On the Approximate Solution of POMDP and the Near-Optimality of Finite-State Controllers

Huizhen (Janey) Yu (janey@mit.edu)

Dimitri Bertsekas (dimitrib@mit.edu)

Lab for Information and Decision Systems, M.I.T.

April 2006

POMDP (Partially Observable Markov Decision Problems)

- Very important in control and other fields

POMDP (Partially Observable Markov Decision Problems)

- Very important in control and other fields
- Hard (because of infinite state space - the space of “beliefs” or conditional state distributions)

POMDP (Partially Observable Markov Decision Problems)

- Very important in control and other fields
- Hard (because of infinite state space - the space of “beliefs” or conditional state distributions)
- Last frontier for Exact DP

POMDP (Partially Observable Markov Decision Problems)

- Very important in control and other fields
- Hard (because of infinite state space - the space of “beliefs” or conditional state distributions)
- Last frontier for Exact DP
- Great challenge for Approximate DP

Approximation/Finite Parametrization Methods for POMDP

Let's focus on finite parametrizations of finite-state POMDP:

- Discretization/interpolation in belief space

Approximation/Finite Parametrization Methods for POMDP

Let's focus on finite parametrizations of finite-state POMDP:

- Discretization/interpolation in belief space
- Problem approximation (use a finite-state MDP to approximate a POMDP/aggregation)

Approximation/Finite Parametrization Methods for POMDP

Let's focus on finite parametrizations of finite-state POMDP:

- Discretization/interpolation in belief space
- Problem approximation (use a finite-state MDP to approximate a POMDP/aggregation)
- Piecewise linear approximations (approximate value and policy iteration)

Approximation/Finite Parametrization Methods for POMDP

Let's focus on finite parametrizations of finite-state POMDP:

- Discretization/interpolation in belief space
- Problem approximation (use a finite-state MDP to approximate a POMDP/aggregation)
- Piecewise linear approximations (approximate value and policy iteration)
- Optimization over a set of finite-state controllers

Outline of the Talk

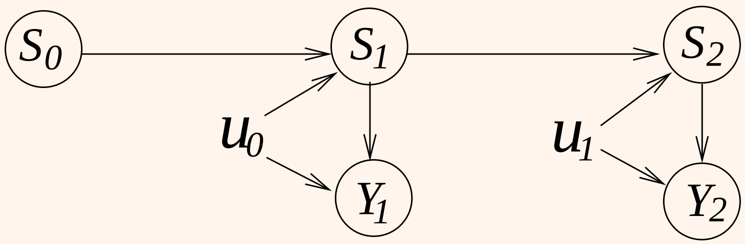
- Introduction to POMDP (finite-state average cost)
- Lower bounds/MDP approximations to POMDP optimal cost function (a brief summary of our work)
- On Near-Optimality of the Set of Finite-State Controllers for Average Cost POMDP

References

- H. Yu, "Approximate Solution Methods for Partially Observable Markov and Semi-Markov Decision Processes," Ph.D. Thesis, Dept of EECS, MIT, Jan. 2006.
- H. Yu and D. P. Bertsekas, "Discretized Approximations for POMDP with Average Cost," The 20th Conference on Uncertainty in Artificial Intelligence, 2004, Banff, Canada.
- H. Yu and D. P. Bertsekas, "Near-Optimality of Finite-State Controllers," LIDS Report 2689, MIT, April 2006.

Introduction to POMDP

Graphical model of discrete-time POMDP:



$\mathcal{S}$ state space, $\mathcal{Y}$ observation space, $\mathcal{U}$ control space

$g(s,u)$ per-stage cost

ξ initial distribution of state

$\pi \in \Pi$ history-dependent randomized policies

$\pi = \{\mu_0, \mu_1, \ldots,\}$ $\quad \mu_0(\cdot), \mu_t(h_t, \cdot) \in \mathcal{P}(\mathcal{U})$

$h_t = (u_0, y_1, u_1, \ldots, y_t)$, observed history up to time t

$\xi, \pi \longrightarrow$ stochastic process $\{S_0, U_0, S_1, Y_1, U_1, \ldots\}$

joint probability distribution $\mathbb{P}^{\xi,\pi}$

Expected Cost Criteria

k-Stage cost

$$J_k^\pi(\xi) = E^{\mathbb{P}^{\xi,\pi}} \left\{ \sum_{t=0}^{k-1} g(S_t, U_t) \right\} \qquad J_k^*(\xi) = \inf_{\pi \in \Pi} J_k^\pi(\xi)$$

Average cost

$$J_-^\pi(\xi) = \liminf_{k \to \infty} \frac{1}{k} J_k^\pi(\xi) \qquad J_+^\pi(\xi) = \limsup_{k \to \infty} \frac{1}{k} J_k^\pi(\xi)$$

$$J_-^*(\xi) = \inf_{\pi \in \Pi} J_-^\pi(\xi) \qquad J_+^*(\xi) = \inf_{\pi \in \Pi} J_+^\pi(\xi)$$

Important Analytical Features

- Sufficient statistic for control is ξ_t:
  belief state/conditional distribution of the state

Important Analytical Features

- Sufficient statistic for control is ξ_t:
  belief state/conditional distribution of the state

- k-stage policy cost function: linear in ξ

Important Analytical Features

- Sufficient statistic for control is ξ_t:
  belief state/conditional distribution of the state

- k-stage policy cost function: linear in ξ

- Lower k-stage optimal cost function: piecewise linear/concave in ξ

Important Analytical Features

- Sufficient statistic for control is ξ_t:
  belief state/conditional distribution of the state

- k-stage policy cost function: linear in ξ

- Lower k-stage optimal cost function: piecewise linear/concave in ξ

- Key questions for average cost: are the lower or upper optimal cost functions flat (constant/independent of the initial belief ξ)? Are they equal?

Important Analytical Features

- Sufficient statistic for control is ξ_t:
  belief state/conditional distribution of the state
- k-stage policy cost function: linear in ξ
- Lower k-stage optimal cost function: piecewise linear/concave in ξ
- Key questions for average cost: are the lower or upper optimal cost functions flat (constant/independent of the initial belief ξ)? Are they equal?
- In POMDP constant cost across beliefs is far less "likely" than in MDP

Average Cost Optimality Equations

Pair of coupled optimality equations

$$J(\xi) = \min_{u \in \mathcal{U}} E\left\{J\left(\phi_u(\xi, Y)\right)\right\}, \qquad U(\xi) \stackrel{def}{=} \arg\min_{u \in \mathcal{U}} E\left\{J\left(\phi_u(\xi, Y)\right)\right\},$$

$$J(\xi) + h(\xi) = \min_{u \in U(\xi)} \left[\bar{g}(\xi, u) + E\left\{h\left(\phi_u(\xi, Y)\right)\right\}\right], \tag{1}$$

Average Cost Optimality Equations

Pair of coupled optimality equations

$$J(\xi) = \min_{u \in \mathcal{U}} E\left\{J\left(\phi_u(\xi, Y)\right)\right\}, \qquad U(\xi) \stackrel{def}{=} \arg\min_{u \in \mathcal{U}} E\left\{J\left(\phi_u(\xi, Y)\right)\right\},$$

$$J(\xi) + h(\xi) = \min_{u \in U(\xi)} \left[\bar{g}(\xi, u) + E\left\{h\left(\phi_u(\xi, Y)\right)\right\}\right], \tag{1}$$

Constant average cost DP equation

$$\lambda + h(\xi) = \min_{u \in \mathcal{U}} \left[\bar{g}(\xi, u) + E\left\{h\left(\phi_u(\xi, Y)\right)\right\}\right]. \tag{2}$$

Average Cost Optimality Equations

Pair of coupled optimality equations

multichain

$$J(\xi) = \min_{u \in \mathcal{U}} E\{J(\phi_u(\xi, Y))\}, \qquad U(\xi) \stackrel{def}{=} \arg\min_{u \in \mathcal{U}} E\{J(\phi_u(\xi, Y))\},$$

$$J(\xi) + h(\xi) = \min_{u \in U(\xi)} [\bar{g}(\xi, u) + E\{h(\phi_u(\xi, Y))\}], \qquad (1)$$

Constant average cost DP equation

unichain

$$\lambda + h(\xi) = \min_{u \in \mathcal{U}} [\bar{g}(\xi, u) + E\{h(\phi_u(\xi, Y))\}]. \qquad (2)$$

Average Cost Optimality Equations

Pair of coupled optimality equations

$$J(\xi) = \min_{u \in \mathcal{U}} E\left\{J\left(\phi_u(\xi, Y)\right)\right\}, \qquad U(\xi) \stackrel{def}{=} \arg\min_{u \in \mathcal{U}} E\left\{J\left(\phi_u(\xi, Y)\right)\right\},$$

multichain

$$J(\xi) + h(\xi) = \min_{u \in U(\xi)} \left[\bar{g}(\xi, u) + E\left\{h\left(\phi_u(\xi, Y)\right)\right\}\right], \tag{1}$$

Constant average cost DP equation

$$\lambda + h(\xi) = \min_{u \in \mathcal{U}} \left[\bar{g}(\xi, u) + E\left\{h\left(\phi_u(\xi, Y)\right)\right\}\right]. \tag{2}$$

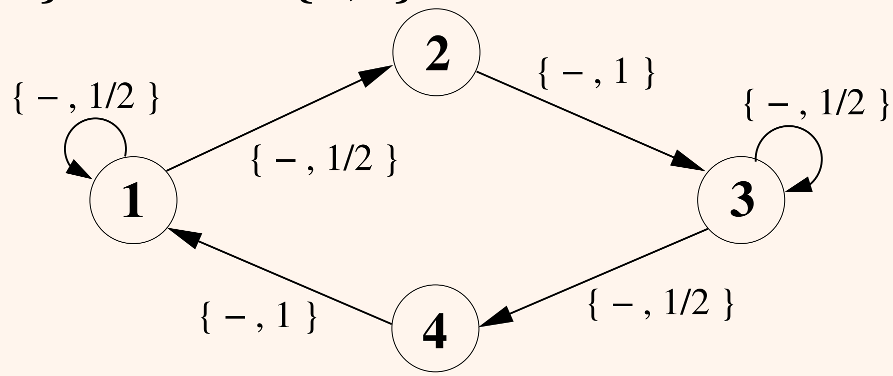
unichain

Bounded Solutions (J^*, h^*) of (1) $\Rightarrow$ $J^*_-(\cdot) = J^*_+(\cdot) = J^*(\cdot)$

Bounded Solutions (λ^*, h^*) of (2) $\Rightarrow$ $J^*_-(\cdot) = J^*_+(\cdot) = \lambda^*$

Example: Non-Constant Optimal Average Cost

States: $\{1, 2, 3, 4\}$, Actions: $\{a, b\}$



- Markov chain recurrent and aperiodic

Observations: $\{c, d\}$, state $1, 3 \rightarrow c$; state $2, 4 \rightarrow d$

Cost: $g(1, a) = 1, g(1, b) = 0$; $g(3, a) = 0, g(3, b) = 1$

$\bar{\xi} : \bar{\xi}(1) = 1$ or $\bar{\xi}(3) = 1$, $J^*(\bar{\xi}) = 0$

$\xi : \xi(1) = \xi(3) = 1/2$, $J^*(\xi) = 1/3 > 0$

- Non-constant optimal average cost in POMDP

Existence of Solutions to Constant AC DP Equation

Sufficient conditions:

- reachability and detectability [Platzman 80]
- interior accessibility, relative interior accessibility [Hsu, Chuang, Arapostathis 05]

Constant AC: not fully understood

Non-constant AC: not much has been done

Outline of the Talk

- Introduction to POMDP and the average cost problem
- Lower bounds for average cost POMDP
- Near-optimality of finite-state controllers

Lower Bounds for Average Cost POMDP

$$\tilde{J}^*(\xi) \leq J^*_{-}(\xi) \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

$\uparrow$

approximating processes (usually finite-state MDPs)

Lower Bounds for Average Cost POMDP

$$\tilde{J}^*(\xi) \leq J^*_-(\xi) \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

$\uparrow$

approximating processes (usually finite-state MDPs)

Notes:

- $\tilde{J}^*$ computable when spaces are finite:
  the bounds involve discretized approximations, finite state and control MDP algorithms

Lower Bounds for Average Cost POMDP

$$\tilde{J}^*(\xi) \leq J^*_-(\xi) \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

$\uparrow$

approximating processes (usually finite-state MDPs)

Notes:

- $\tilde{J}^*$ computable when spaces are finite:
  the bounds involve discretized approximations, finite state and control MDP algorithms
- neither J^* nor $\tilde{J}^*$ have to be constant

Lower Bounds for Average Cost POMDP

$$\tilde{J}^*(\xi) \leq J^*_-(\xi) \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

$\uparrow$

approximating processes (usually finite-state MDPs)

Notes:

- $\tilde{J}^*$ computable when spaces are finite:
  the bounds involve discretized approximations, finite state and control MDP algorithms
- neither J^* nor $\tilde{J}^*$ have to be constant
- use of lower bounds:
  characterize the optimal performance, provide suboptimal control

Lower Bounds for Average Cost POMDP

$$\tilde{J}^*(\xi) \leq J^*_-(\xi) \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

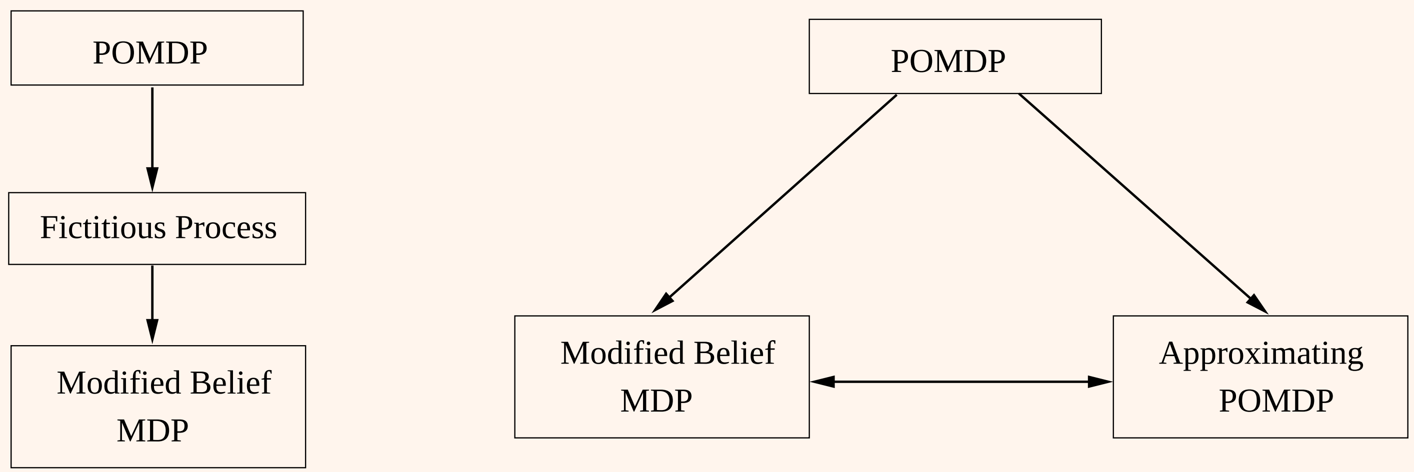
$\uparrow$

approximating processes (usually finite-state MDPs)

Notes:

- $\tilde{J}^*$ computable when spaces are finite:
  the bounds involve discretized approximations, finite state and control MDP algorithms
- neither J^* nor $\tilde{J}^*$ have to be constant
- use of lower bounds:
  characterize the optimal performance, provide suboptimal control
- extension to semi-Markov and constrained cases

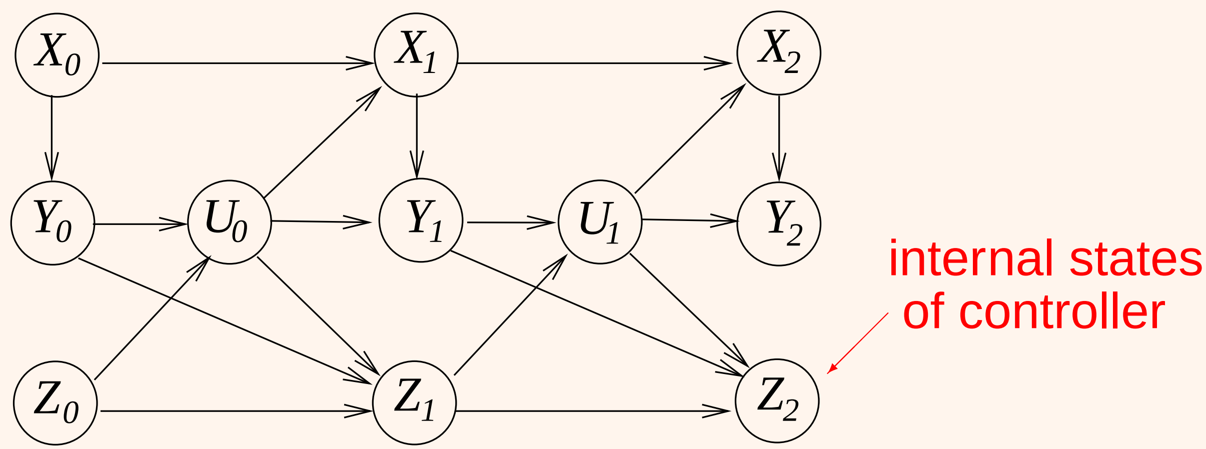
Outline of the Analysis



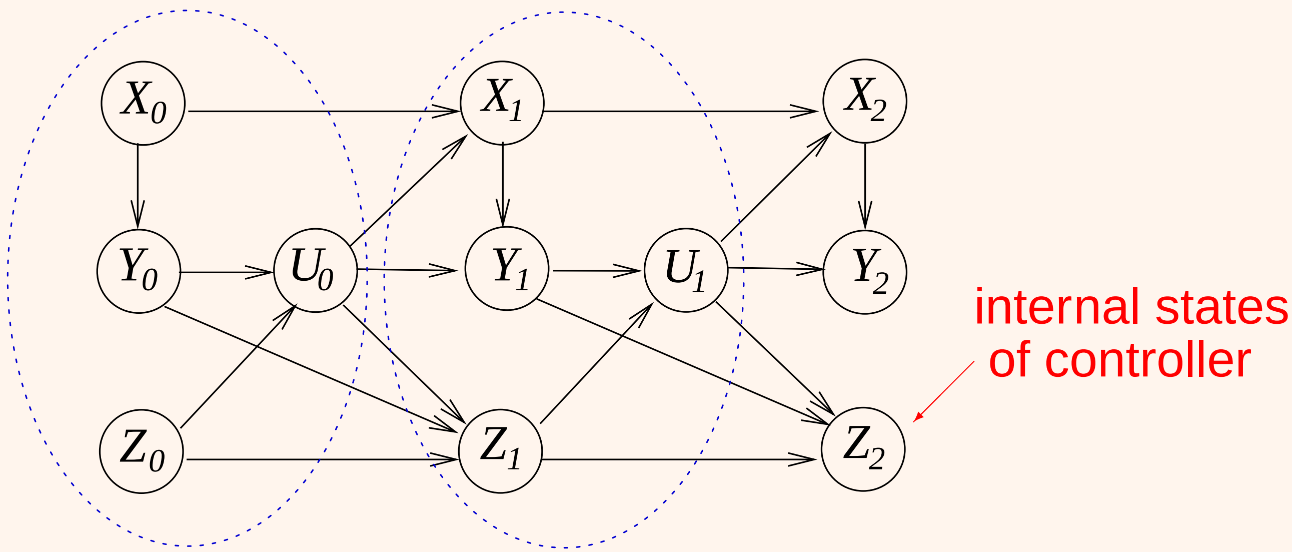
Outline of the Talk

- Introduction to POMDP and the average cost problem
- Lower bounds for average cost POMDP
- Near-optimality of finite-state controllers

Finite-State Controllers



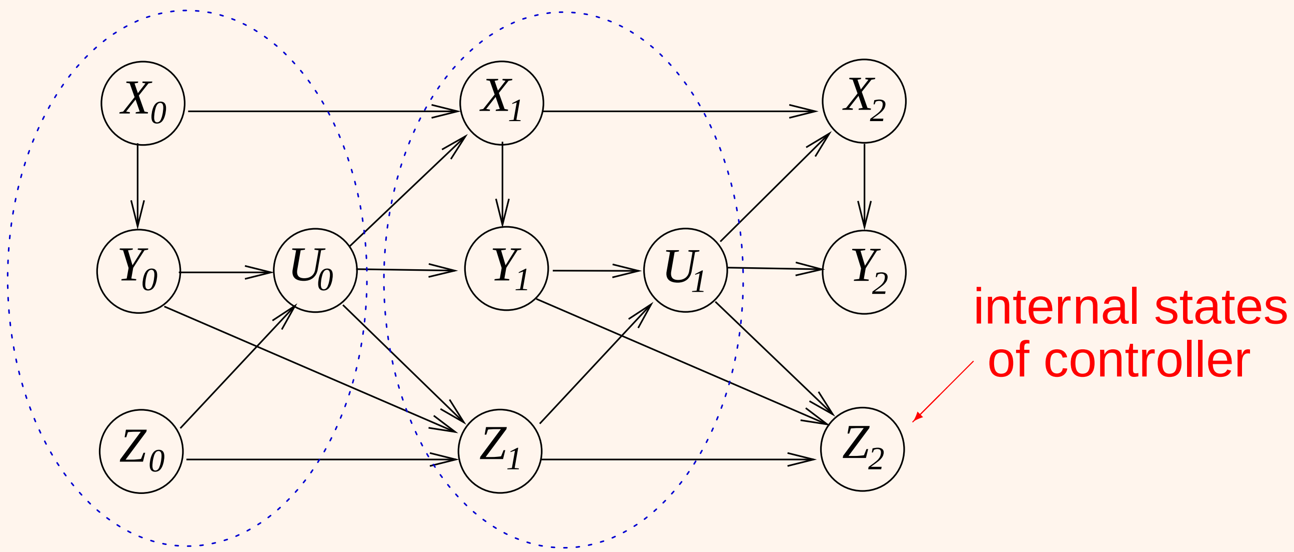
Finite-State Controllers



Key properties and advantages of FSC

$\{(S_t, Y_t, Z_t, U_t)\}$ jointly Markov chain, $\{(S_t, Y_t, Z_t)\}$ marginally Markov chain

Finite-State Controllers

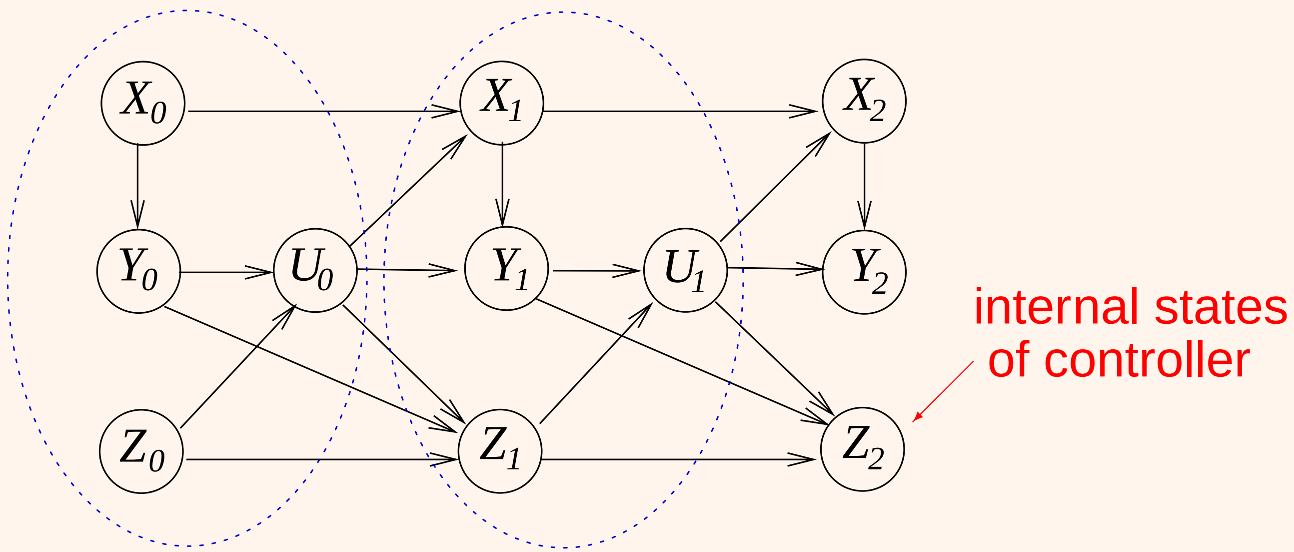


Key properties and advantages of FSC

$\{(S_t, Y_t, Z_t, U_t)\}$ jointly Markov chain, $\{(S_t, Y_t, Z_t)\}$ marginally Markov chain

asymptotic behavior well-understood; MDP theory applies

Finite-State Controllers



Key properties and advantages of FSC

$\{(S_t, Y_t, Z_t, U_t)\}$ jointly Markov chain, $\{(S_t, Y_t, Z_t)\}$ marginally Markov chain

asymptotic behavior well-understood; MDP theory applies

connection with piecewise linear concave approximations (Hansen 1998)

Existence of Near-Optimal FSC/Average Cost

Question: Given $\epsilon > 0$, under what conditions does there exist an ϵ-optimal FSC (with sufficiently large number of internal states)?

Existence of Near-Optimal FSC/Average Cost

Question: Given $\epsilon > 0$, under what conditions does there exist an ϵ-optimal FSC (with sufficiently large number of internal states)?

Notes:

We mean ϵ-optimal simultaneously for all initial beliefs (so J^*_-: flat is a necessary condition)

The existence of an ϵ-optimal FSC for a given single initial belief is an open question

Existence of Near-Optimal FSC/Average Cost

Question: Given $\epsilon > 0$, under what conditions does there exist an ϵ-optimal FSC (with sufficiently large number of internal states)?

Notes:

We mean ϵ-optimal simultaneously for all initial beliefs (so $\underline{J}^*$: flat is a necessary condition)

The existence of an ϵ-optimal FSC for a given single initial belief is an open question

For discounted problems, there are easy answers:

- There is no FSC that is simultaneously ϵ-optimal for all initial beliefs
- There is an ϵ-optimal FSC for a given single initial belief

Near-Optimality of Finite-State Controllers

Assumption: Finite spaces, J_-^* constant

Our Main Theorem: $J_+^* = J_-^*$; *and for all* $\epsilon > 0$, *there exists an* ϵ-*optimal FSC.*

Near-Optimality of Finite-State Controllers

Assumption: Finite spaces, J_-^* constant

Our Main Theorem: $J_+^* = J_-^*$; *and for all* $\epsilon > 0$, *there exists an* ϵ-*optimal FSC.*

Recall definitions:

$$\pi = \{\mu_0, \mu_1, \ldots, \}, \quad \mu_0(\cdot), \mu_t(h_t, \cdot) \in \mathcal{P}(\mathcal{U}) \qquad J_k^\pi(\xi) = E^{\mathbb{P}^{\xi,\pi}} \left\{ \sum_{t=0}^{k-1} g(S_t, U_t) \right\}$$

$$J_-^\pi(\xi) = \liminf_{k \to \infty} \frac{1}{k} J_k^\pi(\xi) \qquad J_-^*(\xi) = \inf_{\pi \in \Pi} J_-^\pi(\xi)$$

Near-Optimality of Finite-State Controllers

Assumption: Finite spaces, J_-^* constant

Our Main Theorem: $J_+^* = J_-^*$*; and for all* $\epsilon > 0$*, there exists an* ϵ*-optimal FSC.*

Recall definitions:

$$\pi = \{\mu_0, \mu_1, \ldots, \}, \quad \mu_0(\cdot), \mu_t(h_t, \cdot) \in \mathcal{P}(\mathcal{U}) \qquad J_k^\pi(\xi) = E^{\mathbb{P}^{\xi,\pi}} \left\{ \sum_{t=0}^{k-1} g(S_t, U_t) \right\}$$

$$J_-^\pi(\xi) = \liminf_{k \to \infty} \frac{1}{k} J_k^\pi(\xi) \qquad J_-^*(\xi) = \inf_{\pi \in \Pi} J_-^\pi(\xi)$$

Lemma: $J_k^\pi(\xi)$ *linear;* $J_-^\pi(\xi)$ *concave;* $J_-^*(\xi)$ *concave.*

Near-Optimality of Finite-State Controllers

Assumption: Finite spaces, J_-^* constant

Our Main Theorem: $J_+^* = J_-^*$; *and for all* $\epsilon > 0$, *there exists an* ϵ*-optimal FSC.*

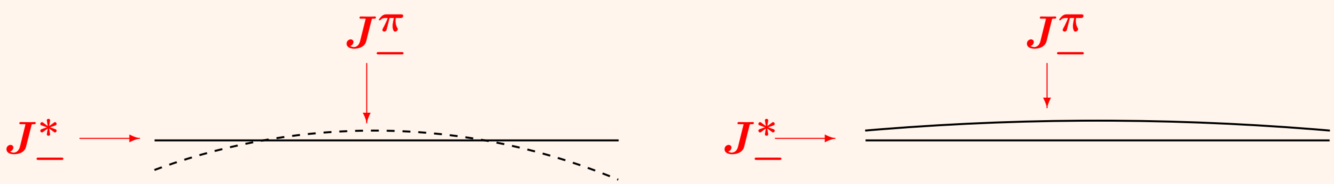
Recall definitions:

$$\pi = \{\mu_0, \mu_1, \ldots, \}, \quad \mu_0(\cdot), \mu_t(h_t, \cdot) \in \mathcal{P}(\mathcal{U}) \qquad J_k^\pi(\xi) = E^{\mathbb{P}^{\xi,\pi}} \left\{ \sum_{t=0}^{k-1} g(S_t, U_t) \right\}$$

$$J_-^\pi(\xi) = \liminf_{k \to \infty} \frac{1}{k} J_k^\pi(\xi) \qquad J_-^*(\xi) = \inf_{\pi \in \Pi} J_-^\pi(\xi)$$

Lemma: $J_k^\pi(\xi)$ *linear;* $J_-^\pi(\xi)$ *concave;* $J_-^*(\xi)$ *concave.*

1st Key Observation: Suppose J_-^* constant. Then, the first case cannot happen ...



Near-Optimality of Finite-State Controllers

Assumption: Finite spaces, J_-^* constant

Our Main Theorem: $J_+^* = J_-^*$*; and for all* $\epsilon > 0$*, there exists an* ϵ*-optimal FSC.*

Recall definitions:

$$\pi = \{\mu_0, \mu_1, \ldots, \}, \quad \mu_0(\cdot), \mu_t(h_t, \cdot) \in \mathcal{P}(\mathcal{U}) \qquad J_k^\pi(\xi) = E^{\mathbb{P}^{\xi,\pi}} \left\{ \sum_{t=0}^{k-1} g(S_t, U_t) \right\}$$

$$J_-^\pi(\xi) = \liminf_{k \to \infty} \frac{1}{k} J_k^\pi(\xi) \qquad J_-^*(\xi) = \inf_{\pi \in \Pi} J_-^\pi(\xi)$$

Lemma: $J_k^\pi(\xi)$ *linear;* $J_-^\pi(\xi)$ *concave;* $J_-^*(\xi)$ *concave.*

1st Key Observation: Suppose J_-^* constant. Then, the first case cannot happen ...

- There must exist π nearly optimal for some interior ξ, with J_-^π almost "flat"

First Result on Independence from Initial Belief

Proposition: For all $\epsilon > 0$, there exists an ϵ-liminf optimal policy π that does not functionally depend on the initial belief ξ, i.e.,

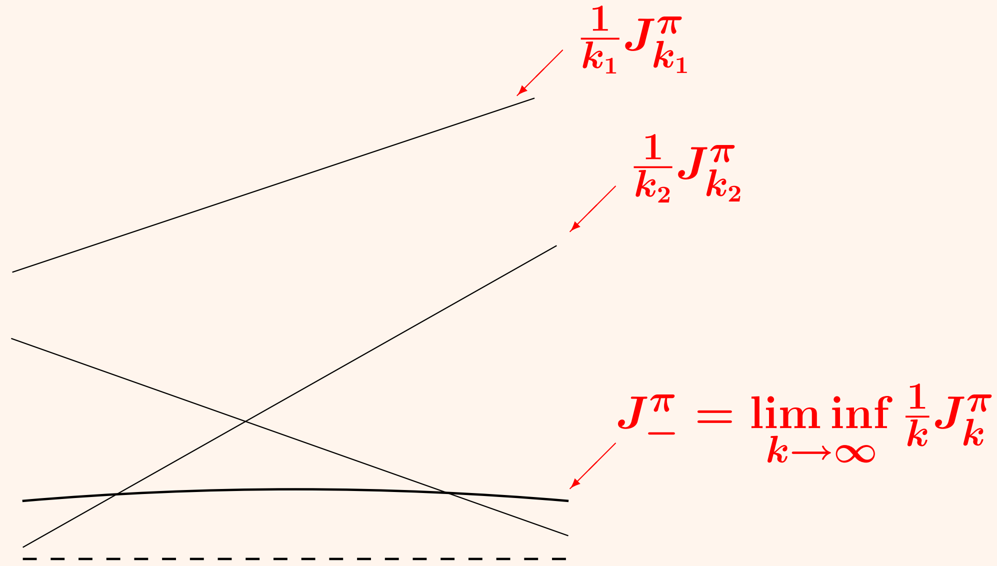
$$J_-^\pi(\xi) \leq J_-^* + \epsilon, \quad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

First Result on Independence from Initial Belief

Proposition: For all $\epsilon > 0$, there exists an ϵ-liminf optimal policy π that does not functionally depend on the initial belief ξ, i.e.,

$$J_{-}^{\pi}(\xi) \leq J_{-}^{*} + \epsilon, \quad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

2nd Key Observation: There must exist k_0, with $\frac{1}{k_0} J_{k_0}^{\pi}$ uniformly "close" to J_{-}^{π}



Finite-Stage Uniform "Closeness" Lemma

Lemma: For all $\epsilon > 0$, there exists $\pi_0 \in \Pi$ and an integer k_0 (depending on π_0) such that

$$\left|\frac{1}{k_0} J_{k_0}^{\pi_0}(\xi) - J_-^*\right| \leq \epsilon, \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

Finite-Stage Uniform "Closeness" Lemma

Lemma: For all $\epsilon > 0$, there exists $\pi_0 \in \Pi$ and an integer k_0 (depending on π_0) such that

$$|\frac{1}{k_0} J_{k_0}^{\pi_0}(\xi) - J_-^*| \leq \epsilon, \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

Final construction: Construct infinite-stage policy by replication of the finite-stage policy:

- Form a new policy π_1 that has a "finite-length history window":

$$\pi_0 = \{\mu_0, \mu_1, \ldots, \mu_{k_0-1}, \ldots\}$$
$$\pi_1 = \{\mu_0, \mu_1, \ldots, \mu_{k_0-1}, \mu'_{k_0}, \mu'_{k_0+1}, \ldots, \mu'_{2k_0-1}, \ldots\}$$

$$\mu'_t(h_t, \cdot) = \mu_{\bar{k}(t)}(\delta_{\bar{k}(t)}(h_t), \cdot), \quad \bar{k}(t) \stackrel{def}{=} \operatorname{mod}(t, k_0),$$

where $\delta_{\bar{k}(t)}$ extracts the last length-$\bar{k}(t)$ segment of a length-t history h_t.

Finite-Stage Uniform "Closeness" Lemma

Lemma: For all $\epsilon > 0$, there exists $\pi_0 \in \Pi$ and an integer k_0 (depending on π_0) such that

$$\left|\frac{1}{k_0} J_{k_0}^{\pi_0}(\xi) - J_-^*\right| \leq \epsilon, \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

Final construction: Construct infinite-stage policy by replication of the finite-stage policy:

- Form a new policy π_1 that has a "finite-length history window":

$$\pi_0 = \{\mu_0, \mu_1, \ldots, \mu_{k_0-1}, \ldots\}$$
$$\pi_1 = \{\mu_0, \mu_1, \ldots, \mu_{k_0-1}, \mu'_{k_0}, \mu'_{k_0+1}, \ldots, \mu'_{2k_0-1}, \ldots\}$$

$$\mu'_t(h_t, \cdot) = \mu_{\bar{k}(t)}(\delta_{\bar{k}(t)}(h_t), \cdot), \quad \bar{k}(t) \stackrel{def}{=} \mathrm{mod}(t, k_0),$$

  where $\delta_{\bar{k}(t)}$ extracts the last length-$\bar{k}(t)$ segment of a length-t history h_t.

- $J_-^{\pi_0}(\cdot)$: uniformly close to $J_-^* \implies J_-^{\pi_1}(\cdot)$: uniformly close to J_-^*

Finite-Stage Uniform "Closeness" Lemma

Lemma: For all $\epsilon > 0$, there exists $\pi_0 \in \Pi$ and an integer k_0 (depending on π_0) such that

$$|\frac{1}{k_0} J_{k_0}^{\pi_0}(\xi) - J_-^*| \leq \epsilon, \qquad \forall\, \xi \in \mathcal{P}(\mathcal{S}).$$

Final construction: Construct infinite-stage policy by replication of the finite-stage policy:

- Form a new policy π_1 that has a "finite-length history window":

$$\pi_0 = \{\mu_0, \mu_1, \ldots, \mu_{k_0-1}, \ldots\}$$
$$\pi_1 = \{\mu_0, \mu_1, \ldots, \mu_{k_0-1}, \mu'_{k_0}, \mu'_{k_0+1}, \ldots, \mu'_{2k_0-1}, \ldots\}$$

$$\mu'_t(h_t, \cdot) = \mu_{\bar{k}(t)}(\delta_{\bar{k}(t)}(h_t), \cdot), \quad \bar{k}(t) \overset{def}{=} \mathrm{mod}(t, k_0),$$

where $\delta_{\bar{k}(t)}$ extracts the last length-$\bar{k}(t)$ segment of a length-t history h_t.

- $J_-^{\pi_0}(\cdot)$: uniformly close to $J_-^* \implies J_-^{\pi_1}(\cdot)$: uniformly close to J_-^*
- For finite spaces: π_1 is FSC $\implies J_-^{\pi_1}(\xi) = J_+^{\pi_1}(\xi), \forall\, \xi \in \mathcal{P}(\mathcal{S})$.

Near-Optimality of FSC

Assumption: Finite spaces, J_-^* constant

Theorem:

- $J_+^* = J_-^*$
- for all $\epsilon > 0$, there is a FSC that is ϵ-optimal, and does not functionally depend on the initial belief ξ.

Note:

The ϵ-optimal FSC is also a finite-history controller

Intuitive implication: If J_-^* does not depend on the initial belief, old information becomes increasingly obsolete