Average-cost Markov decision processes with weakly continuous transition probabilities

Eugene A. Feinberg[1] Pavlo O. Kasyanov[2] Nina V. Zadoianchuk[2]

[1]Stony Brook University, Stony Brook, NY 11794-3600, USA

[2]Institute for Applied System Analysis, Kyiv Polytechnic Institute, Kyiv, Ukraine

This talk presents general sufficient conditions for the existence of stationary optimal policies for discounted and average-reward Markov Decision Process with Borel state sets and with weakly continuous transition probabilities. The results for average costs per unit time extend Scäl's [10] sufficient conditions for the existence of stationary optimal policies to problems with noncompact actions sets. For setwise continuous transition probabilities, similar results were established in Scäl [10] for compact action sets and extended in Hernández-Lerma [5] to general action sets. This talk is based on Feinber, Kasyanov and Zadoianchuk [2].

Consider a *discrete-time MDP* $(\mathbb{X}, \mathbb{Y}, \Phi, q, u)$ with a *state space* $\mathbb{X}$, an *action space* $\mathbb{Y}$, one-step costs u, and transition pobabilities q. The terminology is the same as in [2, 4] and references therein. Assume that $\mathbb{X}$ and $\mathbb{Y}$ are *Borel subsets* of Polish (complete separable metric) spaces. For a topological space $\mathbb{U}$ we denote by $\mathcal{B}(\mathbb{U})$ its Borel σ-field. For all $x \in \mathbb{X}$ a nonempty Borel subset $\Phi(x)$ of $\mathbb{Y}$ represents the *set of actions* available at x. Assume also that $\mathrm{Gr}_{\mathbb{X}}(\Phi) = \{(x, y) : x \in \mathbb{X}, y \in \Phi(x)\}$ is a measurable subset of $\mathbb{X} \times \mathbb{Y}$, that is, $\mathrm{Gr}_{\mathbb{X}}(\Phi) \in \mathcal{B}(\mathbb{X} \times \mathbb{Y})$, where $\mathcal{B}(\mathbb{X} \times \mathbb{Y}) = \mathcal{B}(\mathbb{X}) \otimes \mathcal{B}(\mathbb{Y})$; and there exists a measurable mapping $\phi : \mathbb{X} \to \mathbb{Y}$ such that $\phi(x) \in \Phi(x)$ for all $x \in \mathbb{X}$. The *one step cost*, $u(x, y) \leq +\infty$, for choosing an action $y \in \Phi(x)$ in a state $x \in \mathbb{X}$, is a *bounded below measurable* function on $\mathrm{Gr}_{\mathbb{X}}(\Phi)$. Let $q(B|x, y)$ be the *transition kernel* representing the probability that the next state is in $B \in \mathcal{B}(\mathbb{X})$, given that the action y is chosen in the state x. This means that $q(\cdot|x, y)$ is a probability measure on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ for all $(x, y) \in \mathrm{Gr}_{\mathbb{X}}(\Phi)$; and $q(B|\cdot,\cdot)$ is a Borel function on $(\mathrm{Gr}_{\mathbb{X}}(\Phi), \mathcal{B}(\mathrm{Gr}_{\mathbb{X}}(\Phi)))$ for all $B \in \mathcal{B}(\mathbb{X})$.

Let $\mathrm{Gr}_Z(\Phi) = \{(x, y) \in Z \times \mathbb{Y} : y \in \Phi(x)\}$, where $Z \subseteq \mathbb{X}$. For a topological space $\mathbb{U}$, we denote by $\mathbb{K}(\mathbb{U})$ *the family of all nonempty compact subsets of* $\mathbb{U}$.

For an $\overline{\mathbb{R}}$-valued function f, defined on a nonempty subset U of a topological space $\mathbb{U}$, consider the level sets $\mathcal{D}_f(\lambda; U) = \{y \in U : f(y) \leq \lambda\}$, $\lambda \in \mathbb{R}$. We recall that a function f is *lower semi-continuous (l.s.c.) on* U if all the level sets $\mathcal{D}_f(\lambda; U)$ are closed, and a function f is *inf-compact on* U (lower semi-compact cf. [12]) if all these sets are compact.

Definition 1. *A function* $u : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}}$ *is called* $\mathbb{K}$*-inf-compact on* $\mathrm{Gr}_{\mathbb{X}}(\Phi)$*, if for every* $K \in \mathbb{K}(\mathbb{X})$ *this function is inf-compact on* $\mathrm{Gr}_K(\Phi)$.

We set $\Phi^{\#}(x) = \{y \in \Phi(x) : v(x) = u(x, y)\}$. The first statement of the following theorem extends the well-known Berge's theorem of the minimum [1, Theorem 2, p. 116] or [7, Proposition 3.3, p. 83] to noncompact image (or decision) sets. The proofs and additional details can be found in [3].

Speaker: Eugene A. Feinberg, e-mail: eugene.feinberg@sunysb.edu

Theorem 1. *If the function* $u : \mathbb{X} \times \mathbb{Y} \to \overline{\mathbb{R}}$ *is* $\mathbb{K}$*-inf-compact on* $\mathrm{Gr}_{\mathbb{X}}(\Phi)$*, then the function* $v : \mathbb{X} \to \overline{\mathbb{R}}$ *is l.s.c. If moreover* u *is a continuous function on* $\mathrm{Gr}_{\mathbb{X}}(\Phi)$ *and* $\Phi : \mathbb{X} \to 2^{\mathbb{Y}} \setminus \emptyset$ *is l.s.c., then the function* v *is continuous on* $\mathbb{X}$ *and the solution multifunction* $\Phi^{\#} : \mathbb{X} \to \mathbb{K}(\mathbb{Y})$ *has a closed graph. Additionally, if* Φ *is upper semi-continuous (u.s.c.), then* $\Phi^{\#}$ *is u.s.c.*

The following lemma provides a useful criterium for $\mathbb{K}$-inf-compactness of u on $\mathrm{Gr}_{\mathbb{X}}(\Phi)$, when the spaces $\mathbb{X}$ and $\mathbb{Y}$ are metrizable. In this form the $\mathbb{K}$-inf-compactness assumption is introduced in Feinberg, Kasyanov and Zadoianchuk [2] as Assumption $(\mathbf{W}^*)$(ii).

Lemma 1. *Let* $\mathbb{X}$ *and* $\mathbb{Y}$ *be metrizable spaces. Then* u *is* $\mathbb{K}$*-inf-compact on* $\mathrm{Gr}_{\mathbb{X}}(\Phi)$ *if and only if the following two conditions hold: (i)* u *is l.s.c. on* $\mathrm{Gr}_{\mathbb{X}}(\Phi)$*; (ii) if a sequence* $\{x_n\}_{n\ge1}$ *with values in* $\mathbb{X}$ *converges and its limit* x *belongs to* $\mathbb{X}$ *then any sequence* $\{y_n\}_{n\ge1}$ *with* $y_n \in \Phi(x_n)$*,* $n \ge 1$*, satisfying the condition that the sequence* $\{u(x_n, y_n)\}_{n\ge1}$ *is bounded above, has a limit point* $y \in \Phi(x)$*.*

We also suppose the following assumption that implies the existence of stationary optimal policies for discounted MDPs.

Assumption $(\mathbf{W}^*)$**.** *(i)* u *is bounded below and* $\mathbb{K}$*-inf-compact on* $\mathrm{Gr}_{\mathbb{X}}(\Phi)$*; (ii) the transition probability* $q(\cdot|x, y)$ *is weakly continuous in* $(x, y) \in \mathrm{Gr}_{\mathbb{X}}(\Phi)$*.*

Weak continuity of q in (x, y) means that $\int_{\mathbb{X}} f(z)q(dz|x_k, y_k) \to \int_{\mathbb{X}} f(z)q(dz|x, y)$, $k \to +\infty$, for any sequence $\{(x_k, y_k), k \ge 1\}$ converging to (x, y), where (x_k, y_k), $(x, y) \in \mathrm{Gr}_{\mathbb{X}}(\Phi)$, and for any bounded continuous function $f : \mathbb{X} \to \mathbb{R}$.

Denote the class of all l.s.c. and bounded below functions $\varphi : \mathbb{X} \to \overline{\mathbb{R}}$ with $\mathrm{dom}\,\varphi := \{x \in \mathbb{X} : \varphi(x) < +\infty\} \ne \emptyset$ by $L(\mathbb{X})$. Let $\mathbb{F}$ be a family of Borel mappings $\phi : \mathbb{X} \to \mathbb{Y}$ such that $\phi(x) \in \Phi(x)$ for all $x \in \mathbb{X}$.

An important consequence of Assumption $(\mathbf{W}^*)$ is that it implies that $\mathbb{F}$ contains suitable "minimizers". The following lemma is useful for establishing continuity properties of the value functions; for later relevant results see Feinberg et al. [2]. The proof of this lemma follows from Theorem 1 and from the Arsenin-Kunugui theorem (Kechris [8, p. 297]).

Lemma 2. *If Assumption* $(\mathbf{W}^*)$ *holds and* $\underline{u} \in L(\mathbb{X})$*, then the function* $(x, y) \to u(x, y) + \int_{\mathbb{X}} \underline{u}(z)q(dz|x, y)$ *is* $\mathbb{K}$*-inf-compact on* $\mathrm{Gr}_{\mathbb{X}}(\Phi)$ *and the nonempty sets*

$$\Phi_*(x) = \left\{ y \in \Phi(x) : \underline{u}^*(x) = u(x, y) + \int_{\mathbb{X}} \underline{u}(z)q(dz|x, y) \right\}, \qquad x \in \mathbb{X}, \quad (1)$$

satisfy the following properties: (a) $\mathrm{Gr}_{\mathbb{X}}(\Phi_*)$ *is a Borel subset of* $\mathbb{X} \times \mathbb{Y}$*; (b) if* $\underline{u}^*(x) = +\infty$*, then* $\Phi_*(x) = \Phi(x)$*, and, if* $\underline{u}^*(x) < +\infty$*, then* $\Phi_*(x)$ *is compact.*

As usual a *policy* is a sequence $\pi = \{\pi_n\}_{n=0,1,\dots}$ of decision rules (cf. [2, 4] and references therein), where for each $n = 0, 1, \dots$ $\pi_n(\cdot|h_n)$ is a conditional probability on $(\mathbb{Y}; \mathcal{B}(\mathbb{Y}))$, given the history $h_n = (x_0, y_0, x_1, y_1, \dots, y_{n-1}, x_n)$, satisfying $\pi_n(\Phi(x_n)|h_n) = 1$. The class of *all policies* is denoted by Π. Moreover, π is called *nonrandomized*, if each probability measure $\pi_n(\cdot|h_n)$ is concentrated at one point. A nonrandomized policy is called *Markov*, if all of the decisions depend on the current state and time only. A Markov policy is called *stationary*, if all the decisions depend on the current state only. Thus, a Markov policy π is defined by a sequence

$\phi_0, \phi_1, \dots$ of Borel mappings $\phi_n \in \mathbb{F}$. A stationary policy π is defined by a Borel mapping $\phi \in \mathbb{F}$.

For a policy π, given initial state $x_0 = x \in \mathbb{X}$, for a finite horizon $N \geq 0$ let us define the *expected total discounted costs* $v^\pi_{N,\alpha} := \mathbb{E}^\pi_x \sum_{n=0}^{N-1} \alpha^n u(x_n, y_n)$, $x \in \mathbb{X}$, where $\alpha \geq 0$ is the discount factor and $v^\pi_{0,\alpha}(x) = 0$. When $N = \infty$ and $\alpha \in [0,1)$, $v^\pi_{N,\alpha}$ defines an *infinite horizon expected total discounted cost* denoted by $v^\pi_\alpha(x)$. The *average cost per unit time* is defined as $w^\pi(x) := \limsup_{N\to+\infty} \frac{1}{N} v^\pi_{N,1}(x)$, $x \in \mathbb{X}$. For any function $\triangle^\pi(x)$, including $\triangle^\pi(x) = v^\pi_{N,\alpha}(x)$, $\triangle^\pi(x) = v^\pi_\alpha(x)$, and $\triangle^\pi(x) = w^\pi(x)$, define the *optimal cost* $\triangle(x) := \inf_{\pi\in\Pi} \triangle^\pi(x)$, $x \in \mathbb{X}$. A policy π is called *optimal* for the respective criterion, if $\triangle^\pi(x) = \triangle(x)$ for all $x \in \mathbb{X}$. For $\triangle^\pi = v^\pi_{n,\alpha}$, the optimal policy is called *n-horizon discount-optimal*; for $\triangle^\pi = v^\pi_\alpha$, it is called *discount-optimal*; for $\triangle^\pi = w^\pi$, it is called *average-cost optimal* [2, 4, 5, 6, 10]. These definitions of optimality are standard.

Assumption ($\underline{\mathbf{B}}$). (a) $w^* := \inf_{x\in\mathbb{X}} w(x) < \infty$, (b) $\liminf_{\alpha\uparrow 1} u_\alpha(x) < \infty$ $\forall x \in \mathbb{X}$.

Assumption ($\underline{\mathbf{B}}$)(a) is equivalent to the existence of $x \in \mathbb{X}$ and $\pi \in \Pi$ with $w^\pi(x) < \infty$. If Assumption ($\underline{\mathbf{B}}$)(a) does not hold then the problem is trivial, because $w(x) = \infty$ for all $x \in \mathbb{X}$ and any policy π is average-cost optimal.

To state the main result we also need the following notation [10]: for $\alpha \in [0,1)$: $m_\alpha = \inf_{x\in\mathbb{X}} v_\alpha(x)$, $u_\alpha(x) = v_\alpha(x) - m_\alpha$, $\underline{w} = \liminf_{\alpha\uparrow 1} (1-\alpha) m_\alpha$, $\overline{w} = \limsup_{\alpha\uparrow 1} (1-\alpha) m_\alpha$. Observe that $u_\alpha(x) \geq 0$ for all $x \in \mathbb{X}$. Schäl [10, Lemma 1.2] and Assumption ($\underline{\mathbf{B}}$)(a) implies $0 \leq \underline{w} \leq \overline{w} \leq w^* < +\infty$. According to Schäl [10, Proposition 1.3], under Assumption ($\underline{\mathbf{B}}$)(a), if there exists a measurable function $g : \mathbb{X} \to [0,+\infty)$ and a stationary policy ϕ such that $\underline{w} + g(x) \geq u(x,\phi(x)) + \int_{\mathbb{X}} g(z) q(dz|x,\phi(x))$, $x \in \mathbb{X}$, then ϕ is *average-cost optimal* and $w(x) = w^* = \underline{w} = \overline{w}$ for all $x \in \mathbb{X}$. Here we need a different form of such a statement.

Theorem 2. *Let Assumption ($\underline{\mathbf{B}}$)(a) holds. If there exists a measurable function $g : \mathbb{X} \to [0,+\infty)$ and a stationary policy ϕ such that*

$$\overline{w} + g(x) \geq u(x,\phi(x)) + \int_{\mathbb{X}} g(z) q(dz|x,\phi(x)), \quad x \in \mathbb{X}, \tag{2}$$

then ϕ is average-cost optimal and

$$w(x) = w^\phi(x) = \limsup_{\alpha\uparrow 1} (1-\alpha) v_\alpha(x) = \overline{w} = w^*, \quad x \in \mathbb{X}. \tag{3}$$

Assumption ($\mathbf{W}^*$) and "boundedness" Assumption ($\underline{\mathbf{B}}$) on the function u_α, which is weaker than the boundedness Assumption ($\mathbf{B}$) introduced by Schäl [10], lead to the validity of stationary average-cost optimal inequalities and the existence of stationary policies.

Let us set $\Phi^*(x) := \{ y \in \Phi(x) : \overline{w} + \underline{u}(x) \geq u(x,y) + \int_{\mathbb{X}} \underline{u}(z) q(dz|x,y) \}$, $\underline{u}(x) := \liminf_{\alpha\uparrow 1,\ z\to x} u_\alpha(z)$, $x \in \mathbb{X}$, and let $\Phi_*(x)$, $x \in \mathbb{X}$, be the sets defined in (1) for this function $\underline{u}$; $\Phi_*(x) \subseteq \Phi^*(x)$.

Theorem 3. *Suppose Assumptions* ($\mathbf{W}^*$) *and* ($\underline{\mathbf{B}}$) *hold. There exist a stationary policy* ϕ *satisfying (2). Thus, equalities (3) hold for this policy* ϕ*. Furthermore, the following statements hold:* (a) *the function* $\underline{u} : \mathbb{X} \to \mathbb{R}_+$ *is l.s.c.;* (b) *the nonempty sets* $\Phi^*(x)$, $x \in \mathbb{X}$, *satisfy the following properties:* (b_1) *the graph* $\mathrm{Gr}_{\mathbb{X}}(\Phi^*)$ *is a Borel subset of* $\mathbb{X}\times\mathbb{Y}$*;* ($b_2$) *for each* $x \in \mathbb{X}$ *the set* $\Phi^*(x)$ *is compact;* (c) *a stationary policy* ϕ *is optimal for average costs and satisfies (2), if* $\phi(x) \in \Phi^*(x)$ *for all* $x \in \mathbb{X}$*;* (d) *there exists a stationary policy* ϕ *with* $\phi(x) \in \Phi_*(x) \subseteq \Phi^*(x)$ *for all* $x \in \mathbb{X}$*;* (e) *if, in addition,* u *is inf-compact on* $\mathrm{Gr}_{\mathbb{X}}(\Phi)$*, then the function* $\underline{u}$ *is inf-compact.*

Acknowledgements. This research was partially supported by NSF grants CMMI-0900206 and CMMI-0928490. The authors thank Professor M.Z. Zgurovsky for initiating their research cooperation.

References

[1] C. Berge (1963). *Topological spaces.* Macmillan, New York.

[2] E.A. Feinberg, P.O. Kasyanov, N.V. Zadoianchuk (2012). Average-Cost Markov Decision Processes with Weakly Continuous Transition Probabilities. *arXiv:1202.4122v1*; to appear in *Math. Oper. Res.*

[3] E.A. Feinberg, P. O. Kasyanov, N. V. Zadoianchuk (2012). Berge's Theorem for Noncompact Image Sets. *arXiv:1203.1340v1*; to appear in *J. Math. Anal. Appl.*

[4] E.A. Feinberg, M.E. Lewis (2007). Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem. *Math. Oper. Res.* **32**(4): 769–783.

[5] O. Hernández-Lerma (1991). Averege optimality in dynamic programming on Borel spaces – Unbounded costs and controls. *Systems & Control Lett.* **17**(3): 237–242.

[6] O. Hernández-Lerma, J.B. Lassere (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria.* Springer, New York.

[7] Sh. Hu, N.S. Papageorgiou (1997). *Handbook of multivalued analysis. Volume I: theory.* Kluwer, Dordrecht.

[8] A.S. Kechris (1995). *Classical descriptive set theory.* Springer-Verlag, New York.

[9] F. Luque-Vasques, O. Hernández-Lerma (1995). A counterexample on the semicontinuity lemma. *Proc. Amer. Math. Soc.* **123**(10): 3175-3176.

[10] M. Schäl (1993). Average optimality in dynamic programming with general state space. *Math. Oper. Res.* **18**(1): 163–172.

[11] R. Serfozo (1982). Convergence of Lebesgue integrals with varying measures. *Sankhya: The Indian Journal of Statistics (Series A)* **44**: 380–402.

[12] M.Z. Zgurovsky, V. S. Mel'nik, P. O. Kasyanov (2011). *Evolution Inclusions and Variation Inequalities for Earth Data Processing I.* Springer, Berlin.