Reducing Undiscounted Markov Decision Processes and Stochastic Games with Unbounded Costs to Discounted Ones

Jefferson Huang

School of Operations Research and Information Engineering
Cornell University

December 8, 2016

Northeast Regional Conference on
Optimization and Optimal Control under Uncertainty

IBM Thomas J. Watson Research Center

Yorktown Heights, NY

What This Talk is About

Transformations of certain undiscounted Markov decision processes (MDPs) and zero-sum stochastic games to discounted ones.

- Undiscounted = Total or Average Costs
- For total costs, the "transition rates" may not be substochastic
- Generalization of work done by [Akian & Gaubert, Hoffman & Veinott, Ross]

Lead to reductions of the original model to a (standard) discounted one

- Validity of optimality equations, existence of optimal policies, complexity estimates for algorithms for the original model.

Often easier to study a discounted model than an undiscounted one.

Outline

1. Total-Cost MDPs
   - Transience Assumption
   - Reduction to Discounted MDP
2. Average-Cost MDPs
   - Recurrence Assumption
   - Reduction to Discounted MDP
3. Two-Player Zero-Sum Stochastic Games
   - Reduction of Total Costs to Discounting
   - Reduction of Average Costs to Discounting

Markov Decision Process: Model Definition

$\mathbb{X}$ = state space = countable set

$\mathbb{A}$ = action space = countable set

$A(x)$ = set of available actions at state x = subset of $\mathbb{A}$

$c(x, a)$ = one-step cost when state is x and action a is performed

$q(y|x, a)$ = "transition rate" to state y when current state is x and action a is performed

- Not necessarily substochastic!

For the case of Borel state and action spaces, see [Feinberg & H].

Super-Stochastic Transition Rates

We allow $q(\cdot|x,a)$ to take values greater than one.

Possible Interpretations:

- Controlled Multitype Branching Processes [Eaves, Pliska, Rothblum, Veinott]: $q(y|x,a) =$ expected number of type y individuals born from from a type x individual when action a is applied.
- Multi-Armed Bandits with Risk-Seeking Utilities [Denardo, Feinberg, Rothblum]: $q(y|x,a) = p(y|x,a)e^{\lambda r(x,y)}$, where $\lambda > 0$ and $r(x,y)$ is the payoff earned when bandit a transitions from state x to y.
- Discount Factors Greater Than One [Hinderer, Waldmann]: Equivalently consider discount factors $\alpha(x,a) := \sum_{y\in\mathbb{X}} q(y|x,a)$ and transition probabilities $p(y|x,a) := q(y|x,a)/q(\mathbb{X}|x,a)$.

Optimality Criterion

$\mathbb{F}$ = set of all deterministic stationary policies

For $x \in \mathbb{X}$ and $\phi \in \mathbb{F}$, let $c_\phi(x) := c(x, \phi(x))$ and

$$Q_\phi(x, y) := q(dy|x, \phi(x))$$

Total costs:

$$v^\phi(x) := \sum_{t=0}^{\infty} Q_\phi^t c_\phi(x)$$

$\phi_* \in \mathbb{F}$ is optimal if

$$v^{\phi_*}(x) = \inf_{\phi \in \mathbb{F}} v^\phi(x) \qquad \forall x \in \mathbb{X}.$$

It sufficies to consider deterministic stationary policies [Feinberg & H].

Transience Assumption

Assumption (T)

There is a "weight function" $V : \mathbb{X} \to [1, \infty)$ and a constant $K < \infty$ satsfying

$$V(x)^{-1} \sum_{t=0}^{\infty} Q_{\phi}^{t} V(x) \leqslant K \qquad \forall x \in \mathbb{X},\ \phi \in \mathbb{F}.$$

[Denardo, Hernández-Lerma, Lasserre, Pliska, Rothblum, Veinott]

Implies that for $B \subseteq \mathbb{X}$, the "occupation time"

$$\sum_{t=1}^{\infty} Q_{\phi}^{t} \mathbf{1}_B(x) \leqslant KV(x) \qquad \forall x \in \mathbb{X},\ \phi \in \mathbb{F}.$$

An Equivalent Condition

Theorem (Feinberg & H)

Assumption (T) holds iff there exist functions $V : \mathbb{X} \to [1, \infty)$*,* $\mu : \mathbb{X} \to [1, \infty)$ *and a constant* $K < \infty$ *that satisfy*

$$V(x) \leqslant \mu(x) \leqslant KV(x) \qquad \forall x \in \mathbb{X}$$

and

$$\mu(x) \geqslant V(x) + \sum_{y \in \mathbb{X}} q(y|x, a)\mu(y) \qquad \forall x \in \mathbb{X},\ a \in A(x).$$

Was known to hold under additional compactness-continuity conditions, e.g., [Hernández-Lerma & Lasserre].

Example: Single-Server Arrival and Service Control

At most 1 arrival and 1 service completion per decision epoch.

$\mathbb{X}$ = number of customers in the queue = $\{0, 1, 2, \dots\}$

$\mathbb{A} = A(x) = [a_{\min}, a_{\max}] \times [s_{\min}, s_{\max}] \subseteq (0, 1) \times (0, 1)$, where $a_{\max} < s_{\min}$

- Prob{1 arrival} $= a \in [a_{\min}, a_{\max}]$
- Prob{1 service completion} $= s \in [s_{\min}, s_{\max}]$

$c(x, (s, a)) = c(x) + d_{\text{Arr}}(a) + d_{\text{Serv}}(s)$

- c is polynomially bounded
- d_{Arr} decreasing in a; d_{Serv} increasing in s

Transition rates:

$$q(y|x, (s, a)) := \begin{cases} (1-a)s & x \geqslant 1,\ y = x - 1 \\ as + (1-a)(1-s) & x \geqslant 1,\ y = x \\ a(1-s) & x \geqslant 0,\ y = x + 1 \\ 0 & x = y = 0 \end{cases}$$

Example: Single-Server Arrival and Service Control

$v^{\Phi}(x) =$ expected total cost incurred to empty the queue starting from a queue size of x.

Let $\rho := \frac{a_{\max}(1-s_{\min})}{(1-a_{\max})s_{\min}} < 1$, $r \in (1, \rho^{-1})$, and

$$\gamma := (r-1)r^{-1}a_{\max}(1-s_{\min})(\rho^{-1}-r) > 0$$

For sufficiently large $C > 0$, the functions

$$V(x) := \gamma C r^{x} \qquad \mu(x) := C r^{x}$$

and $K := \gamma^{-1}$ satisfy the hypotheses of the necessary and sufficient condition for Assumption (T).

Transformation to a (Standard) Discounted MDP

$\tilde{\beta} := (K-1)/K$

$\tilde{\mathbb{X}} := \mathbb{X} \cup \{\tilde{x}\}$, and $\tilde{\mathbb{A}} := \mathbb{A} \cup \{\tilde{a}\}$

$\tilde{A}(x) := A(x)$ if $x \in \mathbb{X}$ and $\tilde{A}(\tilde{x}) := \{\tilde{a}\}$.

$$\tilde{p}(y|x,a) := \begin{cases} \frac{1}{\tilde{\beta}\,\mu(x)}\mu(y)q(y|x,a), & x, y \in \mathbb{X}, a \in A(x), \\ 1 - \frac{1}{\tilde{\beta}\,\mu(x)}\sum_{y \in \mathbb{X}} \mu(y)q(y|x,a), & y = \tilde{x}, x \in \mathbb{X}, a \in A(x), \\ 1 & y = \tilde{x}, (x,a) = (\tilde{x}, \tilde{a}). \end{cases}$$

$$\tilde{c}(x,a) := \begin{cases} c(x,a)/\mu(x), & x \in \mathbb{X}, a \in A(x), \\ 0, & (x,a) = (\tilde{x}, \tilde{a}). \end{cases}$$

$$\tilde{v}^{\phi}_{\tilde{\beta}}(x) := \tilde{\mathbb{E}}^{\phi}_{x} \sum_{t=0}^{\infty} \tilde{\beta}^{t} \tilde{c}(x_t, a_t) \qquad x \in \tilde{\mathbb{X}}, \ \phi \in \mathbb{F}$$

Reduction to a Discounted MDP

Theorem (Feinberg & H)

Suppose Assumption (T) holds, and that the constant $\overline{c} < \infty$ satisfies

$$|c(x,a)| \leqslant \overline{c}V(x) \qquad \forall x \in \mathbb{X},\ a \in A(x).$$

Then

$$v^{\phi}(x) = \mu(x)\tilde{v}^{\phi}_{\tilde{\beta}}(x) \qquad \forall x \in \mathbb{X},\ \phi \in \mathbb{F}.$$

Proof. Let $\tilde{c}_{\phi}(x) := \tilde{c}(x,\phi(x))$ and $\tilde{P}_{\phi}(x,y) := \tilde{p}(y|x,\phi(x))$. Then

$$\tilde{\beta}^n \tilde{P}^n_{\phi}\tilde{c}_{\phi}(x) = \mu(x)^{-1} Q^n_{\phi} c_{\phi}(x) \qquad \forall n \in \{0,1,\dots\}$$

□

Implies that to minimize v^{ϕ}, it suffices to minimize $\tilde{v}^{\phi}_{\tilde{\beta}}$.

Leads to results on validity of optimality equation and existence and characterization of optimal policies for the original MDP [Feinberg & H].

Complexity of Policy Iteration

Provides alternative proof of the iteration bound for Howard's policy iteration derived by [Denardo].

- Compute v^{ϕ} for current policy, let ϕ_{+} satisfy $T^{\phi_{+}} v^{\phi} = TV^{\phi}$, replace ϕ with ϕ_{+}, repeat ...

$m :=$ number of state-action pairs (x, a)

Theorem (Denardo)

The number of iterations required by Howard's policy iteration (HPI) algorithm to compute an optimal policy for the original MDP is

$$O(mK \log K).$$

Proof. [Feinberg & H] Reduce the original MDP to a discounted one, show that (HPI) for the discounted one corresponds to (HPI) for the original one, and use the bound derived by [Scherrer] for discounted MDPs. □

Optimality Criterion

$\mathbb{F}$ = set of all deterministic stationary policies

For $x \in \mathbb{X}$ and $\phi \in \mathbb{F}$, let $c_\phi(x) := c(x, \phi(x))$ and

$$Q_\phi(x, y) := q(dy|x, \phi(x))$$

Average costs:

$$w^\phi(x) := \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} Q_\phi^t c_\phi(x)$$

$\phi_* \in \mathbb{F}$ is optimal if

$$w^{\phi_*}(x) = \inf_{\phi\in\mathbb{F}} w^\phi(x) \qquad \forall x \in \mathbb{X}.$$

It sufficies to consider deterministic stationary policies [Feinberg & H].

Recurrence Assumption

Let

$$ {}_\ell Q_\phi(x,y) := \begin{cases} q(y|x,\phi(x)) & y \neq \ell \\ 0 & y = \ell \end{cases} $$

Assumption (HT)

There is a "weight function" $V : \mathbb{X} \to [1,\infty)$ and a constant $K < \infty$ satsfying

$$ V(x)^{-1} \sum_{t=0}^{\infty} {}_\ell Q_\phi^t V(x) \leqslant K \qquad \forall x \in \mathbb{X},\ \phi \in \mathbb{F}. $$

When $\mathbb{X}$ and $\mathbb{A}$ are finite, Assumption (HT) means

- MDP is unichain, and state ℓ is recurrent under all ϕ.
- Hitting time to state ℓ is uniformly bounded in x and ϕ.
- w^ϕ is constant for every ϕ.

Generalizes a condition used by [Ross] to reduce average-cost MDPs to discounted ones.

An Equivalent Assumption

Theorem (Feinberg & H)

Assumption (HT) holds iff there exist functions $V : \mathbb{X} \to [1, \infty)$, $\mu : \mathbb{X} \to [1, \infty)$ *and a constant* $K < \infty$ *that satisfy*

$$V(x) \leqslant \mu(x) \leqslant KV(x) \qquad \forall x \in \mathbb{X}$$

and

$$\mu(x) \geqslant V(x) + \sum_{y \neq \ell} q(y|x, a)\mu(y) \qquad \forall x \in \mathbb{X},\ a \in A(x).$$

Example: Single-Server Arrival and Service Control

Consider the queueing control model from Slide 8, with transition probabilities

$$q(y|x,(s,a)) := \begin{cases} (1-a)s & x \geqslant 1,\ y = x-1 \\ as + (1-a)(1-s) & x \geqslant 1,\ y = x \\ a(1-s) & x \geqslant 0,\ y = x+1 \\ 1 - a(1-s) & x = y = 0 \end{cases}$$

Let $\rho := \frac{a_{\max}(1-s_{\min})}{(1-a_{\max})s_{\min}} < 1$, $r \in (1, \rho^{-1})$, and

$$\gamma := (r-1)r^{-1}a_{\max}(1-s_{\min})(\rho^{-1}-r) > 0$$

Then for sufficiently large C, the functions $V(x) := \gamma C r^x$ and $\mu := C r^x$ and the constant $K := \gamma^{-1}$ satisfy $V \leqslant \mu \leqslant KV$ and

$$\mu(x) \geqslant V(x) + \sum_{y \neq 0} q(y|x,(a,s))\mu(y) \qquad \forall x \in \mathbb{X},\ (a,s) \in \mathbb{A}.$$

and hence satisfies Assumption (HT).

Transformation to a (Standard) Discounted MDP

$\bar{\beta} := (K-1)/K$

$\bar{\mathbb{X}} := \mathbb{X} \cup \{\bar{x}\}$, and $\bar{\mathbb{A}} := \mathbb{A} \cup \{\bar{a}\}$

$\bar{A}(x) := A(x)$ if $x \in \mathbb{X}$ and $\bar{A}(\bar{x}) := \{\bar{a}\}$.

$$\bar{p}(y|x,a) := \begin{cases} \frac{1}{\bar{\beta}\,\mu(x)}\,\mu(y)q(y|x,a), & y \neq \ell, x \in \mathbb{X}; \\ \frac{1}{\bar{\beta}\,\mu(x)}\,[\mu(x) - 1 - \sum_{y \neq \ell} \mu(y)q(y|x,a)] & y = \ell, x \in \mathbb{X}; \\ 1 - \frac{1}{\bar{\beta}\,\mu(x)}\,[\mu(x) - 1], & y = \bar{x}, x \in \mathbb{X}; \\ 1 & y = \bar{x}, (x,a) = (\bar{x}, \bar{a}). \end{cases}$$

$$\bar{c}(x,a) := \begin{cases} c(x,a)/\mu(x), & x \in \mathbb{X}, a \in A(x), \\ 0, & (x,a) = (\bar{x}, \bar{a}). \end{cases}$$

$$\bar{v}_{\bar{\beta}}^{\phi}(x) := \bar{\mathbb{E}}_x^{\phi} \sum_{t=0}^{\infty} \bar{\beta}^t \bar{c}(x_t, a_t) \qquad x \in \bar{\mathbb{X}},\ \phi \in \mathbb{F}.$$

Reduction to a Discounted MDP

Theorem (Feinberg & H)

Suppose Assumption (HT) holds, that $\sum_{y\in\mathbb{X}} q(y|x,a) = 1$ *for all* $x \in \mathbb{X}$ *and* $a \in A(x)$, *and that the constant* $\bar{c} < \infty$ *satisfies*

$$|c(x,a)| \leqslant \bar{c}V(x) \qquad \forall x \in \mathbb{X},\ a \in A(x).$$

Then

$$w^{\phi}(x) = \bar{v}_{\bar{\beta}}^{\phi}(\ell) \qquad \forall x \in \mathbb{X},\ \phi \in \mathbb{F}.$$

Proof. Show that for every ϕ, the function $h^{\phi}(x) := \mu(x)[\bar{v}_{\bar{\beta}}(x) - \bar{v}_{\bar{\beta}}(\ell)]$ satisfies

$$\bar{v}_{\bar{\beta}}^{\phi}(\ell) + h^{\phi}(x) = c_{\phi}(x) + Q_{\phi}h^{\phi}(x) \qquad \forall x \in \mathbb{X},$$

and that

$$\lim_{T\to\infty} \frac{1}{T} Q_{\phi}^{T} h^{\phi}(x) = 0.$$

□

Can be used to verify the validity of the average-cost optimality equation and the existence of stationary optimal policies [Feinberg & H]

Model Definition: Two-Player Zero-Sum Stochastic Game

$\mathbb{X}$ = state space = countable set

$\mathbb{A}^i$ = action space = countable set, $i = 1, 2$

$A^i(x) = \{$set of available actions for player $i = 1, 2$ at state $x\} \subseteq \mathbb{A}^i$

$c(x, a^1, a^2)$ = one-step cost when state is x and player $i = 1, 2$ plays action a^i

$q(y|x, a^1, a^2)$ = "transition rate" to state y when current state is x and player $i = 1, 2$ plays action a^i

- Not necessarily substochastic!

Total-Cost Criterion

Φ^i = set of all randomized stationary policies for player $i = 1, 2$

For $x \in \mathbb{X}$ and $(\phi^1, \phi^2) \in \Phi^1 \times \Phi^2$, let

$$c_{\phi^1,\phi^2}(x) := \sum_{a^1 \in A^1(x)} \sum_{a^2 \in A^2(x)} \phi^1(a^1|x)\phi^2(a^2|x)c(x, a^1, a^2)$$

and

$$Q_{\phi^1,\phi^2}(x, y) := \sum_{a^1 \in A^1(x)} \sum_{a^2 \in A^2(x)} \phi^1(a^1|x)\phi^2(a^2|x)q(y|x, a^1, a^2)$$

Total costs:

$$v^{\phi^1,\phi^2}(x) := \sum_{t=0}^{\infty} Q^t_{\phi^1,\phi^2} c_{\phi^1,\phi^2}(x)$$

Total-Cost Criterion

$\phi_* \in \Phi^1$ is optimal for player 1 if

$$\inf_{\phi^2 \in \Phi^2} v^{\phi_*, \phi^2}(x) = \inf_{\phi^2 \in \Phi^2} \sup_{\phi^1 \in \Phi^1} v^{\phi^1, \phi^2}(x) \qquad \forall x \in \mathbb{X}.$$

$\phi_* \in \Phi^2$ is optimal for player 2 if

$$\sup_{\phi^1 \in \Phi^1} v^{\phi^1, \phi_*}(x) = \sup_{\phi^1 \in \Phi^1} \inf_{\phi^2 \in \Phi^2} v^{\phi^1, \phi^2}(x) \qquad \forall x \in \mathbb{X}.$$

The game has a value v if

$$v(x) := \inf_{\phi^2 \in \Phi^2} \sup_{\phi^1 \in \Phi^1} v^{\phi^1, \phi^2}(x) = \sup_{\phi^1 \in \Phi^1} \inf_{\phi^2 \in \Phi^2} v^{\phi^1, \phi^2}(x) \qquad \forall x \in \mathbb{X}.$$

Transience Assumption

$\mathbb{F}^i$ = set of all deterministic stationary policies for player $i = 1, 2$.

Assumption (T)

There is a "weight function" $V : \mathbb{X} \to [1, \infty)$ and a constant $K < \infty$ satsfying

$$V(x)^{-1} \sum_{t=0}^{\infty} Q^t_{\phi^1,\phi^2} V(x) \leqslant K \qquad \forall x \in \mathbb{X},\ (\phi^1, \phi^2) \in \mathbb{F}^1 \times \mathbb{F}^2.$$

Implies that for $B \subseteq \mathbb{X}$, the "occupation time"

$$\sum_{t=1}^{\infty} Q^t_{\phi^1,\phi^2} \mathbf{1}_B(x) \leqslant KV(x) \qquad \forall x \in \mathbb{X},\ (\phi^1, \phi^2) \in \mathbb{F}^1 \times \mathbb{F}^2.$$

An Equivalent Condition

Theorem (Feinberg & H)

Assumption (T) holds iff there exist functions $V : \mathbb{X} \to [1, \infty)$, $\mu : \mathbb{X} \to [1, \infty)$ *and a constant* $K < \infty$ *that satisfy*

$$V(x) \leqslant \mu(x) \leqslant KV(x) \qquad \forall x \in \mathbb{X}$$

and

$$\mu(x) \geqslant V(x) + \sum_{y \in \mathbb{X}} q(y|x, a^1, a^2)\mu(y) \qquad \forall x \in \mathbb{X},\ a^i \in A^i(x),\ i = 1, 2.$$

Example: Robust Single-Server Service Control

Consider the queueing control model from Slide 8, where the arrival controller wants to maximize the total cost incurred before the queue becomes empty.

Interpretation: Don't know the arrival rate, want to control the service rate the minimize the worst-case total cost incurred before the queue becomes empty.

Using the arguments from Slide 9, this model satisfies Assumption (T).

Reduction to a (Standard) Discounted Zero-Sum Game

$\tilde{\beta} := (K-1)/K$

$\tilde{\mathbb{X}} := \mathbb{X} \cup \{\tilde{x}\}$, and $\tilde{\mathbb{A}}^i := \mathbb{A}^i \cup \{\tilde{a}^i\}$ for $i = 1, 2$

For $i = 1, 2$, $\tilde{A}^i(x) := A^i(x)$ if $x \in \mathbb{X}$ and $\tilde{A}^i(\tilde{x}) := \{\tilde{a}\}$.

$$\tilde{p}(y|x, a^1, a^2) := \begin{cases} \frac{1}{\tilde{\beta}\,\mu(x)}\mu(y)q(y|x, a^1, a^2), & x, y \in \mathbb{X}, (a^1, a^2) \in A^1(x) \times A^2(x), \\ 1 - \frac{1}{\tilde{\beta}\,\mu(x)}\sum_{y\in\mathbb{X}}\mu(y)q(y|x, a^1, a^2), & y = \tilde{x}, x \in \mathbb{X}, (a^1, a^2) \in A^1(x) \times A^2(x), \\ 1 & y = \tilde{x}, (x, a^1, a^2) = (\tilde{x}, \tilde{a}^1, \tilde{a}^2). \end{cases}$$

$$\tilde{c}(x, a^1, a^2) := \begin{cases} c(x, a^1, a^2)/\mu(x), & x \in \mathbb{X}, (a^1, a^2) \in A^1(x) \times A^2(x), \\ 0, & (x, a^1, a^2) = (\tilde{x}, \tilde{a}^1, \tilde{a}^2). \end{cases}$$

Use results for the discounted game (e.g., [Nowak]) to derive the existence of the value and optimal randomized stationary strategies for the original game [Feinberg & H].

Average-Cost Criterion

Φ^i = set of all randomized stationary policies for player $i = 1, 2$

For $x \in \mathbb{X}$ and $(\phi^1, \phi^2) \in \Phi^1 \times \Phi^2$, let

$$c_{\phi^1,\phi^2}(x) := \sum_{a^1 \in A^1(x)} \sum_{a^2 \in A^2(x)} \phi^1(a^1|x)\phi^2(a^2|x)c(x, a^1, a^2)$$

and

$$Q_{\phi^1,\phi^2}(x, y) := \sum_{a^1 \in A^1(x)} \sum_{a^2 \in A^2(x)} \phi^1(a^1|x)\phi^2(a^2|x)q(y|x, a^1, a^2)$$

Total costs:

$$w^{\phi^1,\phi^2}(x) := \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} Q^t_{\phi^1,\phi^2} c_{\phi^1,\phi^2}(x)$$

Average-Cost Criterion

$\phi_* \in \Phi^1$ is optimal for player 1 if

$$\inf_{\phi^2 \in \Phi^2} w^{\phi_*,\phi^2}(x) = \inf_{\phi^2 \in \Phi^2} \sup_{\phi^1 \in \Phi^1} w^{\phi^1,\phi^2}(x) \qquad \forall x \in \mathbb{X}.$$

$\phi_* \in \Phi^2$ is optimal for player 2 if

$$\sup_{\phi^1 \in \Phi^1} w^{\phi^1,\phi_*}(x) = \sup_{\phi^1 \in \Phi^1} \inf_{\phi^2 \in \Phi^2} w^{\phi^1,\phi^2}(x) \qquad \forall x \in \mathbb{X}.$$

The game has a value w if

$$w(x) := \inf_{\phi^2 \in \Phi^2} \sup_{\phi^1 \in \Phi^1} w^{\phi^1,\phi^2}(x) = \sup_{\phi^1 \in \Phi^1} \inf_{\phi^2 \in \Phi^2} w^{\phi^1,\phi^2}(x) \qquad \forall x \in \mathbb{X}.$$

Recurrence Assumption

Let

$$
{}_{\ell}Q_{\phi^1,\phi^2}(x,y) := \begin{cases} Q_{\phi^1,\phi^2}(x,y) & y \neq \ell \\ 0 & y = \ell \end{cases}
$$

Assumption (HT)

There is a "weight function" $V : \mathbb{X} \to [1,\infty)$ and a constant $K < \infty$ satsfying

$$
V(x)^{-1} \sum_{t=0}^{\infty} {}_{\ell}Q^t_{\phi^1,\phi^2} V(x) \leqslant K \qquad \forall x \in \mathbb{X},\ (\phi^1,\phi^2) \in \mathbb{F}^1 \times \mathbb{F}^2.
$$

When $\mathbb{X}$, $\mathbb{A}^1$, and $\mathbb{A}^2$ are finite, Assumption (HT) means

- state ℓ is recurrent under all $(\phi^1,\phi^2) \in \mathbb{F}^1 \times \mathbb{F}^2$.
- Hitting time to state ℓ is uniformly bounded in x and $(\phi^1,\phi^2) \in \mathbb{F}^1 \times \mathbb{F}^2$.
- w^{ϕ^1,ϕ^2} is constant for every $(\phi^1,\phi^2) \in \mathbb{F}^1 \times \mathbb{F}^2$.

Generalizes the assumption used by [Akian & Gaubert] to reduce the original average-cost game to a discounted one.

Example: Robust Single-Server Service Control

Consider the version of the queueing control model described on Slide 16, where the arrival controller wants to maximize the average cost incurred.

Interpretation: Don't know the arrival rate, want to control the service rate the minimize the worst-case average cost.

Using the arguments from Slide 16, this model satisfies Assumption (HT).

Reduction to a (Standard) Discounted Zero-Sum Game

$\bar\beta := (K-1)/K$

$\bar{\mathbb{X}} := \mathbb{X} \cup \{\bar x\}$, and $\bar{\mathbb{A}}^i := \mathbb{A}^i \cup \{\bar a^i\}$ for $i = 1, 2$

For $i = 1, 2$, $\bar A^i(x) := A^i(x)$ if $x \in \mathbb{X}$ and $\bar A^i(\bar x) := \{\bar a\}$.

$$\bar p(y|x, a^1, a^2) := \begin{cases} \frac{1}{\bar\beta\,\mu(x)}\mu(y)q(y|x, a^1, a^2), & y \neq \ell, x \in \mathbb{X}; \\ \frac{1}{\bar\beta\,\mu(x)}[\mu(x) - 1 - \sum_{y\neq\ell}\mu(y)q(y|x, a^1, a^2)] & y = \ell, x \in \mathbb{X}; \\ 1 - \frac{1}{\bar\beta\,\mu(x)}[\mu(x) - 1], & y = \bar x, x \in \mathbb{X}; \\ 1 & y = \bar x, (x, a^1, a^2) = (\bar x, \bar a^1, \bar a^2). \end{cases}$$

$$\bar c(x, a^1, a^2) := \begin{cases} c(x, a^1, a^2)/\mu(x), & x \in \mathbb{X}, (a^1, a^2) \in A^1(x) \times A^2(x), \\ 0, & (x, a^1, a^2) = (\bar x, \bar a^1, \bar a^2). \end{cases}$$

Use results for the discounted game (e.g., [Nowak]) to derive the existence of the value and optimal randomized stationary strategies for the original game [Feinberg & H].

Summary

1. Conditions under which undiscounted MDPs and stochastic games can be reduced to discounted ones.
   - Total Costs: Transience
   - Average Costs: Recurrence

2. Lead to validity of optimality equations, existence of optimal policies, complexity estimates for computing an optimal policy.

Future Work:

- Consequences for specific models? (e.g., queueing control, replacement & maintenance) [Feinberg & H]
- More general conditions under which a reduction holds?
  - Complexity estimates for average-cost problems