The Value Functions of Markov Decision Problems

Eilon Solan, Tel Aviv University

with

Ehud Lehrer

Omri N. Solan

Tel Aviv University

Markov Decision Problems

- **S = a finite set of states.**
- **μ_0 in $\Delta(S)$ = initial probability distribution.**
- **$A(s)$ = a finite set of actions available at state s.**

$SA := \{ (s,a) : s \text{ in } S, a \text{ in } A(s) \}$.

- **$r : SA \to \Re$ = payoff function.**
- **$q : SA \to \Delta(S)$ = transitions.**

Initial state s_0 is chosen according to μ_0.
At every stage $n=0,1,2,\ldots$ the DM chooses an action a_n in $A(s_n)$, receives payoff $r(s_n,a_n)$, and state s_{n+1} is chosen according to $q(s_n,a_n)$.

Markov Decision Problems

A **(pure) strategy** σ is a function that assigns an action in $A(s_n)$ to every finite history $h=(s_1,a_1,\ldots,s_{n-1},a_{n-1},s_n)$.

A **behavior strategy** assigns a mixed action in $\Delta(A(s_n))$ to every such finite history.

A strategy is **stationary** if $\sigma(h)$ depends only on the current state s_n, and not on past play.

For every strategy σ and every discount factor λ in $[0,1)$, the **λ-discounted payoff** is:

$$\gamma_\lambda(\mu_0,\sigma) := E_{\mu_0,\sigma}\left[\ \sum_{n=0}^{\infty} \lambda^n r(s_n,a_n)\ \right]$$

Markov Decision Problems

The **λ-discounted value**: $v_\lambda(\mu_0) := \max_\sigma \gamma_\lambda(\mu_0,\sigma)$

A strategy that attains the maximum is **λ-discounted optimal** at μ_0.

Theorem (Blackwell, 1962): The λ-discounted value exists. Moreover, there is a λ-discounted optimal pure stationary strategy.

The **value function**: $\lambda \to v_\lambda(\mu_0)$.

Question: What is the set of all possible value functions?

Stationary Strategies

For every pure stationary strategy σ and every discount factor λ, $(\gamma(s,\sigma))_{s \text{ in } S}$ is the solution of a set of linear equations in λ.

$$\gamma_\lambda(s,\sigma) = r(s,\sigma(s)) + \lambda\, \Sigma_{\{s' \text{ in } S\}}\, q(s' | s, \sigma(s))\, \gamma_\lambda(s',\sigma)$$

$$\gamma_\lambda(\cdot,\sigma) = (I - \lambda\, q(\cdot \mid \cdot, \sigma(\cdot))^{-1}\, r(\cdot,\sigma(\cdot))$$

Corollary: $\gamma_\lambda(s,\sigma) = P(\lambda)/Q(\lambda)$ **is a rational function of λ. If a root λ of Q satisfies $|\lambda|=1$, then it is a unit root.**

Observation: **The roots of Q are not in the interior of the unit ball in the complex plane, and if they are on the boundary of the unit ball, they have multiplicity 1.**

Main Result

$\mathcal{V}$ = all functions that are the value of some MDP.
$\mathcal{V}_D$ = all functions that are the value of degenrate MDP's (the DM has one action in each state).
$\mathcal{F}$ = all rational functions $P(\lambda)/Q(\lambda)$ in which the roots of the denominator are either (a) outside the unit ball in the complex plane, or (b) unit roots with multiplicity 1.

Theorem: $\mathcal{F} = \mathcal{V}_D$. Consequently, a function f is in $\mathcal{V}$ if and only if it is the maximum of finitely many functions in $\mathcal{F}$.

Proof of "consequently": $\mathcal{V} = \max \mathcal{V}_D = \max \mathcal{F}$

Proof

Lemma: If f,g are in $\mathcal{V}_D$ then:

a) $af(\lambda)$ is in $\mathcal{V}_D$ for every real number a.

b) $\lambda f(\lambda)$ is in $\mathcal{V}_D$.

c) $f+g$ is in $\mathcal{V}_D$.

It remains to show that for any polynomial Q such that $1/Q$ is in $\mathcal{F}$, we have that $1/Q$ is in $\mathcal{V}_D$.

Corollary: If Q' divides Q, and $1/Q$ is in $\mathcal{V}_D$, so is $1/Q'$.

Because $1/Q' = (Q/Q')/Q$.

Proof

Lemma: If f,g are in $\mathcal{V}_D$ then:

a) $af(\lambda)$ is in $\mathcal{V}_D$ for every real number a.

b) $\lambda f(\lambda)$ is in $\mathcal{V}_D$.

c) $f+g$ is in $\mathcal{V}_D$.

It remains to show that for any polynomial Q such that $1/Q$ is in $\mathcal{F}$, we have that $1/Q$ is in $\mathcal{V}_D$.

Corollary: If Q' divides Q, and $1/Q$ is in $\mathcal{F}$, so is $1/Q'$.

Lemma: If f is in $\mathcal{V}_D$ then

a) $f(\lambda^n)$ is in $\mathcal{V}_D$ for every natural number n.

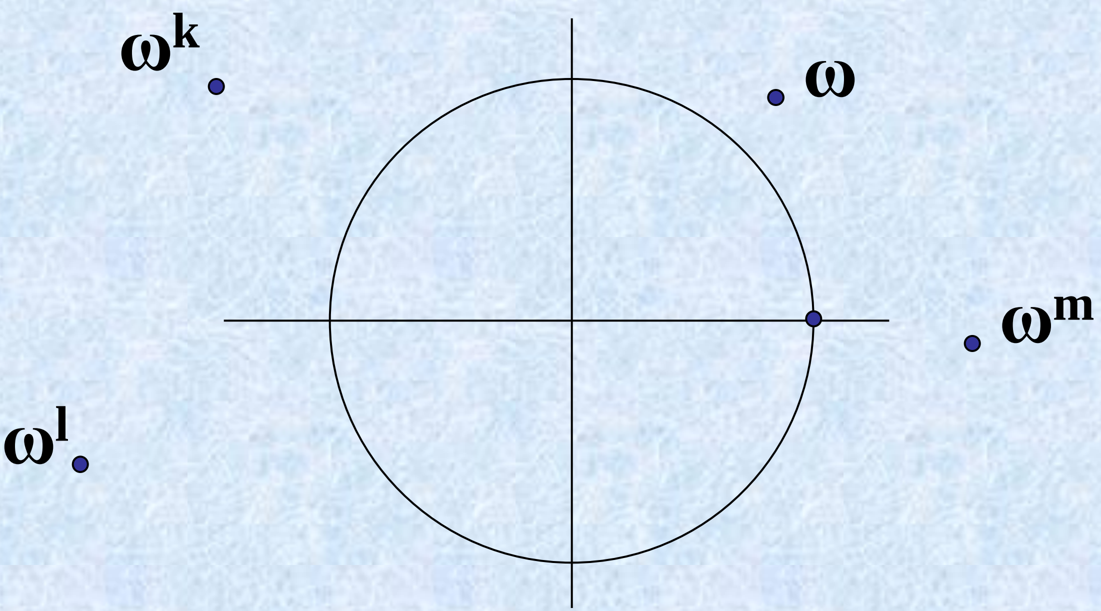
b) $1/(1- \lambda)$ is in $\mathcal{V}_D$.

Corollary: If all roots of Q are unit roots with multiplicity 1, then $1/Q$ is in $\mathcal{V}_D$.

Proof - Continued

Observation: **For every complex number ω not in the unit ball there are natural numbers $k<l<m$ and nonnegative reals $\alpha_1, \alpha_2, \alpha_3$ that sum to 1 such that**

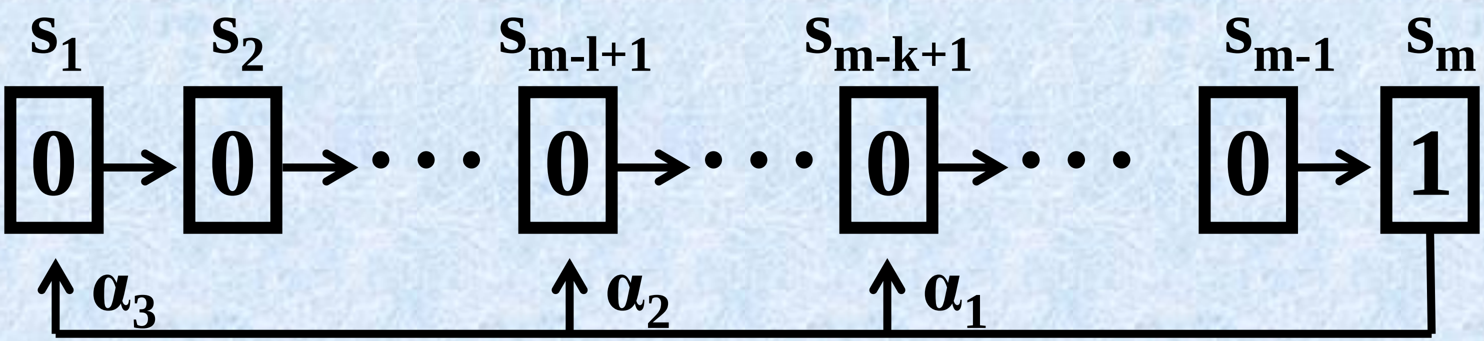
$$1 = \alpha_1 \omega^k + \alpha_2 \omega^l + \alpha_3 \omega^m.$$



Proof - Continued

Observation: For every complex number ω not in the unit ball there are natural numbers $k<l<m$ and nonnegative reals $\alpha_1, \alpha_2, \alpha_3$ that sum to 1 such that $1 = \alpha_1 \omega^k + \alpha_2 \omega^l + \alpha_3 \omega^m$.

Observation: The value of the following degenerate MDP is $1/(1 - \alpha_1 \lambda^k - \alpha_2 \lambda^l - \alpha_3 \lambda^m)$.



Corollary: For every complex number ω not in the unit ball, $1/((1- \omega)(1- \overline{\omega}))$ is in $\mathcal{V}_D$.

Proof - Continued

Lemma: If f,g are in $\mathcal{V}_D$ then $f(\lambda)g(c\lambda)$ is in $\mathcal{V}_D$, for every $0 \leq c < 1$.

Let ω be a complex number not in the unit ball, $1/|\omega| < c < 1$.

Then $\dfrac{1}{(c\omega - \lambda)(c\bar{\omega} - \lambda)}$ in $\mathcal{V}_D$.

Therefore

$$\frac{f(\lambda)}{(c\omega - c\lambda)(c\bar{\omega} - c\lambda)} = \frac{(1/c)^2 f(\lambda)}{(\omega - \lambda)(\bar{\omega} - \lambda)} \quad \text{in } \mathcal{V}_D.$$

Merci

dziękuję

شكرا

謝謝

Dank

Hvala

धन्यवाद

고맙습니다

Thank you

תודה רבה

Дякую

ありがとう

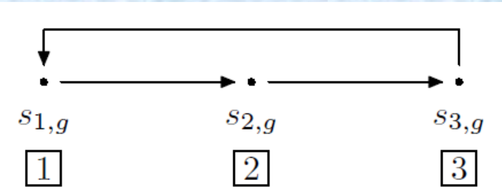
Köszönöm

ขอขอบคุณ

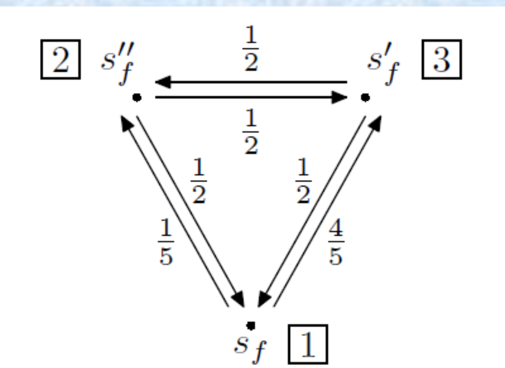
Danke

Proof of the Last Lemma

Lemma: **If f,g are in $\mathcal{V}_D$ then $f(\lambda)g(c\lambda)$ is in $\mathcal{V}_D$, for every $0 \leq c < 1$.**



MDP M_g



MDP M_f

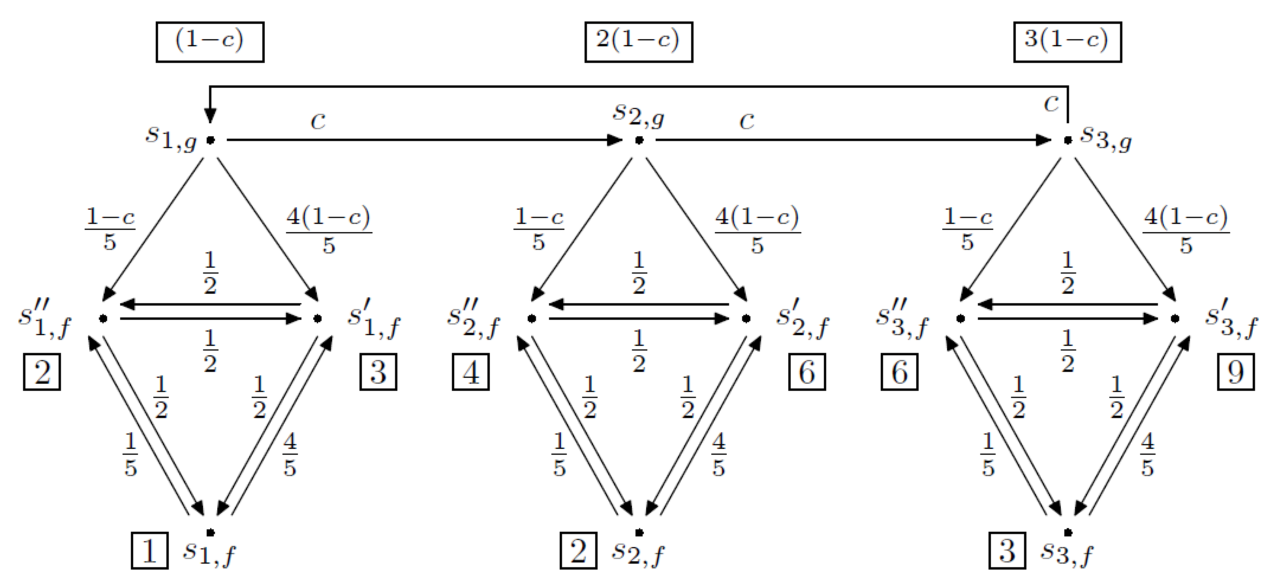


Figure 2: The degenerate MDP M.