Infinite-State Discrete-time Markov Decision Processes

Eugene A. Feinberg

Stony Brook University

Eurandom, Eindhoven
September 5, 2012

Motivation

Early studies of inventory control problems by Arrow, Harris, and Marschak (1951), Dovertzky, Kiefer, and Wolfowitz (1952), Arrow, Karlin, and Scarf (1958), Scarf (1960) were one of the major motivating factors for the development of the theory of Markov Decision Processes (MDPs).

Inventory control models lead to MDPs with infinite state sets, noncompact action sets, and unbounded costs. We describe recent developments for such MDPs.

Plan of the talk

1. Finite state-action MDPs
2. Discounted MDPs with Borel state and compact action sets
3. Weak and setwise continuity assumptions for transition probabilities
4. Countable state MDPs with average costs
5. Average-costs MDPs with Borel state and compact action sets
6. MDPs with setwise continuous transition probabilities
7. MDPs with weakly continuous transition probabilities
8. Berge's theorem
9. Fatou's lemma

Finite state and action MDPs

(i) X is a state set;

(ii) A is an action set;

(iii) $A(x)$ is the set of actions available at state $x \in X$;

(iv) $c(x, a)$ is the one step cost;

(v) $p(y \mid x, a)$ is the transition probability;

At time $n = 0, 1, \ldots$, the history $h_n = x_0, a_0, \ldots x_{n-1}, a_{n-1}, x_n$ is observed and the action a_n is selected.

A policy is a sequence $\pi_0, \pi_1, \ldots$ of regular transition probabilities π_n from $H_n = (X \times A)^{n-1} \times X$ to A such that $\pi_n(A(x_n) \mid x_0, a_0, \ldots x_{n-1}, a_{n-1}, x_n) = 1$.

According to the Ionescu Tulcea theorem, any initial state x and any policy π define a probability measure $\mathbb{P}_x^\pi$ on $(X \times A)^\infty, \mathfrak{B}((X \times A)^\infty)$ and the expectation with respect to $\mathbb{P}_x^\pi$ is denoted by E_x^π.

A policy φ is called *Markov*, if decisions are non-randomized and depend only on the current state and time, $a_n = \varphi_n(x_n)$. A policy φ is called *stationary*, if decisions are non-randomized and depend only on the current state, $a_n = \varphi(x_n)$.

Let Π be the set of all policies.

Cost criterion

Finite-horizon expected total cost:

$$v^{\pi}_{\alpha,n}(x) = E^{\varphi}_{x} \sum_{t=1}^{n-1} \alpha^{t} c(x_t, a_t),$$

where $\alpha \geq 0$ is a discount factor.

Infinite-horizon expected total cost:

$$v^{\pi}_{\alpha}(x) = E^{\varphi}_{x} \sum_{t=1}^{\infty} \alpha^{t} c(x_t, a_t),$$

where $0 \leq \alpha < 1$.

Average costs per unit time:

$$w^{\pi}(x) = \limsup_{n\to\infty} \frac{1}{n} V^{\varphi}_{1,n}(x).$$

For any objective function $g^\pi(x)$, let $g(x) = \inf_{\pi \in \Pi} g^\pi(x)$ be the value function. A policy π is called *optimal*, if $g^\pi(x) = g(x)$ for all $x \in X$.

Value functions $v_{\alpha,n}, v_\alpha$ satisfy the optimality equations

$$v_{\alpha,n+1}(x) = \min_{a \in A(x)} \left\{ v_{\alpha,n}(x) + \sum_{y \in X} p(y \mid x, a) v_{\alpha,n}(y) \right\}, \ x \in X, \tag{1}$$

$$v_\alpha(x) = \min_{a \in A(x)} \left\{ v_\alpha(x) + \sum_{y \in X} p(y \mid x, a) v_\alpha(y) \right\}, \ x \in X. \tag{2}$$

A stationary policy φ is discount optimal if and only if, for all $x \in X$, the minimum in (2) is achieved for $a = \varphi(x)$.

Value function w for average costs per unit time satisfies the cannonical equations

$$
\begin{aligned}
w(x) &= \min_{a \in A(x)} \sum_{y \in X} p(y \mid x, a) w(y), & (3)\\
w(y) + u(y) &= \min_{a \in A(x)} \left\{ c(x, a) + \sum_{y \in X} p(y \mid x, a) u(x) \right\}. & (4)
\end{aligned}
$$

The canonical equations define a stationary optimal policy. For many problems $w(x) =$ const. Then (3) and (4) lead to

$$
w + u(y) = \min_{a \in A(x)} \left\{ c(x, a) + \sum_{y \in X} p(y \mid x, a) u(x) \right\}. \tag{5}
$$

Vanishing discount optimality approach

The existence of stationary discount optimal policy follows from the optimal equation.

The existence of a stationary average cost optimal policy follows from

$$w_\alpha^\varphi(x) = \lim_{n\to\infty}(1-\alpha)v_\alpha^\varphi(x), \qquad x \in X,$$

for any stationary policy φ. In general, from the Tauberian theorem, for any policy π,

$$\liminf_{n\to\infty}\frac{1}{n}v_{1,n}^\pi(x) \le \liminf_{\alpha\uparrow 1}(1-\alpha)v_\alpha^\pi(x) \le \limsup_{\alpha\uparrow 1}(1-\alpha)v_\alpha^\pi(x) \le w^\pi(x).$$

Infinite state space

(i) X and A are standard Borel spaces; $(X \times \mathfrak{B}(X))$, $(A \times \mathfrak{B}(A))$.

(ii) The graph $Gr(A) = \{x \in X, a \in A(x)\}$ is a Borel subset of $(X \times A)$ such that there exists a stationary policy φ (the measurable mapping $\varphi : X \to A$ such that $\varphi(x) \in A(x)$.

(iii) The cost function c is measurable and bounded above.

(iv) p is the regular transition probability from $X \times A$ to X;
   - $p(\cdot \mid x, a)$ is a probability measure on $(X \times \mathfrak{B}(X))$ for all $(x, a) \in (X \times A)$
   - $p(E \mid x, a)$ is a measurable function on $(X \times A)$ for each $E \in \mathfrak{B}(X)$.

Discounted costs

For $x \in X$ and $n = 0, 1, \ldots$, the optimality equations can be written as

$$v_{\alpha,n+1}(x) = \inf_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X v_{\alpha,n}(y) p(dy \mid x,a) \right\},$$

$$v_{\alpha}(x) = \inf_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X v_{\alpha}(y) p(dy \mid x,a) \right\}.$$

The relevant questions are:
(i) Can "inf" be replaced with min?
(ii) Is it true that $v_{\alpha,n}(x) \to v_{\alpha}(x)$?
(iii) Does an optimal policy exist?

Conditions for discount optimality for compact action sets (Schäl, 1975)

Let $\mathbb{P}(X)$ be the set of probability measures on X and $\mathbb{C}(A)$ be the set of nonempty compact subsets of A.

Assumption (S):

1. $A(x) \in \mathbb{C}(A)$ for all $x \in X$,
2. For $x \in X$, the function $c(x, a)$ is lower semi-continuous in a,
3. For each $x \in X$, the transition probability $p(\cdot \mid x, a)$ is continuous in a with respect to setwise convergence in $\mathbb{P}(x)$; that is if $a_n \to a$, then

$$\int_X f(y) p(dy \mid x, a_n) \to \int_X f(y) p(dy \mid x, a)$$

for any bounded measurable function f on X.

Conditions for discount optimality for compact action sets (Schäl, 1975)

Assumption (W):

1. $A(x) \in \mathbb{C}(A)$ for all $x \in X$,
2. c is lower semi-continuous on $Gr(A)$,
3. $A : X \to \mathbb{C}(A)$ is upper semi-continuous ($A^{-1}(B)$ is closed in X for every closed set $B \subset A$),
4. $p : Gr(A) \to \mathbb{P}(X)$ is continuous with respect to weak convergence in $\mathbb{P}(X)$, that is, if $x_n \to x$ and $a_n \to a$ then

   $$\int_X f(y)p(dy \mid x_n, a_n) \to \int_X f(y)p(dy \mid x, a)$$

   for any bounded continuous function f on X.

Why $A(x) : X \to A$ is assumed to be upper semi-continuous?

Answer: Berge's Theorem

Let X and Y be Hausdorff topological spaces. *Berge's Theorem:* If $u : X \times Y \to \bar{R}$ is a lower semi-continuous function and $\phi : X \to \mathbb{C}(Y)$ is upper semi-continuous set-valued mapping, then the function

$$v(x) = \inf_{y \in \phi(x)} u(x, y)$$

is lower semi-continuous on X.

$$v_{\alpha,n+1}(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X v_{\alpha,n}(y) p(dy \mid x, a) \right\} \qquad (6)$$

Both summands should be lower semi-continuous for the minimum to be achieved. Thus, if $v_{\alpha,n}$ is lower semi-continuous then the minimum is achieved. But, we also need lower semi-continuity of $v_{\alpha,n+1}$ for (6) with n increased by 1.

Conditions for average cost optimality

Sennott (1986-2002) introduced conditions for average-cost optimality when X is countable and A is finite.

Sennott's assumptions: There exists a "distinguished state" z such that:

(1) The function $(1-\alpha)v_\alpha(z)$ is bounded for $\alpha \in (0,1)$;

(2) There exists a non-negative finite constant L and a non-negative finite function $M(x)$ such that

$$-L \le v_\alpha(x) - v_\alpha(z) \le M(x),$$

for all $x \in X$ and for all $\alpha \in (0,1)$.

Sennott used the vanishing discount optimality approach and showed that the optimality inequality

$$w + u(x) \geq \min_{a \in A(x)} \left\{ c(x,a) + \int_X u(y) p(dy \mid x, a) \right\}, \ x \in X \quad (7)$$

holds. Then, a stationary policy φ such that

$$w + u(x) \geq c(x, \varphi(x)) + \int_X u(y) p(dy \mid x, \varphi(x)), \ x \in X$$

is optimal.

Cavozos-Cadena (1991) constructed an example when the optimality inequality holds but the optimality equality does not.

Conditions for average-cost optimality for Borel state and compact action sets (Schäl, 1993)

Assumptions (G) (general): $g := \inf_{x \in X} w(x) < \infty$.

Assumptions (B): Assumption (G) holds and

$$\sup_{\alpha \in [0,1)} u_\alpha(x) < \infty,$$

where, $u(x) = v_\alpha(x) - m_\alpha$ for
$$m_\alpha = \inf_{x \in X} v_\alpha(x)$$

[Assumption B] and [either Assumption S or Assumption W] implies the validity of the optimality inequality, existence of stationary optimal policies, and convergence of discount optimal policies to average-cost optimal policies.

Conditions for average-cost optimality for general action sets (Hernandez-Lerma, 1991)

Condition S*:

(i) For each $x \in X$, the function $c(x, a)$ is inf-compact in A, that is, the set $A_r(x) = \{a \in A(x) \mid c(x, a) \leq r\}$ is compact for each finit constant r.

(ii) For each $x \in X$, the transition probability $p(\cdot \mid x, a)$ is continuous with respect to setwise convergence on $\mathbb{P}(X)$.

Assumptions (B) and (W) imply the validity of the optimality inequality, existence of stationary optimal policies, and convergence of discount optimal policies to average-cost optimal policies.

Conditions for discounted optimality for general action sets (Feinberg and Lewis, 2007)

Assumption (Wu):

1. c is inf-compact on $Gr(A)$; that is the set

$$\{(x,a) \in Gr(A) : c(x,a) \leq r\}$$

   is compact for any finite constant r.
2. $p : Gr(A) \to \mathbb{P}(X)$ is continuous with respect to weak convergence on $\mathbb{P}(X)$.

Under assumption (Wu), there exist stationary discount optimal policies that satisfy optimality equations and $v_{n,\alpha} \to v_\alpha$.

In fact, F. and Lewis introduced a new version of Berge's theorem.

Conditions for average-cost optimality (Feinberg and Lewis, 2007)

Let $r_{\alpha_0}(x) := \sup_{\alpha_0 \le \alpha < 1} u_\alpha(x)$, where $\alpha \in [0,1)$ and

$$u_\alpha(x) = v_\alpha(x) - m_\alpha.$$

Assumption (LB): Assumption (B) holds and there exists $\alpha_0 \in [0,1)$ such that the function $r_{\alpha_0}(x)$ is locally bounded on X, that is, for any $x \in X$ there is a neighbourhood of x on which $r_{\alpha_0}(x)$ is bounded.

Assumptions (LB) and (Wu) imply the validity of the optimality inequality, existence of stationary optimal policies, and convergence of discount-optimal policies to average cost optimal policies.

Applications: Inventory control, cash management problem.

Average-cost optimality (Feinberg, Kasyanov, Zadoianchuk 2012)

Assumption (W*):

1. c is lower semi-continuous on $Gr(A)$
2. If a sequence $\{x_n\}_{n=1,2,\dots}$ with values in X converges and its limit x belongs to X, then any sequence $\{a_n\}_{n=1,2,\dots}$ with $a_n \in A(x_n), n = 1, 2, \dots,$ satisfying the condition that the sequence $\{c(x_n, a_n)\}_{n=1,2,\dots}$ is bounded above has a limit point $a \in A(x)$
3. $p : Gr(A) \to \mathbb{P}(X)$ is continuous with respect to weak convergence on $\mathbb{P}(X)$.

<u>Lemma:</u>

(i) Assumption (W) implies Assumption (W*),

(ii) Assumption (W$_u^*$) implies Assumption (W*).

Average-cost optimality

Let $\bar{w} = \limsup_{\alpha \uparrow 1} (1-\alpha) m_\alpha, \quad$ where $m_\alpha = \inf_{x \in X} v_\alpha(x)$.

Recall that,
Assumption (G): $g := \inf_{x \in X} w(x) < \infty$.

Optimality inequality: Let Assumption (G) hold. If there exists a measurable function $u : X \to [0, \infty)$ and a stationary policy ϕ such that

$$\bar{w} + u(x) \geq c(x, \phi(x)) + \int_X u(y) p(dy \mid x, \phi(x))$$

then ϕ is average-cost optimal and

$$w(x) = w^\phi(x) = \limsup_{\alpha \uparrow 1} v_\alpha(x) = \bar{w} = g.$$

Average-cost optimality

Assumption ($\underline{\text{B}}$): Assumption (G) holds and $\liminf_{\alpha\uparrow 1} u_\alpha(x) < \infty$ for all $x \in X$.

Assumption ($\underline{\text{B}}$) is weaker than assumption (B) that requires $\limsup_{\alpha\uparrow 1} u_\alpha(x) < \infty$.

Define,

$$U_\beta(x) = \inf_{\alpha\in[\beta,1)} u_\alpha(x), \quad \underline{u}_\beta(x) = \liminf_{y\to\infty} U_\beta(y),\ \beta \in [0,1), x \in X,$$

and

$$u(x) = \lim_{\beta\uparrow 1} \underline{u}_\beta(x).$$

Average-cost optimality

Theorem (1)

Suppose assumptions (W^*) and ($\underline{B}$) hold. There exists a stationary policy satisfying the optimality inequality and therefore, this stationary policy is optimal and

$$w(x) = w^{\Phi}(x) = \limsup_{\alpha \uparrow 1} v_\alpha(x) = \limsup_{\alpha \uparrow 1} m_\alpha = g.$$

Define, for all $x \in X$,

$$A^*(x) := \left\{ a \in A(x) : \bar{w} + u(x) \geq c(x,a) + \int_X u(y) p(dy \mid x, a) \right\}.$$

Any stationary policy φ such that $\Phi(x) \in A^*(x)$ for all $x \in X$ is optimal.

Theorem (2)

Suppose assumption (W^*) and (B) hold. Then, for the optimal policy φ satisfying the optimality inequality, the following additional properties hold

$$w^{\varphi}(x) = \liminf_{\alpha \uparrow 1} m_\alpha = \lim_{\alpha \uparrow 1} (1-\alpha) v_\alpha(x) = \lim_{n \to \infty} \frac{1}{n} v^{\varphi}_{1,n}(x).$$

There are many open questions: Theorem (2) implies that the any policy φ and for any policy π

$$w^{\varphi}(x) = \lim_{n \to \infty} \frac{1}{n} v^{\varphi}_{1,n}(x) \geq \limsup_{n \to \infty} \frac{1}{n} v^{\pi}_{1,n}(x).$$

But we do not know whether

$$w^{\varphi}(x) \geq \liminf_{n \to \infty} \frac{1}{n} v^{\pi}_{1,n}(x).$$

Discount-optimality

Theorem (Feinberg, Kasyanov, Zadoianchuk)

Let assumption (W) hold. Then:*

(i) *The function* $v_{\alpha,n}, n = 0, 1, \ldots$ *and* v_α *are lower semi-continuous and* $v_{\alpha,n}(x) \to v_\alpha(x)$ *as* $n \to \infty$ *for all* $x \in X$.

(ii) $v_{\alpha,n+1}(x) = \min\limits_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X v_{\alpha,n}(y) p(dy \mid x,a) \right\},\ x \in X, n = 0, 1, \ldots$

(iii) $v_\alpha(x) = \min\limits_{a \in A(x)} \left\{ c(x,a) + \alpha \int_X v_\alpha(y) p(dy \mid x,a) \right\}$

(iv) *Let*

$$A_\alpha(x) = \{a \in A(x) : v_\alpha(x) = c(x,a) + \alpha \int_X v_\alpha(y) p(dy \mid x,a)\}.$$

A stationary optimal policy is discount optimal, if and only if, $\varphi(x) \in A_\alpha(x)$ *for all* $x \in X$.

Approximations of average cost optimal policies

Consider the upper topological limit of the sets $A_\alpha(x)$ as $\alpha \uparrow 1$.

$$\overline{\lim_{\alpha\uparrow 1}} Gr(A_\alpha) = \left\{ (x,a) \in X \times A : \begin{array}{l} \exists\ \alpha_n \uparrow 1 \text{ as } n \to \infty, \\ \exists\ (x_n, a_n) \in Gr(A_{\alpha_n}), n \geq 1 \\ \text{such that } (x,a) = \lim_{n\to\infty} (x_n, a_n) \end{array} \right\}$$

Let $A^{app}(x) = \{a \in A^*(x) : (x,a) \in \overline{\lim_{\alpha\uparrow 1}} Gr(A_\alpha), x \in X\}$.

Theorem

Under assumptions (W) and (B̲), the set $A^{app}(x)$ are nonempty and compact. Furthermore, there exists a stationary policy φ^{aap} such that $\varphi^{aap}(x) \in A^{app}(x)$ and any such policy is average-cost optimal.*

Extension of Berge's theorem to non-compact action sets

Why

$$v_{\alpha,n+1} = \min_{a \in A(x)} \left\{ c(x,a) + \int_X v_{\alpha,n}(y) p(dy \mid x,a) \right\}$$

is upper semi-continuous? This does not follow from Berge's theorem that states the continuity of the function

$$v(x) := \inf_{y \in \phi(x)} u(x,y)$$

for compact action sets. But, we do not assume that $A(x)$ is compact.

Berge's theorem

Definition

A function $u : X \times Y \to \bar{R}$ is called K-inf-compact on $Gr_X(\phi)$, if for every $K \in \mathbb{C}(X)$ this function is inf-compact on $Gr_K(\phi)$

Theorem (Generalization of Berge's theorem, F., Kasyanov, Zadoianchuk, 2012)

If the function $u : X \times Y \to R$ is K-inf-compact on $Gr(\phi)$, then the function $v(x) = \inf_{y \in \phi(x)} u(x, y)$ is inf-compact on $Gr_K(\phi)$.

Lemma

(i) *Under the assumptions of Berge's theorem, the function $u(x, y)$ is K-inf-compact*

(ii) *If $u(x, y)$ is inf-compact on $Gr_X(\phi)$ then it is K-inf-compact.*

Berge's theorem

Lemma

(i) *If* $u(\cdot,\cdot)$ *is a* K*-inf-compact function on* $Gr_X(\phi)$*, then, for every* $x \in X$*, the function* $u(x,a)$ *is inf-compact on* $\phi(x)$

(ii) *A* K*-inf-compact function* $u(\cdot,\cdot)$ *on* $Gr_X(\phi)$ *is lower semi-continuous on* $Gr_X(\phi)$*.*

Properties (i) and (ii) typically hold for inventory control problems

Lemma

Let X *and* Y *are metrizable space. Then* $Gr_X(\phi)$ *is* K*-inf-compact on* $Gr_X(\phi)$ *if and only if conditions (W* 1,2) hold.*

Fatou's lemma for weakly converging measures

Fatou's Lemma: For any $\mu \in \mathbb{P}(X)$ and for any sequence of nonnegative measurable functions $f_1, f_2, \ldots$

$$\int\limits_X \liminf_{n\to\infty} f_n(x)\mu(dx) \le \liminf_{n\to\infty} \int\limits_X f_n(x)\mu(dx).$$

(Royden, 1968) If $\{\mu_n\} \in \mathbb{P}(X)$ setwise converges to $\mu \in \mathbb{P}(X)$ then,

$$\int\limits_X \liminf_{n\to\infty} f_n(x)\mu(dx) \le \liminf_{n\to\infty} \int\limits_X f_n(x)\mu_n(dx).$$

(Schäl, 1993, Feinberg, Kasyanov, Zadoianchuk, 2012) If μ_n converge weakly to μ then

$$\int_X \underline{f}(x)\mu(dx) \leq \liminf_{n\to\infty} \int_X f_n(x)\mu_n(dx),$$

where $\underline{f}(x) = \liminf_{n\to\infty} \inf_{s'\to s} f_n(x), \quad x \in X$.

Conclusions

This talk covered:

(i) General conditions of optimality for discounted and average-reward MDPs with Borel state and action sets, unbounded costs, and weakly continuous transition probabilities. The results are applicable to inventory control problems, workload control in queues, ...

(ii) Extension of Berge's theorem to noncompact action sets

(iii) Proof and extensions of Fatou's theorem for weakly converging measures formulated by Schäl (1993); relevant results were established by Serfozo (1982).