Optimal decentralized control of coupled subsystems with control sharing

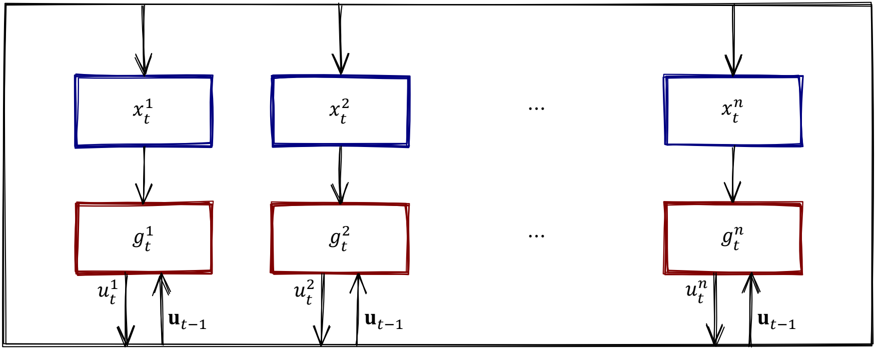
Aditya Mahajan
McGill University

IEEE Conference on Decision and Control, 2011

Notation

- Random variables: X, realizations: x, state spaces: $\mathcal{X}$.
- a_t^i means that variable a belongs to subsystem i at time t.
- $a_{1:t} = (a_1, a_2, \dots, a_t)$
- $\mathbf{a} = (a^1, a^2, \dots, a^n)$.

System Model



Control-coupled subsystems

$$x_{t+1}^i = f_t^i(x_t^i, \mathbf{u}_t, w_t^i)$$

Controller with control sharing

$$u_t^i = g_t^i(x_{1:t}^i, \mathbf{u}_{1:t-1})$$

Objective

$$\min_{\text{all policies } \mathbf{g}} \mathbb{E}\Big[\sum_{t=1}^{T} c_t(\mathbf{x}_t, \mathbf{u}_t)\Big]$$

Some applications

- Feedback communication systems (physical layer)

  Point-to-point real-time source coding, multi-terminal source coding with feedback, some classes of multiple access channel with feedback

- Queueing networks (media access layer)

  Multi-access broadcast, some classes of decentralized scheduling and routing.

- Cellular networks

  Paging and registration in cellular networks

Conceptual difficulties

- The system has non-classical information structure
- Data at each controller is increasing with time

$$u_t^i = g_t^i(x_{1:t}^i, \mathbf{u}_{1:t-1})$$

  - Is part of this data redundant?
  - Can part of this data be compressed to a sufficient statistic?

- Multi-stage decision making
  - How does current control action affect future estimation?
  - What information does controller i communicate to controller j via its control action?

Literature Overview

- Control sharing info-structure (Bismut, 1972, Sandell and Athans, 1974)
  - Considered the LQG version of the problem
  - Exploit the fact that the action space is continuous and compact to embed the observations in control
  - Reduces to one-step delayed sharing pattern
- Other non-classical info-structures with sharing
  - Delayed state sharing: Aicadri, Davoli, and Minciardi, 1987
  - Delayed (observation) sharing: Witsenhausen 1971, Varaiya and Walrand, 1979, Nayyar, Mahajan, and Teneketzis, 2011
  - Periodic sharing: Ooi, Verbout, Ludwig, Wornell, 1997
  - Belief sharing: Yüksel, 2009
  - Partial history sharing: Mahajan, Nayyar, Teneketzis, 2008

Outline of the results

- First structural result (based on person-by-person opt.)

  $x^i_{1:t-1}$ is redundant for optimal performance.

  $$\text{wlo,}\quad u^i_t = g^i_t(x^i_t, \mathbf{u}_{1:t-1})$$

- Second structural result (based on common info approach of MNT 2008)

  Define $\Pi^i_t(x) = \mathbb{P}(X^i_t = x \mid \mathbf{U}_{1:t-1})$ and $\mathbf{\Pi}_t = (\Pi^1_t, \dots, \Pi^n_t)$.

  $\boldsymbol{\pi}_t$ is a sufficient statistic of $\mathbf{u}_{1:t-1}$ for optimal performance.

  $$\text{wlo,}\quad u^i_t = g^i_t(x^i_t, \boldsymbol{\pi}_t)$$

- Dynamic programming decomposition

Structural result based on person-by-person optimality

Main lemma

The states processes are conditionally independent given the past control actions.

$$\mathbb{P}(\mathbf{X}_{1:t} = \mathbf{x}_{1:t} \mid \mathbf{U}_{1:t}) = \prod_{i=1}^{n} \mathbb{P}(X^i_{1:t} = x^i_{1:t} \mid \mathbf{U}_{1:t})$$

Implications

Fix g^{-i} and consider optimal design of g^i. Let $R^i_t = (X^i_t, \mathbf{U}_{1:t-1})$. Then $\{R^i_t, t = 1, \dots\}$ is a controlled MDP with control action U^i_t.

- $\mathbb{P}(r^i_{t+1} \mid r^i_{1:t}, u^i_{1:t}) = \mathbb{P}(r^i_{t+1} \mid r^i_t, u^i_t)$
- $\mathbb{E}[c_t(\mathbf{x}_t, \mathbf{u}_t) \mid r^i_{1:t}, u^i_{1:t}] = \mathbb{E}[c_t(\mathbf{x}_t, \mathbf{u}_t) \mid r^i_t, u^i_t]$

Structural result . . . (cont.)

- Original model

$$u_t^i = g_t^i(x_{1:t}^i, \mathbf{u}_{1:t-1})$$

- Implication of person-by-person optimality argument

$$u_t^i = g_t^i(r_t^i) = g_t^i(x_t^i, \mathbf{u}_{1:t-1})$$

- Design difficulty

  Data at the controller is still increasing with time

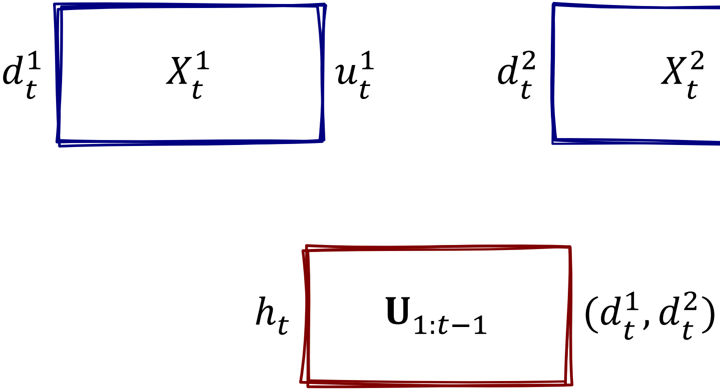
A coordinator based on common information

General idea proposed in (Mahajan, Nayyar, and Teneketzis 2008)

g_t^1 $X_t^1, \mathbf{U}_{1:t-1}$ u_t^1

g_t^2 $X_t^2, \mathbf{U}_{1:t-1}$ u_t^2

A coordinator based on common information (cont.)



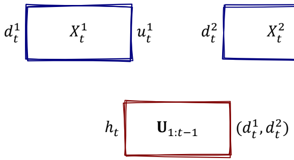
where $d_t^i(\cdot) = g_t^i(\cdot, \mathbf{u}_{1:t-1})$

A coordinator based on common information (cont.)

- Solution approach
  - The coordinated system is a POMDP
  - Identify the structure of optimal coordination strategies for the coordinated system
  - Show that the coordinated system is equivalent to the original model
  - Translate the structure of optimal coordination strategies to the original model

The coordinated system

- State: $\mathbf{x}_t = (x^1_t, \dots, x^n_t)$
- Observations: $\mathbf{u}_{t-1} = (u^1_{t-1}, \dots, u^n_{t-1})$
- Control actions: $\mathbf{d}_t = (d^1_t, \dots, d^n_t)$,



- Coordination rule: $h_t : \Big(\prod_{i=1}^{n} \mathcal{U}^i \Big)^{t-1} : \prod_{i=1}^{n} (\mathcal{X}^i \to \mathcal{U}^i_t)$

$$\mathbf{d}_t = h_t(\mathbf{u}_{1:t-1})$$

- Structure of optimal coordination strategy

Define $\Xi_t = \mathbb{P}(\text{state} \mid \text{history of observations}) = \mathbb{P}(\mathbf{x} \mid \mathbf{U}_{1:t-1})$. Then,

$$\text{wlo,} \quad \mathbf{d}_t = h_t(\xi_t)$$

The coordinated system (cont.)

- Dynamic programming decomposition

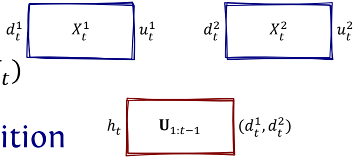
$$V_t(\xi) = \min_{\mathbf{d}} \mathbb{E}\left[c_t(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\Xi_{t+1}) \mid \Xi_t = \xi\right]$$

- Salient features
  - The optimization at each step is a functional optimization problem.
  - (In our opinion) functional optimization at each step is the only way to circumvent the issue of signaling.

Translation of results back to the original system

- Structural result

  wlo, $u_t^i = d_t^i(x_t^i) = h_t^i(\xi_t)(x_t^i) = g_t^i(x_t^i, \xi_t)$



- Dynamic programming decomposition
  - Solve the DP for coordinated system.
  - Choose $g_t^i(x_t^i, \xi_t) = h_t^i(\xi_t)(x_t^i)$

Further simplification of structural result

- Recall main lemma:

  The states processes are conditionally independent given the past control actions.

$$\mathbb{P}(\mathbf{X}_{1:t} = \mathbf{x}_{1:t} \mid \mathbf{U}_{1:t}) = \prod_{i=1}^{n} \mathbb{P}(X^i_{1:t} = x^i_{1:t} \mid \mathbf{U}_{1:t})$$

- Implication

$$\xi_t(\mathbf{x}) = \mathbb{P}(\mathbf{X}_t = \mathbf{x} \mid \mathbf{U}_{1:t-1}) = \prod_{i=1}^{n} \pi^i_t(x^i_t)$$

Further simplification of structural result (cont.)

- Simplified structural result

$$\text{wlo,} \quad u_t^i = g_t^i(x_t^i, \xi_t) = g_t^i(x_t^i, \boldsymbol{\pi}_t)$$

Significant reduction is size.

$$\xi_t \in \Delta(\mathcal{X}^1 \times \cdots \times \mathcal{X}^n) \quad \text{while} \quad \boldsymbol{\pi}_t \in \Delta(\mathcal{X}^1) \times \cdots \times \Delta(\mathcal{X}^n)$$

- Simplified dynamic programming decomposition

$$V_t(\boldsymbol{\pi}) = \min_{\mathbf{d}} \mathbb{E}\left[c_t(\mathbf{X}_t, \mathbf{U}_t) + V_{t+1}(\mathbf{\Pi}_{t+1}) \mid \mathbf{\Pi}_t = \boldsymbol{\pi}\right]$$

Recap of structural results

- Original:

$$u_t^i = g_t^i(x_{1:t}^i, \mathbf{u}_{1:t-1})$$

- Using person-by-person approach

$$u_t^i = g_t^i(x_t^i, \mathbf{u}_{1:t-1})$$

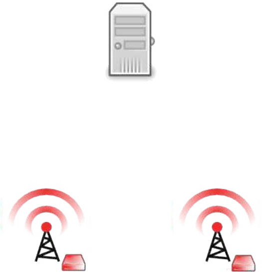
- Using the common information approach of (NMT 2008, 2011)

$$u_t^i = g_t^i(x_t^i, \xi_t), \quad \xi_t = \mathbb{P}(\mathbf{X}_t \mid \mathbf{u}_{1:t-1})$$

- Using specific conditional independence due to the dynamics

$$u_t^i = g_t^i(x_t^i, \boldsymbol{\pi}_t), \quad \pi_t^i = \mathbb{P}(X_t^i \mid \mathbf{u}_{1:t-1})$$

An Example: Two-user multiple access broadcast



- Two-user with single slot buffer
  - $x_t^i \in \{0, 1\}$: # of packets in queue
  - $w_t^i \in \{0, 1\}$: # of arrivals $\sim \text{Ber}(p_i)$
  - $u_t^i \in \{0, 1\}$: # of transmitted packets
- Multiple-access channel
  - Throughput: $r_t = u_t^1(1 - u_t^2) + (1 - u_t^1)u_t^2$
  - r_t available to both users after one-step delay
  - State update: $x_{t+1}^i = \max\big((x_t^i - u_t^i r_t) + w_t^i, 1\big)$

An Example: Two-user multiple access broadcast (cont.)

- Literature overview
  - Symmeric arrivals: Hlyuchj and Gallager, 1981 feasible lower bound
  - Symmeric arrivals: Ooi and Wornell, 1996 genie aided upper bound that numerically matched lower bound.
  - Asymmetric arrivals: Used as benchmark problem in AI community (Hansen et al, 2004, Bernstein et al, 2005, Shez Charpillet, 2006) for numerical algorithms for DEC-POMDPs.

An Example: Two-user multiple access broadcast (cont.)

- Structure of optimal control policy
  - π_t^i is equivalent to $\pi_t^i(1) =: q_t^i \in \{0, 1\}$
  - d_t^i is equivalent to $d_t^i(1) =: s_t^i \in \{0, 1\}$
  - Structure of optimal policy

$$u_t^i = x_t^i \cdot s_t^i, \quad \text{where} \quad (s_t^1, s_t^2) = h_t^i(q_t^1, q_t^2)$$

An Example: Two-user multiple access broadcast (cont.)

Optimal policy for symmetric arrivals

- Notation: for any $q \in [0,1]$, let $Aq = 1 - (1-p)(1-q)$
- Characteristic polynomial: $\varphi_n(x) = 1 + (1-x)^2 - (3+x)(1-x)^{n+1}$.
- Let α_n be the root of φ_n in $[0,1]$ and τ be the root of $x = (1-x)^2$
- Optimal performance:

$$J^* = \begin{cases} 1 - (1-p)^2, & \text{if } p \geq \alpha_1 \\ p(1 - (2p^2 - 1)) \,/\, (1 + p^2 + p^3), & \text{otherwise} \end{cases}$$

An Example: Two-user multiple access broadcast (cont.)

- Optimal policy
  - When $p > \tau$

$$h^*(q^1, q^2) = \begin{cases} (1,0) & \text{if } q^1 > q^2 \\ (0,1) & \text{if } q^1 < q^2 \\ (1,0) \text{ or } (0,1) & \text{if } q^1 = q^2 \end{cases}$$

  - When $p < \tau$, let $n \in \mathbb{N}$ be such that $\alpha_{n+1} < p \leq \alpha_n$.

$$h^*(q^1, q^2) = \begin{cases} (1,1) & \text{if } q^1 \leq A^n \text{ and } q^2 \leq A^n p \\ (1,0) & \text{if } q^1 > \max(A^n p, q^2) \\ (0,1) & \text{if } q^2 > \max(A^n p, q^1) \\ (1,0) \text{ or } (0,1) & \text{if } q^1 = q^2 = 1 \end{cases}$$

- Analytic proof of optimality of the policy proposed by Hlyuchj and Gallager, 1981.

Conclusion

- Coupled subsystems with control-sharing
  - Non-classical information structure
  - Use properties of the system dynamics and the common information approach of (Mahajan, Nayyar, Teneketzis 2008) to find structure of optimal controller and a dynamic programming decomposition.
  - Allows using standard tools from stochastic control to analyze specific applications.
- Key take-home points
  - Subclasses of decentralized control problems with signaling are solvable!
  - Each step of the DP is a functional optimization problem.