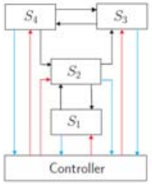
Control of Networked Markov Decision Systems with Delays

Sachin Adlakha, Sanjay Lall and Andrea Goldsmith

Introduction



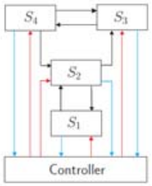
- Feedback control of networked systems.
- Controller has *delayed* access to system states.
- Each subsystem a *Markov Decision Process*.

We show

> An optimal controller exists with *finite memory*

- We call the controller *banded* by analogy with matrix case.
- Resulting dynamic program is computationally tractable.

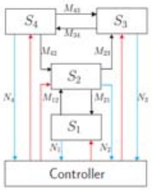
Motivation



- Networked Markov decision processes are used to model
  - Distributed vehicle coordination.
  - Scheduling over multiple servers.
  - Network of interacting queues.
- Delays cause the system to be *partially observable*.

> Provide sufficient information state for such systems.

Model



$$x^1_{t+1} = f^1\left(x^1_t, u^1_t, w^1_t, x^2_{t-M_{21}}\right)$$
$$x^2_{t+1} = f^2\left(x^2_t, u^2_t, w^2_t, x^1_{t-M_{12}}, x^4_{t-M_{42}}\right)$$
$$x^3_{t+1} = f^3\left(x^3_t, u^3_t, w^3_t, x^2_{t-M_{23}}, x^4_{t-M_{43}}\right)$$
$$x^4_{t+1} = f^4\left(x^4_t, u^4_t, w^4_t, x^3_{t-M_{34}}\right)$$

- x^i_t is the state of subsystem S_i at time t
- M_{ij} is the delay from subsystem S_i to subsystem S_j.
- N_i is the delay in receiving observations from subsystem S_i.
- The controller gets states from every subsystem.

Main Result

The optimal controller is of the form $u_t = \mu^{\text{opt}}_t(i^{\text{mem}}_t)$ where

$$i^{\text{mem}}_t = \left(u^i_{t-d_i:t-1},\ x^i_{t-N_i-b_i:t-N_i} \mid i = 1, 2, \dots, n\right)$$

- Optimal controller only needs *part* of the observation history
- Required history is i^{mem}_t
- Required history is *finite*, length does not depend on t
- Constants b_k, d_k, called *bandwidths* depend on graph structure and delays, not on dynamics

Main Result

The optimal controller is of the form $u_t = \mu^{\text{opt}}_t(i^{\text{mem}}_t)$ where

$$i^{\text{mem}}_t = \left(u^i_{t-d_i:t-1},\ x^i_{t-N_i-b_i:t-N_i} \mid i = 1, 2, \dots, n\right)$$

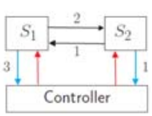
The bandwidths b_k and d_k are

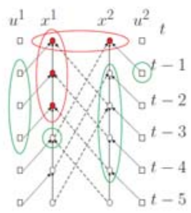
$$b_k = \max\left\{d_k, d_s + M_{ks} \mid s \in \mathcal{O}^k\right\} - N_k$$

$$d_k = \max\left\{N_k, N_s - 1 - M_{sk} \mid s \in \mathcal{I}^k\right\}$$

- $\mathcal{I}^k$ is the set of vertices with incoming edges into vertex k.
- $\mathcal{O}^k$ is the set of vertices with outgoing edges from vertex k.

Connections to Bayesian Networks



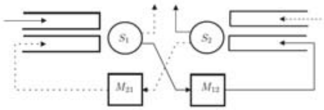


It is easy to check that

> i^{mem}_t forms the Markov blanket.

Example : Interconnected Queues

$$x^1(t+1) = f^1(x^1(t), x^2(t - M_{21}), u^1(t), w^1(t)),$$
$$x^2(t+1) = f^1(x^2(t), x^1(t - M_{12}), u^2(t), w^2(t))$$

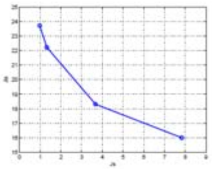


- The state cost is

$$g_s(x^1(t), x^2(t)) = \left(x^1_R(t) + x^1_B(t) + x^2_R(t) + x^2_B(t)\right)^2.$$

- The action cost is

$$g_a(u^1(t), u^2(t)) = \left(u^1(t) + 1 + u^2(t) + 1\right)^2,$$



- Minimize the infinite horizon discounted cost

$$J = \mathbb{E}\left(\sum_{t=0}^{\infty} \beta^t \left((1-\alpha) g_s + \alpha g_a\right)\right),$$
$$= (1-\alpha) J_s + \alpha J_a$$

- The propagation and measurement delays are

$$M_{12} = 2, \quad M_{21} = 1, \quad \text{and} \quad N_1 = 2, \quad N_2 = 1$$

The bands for the optimal controller are

> $b_1 = 1, \quad b_2 = 2 \quad \text{and} \quad d_1 = 2, \quad d_2 = 1.$

Conclusions

Networked MDPs

- Optimal control synthesis
- Centralized control, distributed delayed state feedback
- Finite state systems

> An optimal controller exists with finite memory.

- Bandwidths depend only on the graph structure and the delays.
- The bands forms a Markov blanket for unknown states.