010} = p_{100} \quad \text{and} \quad p_{110} = p_{101} = p_{011}.$$

Exchangeable Markov process

Definition A (discrete-time) Markov process $\{\mathbf{X}_t\}_{t=1}^{\infty}$ defined on $\mathcal{X}^n$ is exchangeable if
- The initial state $\mathbf{X}_1$ is exchangeable
- The transition matrix is invariant under permutations, i.e., for any permutation σ,

$$\mathbb{P}(\mathbf{X}_{t+1} = \sigma \boldsymbol{y} \mid \mathbf{X}_t = \sigma \boldsymbol{x}) = \mathbb{P}(\mathbf{X}_{t+1} = \boldsymbol{y} \mid \mathbf{X}_t = \boldsymbol{x}).$$

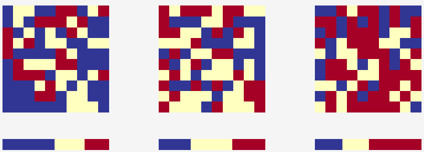
Note Let $\{\mathbf{X}_t\}_{t=1}^{\infty}$ be an exchangeable Markov process. Then, for every t, $\mathbf{X}_t$ is an exchangeable random vector.

Example Interacting particle system

Mean-field projection of an exchangeable Markov process

Definition

Let $\{\mathbf{X}_t\}_{t=1}^{\infty}$ be a $\mathcal{X}^n$-valued exchangeable Markov process. Its mean-field projection is the process $\{Z_t\}_{t=1}^{\infty}$, where $Z_t = \xi(\mathbf{X}_t)$.



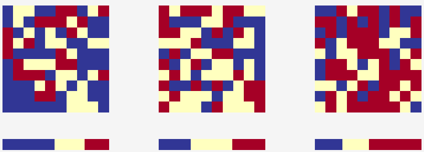
Mean-field projection of an exchangeable Markov process

Definition Let $\{\mathbf{X}_t\}_{t=1}^{\infty}$ be a $\mathcal{X}^n$-valued exchangeable Markov process. Its mean-field projection is the process $\{Z_t\}_{t=1}^{\infty}$, where $Z_t = \xi(\mathbf{X}_t)$.

Proposition

- The mean-field projection is a Markov process, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_{1:t} = z_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = z_t).$$



- The mean-field is a sufficient statistic for predicting the mean-field projection, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid \mathbf{X}_{1:t} = \boldsymbol{x}_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = \xi(\boldsymbol{x}_t)).$$

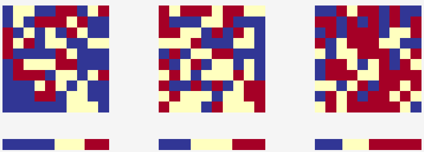
Mean-field projection of an exchangeable Markov process

Definition Let $\{\mathbf{X}_t\}_{t=1}^{\infty}$ be a $\mathcal{X}^n$-valued exchangeable Markov process. Its mean-field projection is the process $\{Z_t\}_{t=1}^{\infty}$, where $Z_t = \xi(\mathbf{X}_t)$.

Proposition

- The mean-field projection is a Markov process, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_{1:t} = z_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = z_t).$$



- The mean-field is a sufficient statistic for predicting the mean-field projection, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid \mathbf{X}_{1:t} = \boldsymbol{x}_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = \xi(\boldsymbol{x}_t)).$$

Theorem Conditioned on the mean-field, all feasible realizations are equally likely, i.e.,

$$\begin{aligned}\mathbb{P}(\mathbf{X}_t = \boldsymbol{x}_t \mid Z_{1:t} = z_{1:t}) &= \mathbb{P}(\mathbf{X}_t = \boldsymbol{x}_t \mid Z_t = z_t) \\ &= \mathbb{P}(\mathbf{X}_t = \sigma \boldsymbol{x}_t \mid Z_t = z_t) \\ &= \frac{\mathbb{1}\{\xi(\boldsymbol{x}_t) = z_t\}}{|\Xi(z_t)|}\end{aligned}$$

Proof Outline

Step 1 Follow the common information approach to convert the decentralized control problem into a centralized one.

Preliminaries Exchangeable Markov processes

Step 2 Exploit exchangeability to identify a simpler information state

A simpler information state

Theorem The mean-field $Z_t = \xi(\mathbf{X}_t)$ is an information state for the (centralized) coordinated system.

A simpler information state

Theorem The mean-field $Z_t = \xi(\mathbf{X}_t)$ is an information state for the (centralized) coordinated system.

1. Z_t is sufficient to compute the belief state π_t

$$\pi_t = \mathbb{P}(\mathbf{X}_t = \mathbf{x}_t \mid Z_{1:t} = z_{1:t}, \gamma_{1:t}) = \mathbb{P}(\mathbf{X}_t = \mathbf{x}_t \mid Z_t = z_t) \frac{\mathbb{1}\{\xi(\mathbf{x}_t) = z_t\}}{|\Xi(z_t)|}$$

2. Z_t evolves in a controlled Markov manner

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_{1:t} = z_{1:t}, \gamma_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = z_t, \gamma_t)$$

A simpler information state

Theorem The mean-field $Z_t = \xi(\mathbf{X}_t)$ is an information state for the (centralized) coordinated system.

Dynamic Program

$$V_t(z) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(Z_{t+1}) \mid Z_t = z, \{U^n_t = \gamma_t(X^n_t)\}_{n \in \mathcal{N}}\right]$$

A simpler information state

Theorem

The mean-field $Z_t = \xi(\mathbf{X}_t)$ is an information state for the (centralized) coordinated system.

Dynamic Program

$$V_t(z) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(Z_{t+1}) \mid Z_t = z, \{U_t^n = \gamma_t(X_t^n)\}_{n \in \mathcal{N}}\right]$$

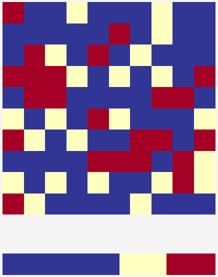
Complexity

Recall that $|\Delta_n| \leqslant (n+1)^{|\mathcal{X}|}$.

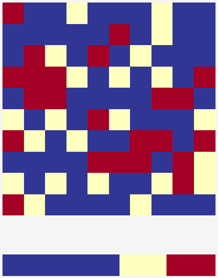
Thus, the size of the information state increases polynomially with number of agents.

The size of the action space does not depend on the number of agents.

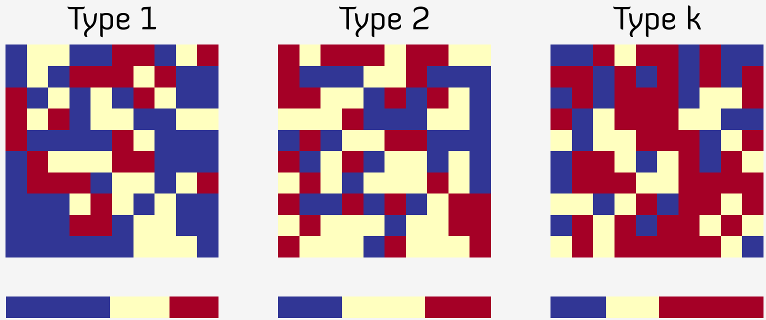
Generalizations



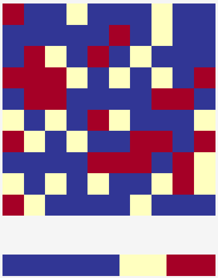
Homogeneous agents



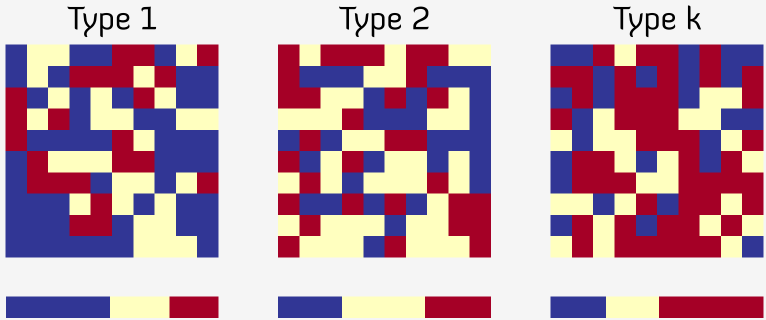
Homogeneous agents



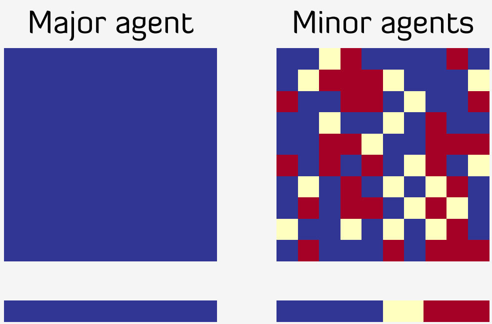
Homogeneous agents with multiple types



Homogeneous agents



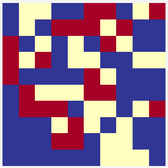
Homogeneous agents with multiple types



Major-minor agents

Recap

Mean-field coupled dynamical systems



Model A finite population $\mathcal{N}$ of homogeneous agents.

X^n_t : State of agent n at time t

U^n_t : Control action of agent n at time t

$\xi(\mathbf{X}_t)$: Mean-field of the population at time t

Dynamics $X^n_{t+1} = f(X^n_t, \xi(\mathbf{X}_t), U^n_t, W^n_t)$

Per-stage Cost Arbitrary coupled cost: $\ell(\mathbf{X}_t, \mathbf{U}_t)$

Information Structure $U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t}))$.

Recap

The main results

Structure of optimal strategies

Under (A1) and (A2), there is no loss of optimality is using control strategies of the form:

$$U^n_t = g_t(X^n_t, \xi(\mathbf{X}_t))$$

instead of

$$U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t})).$$

Dynamic programming decomposition

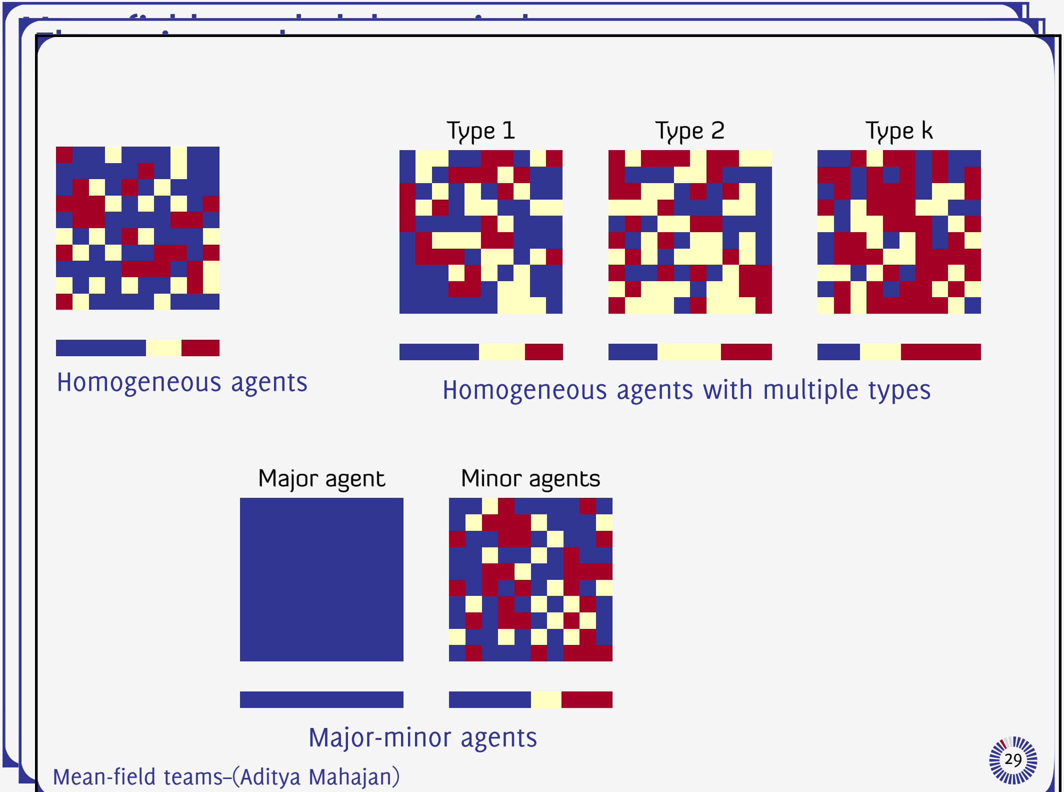
For a given $z \in \Delta_n$, define $\gamma_t \colon \mathcal{X} \to \mathcal{U}$ as $\gamma_t(\cdot) = g_t(\cdot, z)$.

Recursively define $V_t \colon \Delta_n \to \mathbb{R}$ as follows:
$V_{T+1}(z) = 0$, and for $t = T, \dots, 1$,

$$V_t(z) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\xi(\mathbf{X}_{t+1})) \mid \xi(\mathbf{X}_t) = z, \{U^n_t = \gamma_t(X^n_t)\}_{n \in \mathcal{N}}\right]$$

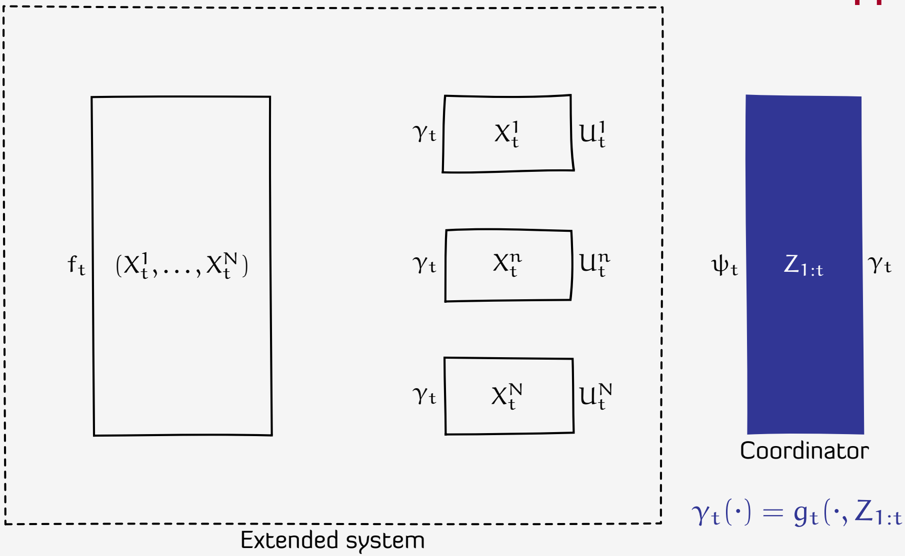
Let $\psi_t(z)$ denote the arg min of the right hand side. Then, the strategy $g_t(x^n_t, z) = \psi_t(z)(x^n_t)$ is optimal.

Recap


Type 1
Type 2
Type k
Homogeneous agents
Homogeneous agents with multiple types
Major agent
Minor agents
Major-minor agents
Mean-field teams–(Aditya Mahajan)
29

Recap

From decentralized to centralized control: the common information approach



$$\gamma_t(\cdot) = g_t(\cdot, Z_{1:t})$$

Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Recap

Mean-field projection of an exchangeable Markov process

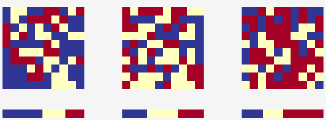
Definition

Let $\{\mathbf{X}_t\}_{t=1}^{\infty}$ be a $\mathcal{X}^n$-valued exchangeable Markov process. Its mean-field projection is the process $\{Z_t\}_{t=1}^{\infty}$, where $Z_t = \xi(\mathbf{X}_t)$.

Proposition

- The mean-field projection is a Markov process, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_{1:t} = z_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = z_t).$$



- The mean-field is a sufficient statistic for predicting the mean-field projection, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid \mathbf{X}_{1:t} = \mathbf{x}_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = \xi(\mathbf{x}_t)).$$

Theorem

Conditioned on the mean-field, all feasible realizations are equally likely, i.e.,

$$\begin{aligned}\mathbb{P}(\mathbf{X}_t = \mathbf{x}_t \mid Z_{1:t} = z_{1:t}) &= \mathbb{P}(\mathbf{X}_t = \mathbf{x}_t \mid Z_t = z_t) \\ &= \mathbb{P}(\mathbf{X}_t = \sigma \mathbf{x}_t \mid Z_t = z_t) \\ &= \frac{\mathbb{1}\{\xi(\mathbf{x}_t) = z_t\}}{|\Xi(z_t)|}\end{aligned}$$

Conclusion

Features of the solution

- State space polynomial in # of agents.
- Action space does not depend on # of agents.
- State and action spaces do not depend on time; hence, the results extend naturally to infinite horizon.

Conclusion

Features of the solution

- State space polynomial in # of agents.
- Action space does not depend on # of agents.
- State and action spaces do not depend on time; hence, the results extend naturally to infinite horizon.

Appropriateness of the model

- Assume that mean-field is observed by all users.
- Happens naturally in some applications (e.g., EV charging, communication networks).
- Distributed computation using consensus protocols.

Conclusion

Features of the solution

- State space polynomial in # of agents.
- Action space does not depend on # of agents.
- State and action spaces do not depend on time; hence, the results extend naturally to infinite horizon.

Appropriateness of the model

- Assume that mean-field is observed by all users.
- Happens naturally in some applications (e.g., EV charging, communication networks).
- Distributed computation using consensus protocols.

Current work

- LQG setup. Same results; easier proof. (CDC 2015)
- Asymptotics as population size goes to infinity. Interest relationships with exchageability and theory of types.
- Mean-field approximations: How good is the infinite population strategy for finite large populations?

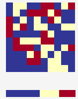
Mean-field teams

Aditya Mahajan
McGill University

Joint work with Jalal Arabneydi

CommNets Seminar
University of Southern California, 26 Aug 2015

Mean-field coupled dynamical systems



Model A finite population $\mathcal{N}$ of homogeneous agents.

X^n_t : State of agent n at time t
U^n_t : Control action of agent n at time t
$\xi(\mathbf{X}_t)$: Mean-field of the population at time t

Dynamics $X^n_{t+1} = f(X^n_t, \xi(\mathbf{X}_t), U^n_t, W^n_t)$

Per-stage Cost Arbitrary coupled cost: $\ell(\mathbf{X}_t, \mathbf{U}_t)$

Information Structure $U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t}))$.

Identical control strategies are not optimal in a system with homogeneous agents

A counterexample

- N homogeneous agents and horizon $T = 2$.
- $\mathcal{X} = \mathcal{U} = \{1, \dots, N\}$
- $X^n_1 \sim \text{Unif}\{1, \dots, N\}$.
- Dynamics: $X^n_2 = U^n_1$

$$\ell_1(x_1, u_1) = 0 \quad \text{and} \quad \ell_2(x_2, u_2) = k\,\mathbb{1}[\xi(x_2) \neq \{\tfrac{1}{N}, \dots \tfrac{1}{N}\}]$$

If all agents use identical strategy Total cost $= k\,\mathbb{E}[\xi(\mathbf{X}_2) \neq \{\tfrac{1}{N}, \dots \tfrac{1}{N}\}] > 0$

Optimal strategy $g^n_1(x^n_1, \xi(x_1)) = n$. Total cost $= 0$.

The main results

Structure of optimal strategies Under (A1) and (A2), there is no loss of optimality is using control strategies of the form:

$$U^n_t = g_t(X^n_t, \xi(\mathbf{X}_t))$$

instead of

$$U^n_t = g^n_t(X^n_t, \xi(\mathbf{X}_{1:t})).$$

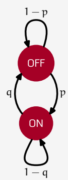
Dynamic programming decomposition For a given $z \in \Delta_n$, define $\gamma_t : \mathcal{X} \to \mathcal{U}$ as $\gamma_t(\cdot) = g_t(\cdot, z)$.

Recursively define $V_t : \Delta_n \to \mathbb{R}$ as follows:
$V_{T+1}(z) = 0$, and for $t = T, \dots, 1$,

$$V_t(z) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\xi(\mathbf{X}_{t+1})) \mid \xi(\mathbf{X}_t) = z, \{U^n_t = \gamma_t(X^n_t)\}_{n \in \mathcal{N}}\right]$$

Let $\psi_t(z)$ denote the arg min of the right hand side. Then, the strategy $g_t(x^n_t, z) = \psi_t(z)(x^n_t)$ is optimal.

An example: demand response in smart grids



$$Q = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$

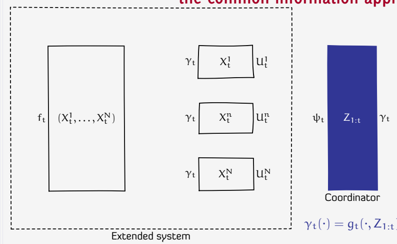
Three actions

Action	Transition Matrix	Cost
DO NOTHING	Q	0
FORCE OFF	$(1-\varepsilon)\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + \varepsilon Q$	c
FORCE ON	$(1-\varepsilon)\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} + \varepsilon Q$	c

Reference Signal $\{\zeta_t\}_{t=1}^{T}$

Per stage cost $\ell(\mathbf{X}_t, \mathbf{U}_t) = D(\xi(\mathbf{X}_t) \| \zeta_t) + \dfrac{1}{|\mathcal{N}|} \sum_{n \in \mathcal{N}} c(U^i_t)$

From decentralized to centralized control: the common information approach



$\gamma_t(\cdot) = g_t(\cdot, Z_{1:t})$

- Nayyar, Mahajan, Teneketzis, "Decentralized stochastic control with partial history sharing: A common information approach," IEEE TAC 2013.

Mean-field projection of an exchangeable Markov process

Definition Let $\{\mathbf{X}_t\}_{t=1}^{\infty}$ be a $\mathcal{X}^n$-valued exchangeable Markov process. Its mean-field projection is the process $\{Z_t\}_{t=1}^{\infty}$, where $Z_t = \xi(\mathbf{X}_t)$.

Proposition

- The mean-field projection is a Markov process, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_{1:t} = z_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = z_t).$$

- The mean-field is a sufficient statistic for predicting the mean-field projection, i.e.,

$$\mathbb{P}(Z_{t+1} = z_{t+1} \mid \mathbf{X}_{1:t} = \mathbf{x}_{1:t}) = \mathbb{P}(Z_{t+1} = z_{t+1} \mid Z_t = \xi(\mathbf{x}_t)).$$

Theorem Conditioned on the mean-field, all feasible realizations are equally likely, i.e.,

$$\begin{aligned}\mathbb{P}(\mathbf{X}_t = \mathbf{x}_t \mid Z_{1:t} = z_{1:t}) &= \mathbb{P}(\mathbf{X}_t = \mathbf{x}_t \mid Z_t = z_t) \\ &= \mathbb{P}(\mathbf{X}_t = \sigma \mathbf{x}_t \mid Z_t = z_t) \\ &= \frac{\mathbb{1}\{\xi(\mathbf{x}_t) = z_t\}}{|\Xi(z_t)|}\end{aligned}$$

A simpler information state

Theorem The mean-field $Z_t = \xi(\mathbf{X}_t)$ is an information state for the (centralized) coordinated system.

Dynamic Program

$$V_t(z) = \min_{\gamma_t} \mathbb{E}\left[\ell(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(Z_{t+1}) \mid Z_t = z, \{U^n_t = \gamma_t(X^n_t)\}_{n \in \mathcal{N}}\right]$$

Complexity Recall that $|\Delta_n| \leqslant (n+1)^{|\mathcal{X}|}$.

Thus, the size of the information state increases polynomially with number of agents.

The size of the action space does not depend on the number of agents.