Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2013, Article ID 271279, 5 pages
http://dx.doi.org/10.1155/2013/271279

*Research Article*

# Nonuniqueness versus Uniqueness of Optimal Policies in Convex Discounted Markov Decision Processes

**Raúl Montes-de-Oca,[1] Enrique Lemus-Rodríguez,[2] and Francisco Sergio Salem-Silva[3]**

[1] *Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa, Avenida San Rafael Atlixco 186, Col. Vicentina, 09340 México, DF, Mexico*
[2] *Universidad Anáhuac México-Norte, Avenida Universidad Anáhuac 46, Lomas Anáhuac, 52786 Huixquilucan, MEX, Mexico*
[3] *Facultad de Matemáticas, Universidad Veracruzana, Circuito Gonzalo Aguirre Beltrán s/n, Zona Universitaria, 91000 Xalapa, VER, Mexico*

Correspondence should be addressed to Raúl Montes-de-Oca; momr@xanum.uam.mx

Received 16 October 2012; Accepted 12 February 2013

Academic Editor: Debasish Roy



From the classical point of view, it is important to determine if in a Markov decision process (MDP), besides their existence, the uniqueness of the optimal policies is guaranteed. It is well known that uniqueness does not always hold in optimization problems (for instance, in linear programming). On the other hand, in such problems it is possible for a slight perturbation of the functional cost to restore the uniqueness. In this paper, it is proved that the value functions of an MDP and its cost perturbed version stay close, under adequate conditions, which in some sense is a priority. We are interested in the stability of Markov decision processes with respect to the perturbations of the cost-as-you-go function.

## 1. Introduction

From the classical point of view (for instance, in Hadamard's concept of well-posedness [1]) in a mathematical modeling problem, it is crucial that both the existence and the uniqueness are secured. But, in optimization, neither of these is guaranteed, and even if extra conditions ensure the existence of optimizers, their uniqueness will not automatically follow. For instance, in linear programming, we even have the extreme case that when there are two different optimal vectors all of their convex linear combinations become optimal automatically. But a slight perturbation of the cost functional will "destroy" most of the optimizers. In this sense, nonuniqueness in linear programming is highly unstable. This question is of interest with respect to the standard discounted Markov decision model, as in [2], which presents conditions that guarantee the uniqueness of the optimal policies.

In this paper, we study a family of perturbations of the cost of an MDP and establish that, under convexity and adequate bounds, the value functions of both the original and the cost-perturbed Markov decision processes (MDPs) are uniformly close. This result will eventually help us determine whether both the uniqueness and the nonuniqueness are stable with respect to this kind of perturbation.

The structure of this paper is simple. Firstly, the preliminaries and assumptions of the model are outlined. Secondly, the main theorem is stated and proved, followed by the main example. A brief section with the concluding remarks closes the paper.

## 2. Preliminaries: Discounted MDPs and Convexity Assumptions

Let $(X, A, \{A(x) : x \in X\}, Q, c)$ be a Markov control model (see [3] for details and terminology) which consists of the state space $X$, the control (or action) set $A$, the transition law $Q$, and the cost-per-stage $c$. It is assumed that both $X$ and $A$ are subsets of $\mathbb{R}$ (this is supposed for simplicity, but it is also possible to present the theory of this paper considering that $X$ and $A$ are subsets of Euclidean spaces of the dimension greater than one). For each $x \in X$, there is a nonempty measurable set $A(x) \subset A$ whose elements are the feasible

actions when the state of the system is $x$. Define $\mathbb{K} := \{(x, a) : x \in X, a \in A(x)\}$. Finally, the cost-per-stage $c$ is a nonnegative and measurable function on $\mathbb{K}$.

Let $\Pi$ be the set of all (possibly randomized, history-dependent) admissible policies. By standard convention, a *stationary policy* is identified with a measurable function $f : X \to A$ such that $f(x) \in A(x)$ for all $x \in X$. The set of stationary polices is denoted by $\mathbb{F}$. For every $\pi \in \Pi$ and an initial state $x \in X$, let

$$V(\pi, x) = E_x^{\pi}\left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)\right] \tag{1}$$

be the *total expected discounted cost* when using the policy $\pi$, given the initial state $x$. The number $\alpha \in (0, 1)$ is called the *discount factor* ($\alpha$ is assumed to be fixed). Here $\{x_t\}$ and $\{a_t\}$ denote the state and the control sequences, respectively, and $E_x^{\pi}$ is the expectation operator. A policy $\pi^*$ is said to be *optimal* if $V(\pi^*, x) = V^*(x)$ for all $x \in X$, where $V^*(x) = \inf_{\pi \in \Pi} V(\pi, x)$, $x \in X$. $V^*$ is called the *optimal value function*. The following assumption will also be taken into consideration.

*Assumption 1.* (a) $c$ is lower semicontinuous and inf-compact on $\mathbb{K}$ (i.e., for every $x \in X$ and $r \in \mathbb{R}$ the set $\{a \in A(x) : c(x, a) \le r\}$ is compact).

(b) The transition law $Q$ is strongly continuous, that is, $w(x, a) = \int u(y) Q(dy \mid x, a)$, $(x, a) \in \mathbb{K}$ is continuous and bounded on $\mathbb{K}$, for every measurable bounded function $u$ on $X$.

(c) There exists a policy $\pi$ such that $V(\pi, x) < \infty$, for each $x \in X$.

*Remark 2.* The following consequences of Assumption 1 are well known (see Theorem 4.2.3 and Lemma 4.2.8 in [3]).

(a) The optimal value function $V^*$ is the solution of the *optimality equation* (OE), that is, for all $x \in X$,

$$V^*(x) = \min_{a \in A(x)} \left\{c(x, a) + \alpha \int V^*(y) Q(dy \mid x, a)\right\}. \tag{2}$$

There is also $f^* \in \mathbb{F}$ such that

$$V^*(x) = c(x, f^*(x)) + \alpha \int V^*(y) Q(dy \mid x, f^*(x)), \quad x \in X, \tag{3}$$

and $f^*$ is optimal.

(b) For every $x \in X$, $v_n(x) \uparrow V^*(x)$, with $v_n$ defined as

$$v_n(x) = \min_{a \in A(x)} \left\{c(x, a) + \alpha \int v_{n-1}(y) Q(dy \mid x, a)\right\}, \tag{4}$$

$x \in X, n = 1, 2, \ldots$, and $v_0(x) = 0$. Moreover, for each $n$, there is $f_n \in \mathbb{F}$ such that for each $x \in X$,

$$\min_{a \in A(x)} \left\{c(x, a) + \alpha \int v_{n-1}(y) Q(dy \mid x, a)\right\} = c(x, f_n(x)) + \alpha \int v_{n-1}(y) Q(dy \mid x, f_n(x)). \tag{5}$$

Let $(X, A, \{A(x) : x \in X\}, Q, c)$ be a fixed Markov control model. Take $M$ as the MDP with the Markov control model $(X, A, \{A(x) : x \in X\}, Q, c)$. The optimal value function, the optimal policy which comes from (3), and the minimizers in (5) will be denoted for $M$ by $V^*$, $f^*$, and $f_n$, $n = 1, 2, \ldots$, respectively. Also let $v_n$, $n = 1, 2, \ldots$, be the value iteration functions for $M$. Let $G(x, a) := c(x, a) + \alpha \int V^*(y) Q(dy \mid x, a)$, $(x, a) \in \mathbb{K}$.

It will be also supposed that the MDPs taken into account satisfy one of the following Assumptions 3 or 4.

*Assumption 3.* (a) $X$ and $A$ are convex.

(b) $(1 - \lambda)a + a' \in A((1 - \lambda)x + x')$ for all $x, x' \in X$, $a \in A(x)$, $a' \in A(x')$, and $\lambda \in [0, 1]$. Besides, it is assumed that if $x$ and $y \in X$, $x < y$, then $A(y) \subseteq A(x)$, and $A(x)$ are convex for each $x \in X$.

(c) $Q$ is induced by a difference equation $x_{t+1} = F(x_t, a_t, \xi_t)$, with $t = 0, 1, \ldots$, where $F : X \times A \times S \to X$ is a measurable function and $\{\xi_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with values in $S \subseteq \mathbb{R}$, and with a common density $\Delta$. In addition, we suppose that $F(\cdot, \cdot, s)$ is a convex function on $\mathbb{K}$, for each $s \in S$, and if $x$ and $y \in X$, $x < y$, then $F(x, a, s) \le F(y, a, s)$ for each $a \in A(y)$ and $s \in S$.

(d) $c$ is convex on $\mathbb{K}$, and if $x$ and $y \in X$, $x < y$, then $c(x, a) \le c(y, a)$, for each $a \in A(y)$.

*Assumption 4.* (a) The same as Assumption 3(a).

(b) $(1 - \lambda)a + a' \in A((1 - \lambda)x + x')$ for all $x, x' \in X$, $a \in A(x)$, $a' \in A(x')$, and $\lambda \in [0, 1]$. Besides, $A(x)$ is assumed to be convex for each $x \in X$.

(c) $Q$ is given by the relation $x_{t+1} = \gamma x_t + \delta a_t + \xi_t$, $t = 0, 1, \ldots$, where $\{\xi_t\}$ are i.i.d. random variables taking values in $S \subseteq \mathbb{R}$ with the density $\Delta$, $\gamma$ and $\delta$ are real numbers.

(d) $c$ is convex on $\mathbb{K}$.

*Remark 5.* Assumptions 3 and 4 are essentially the same as assumptions C1 and C2 in pages 419–420 of reference [2], with the difference that we are now able to assume that the function $c(\cdot, \cdot)$ is convex and not necessarily strictly convex. (in fact, in [2], Conditions C1 and C2 take into account the more general situation in which both $X$ and $A$ are subsets of Euclidean spaces of the dimension greater than one). Also note that it is possible to obtain that each of Assumptions 3 and 4 implies that, for each $x \in X$, $G(x, \cdot)$ is convex but not necessarily strictly convex (hence, $M$ does not necessarily have a unique optimal policy). The proof of this fact is a direct consequence of the convexity of the cost function $c$ and of the proof of Lemma 6.2 in [2].

## 3. Main Result and an Example

For $\epsilon > 0$, consider the following MDP denoted by $M_\epsilon$ with the Markov control model $(X, A, \{A(x) \mid x \in X\}, Q, c^*)$, where $c^*(x, a) = c(x, a) + \epsilon a^2$, $(x, a) \in \mathbb{K}$, where $c$ is the cost function for $M$. Observe that both MDPs $M$ and $M_\epsilon$ coincide in the components of the Markov control model except for the cost function; moreover, $\mathbb{F}$ is the same set in both models. Additionally we suppose that.

*Assumption 6.* There is a policy $\phi$ such that $E_x^{\phi}[\sum_{t=0}^{\infty} \alpha^t c^*(x_t, a_t)] < \infty$, for each $x \in X$.

*Remark 7.* Suppose that, for $M$, Assumption 1 holds. Then, it is direct to verify that if $M_\epsilon$ satisfies Assumption 6, then it also satisfies Assumption 1.

For $M_\epsilon$, let $W^*$, $g^*$, and $g_n$, $n = 1, 2, \ldots$, denote the optimal value function, the optimal policy which comes from (3), and the minimizers in (5), respectively. Moreover, let $w_n$, $n = 1, 2, \ldots$, be the corresponding value iteration functions for $M_\epsilon$.

*Remark 8.* Suppose that, for $M$, one of Assumptions 3 or 4 holds. Then, notice that as $c$ is a convex function, it is trivial to prove that $c^*$ is strictly convex. Then, under Assumption 6, it follows that $M_\epsilon$ satisfies C1 or C2 in [2] and that $G^*$ is strictly convex, where $G^*(x,a) = c^*(x,a) + \alpha \int W^*(y) Q(dy \mid x, a)$, $(x,a) \in \mathbb{K}$, so $g^*$ is unique.

Let $\Theta : X \to \mathbb{R}$, so that $\Theta(x) \geq W^*(x)$ and $\Theta(x) \geq V^*(x)$, $x \in X$, and take, for each $x \in X$, $B(x) = \{a \in A(x) : c(x,a) \leq \Theta(x)\}$ and $B^*(x) = \{a \in A(x) : c^*(x,a) \leq \Theta(x)\}$.

*Remark 9.* It is easy to verify, using Assumption 1, that for each $x \in X$, $B(x)$ and $B^*(x)$ are nonempty and compact. Moreover, since $c \geq 0$ and from Remark 2, $f_n(x)$, $f^*(x) \in B(x)$; $g_n(x)$, $g^*(x) \in B^*(x)$ for each $x \in X$ and $n \geq 1$. It is also trivial to prove that, for each $x \in X$, $B^*(x) \subseteq B(x)$; hence $f_n(x)$, $f^*(x)$, $g_n(x)$, $g^*(x) \in B(x)$, for each $x \in X$ and $n \geq 1$.

*Condition 10.* There exists a measurable function $Z : X \to \mathbb{R}$, which may depend on $\alpha$, such that $c^*(x,a) - c(x,a) = \epsilon a^2 \leq \epsilon Z(x)$, and $\int Z(y) Q(dy \mid x, a) \leq Z(x)$ for each $x \in X$ and $a \in B(x)$.

*Remark 11.* With respect to the existence of the function $Z$ mentioned in Condition 10 that satisfies that $\int Z(y) Q(dy \mid x, a) \leq Z(x)$ for each $x \in X$ and $a \in B(x)$, it is important to note that this kind of requirement has been previously used in the unbounded MDPs literature (see, for instance, the Remarks presented on page 578 of [4]).

**Theorem 12.** *Suppose that Assumptions 1 and 6 hold, and that, for $M$, one of Assumptions 3 or 4 holds. Let $\epsilon$ be a positive number. Then,*

(a) *if $A$ is compact, $|W^*(x) - V^*(x)| \leq \epsilon K^2/(1-\alpha)$, $x \in X$, where $K$ is the diameter of a compact set $D$ such that $0 \in D$ and $A \subseteq D$;*

(b) *under Condition 10, $|W^*(x) - V^*(x)| \leq \epsilon Z(x)/(1-\alpha)$, $x \in X$.*

*Proof.* The proof of case (a) follows from the proof of case (b) given that $Z(x) = K^2$, $x \in X$ (observe that in this case, if $a \in A$, then $a^2 = (a-0)^2 \leq K^2$).

(b) Assume that $M$ satisfies Assumption 3 (the proof for the case in which $M$ satisfies Assumption 4 is similar).

Firstly, for each $x \in X$,

$$\begin{aligned} w_1(x) - v_1(x) &= \min_{a \in A(x)} c^*(x,a) - \min_{a \in A(x)} c(x,a) \\ &\geq c^*\left(x, g_1(x)\right) - c\left(x, g_1(x)\right) \\ &= \epsilon \left(g_1(x)\right)^2 \\ &\geq 0, \end{aligned} \tag{6}$$

$$\begin{aligned} \left|w_1(x) - v_1(x)\right| &= w_1(x) - v_1(x) \\ &= \min_{a \in A(x)} c^*(x,a) - \min_{a \in A(x)} c(x,a) \end{aligned} \tag{7}$$

$$\leq c^*\left(x, f_1(x)\right) - c\left(x, f_1(x)\right) \tag{8}$$

$$= \epsilon \left(f_1(x)\right)^2 \tag{9}$$

$$\leq \epsilon Z(x). \tag{10}$$

Secondly, assume that for some positive integer $n$ and for each $x \in X$,

$$\begin{aligned} 0 \leq w_n(x) - v_n(x) &= \left|w_n(x) - v_n(x)\right| \\ &\leq \left(1 + \alpha + \alpha^2 + \cdots + \alpha^n\right) \epsilon Z(x). \end{aligned} \tag{11}$$

Consequently, using Condition 10, for each $x \in X$,

$$w_{n+1}(x) - v_{n+1}(x) \tag{12}$$

$$\begin{aligned} &= \min_{a \in A(x)} \left[c^*(x,a) + \alpha \int w_n(y) Q(dy \mid x, a)\right] \\ &\quad - \min_{a \in A(x)} \left[c(x,a) + \alpha \int v_n(y) Q(dy \mid x, a)\right] \\ &\leq \left[c^*\left(x, f_n(x)\right) + \alpha \int w_n(y) Q\left(dy \mid x, f_n(x)\right)\right] \\ &\quad - \left[c\left(x, f_n(x)\right) + \alpha \int v_n(y) Q\left(dy \mid x, f_n(x)\right)\right] \end{aligned} \tag{13}$$

$$\begin{aligned} &= c^*\left(x, f_n(x)\right) - c\left(x, f_n(x)\right) \\ &\quad + \alpha \int \left[w_n(y) - v_n(y)\right] Q\left(dy \mid x, f_n(x)\right) \\ &\leq \epsilon Z(x) + \alpha \left(1 + \alpha + \alpha^2 + \cdots + \alpha^n\right) \epsilon Z(x) \\ &= \left(1 + \alpha + \alpha^2 + \cdots + \alpha^{n+1}\right) \epsilon Z(x). \end{aligned}$$

On the other hand, from (11) and the fact that $c^*(x,a) - c(x,a) = \epsilon a^2$, $a \in A(x)$, for each $x \in X$,

$$\begin{aligned} & w_{n+1}(x) - v_{n+1}(x) \\ &= \min_{a\in A(x)} \left[ c^*(x,a) + \alpha \int w_n(y)\, Q(dy \mid x,a) \right] \\ &\quad - \min_{a\in A(x)} \left[ c(x,a) + \alpha \int v_n(y)\, Q(dy \mid x,a) \right] \\ &\geq \left[ c^*(x, g_n(x)) + \alpha \int w_n(y)\, Q(dy \mid x, g_n(x)) \right] \\ &\quad - \left[ c(x, g_n(x)) + \alpha \int v_n(y)\, Q(dy \mid x, g_n(x)) \right] \\ &= c^*(x, g_n(x)) - c(x, g_n(x)) \\ &\quad + \alpha \int [w_n(y) - v_n(y)]\, Q(dy \mid x, f_n(x)) \\ &\geq 0. \end{aligned} \tag{14}$$

In conclusion, combining (10), (13), and (14), it is obtained that, for each $x \in X$, (11) holds for all $n \geq 1$. Now, letting $n \to +\infty$ in (11), we get $|W^*(x) - V^*(x)| \leq \epsilon Z(x)/(1-\alpha)$, $x \in X$. □

The following corollary is immediate.

**Corollary 13.** *Suppose that Assumptions 1 and 6 hold. Suppose that for M one of Assumptions 3 or 4 holds (hence, M does not necessarily have a unique optimal policy). Let $\epsilon$ be a positive number. If A is compact or Condition 10 holds, then there exists an MDP $M_\epsilon$ with a unique optimal policy $g^*$, such that inequalities in Theorem 12 (a) or (b) hold, respectively.*

*Example 14.* Let $X = (0,\infty)$, $A = A(x) = (-\infty, 0]$, for all $x \in X$. The dynamic of the system is given by

$$x_{t+1} = x_t e^{a_t + \xi_t}, \tag{15}$$

$t = 0, 1, \ldots$. Here, $\xi_0, \xi_1, \ldots$ are i.i.d. random variables with values in $S = (-\infty, 0]$ and with a common continuous bounded density denoted by $\Delta$. The cost function is given by $c(x,a) = x + |a|$, $x \in X$ (observe that $c$ is convex but not strictly convex).

**Lemma 15.** *Example 14 satisfies Assumptions 1, 3, and 6, and Condition 10.*

*Proof.* Assumption 1 (a) trivially holds. The proof of the strong continuity of $Q$ is as follows: if $u : X \to \mathbb{R}$ is a measurable and bounded function, then, using the change of variable theorem, a simple computation shows that

$$\begin{aligned} \int_{(0,\infty)} u(y)\, Q(dy \mid x,a) &= \int_{(-\infty,0]} u\left(xe^{a+s}\right) \Delta(s)\, ds \\ &= \int_{(0,xe^a]} u(w)\, \Delta\left(\ln \frac{w}{x} - a\right) \frac{1}{w} dw, \end{aligned} \tag{16}$$

$(x,a) \in \mathbb{K}$. As $u$ is a bounded function and $\Delta$ is a bounded continuous function, it follows directly, using the convergence dominated theorem, that

$$\int_{(0,xe^a]} u(w)\, \Delta\left(\ln \frac{w}{x} - a\right) \frac{1}{w} dw, \tag{17}$$

$(x,a) \in \mathbb{K}$ is a continuous function on $\mathbb{K}$. Hence,

$$\int_{(0,\infty)} u(y)\, Q(dy \mid \cdot, \cdot) \tag{18}$$

is a continuous function on $\mathbb{K}$.

By direct computations we get, for the stationary policy $g(x) = 0$, $x \in X$, both $V(g,x) = E_x^g[\sum_{t=0}^\infty \alpha^t c(x_t, a_t)]$ and $W(g,x) = E_x^g[\sum_{t=0}^\infty \alpha^t c^*(x_t, a_t)]$ are less or equal to $\Theta(x) = x/(1-\alpha)$ for all $x \in X$ (observe that, in this case, $\Theta(x) \geq W^*(x)$ and $\Theta(x) \geq V^*(x)$, $x \in X$); consequently, Assumptions 1 and 6 hold.

On the other hand, Assumptions 3(a), (b), and (d) are immediate. Let $F(x,a,s) = xe^{a+s}$, $x \in (0,\infty)$, $a, s \in (-\infty, 0]$. Clearly, $F(\cdot, a, s)$ is nondecreasing in the first variable.

Now, take $\lambda \in [0,1]$ and $(x,a)$, $(x', a') \in \mathbb{k}$ and $s \in S$. Then, considering that $e^s$, $e^{\lambda a}$, and $e^{(1-\lambda)a'}$ are less or equal than one,

$$\begin{aligned} & F\left(\lambda x + (1-\lambda)x', \lambda a + (1-\lambda)a', s\right) \\ &= \left(\lambda x + (1-\lambda)x'\right) e^{(\lambda a + (1-\lambda)a' + s)} \\ &= e^s \left(\lambda x + (1-\lambda)x'\right) e^{\lambda a} e^{(1-\lambda)a'} \\ &= e^s \left[\lambda x e^{\lambda a} e^{(1-\lambda)a'} + (1-\lambda)x' e^{\lambda a} e^{(1-\lambda)a'}\right] \\ &\leq e^s \left[\lambda x e^{\lambda a} + (1-\lambda)x' e^{(1-\lambda)a'}\right] \\ &= \lambda F(x,a,s) + (1-\lambda) F\left(x', a', s\right), \end{aligned} \tag{19}$$

hence, $F(\cdot, \cdot, s)$ is convex, that is, Assumption 3(c) holds.

Now, for each $x \in X$,

$$\begin{aligned} B(x) &= \left\{ a \in \mathbb{R} : x + |a| \leq \frac{x}{1-\alpha} \right\} \\ &= \left\{ a \in \mathbb{R} : |a| \leq \frac{\alpha x}{1-\alpha} \right\}. \end{aligned} \tag{20}$$

Hence, taking $Z(x) = [\alpha x/(1-\alpha)]^2$, $x \in X$, using (20) and, again, that $e^{a+s} \leq 1$, it is possible to obtain that for each $x \in X$ and $a \in B(x)$,

$$c^*(x,a) - c(x,a) = \epsilon a^2 \leq \epsilon \left[\frac{\alpha x}{1-\alpha}\right]^2, \tag{21}$$

and that $\int Z(y) Q(dy \mid x,a) = [\alpha/(1-\alpha)]^2 \int y^2 Q(dy \mid x,a) = [\alpha/(1-\alpha)]^2 x^2 \int [e^{(a+s)}]^2 \Delta(s) ds \leq [\alpha/(1-\alpha)]^2 x^2 = Z(x)$. □

## 4. Concluding Remarks

The specific form of the perturbation used in this paper is taken from [5, Exercise 28, page 81], where it is established that a convex function $f$ perturbed by a suitable quadratic positive function becomes strictly convex and coercive. In fact, this kind of perturbation is very much related to the one Tanaka et al. propose in their paper [6], and further research in this direction is being conducted.

Both state and action spaces are considered to be subsets of $\mathbb{R}$, just for simplicity of exposition. All the results hold in $\mathbb{R}^n$. In this case, if $A \subseteq \mathbb{R}^n$ $(n > 1)$, then it is possible to take

$$c^*(x,a) = c(x,a) + \epsilon\|a\|^2, \tag{22}$$

$(x,a) \in \mathbb{K}$, where $\|a\|^2 = a_1^2 + \cdots + a_n^2$, $a = (a_1, \ldots, a_n)$ (see [5, Exercise 28, page 81]), and all the results on this article remain valid.

Theorem 12, on the closeness of the value functions of the original and the perturbed MDPs, requires conditions that are all very common in the MDPs technical literature. The importance of the result lies in the fact that it is a crucial step to the study of the problem of stability under the cost perturbation of the uniqueness or nonuniqueness of optimal policies.

Finally, we should mention that this research was motivated by our interest in understanding the relationship between nonuniqueness and robustness in several statistical procedures based on optimization.

## References


[1] J. Hadamard, *Sur les Problemes aux Derivees Partielles et Leur Signification Physique*, Princeton University Bulletin, 1902.

[2] D. Cruz-Suárez, R. Montes-de-Oca, and F. Salem-Silva, "Conditions for the uniqueness of optimal policies of discounted Markov decision processes," *Mathematical Methods of Operations Research*, vol. 60, no. 3, pp. 415–436, 2004.

[3] O. Hernández-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes*, vol. 30, Springer, New York, NY, USA, 1996.

[4] J. A. E. E. Van Nunen and J. Wessels, "A note on dynamic programming with unbounded rewards," *Management Sciences*, vol. 24, no. 5, pp. 576–580, 1978.

[5] A. L. Peressini, F. E. Sullivan, and J. J. Uhl, Jr., *The Mathematics of Nonlinear Programming*, Springer, New York, NY, USA, 1988.

[6] K. Tanaka, M. Hoshino, and D. Kuroiwa, "On an $\epsilon$-optimal policy of discrete time stochastic control processes," *Bulletin of Informatics and Cybernetics*, vol. 27, no. 1, pp. 107–119, 1995.