A systematic process for evaluating structured equilibria in dynamic games with asymmetric information

Achilleas Anastasopoulos
anastas@umich.edu

EECS Department
University of Michigan

February 1, 2016

- Joint work with Deepanshu Vasal (PhD student graduating May 2016) and Prof. Vijay Subramanian

Overview

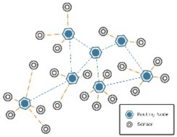
1. Introduction, Motivation, Examples

2. Decentralized teams

3. Games with asymmetric information

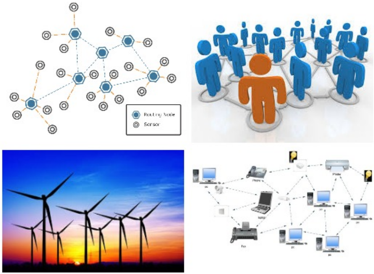
Decentralized decision making in dynamic systems

- Communication networks
- Sensor networks
- Social networks
- Queuing systems
- Energy markets
- Wireless resource sharing
- Repeated online advertisement auctions
- Competing sellers/buyers



Salient features

- Multiple agents (cooperative or strategic)
- Objective: Maximize expected (social or self) reward
- Underlying system state (not perfectly observed)
- Agents make observations (asymmetric information) and take actions partially affecting future state



Classification of problems

	Teams	Games
Symmetric Information	Markov decision processes (MDP) or partially observed MDP (POMDP)	subgame-perfect equilibrium (SPE) Markov-perfect equilibrium (MPE)

Classification of problems

	Teams	Games
Symmetric Information	Markov decision processes (MDP) or partially observed MDP (POMDP)	subgame-perfect equilibrium (SPE) Markov-perfect equilibrium (MPE)
Asymmetric Information	Common information approach [1]	

[1] 2015 IEEE Control Theory Axelby paper award [Nayyar, Mahajan, Teneketzis, 2013]

Classification of problems

	Teams	Games
Symmetric Information	Markov decision processes (MDP) or partially observed MDP (POMDP)	subgame-perfect equilibrium (SPE) Markov-perfect equilibrium (MPE)
Asymmetric Information	Common information approach [1]	Perfect Bayesian (PBE) Sequential eq. (SE) and refinements **No methodology!** ?

[1] 2015 IEEE Control Theory Axelby paper award [Nayyar, Mahajan, Teneketzis, 2013]

Model

- Discrete-time dynamical system with N strategic agents over finite horizon T
- Player i privately observes her (static[2]) type $X^i \in \mathcal{X}^i$ where

$$P(X) = \prod_{i=1}^{N} Q^i(X^i), \qquad X = (X^1, X^2, \dots X^N) \in \mathcal{X}$$

- Player i takes action $A^i_t \in \mathcal{A}^i$ which is publicly observed
- Player i's observations: Private: X^i,
  Common: $A_{1:t-1} = (A_1, A_2, \dots, A_{t-1}) = (A^j_k)^{j \in \mathcal{N}}_{k \leq t-1}$
- Action (randomized) $A^i_t \sim \sigma^i_t(\cdot | X^i, A_{1:t-1})$
- Instantaneous reward $R^i(X, A_t)$
- Player i's objective

$$\max_{\sigma^i} \ \mathbb{E}^{\sigma} \left\{ \sum_{t=1}^{T} R^i(X, A_t) \right\}$$

[2] Generalization to dynamic types straightforward.

Concrete example: A public goods game[3]

- Two players take action to either contribute ($A_t^i = 1$) or not contribute ($A_t^i = 0$) to the production of a public good
- Player i's type (private information) is her cost of contributing: $X^i \in \{L, H\}$, where X^i's are i.i.d. with $P(X^i = H) = q$
- If either player contributes, the public good is produced and the utility enjoyed is 1 for both users (free riding)
- Per-period rewards $(R^1(X^1, A_t), R^2(X^2, A_t))$ are

	contribute($A_t^2 = 1$)	don't contribute($A_t^2 = 0$)
contribute($A_t^1 = 1$)	$(1 - X^1, 1 - X^2)$	$(1 - X^1, 1)$
don't contribute($A_t^1 = 0$)	$(1, 1 - X^2)$	$(0,0)$

- Each player's action $A_t^i \sim \sigma_t^i(\cdot|X^i, A_{1:t-1})$.

[3] Adapted from [Fudenberg and Tirole, 1991, Example 8.3]

Overview

Team problem

- Same information structure but common (team) objective
- Design objective for entire team

$$\max_{\sigma} \; \mathbb{E}^{\sigma}\left\{ \sum_{t=1}^{T} \underbrace{R(X, A_t)}_{\text{e.g., } \sum_{i\in\mathcal{N}} R^i(X, A_t)} \right\}$$

Team problem

- Same information structure but common (team) objective
- Design objective for entire team

$$\max_{\sigma} \; \mathbb{E}^{\sigma}\left\{ \sum_{t=1}^{T} \underbrace{R(X, A_t)}_{\text{e.g., } \sum_{i\in\mathcal{N}} R^i(X, A_t)} \right\}$$

- Problems to be addressed[4]
  1. Presence of common $A_{1:t-1}$ and private X^i information for agent i
  2. Decentralized, non-classical information structure (this is **not** a MDP/POMDP-like problem!)
  3. Domain of policies $A_t^i \sim \sigma_t^i(\cdot|X^i, A_{1:t-1})$ increases with time.

[4] All these have been addressed in [Nayyar, Mahajan, Teneketzis, 2013]

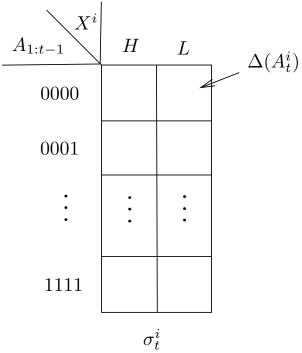
A simple but powerful idea

A policy $\sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be interpreted in two equivalent ways:

A simple but powerful idea

A policy $\sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be interpreted in two equivalent ways:

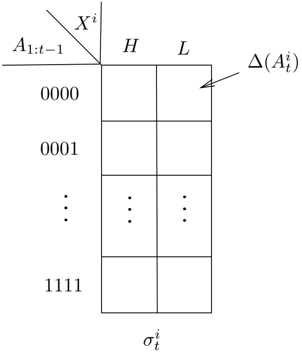
1) A function of $A_{1:t-1}$ and X^i to $\Delta(\mathcal{A}^i)$



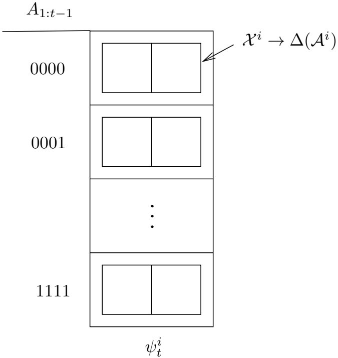
A simple but powerful idea

A policy $\sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be interpreted in two equivalent ways:

1) A function of $A_{1:t-1}$ and X^i to $\Delta(\mathcal{A}^i)$

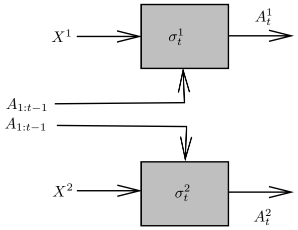


2) A function of $A_{1:t-1}$ to **mappings** from $\mathcal{X}^i$ to $\Delta(\mathcal{A}^i)$



A simple but powerful idea

In the first interpretation, the policies to be designed $(\sigma^i)_{i \in \mathcal{N}}$ have inherent **asymmetric** information structure



A simple but powerful idea

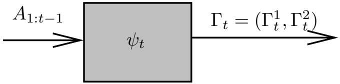
In the second interpretation, each agent's action $A_t^i \sim \sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be thought of as a **two-stage** process

A simple but powerful idea

In the second interpretation, each agent's action $A_t^i \sim \sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be thought of as a **two-stage** process

1. Based on common info $A_{1:t-1}$ select **"prescription"** functions $\Gamma_t^i : \mathcal{X}^i \to \Delta(\mathcal{A}^i)$ through the pre-encoder mapping ψ^i

$$\Gamma_t^i = \psi_t^i[A_{1:t-1}]$$



A simple but powerful idea

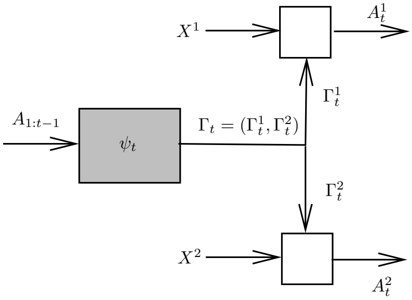
In the second interpretation, each agent's action $A^i_t \sim \sigma^i_t(\cdot|X^i, A_{1:t-1})$ can be thought of as a **two-stage** process

1. Based on common info $A_{1:t-1}$ select **"prescription"** functions $\Gamma^i_t : \mathcal{X}^i \to \Delta(\mathcal{A}^i)$ through the pre-encoder mapping ψ^i

$$\Gamma^i_t = \psi^i_t[A_{1:t-1}]$$

2. The actions A^i_t are determined by "evaluating" Γ^i_t at the private information X^i, i.e.,

$$A^i_t \sim \Gamma^i_t(\cdot|X^i)$$



A simple but powerful idea

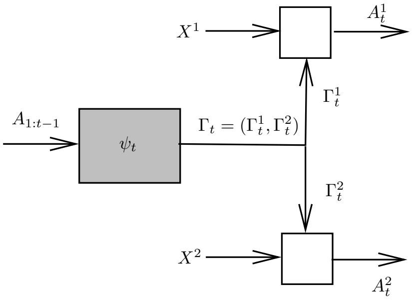
In the second interpretation, each agent's action $A_t^i \sim \sigma_t^i(\cdot|X^i, A_{1:t-1})$ can be thought of as a **two-stage** process

1. Based on common info $A_{1:t-1}$ select **"prescription"** functions $\Gamma_t^i : \mathcal{X}^i \to \Delta(\mathcal{A}^i)$ through the pre-encoder mapping ψ^i

$$\Gamma_t^i = \psi_t^i[A_{1:t-1}]$$

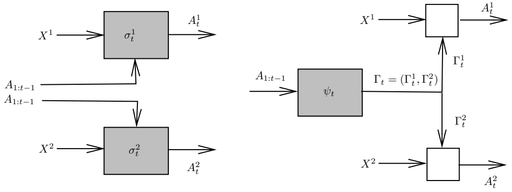
2. The actions A_t^i are determined by "evaluating" Γ_t^i at the private information X^i, i.e.,

$$A_t^i \sim \Gamma_t^i(\cdot|X^i)$$



Overall $\quad A_t^i \sim \Gamma_t^i(\cdot|X^i) = \psi_t^i[A_{1:t-1}](\cdot|X^i) = \sigma_t^i(\cdot|X^i, A_{1:t-1})$

Transformation to a centralized problem



- Generation of A_t^i is a "dumb" evaluation $A_t^i \sim \Gamma_t^i(\cdot|X^i)$ (nothing to be designed here)
- The control problem boils down to selecting prescription functions $\Gamma_t^i = \psi_t^i[A_{1:t-1}]$ through policy $\psi = (\psi_t^i)_{t\in\mathcal{T}}^{i\in\mathcal{N}}$
- All agents can evaluate each-other's prescription functions (think of a *fictitious* common agent with actions Γ_t)
- The decentralized control problem has been transformed to a **centralized control** problem
- Last issue to address: increasing domain $\mathcal{A}^{t-1}$ of the pre-encoder mappings ψ_t.

Introduction of information state

- We would like to summarize $A_{1:t-1}$ in a quantity (state) with time invariant domain

Introduction of information state

- We would like to summarize $A_{1:t-1}$ in a quantity (state) with time invariant domain
- Consider the dynamical system with
  state: (X, A_{t-1})
  observation: A_{t-1}
  action: Γ_t
  reward: $\mathbb{E}\{R(X, A_t)|X, A_{1:t-1}, \Gamma_{1:t}\} = \sum_{a_t} \Gamma_t(a_t|X)R(X, a_t) := \tilde{R}(X, \Gamma_t)$

Introduction of information state

- We would like to summarize $A_{1:t-1}$ in a quantity (state) with time invariant domain
- Consider the dynamical system with
  state: (X, A_{t-1})
  observation: A_{t-1}
  action: Γ_t
  reward: $\mathbb{E}\{R(X, A_t)|X, A_{1:t-1}, \Gamma_{1:t}\} = \sum_{a_t} \Gamma_t(a_t|X)R(X, a_t) := \tilde{R}(X, \Gamma_t)$
- This is a POMDP! Define the posterior belief $\Pi_t \in \Delta(\mathcal{X})$

$$\Pi_t(x) := P(X = x|A_{1:t-1}, \Gamma_{1:t-1}) \qquad \text{for all } x \in \mathcal{X}$$

- Can show that Π_t can be updated using common information

$$\Pi_{t+1} = F(\Pi_t, \Gamma_t, A_t)$$

(*) for this problem it also factors into its marginals

$$\Pi_t(x) = \prod_{i \in \mathcal{N}} \Pi_t^i(x^i) \qquad \text{with} \qquad \Pi_{t+1}^i = F(\Pi_t^i, \Gamma_t^i, A_t^i)$$

Characterization of optimal team policy

- From standard POMDP results, optimal policy is Markovian, i.e.,

$$\Gamma_t = (\Gamma_t^i)_{i \in \mathcal{N}} = \psi_t[A_{1:t-1}] = \theta_t[\Pi_t]$$

$$A_t^i \sim \Gamma_t^i(\cdot|X^i) = \theta_t^i[\Pi_t](\cdot|X^i) = m_t^i(\cdot|X^i, \Pi_t)$$

and can be obtained using backward dynamic programming (DP)

$$\boxed{\theta_t[\pi_t] = \gamma_t^* = \arg\max_{\gamma_t} \mathbb{E}\left\{R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t, A_t))|\pi_t, \gamma_t\right\}}$$

$$\boxed{V_t(\pi_t) = \max_{\gamma_t} \mathbb{E}\left\{R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t, A_t))|\pi_t, \gamma_t\right\}}$$

on the space of beliefs $\pi_t \in \Delta(\mathcal{X})$ over prescriptions $\gamma_t \in \underset{i \in \mathcal{N}}{\times}(\mathcal{X}^i \to \mathcal{A}^i)$

Characterization of optimal team policy

- From standard POMDP results, optimal policy is Markovian, i.e.,

$$\Gamma_t = (\Gamma_t^i)_{i\in\mathcal{N}} = \psi_t[A_{1:t-1}] = \theta_t[\Pi_t]$$

$$A_t^i \sim \Gamma_t^i(\cdot|X^i) = \theta_t^i[\Pi_t](\cdot|X^i) = m_t^i(\cdot|X^i,\Pi_t)$$

and can be obtained using backward dynamic programming (DP)

$$\boxed{\theta_t[\pi_t] = \gamma_t^* = \arg\max_{\gamma_t} \mathbb{E}\left\{R(X,A_t) + V_{t+1}(F(\pi_t,\gamma_t,A_t))|\pi_t,\gamma_t\right\}}$$

$$\boxed{V_t(\pi_t) = \max_{\gamma_t} \mathbb{E}\left\{R(X,A_t) + V_{t+1}(F(\pi_t,\gamma_t,A_t))|\pi_t,\gamma_t\right\}}$$

on the space of beliefs $\pi_t \in \Delta(\mathcal{X})$ over prescriptions $\gamma_t \in \underset{i\in\mathcal{N}}{\times}(\mathcal{X}^i \to \mathcal{A}^i)$

- In the public goods example:
$\pi_t \equiv (\pi_t^1(H), \pi^2(H)) \in [0,1]^2$ and
$\gamma_t \equiv (\gamma_t^1(0|H), \gamma_t^1(0|L), \gamma_t^2(0|H), \gamma_t^2(0|L)) \in [0,1]^4$

Summary of team problem

- Introduction of prescription functions was crucial
- We gained:
  - Decentralized non-classical information structure $\Rightarrow$ POMDP
    $\Rightarrow A_t^i \sim \theta_t^i[\Pi_t](\cdot|X^i)$ and θ can be obtained using DP

Summary of team problem

- Introduction of prescription functions was crucial

- We gained:
  - Decentralized non-classical information structure $\Rightarrow$ POMDP
  $\Rightarrow A_t^i \sim \theta_t^i[\Pi_t](\cdot|X^i)$ and θ can be obtained using DP

- We gave up:
  - *Fictitious common* agent does not observe X^i.
  - Can only maximize average reward-to-go $\mathbb{E}\{\sum_{t'=t}^{T} R(X, A_{t'})|A_{1:t-1}\}$ **before** seeing private information,
  - This is not a problem in teams since we are interested in maximizing the average reward

Overview

Perfect Bayesian equilibria (PBE)

- A PBE is an assessment (σ^*, μ^*) of strategy profiles σ^* and beliefs μ^* satisfying (a) sequential rationality and (b) consistency

(a) For every $t \in \mathcal{T}$, agent $i \in \mathcal{N}$, information set $(A_{1:t-1}, X^i)$, and unilateral deviation σ^i

$$\mathbb{E}^{\mu^*, \sigma^{*i}\sigma^{*-i}}\{\sum_{t'=t}^{T} R^i(X, A_{t'})|A_{1:t-1}, X^i\} \geq \mathbb{E}^{\mu^*, \sigma^i\sigma^{*-i}}\{\sum_{t'=t}^{T} R^i(X, A_{t'})|A_{1:t-1}, X^i\}$$

(b) Beliefs μ^* should be updated by Bayes law (whenever possible) given σ^* and satisfy further consistency conditions [Fudenberg and Tirole, 1991, ch. 8]

- Due to the circular dependence of μ^* and σ^* finding PBE is a large fixed-point problem (no time decomposition)

Ideas from teams: structured equilibrium strategies σ^*

- Useful idea from teams:
  Instead of considering equilibria with general strategies $\sigma^* = (\sigma_t^{*i})_{t\in\mathcal{T}}^{i\in\mathcal{N}}$ of the form
  $$A_t^i \sim \sigma_t^{*i}(\cdot|X^i, A_{1:t-1})$$
  consider equilibria with **structured** strategies $\theta = (\theta_t^i)_{t\in\mathcal{T}}^{i\in\mathcal{N}}$ of the form
  $$A_t^i \sim \Gamma_t^i(\cdot|X^i) = \theta_t^i[\Pi_t](\cdot|X^i) = m_t^i(\cdot|X^i, \Pi_t)$$
  where
  $$\Pi_{t+1} = F(\Pi_t, \Gamma_t, A_t) = F(\Pi_t, \theta_t[\Pi_t], A_t) = F_t^\theta(A_{1:t})$$
- $\sigma^* \Leftrightarrow \theta$ (clarification: unilateral deviations need not be structured!)

Ideas from teams: structured equilibrium strategies σ^*

- Useful idea from teams:
  Instead of considering equilibria with general strategies $\sigma^* = (\sigma_t^{*i})_{t\in\mathcal{T}}^{i\in\mathcal{N}}$ of the form

$$A_t^i \sim \sigma_t^{*i}(\cdot|X^i, A_{1:t-1})$$

  consider equilibria with **structured** strategies $\theta = (\theta_t^i)_{t\in\mathcal{T}}^{i\in\mathcal{N}}$ of the form

$$A_t^i \sim \Gamma_t^i(\cdot|X^i) = \theta_t^i[\Pi_t](\cdot|X^i) = m_t^i(\cdot|X^i, \Pi_t)$$

  where

$$\Pi_{t+1} = F(\Pi_t, \Gamma_t, A_t) = F(\Pi_t, \theta_t[\Pi_t], A_t) = F_t^\theta(A_{1:t})$$

- $\sigma^* \Leftrightarrow \theta$ (clarification: unilateral deviations need not be structured!)

- This is the parallel to MPE, although no equilibrium claim is made yet.

Parenthesis: are structured strategies restrictive?

Lemma ([Vasal, Subramanian, A, 2015a])

For any given strategy profile $\sigma = (\sigma^i)_{i \in \mathcal{N}}$*, there exists a structured strategy profile* $\theta \leftrightarrow m = (m^i)_{i \in \mathcal{N}}$ *with the players receiving the same average rewards for both* σ *and* m*.*

Parenthesis: are structured strategies restrictive?

Lemma ([Vasal, Subramanian, A, 2015a])

For any given strategy profile $\sigma = (\sigma^i)_{i\in\mathcal{N}}$, there exists a structured strategy profile $\theta \leftrightarrow m = (m^i)_{i\in\mathcal{N}}$ with the players receiving the same average rewards for both σ and m.

Proof: Every σ strategy is equivalent to a ψ strategy (common agent viewpoint).
Every ψ strategy induces a distribution $P^\psi(X^i = x^i | A_{1:t-1}) =: \Pi_t(x^i)$.
Π_t can be factored and updated as $\Pi_{t+1} = F(\Pi_t, \Gamma_t, A_t)$.
Every ψ strategy induces a distribution $P^\psi(d\gamma_t^i | \Pi_t)$.
Set $m_t^i(\cdot|X^i, \Pi_t) := \int \gamma_t^i(\cdot|X^i) P^\psi(d\gamma_t^i|\Pi_t)$ and proceed with forward induction. ■

Parenthesis: are structured strategies restrictive?

Lemma ([Vasal, Subramanian, A, 2015a])

For any given strategy profile $\sigma = (\sigma^i)_{i \in \mathcal{N}}$, there exists a structured strategy profile $\theta \leftrightarrow m = (m^i)_{i \in \mathcal{N}}$ with the players receiving the same average rewards for both σ and m.

Proof: Every σ strategy is equivalent to a ψ strategy (common agent viewpoint).
Every ψ strategy induces a distribution $P^{\psi}(X^i = x^i | A_{1:t-1}) =: \Pi_t(x^i)$.
Π_t can be factored and updated as $\Pi_{t+1} = F(\Pi_t, \Gamma_t, A_t)$.
Every ψ strategy induces a distribution $P^{\psi}(d\gamma_t^i | \Pi_t)$.
Set $m_t^i(\cdot | X^i, \Pi_t) := \int \gamma_t^i(\cdot | X^i) P^{\psi}(d\gamma_t^i | \Pi_t)$ and proceed with forward induction. ■

- Bottom line: Structured strategy profiles m are a sufficiently rich class so that we can concentrate on equilibria within this class.
- **Caveat:** Each m^i depends on the entire $\sigma = (\sigma^i)_{i \in \mathcal{N}}$, so unilateral deviations in σ^i result in multilateral deviations in m

Ideas from teams: beliefs μ^*

- Recall that in PBE, μ^* is a set of beliefs on unobserved types X^{-i} for each agent i and for each private history (information set) $(A_{1:t-1}, X^i)$
- Consider beliefs that are:
  (a) only functions of the common history $A_{1:t-1}$ and
  (b) are generated from a common belief in product form

$$\mu_t^*[A_{1:t-1}](X) = \prod_{j \in \mathcal{N}} \mu_t^{*j}[A_{1:t-1}](X^j)$$

- So, for each agent i and for each history $(A_{1:t-1}, X^i)$ belief on X^{-i} is

$$\prod_{j \in \mathcal{N} \setminus \{i\}} \mu_t^{*j}[A_{1:t-1}](X^j)$$

- In addition, with structured (equilibrium) strategies $\sigma^* \Leftrightarrow \theta$, these beliefs are updated as

$$\underbrace{\mu_{t+1}^{*i}[A_{1:t}]}_{\Pi_{t+1}^i} = F(\underbrace{\mu_t^{*i}[A_{1:t-1}]}_{\Pi_t^i}, \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}, A_t^i)$$

Ideas from teams: beliefs μ^*

- Recall that in PBE, μ^* is a set of beliefs on unobserved types X^{-i} for each agent i and for each private history (information set) $(A_{1:t-1}, X^i)$
- Consider beliefs that are:
  (a) only functions of the common history $A_{1:t-1}$ and
  (b) are generated from a common belief in product form

$$\mu_t^*[A_{1:t-1}](X) = \prod_{j \in \mathcal{N}} \mu_t^{*j}[A_{1:t-1}](X^j)$$

- So, for each agent i and for each history $(A_{1:t-1}, X^i)$ belief on X^{-i} is

$$\prod_{j \in \mathcal{N} \setminus \{i\}} \mu_t^{*j}[A_{1:t-1}](X^j)$$

- In addition, with structured (equilibrium) strategies $\sigma^* \Leftrightarrow \theta$, these beliefs are updated as

$$\underbrace{\mu_{t+1}^{*i}[A_{1:t}]}_{\Pi_{t+1}^i} = F(\underbrace{\mu_t^{*i}[A_{1:t-1}]}_{\Pi_t^i}, \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}, A_t^i)$$

- Bottom line: all "consistency" conditions are satisfied automatically.

Summary so far

- We have motivated the use of structured (equilibrium) strategies $\sigma^* \Leftrightarrow \theta$

$$A_t^i \sim \sigma_t^{*i}(\cdot|A_{1:t-1}, X^i) = \underbrace{\theta_t^i[\overbrace{\mu_t^*[A_{1:t-1}]}^{\Pi_t}]}_{\Gamma_t^i}(\cdot|X^i)$$

- We have restricted attention to a class of beliefs μ^* that are updated as

$$\underbrace{\mu_{t+1}^{*i}[A_{1:t}]}_{\Pi_{t+1}^i} = F(\underbrace{\mu_t^{*i}[A_{1:t-1}]}_{\Pi_t^i}, \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}, A_t^i)$$

- PBE equilibrium $(\sigma^*, \mu^*) \equiv (\theta, \mu^*)$ even in this restricted class is still the solution of a large fixed point equation. Circularity between θ and μ^* still present

Summary so far

- We have motivated the use of structured (equilibrium) strategies $\sigma^* \Leftrightarrow \theta$

$$A_t^i \sim \sigma_t^{*i}(\cdot|A_{1:t-1}, X^i) = \underbrace{\theta_t^i[\overbrace{\mu_t^*[A_{1:t-1}]}^{\Pi_t}]}_{\Gamma_t^i}(\cdot|X^i)$$

- We have restricted attention to a class of beliefs μ^* that are updated as

$$\underbrace{\mu_{t+1}^{*i}[A_{1:t}]}_{\Pi_{t+1}^i} = F(\underbrace{\mu_t^{*i}[A_{1:t-1}]}_{\Pi_t^i}, \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}, A_t^i)$$

- PBE equilibrium $(\sigma^*, \mu^*) \equiv (\theta, \mu^*)$ even in this restricted class is still the solution of a large fixed point equation. Circularity between θ and μ^* still present
- How can we find θ with a simple algorithm?
- Same idea as in POMDPs: beliefs and policies are decomposed by considering the policies for all possible beliefs π; not just for μ^*

First erroneous attempt

- Recall DP equation from team problem

$$\theta_t[\pi_t] = \gamma_t^* = \arg\max_{\gamma_t^i \gamma_t^{-i}} \mathbb{E}\left\{R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t^i \gamma_t^{-i}, A_t)) | \pi_t, \gamma_t^i \gamma_t^{-i}\right\}$$

- What is the logical extension in games?

First erroneous attempt

- Recall DP equation from team problem

$$\theta_t[\pi_t] = \gamma_t^* = \arg\max_{\gamma_t^i \gamma_t^{-i}} \mathbb{E}\left\{R(X, A_t) + V_{t+1}(F(\pi_t, \gamma_t^i \gamma_t^{-i}, A_t))|\pi_t, \gamma_t^i \gamma_t^{-i}\right\}$$

- What is the logical extension in games?

$$\text{for all } i \in \mathcal{N}$$

$$\gamma_t^{*i} \in \arg\max_{\gamma_t^i} \mathbb{E}\left\{R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, A_t))|\pi_t, \gamma_t^i \gamma_t^{*-i}\right\}$$

where expectation is explicitly given by

$$\mathbb{E}\{\cdot|\cdot\} = \sum_{a_t, x} \gamma_t^i(a_t^i|x^i)\gamma_t^{*-i}(a_t^{-i}|x^{-i})\pi_t(x) \times \left(R^i(x, a_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, a_t))\right)$$

Once this per-stage FP equation is solved $\gamma_t^* = \theta_t[\pi_t]$, update

$$V_t^i(\pi_t) = \mathbb{E}\left\{R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^*, A_t))|\pi_t, \gamma_t^*\right\}$$

First erroneous attempt: what is the catch?

$$\text{for all } i \in \mathcal{N}$$
$$\gamma_t^{*i} \in \arg\max_{\gamma_t^i} \mathbb{E}\left\{R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i\gamma_t^{*-i}, A_t))|\pi_t, \gamma_t^i\gamma_t^{*-i}\right\}$$

- Why erroneous?

First erroneous attempt: what is the catch?

for all $i \in \mathcal{N}$

$$\gamma_t^{*i} \in \arg\max_{\gamma_t^i} \mathbb{E}\left\{R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, A_t)) | \pi_t, \gamma_t^i \gamma_t^{*-i}\right\}$$

- Why erroneous?
- **Explanation:** reward-to-go is not conditioned on the entire history $(A_{1:t-1}, X^i)$ for user i but only on part of it $A_{1:t-1} \leftrightarrow \Pi_t$.
  This was OK in teams but is not sufficient to prove sequential rationality in games!

$$\mathbb{E}^{\mu^*, \sigma^{*i} \sigma^{*-i}}\{\sum_{t'=t}^{T} R^i(X, A_{t'}) | A_{1:t-1}, X^i\} \geq \mathbb{E}^{\mu^*, \tilde{\sigma}^i \sigma^{*-i}}\{\sum_{t'=t}^{T} R^i(X, A_{t'}) | A_{1:t-1}, X^i\}$$

Special case[5]

- Consider dynamical systems for which belief update is prescription-independent, i.e., $\Pi_{t+1} = F(\Pi_t, A_t)$
- In that case the backward process decomposes and conditioning on X^i is irrelevant
- A strong statement can be made for this special case:
  "For every PBE there exists a structured PBE that corresponds to a SPE of an equivalent symmetric-information game"

[5] [Nayyar, Gupta, Langbort, Başar, 2014], [Gupta, Nayyar, Langbort, Başar, 2014]

Second erroneous attempt

Condition on X^i in the backward induction step to be consistent with sequential rationality condition

- For each $t = T, T-1, \ldots, 1$ and for every $\pi_t \in \Delta(\mathcal{X})$ solve the following one-step fixed-point equation

  $$\text{for all } i \in \mathcal{N} \text{ and for all } x^i \in \mathcal{X}^i$$

  $$\gamma_t^{*i} \in \arg\max_{\gamma_t^i} \mathbb{E}\left\{R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, A_t), x^i) | x^i, \pi_t, \gamma_t^i \gamma_t^{*-i}\right\}$$

  where expectation is explicitly given by

  $$\mathbb{E}\{\cdot|\cdot\} = \sum_{a_t, x^{-i}} \gamma_t^i(a_t^i|x^i)\gamma_t^{*-i}(a_t^{-i}|x^{-i})\pi_t^{-i}(x^{-i})\times$$
  $$\left(R^i(x^i x^{-i}, a_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, a_t), x^i)\right)$$

- Note in this case reward-to-go is $V_t^i(\pi_t, x^i)$

Second erroneous attempt: explanation

$$\mathbb{E}\{\cdot|\cdot\} = \sum_{a_t, x^{-i}} \gamma_t^i(a_t^i|x^i)\gamma_t^{*-i}(a_t^{-i}|x^{-i})\pi^{-i}(x^{-i})\times$$

$$\big(R^i(x^i x^{-i}, a_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i \gamma_t^{*-i}, a_t), x^i)\big)$$

- This is an unusual fixed point equation: dependence on $\gamma_t^i(\cdot|x^i)$ but also on the entire $\gamma_t^i(\cdot|\cdot)$ (inside the belief update)

Second erroneous attempt: explanation

$$\mathbb{E}\{\cdot|\cdot\} = \sum_{a_t, x^{-i}} \gamma_t^i(a_t^i|x^i)\gamma_t^{*-i}(a_t^{-i}|x^{-i})\pi^{-i}(x^{-i})\times$$

$$\left(R^i(x^i x^{-i}, a_t) + V_{t+1}^i(F(\pi_t, \gamma_t^i\gamma_t^{*-i}, a_t), x^i)\right)$$

- This is an unusual fixed point equation: dependence on $\gamma_t^i(\cdot|x^i)$ but also on the entire $\gamma_t^i(\cdot|\cdot)$ (inside the belief update)

- Unfortunately this results in an "equilibrium generating" mapping θ with $\gamma_t^* = \theta_t[\pi_t, x]$ so resulting policy is of the form

$$A_t^i \sim \Gamma_t^{*i}(\cdot|X^i) = \theta_t^i[\Pi_t, X](\cdot|X^i)$$

which is **not implementable** (requires unknown private information X^{-i} for the strategy of i).

An algorithm for PBE evaluation: backward recursion

- For each $t = T, T-1, \ldots, 1$ and for every $\pi_t \in \Delta(\mathcal{X})$ solve the following one-step fixed-point equation

for all $i \in \mathcal{N}$ and for all $x^i \in \mathcal{X}^i$

$$\gamma_t^{*i}(\cdot|x^i) \in \arg\max_{\gamma_t^i(\cdot|x^i)} \mathbb{E}\left\{R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \boxed{\gamma_t^{*i}\gamma_t^{*-i}}, A_t), x^i)|x^i, \pi_t, \gamma_t^i\gamma_t^{*-i}\right\}$$

where expectation is explicitly given by

$$\mathbb{E}\{\cdot|\cdot\} = \sum_{a_t, x^{-i}} \gamma_t^i(a_t^i|x^i)\gamma_t^{*-i}(a_t^{-i}|x^{-i})\pi^{-i}(x^{-i}) \times$$

$$\left(R^i(x^i x^{-i}, a_t) + V_{t+1}^i(F(\pi_t, \boxed{\gamma_t^{*i}\gamma_t^{*-i}}, a_t), x^i)\right)$$

- This results in an "equilibrium generating" mapping θ with $\gamma_t^* = \theta_t[\pi_t]$ for all $\pi_t \in \Delta(\mathcal{X})$

Special backward induction step

for all $i \in \mathcal{N}$ and for all $x^i \in \mathcal{X}^i$

$$\gamma_t^{*i}(\cdot|x^i) \in \arg\max_{\gamma_t^i(\cdot|x^i)} \mathbb{E}\left\{R^i(X, A_t) + V_{t+1}^i(F(\pi_t, \boxed{\gamma_t^{*i}\gamma_t^{*-i}}, A_t), x^i)|x^i, \pi_t, \gamma_t^i\gamma_t^{*-i}\right\}$$

- This is **not a best-response** type function: γ_t^{*i} present on left/right hand side
- Find $\gamma_t^i(\cdot|x^i)$ that is optimal under unperturbed belief update!

An algorithm for PBE evaluation: forward recursion

- From backard recursion we have obtained $\theta = (\theta_t^i)_{t \in \mathcal{T}}^{i \in \mathcal{N}}$.
- For each $t = 1, 2, \ldots, T$ and for every $i \in \mathcal{N}$, $A_{1:t}$, and X^i

$$\sigma_t^{*i}(A_t^i | A_{1:t-1}, X^i) := \underbrace{\theta_t^i[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t^i}(A_t^i | X^i)$$

$$\underbrace{\mu_{t+1}^*[A_{1:t}]}_{\Pi_{t+1}} := F(\underbrace{\mu_t^*[A_{1:t-1}]}_{\Pi_t}, \underbrace{\theta_t[\mu_t^*[A_{1:t-1}]]}_{\Gamma_t}, A_t)$$

- In fact we can obtain a family of PBEs for any type distribution $\prod_{i \in \mathcal{N}} Q^i(X^i)$ with appropriate initialization of μ_1^*

Main Result

Theorem ([Vasal, Subramanian, A, 2015a])

(σ^*, μ^*) generated by the backward/forward algorithm (whenever it exists) is a PBE, i.e. for all $i, t, A_{1:t-1}, X^i, \sigma^i$,

$$\mathbb{E}^{\sigma^{*i}_{t:T}\sigma^{*-i}_{t:T}\mu^*_t}\left\{\sum_{n=t}^{T} R^i(X, A_n)\Big| A_{1:t-1}X^i\right\}$$
$$\geq \mathbb{E}^{\sigma^{i}_{t:T}\sigma^{*-i}_{t:T}\mu^*_t}\left\{\sum_{n=t}^{T} R^i(X, A_n)\Big| A_{1:t-1}X^i\right\}$$

and μ^* satisfies the consistency conditions.

Sketch of the proof

- Independence of types and specific DP equation are crucial in proving the result
- Modified comparison principle (backward induction)

- Specific DP guarantees that unperturbed reward-to-go (LHS) at time t is the obtained value function $V_t^i = R^i + V_{t+1}^i$
- Specific DP guarantees that unilateral deviations with fixed belief update reduce V_t^i
- Induction step reduces V_{t+1}^i to (perturbed) reward-to-go at time $t+1$
- Independence of types guarantees that resulting expression is exactly the (perturbed) reward-to-go at time t (RHS)

Comments on per-stage fixed point equation

- This is not a best-response type of FP equation (due to presence of γ^{*i} on both the LHS and RHS of equation)
- Standard tools for existence of solution (e.g., Brouwer, Kakutani) do not apply (problem with continuity of $V(\cdot)$ functions)

Comments on per-stage fixed point equation

- This is not a best-response type of FP equation (due to presence of γ^{*i} on both the LHS and RHS of equation)
- Standard tools for existence of solution (e.g., Brouwer, Kakutani) do not apply (problem with continuity of $V(\cdot)$ functions)

- Existence can be shown for a special case[6] where $R^i(X, A_t)$ does not depend on its own type X^i
- In that case prescriptions $\Gamma^i_t(\cdot|X^i) = \Gamma^i_t(\cdot)$ do not depend on private type X^i and FP equation reduces to best response.
  No signaling!
  Essentially reduces to the model $\Pi_{t+1} = F(\Pi_t, A_t)$

[6] [Ouyang, Tavafoghi, Teneketzis, 2015]

Current/Future work

- Model generalizations:
  - Types are independent controlled Markov processes (controlled by **all** actions) $P(X_t|X_{1:t-1}, A_{1:t-1}) = \prod_{i\in\mathcal{N}} Q^i(X_t^i|X_{t-1}^i, A_{t-1})$[7]
  - Dependence types with "strategic independence"[8]
  - Types are observed through a noisy channel (even by same user) $Q(Y_t^i|X_t^i)$. Example: "informational cascades" literature
  - Infinite horizon and continuous action spaces
- Existence results: prove existence for the simplest non-trivial class of problems. Core issue: the per-stage FP equation is not a best response
- Dynamic mechanism design (indirect mechanisms with message space smaller than type space)

[7] [Vasal, Subramanian, A, 2015b]
[8] [Battigalli, 1996]

Thank you!

Battigalli, P. (1996).
Strategic independence and perfect Bayesian equilibria.
Journal of Economic Theory, 70(1):201–234.

Fudenberg, D. and Tirole, J. (1991).
Game Theory.
MIT Press, Cambridge, MA.

Gupta, A., Nayyar, A., Langbort, C., and Başar, T. (2014).
Common information based markov perfect equilibria for linear-gaussian games with asymmetric information.
SIAM Journal on Control and Optimization, 52(5):3228–3260.

Nayyar, A., Gupta, A., Langbort, C., and Başar, T. (2014).
Common information based markov perfect equilibria for stochastic games with asymmetric information: Finite games.
IEEE Trans. Automatic Control, 59(3):555–570.

Nayyar, A., Mahajan, A., and Teneketzis, D. (2013).
Decentralized stochastic control with partial history sharing: A common information approach.
Automatic Control, IEEE Transactions on, 58(7):1644–1658.

Ouyang, Y., Tavafoghi, H., and Teneketzis, D. (2015).
Dynamic oligopoly games with private markovian dynamics.
Available at www-personal.umich.edu/~tavaf/Oligopolygames.pdf.

Vasal, D., Subramanian, V., and Anastasopoulos, A. (2015a).
A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information.
Technical report.

Vasal, D., Subramanian, V., and Anastasopoulos, A. (2015b).
A systematic process for evaluating structured perfect Bayesian equilibria in dynamic games with asymmetric information.
In *American Control Conference*.
(Accepted for publication/presentation).