Adv. Appl. Prob. **44,** 774–793 (2012)
Printed in Northern Ireland

THE EXPECTED TOTAL COST CRITERION FOR MARKOV DECISION PROCESSES UNDER CONSTRAINTS: A CONVEX ANALYTIC APPROACH

FRANÇOIS DUFOUR,* *Université Bordeaux, IMB and INRIA Bordeaux Sud-ouest*
M. HORIGUCHI,** *Kanagawa University*
A. B. PIUNOVSKIY,*** *University of Liverpool*

Abstract

This paper deals with discrete-time Markov decision processes (MDPs) under constraints where all the objectives have the same form of expected total cost over the infinite time horizon. The existence of an optimal control policy is discussed by using the convex analytic approach. We work under the assumptions that the state and action spaces are general Borel spaces, and that the model is nonnegative, semicontinuous, and there exists an admissible solution with finite cost for the associated linear program. It is worth noting that, in contrast to the classical results in the literature, our hypotheses do not require the MDP to be transient or absorbing. Our first result ensures the existence of an optimal solution to the linear program given by an occupation measure of the process generated by a randomized stationary policy. Moreover, it is shown that this randomized stationary policy provides an optimal solution to this Markov control problem. As a consequence, these results imply that the set of randomized stationary policies is a sufficient set for this optimal control problem. Finally, our last main result states that all optimal solutions of the linear program coincide on a special set with an optimal occupation measure generated by a randomized stationary policy. Several examples are presented to illustrate some theoretical issues and the possible applications of the results developed in the paper.

Keywords: Markov decision process; expected total cost criterion; constraint; linear programming; occupation measure

2010 Mathematics Subject Classification: Primary 90C40
Secondary 60J10; 90C90

1. Introduction

The objective of this work is to study time-homogeneous Markov decision processes (MDPs) with constraints when all the objectives have the same form of expected total cost over the infinite time horizon. The class of MDPs is a general family of controlled stochastic processes suitable for the modeling of sequential decision-making problems. The convex analytic approach has proved to be a very efficient method for solving MDPs with constraints. We do not attempt to present an exhaustive survey on this topic, but refer the interested reader to [1], [5], [13], and the references therein for a detailed exposition of this technique. While the convex analytic

Received 10 October 2011; revision received 30 January 2012.
* Postal address: INRIA Bordeaux Sud-ouest, CQFD Team, 351 cours de la Libération, F-33400 Talence, France. Email address: dufour@math.u-bordeaux1.fr
** Postal address: Department of Mathematics, Faculty of Engineering, Kanagawa University, 3-27-1 Rokkakubashi, Kanagawa-ku, Yokohama 221-8686, Japan. Email address: horiguchi@kanagawa-u.ac.jp
*** Postal address: Department of Mathematical Sciences, University of Liverpool, Liverpool L69 7ZL, UK. Email address: piunov@liverpool.ac.uk

formulation is available for a large variety of cost criteria, the expected total cost (ETC) criterion has received less attention.

Most of the works on the ETC criterion deal with the dynamic programming approach and, consequently, do not consider the cases with constraints; see, for example, the books [1], [3], [9], and [14], and the survey [7]. Indeed, when the convex analytic technique is applied to the ETC criterion, one mainly encounters two important difficulties, as pointed out in [5, pp. 357–358] and [9, pp. 92–94]: the expected state-action frequency may not be finite for some or all policies, and an admissible solution for the linear program (LP) may not correspond to any expected state-action frequency of the process. By imposing suitable conditions on the model such as the so-called transient and absorbing conditions, we can ensure that the expected state-action frequency is finite for all stationary policies. Discounted models form a special class in this area. It is important to point out that even in the transient case, some occupation measures may not be generated by a stationary control policy as explained in [5, p. 358], meaning that the sufficiency of stationary policies is under question. To the best of our knowledge, the book [1] is the only reference in the literature in which the ETC criterion with constraints is analyzed using the convex analytic approach under the hypotheses that the state and action spaces are discrete and the model is transient or absorbing. We also mention the papers by Dufour and Piunovskiy [6] and Horiguchi [10], [11], who studied the optimal stopping problems through the ETC criterion with constraints.

In this paper we investigate the ETC criterion with constraints using the convex analytic approach. We work under the assumptions that the state and action spaces are general Borel spaces, and that the model is nonnegative, semicontinuous, and there exists an admissible solution for the LP with a finite cost. We do not require the MDP to be transient or absorbing. It is important to point out that our hypotheses are very weak and do not exclude the pathologies previously described. In particular, the (optimal) occupation measures are not necessarily finite and an admissible solution for the LP may not correspond to any occupation measure of the controlled process. Moreover, it appears necessary to impose the condition that the cost functions be nonnegative. Indeed, the first example in Section 5 shows that if the cost functions can take negative values then the optimal solution of the LP may make no sense. Consequently, our results appear to be very general compared to those in the existing literature. We show in Theorem 4.1 that there exists a randomized stationary policy φ^* having the following properties: there exists an optimal solution to the LP given by an occupation measure of the process generated by the policy φ^*, and the policy φ^* is optimal for the constrained control problem. As a consequence, we show in Corollary 4.1 that the set of randomized stationary policies is a sufficient set for solving the control problem under consideration. Although the occupation measures are not necessarily finite, we prove the remarkable property that there exists a special set on which the occupation measures are σ-finite. Finally, our last main result (Theorem 4.2) states that all optimal solutions of the LP coincide on this special set with an optimal occupation measure generated by a randomized stationary policy. A related approach has been used in [6] in a considerably simpler context given by an optimal stopping problem. As a result, the transition distribution of the process does not depend on the control, contrary to the present work. This difference imposes the development of a radically different approach to deal with this general framework.

The rest of the paper is organized as follows. In Section 2 we introduce some notation, basic assumptions, and present the control problem that will be considered throughout this work. Preliminary results are derived in Section 3. In particular, a special set is constructed that will be crucial for the analysis of the constrained control problem. The LP is studied

in Section 4 where we derive the main results of our paper. Finally, Section 5 is dedicated to the presentation of several examples illustrating some theoretical issues and the possible applications of the results developed in the paper.

2. Problem formulation

The purpose of this section is to present some standard notation and some basic definitions as well as the discrete-time Markov control model that will be considered throughout the paper.

The following notation will be used in the paper: $\mathbb{N}$ denotes the set of natural numbers, $\mathbb{N}^0 = \mathbb{N} \cup \{0\}$, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}_+$ denotes the set of nonnegative real numbers, and $\overline{\mathbb{R}}_+$ denotes $\mathbb{R}_+ \cup \{+\infty\}$. For any $q \in \mathbb{N}$, $\mathbb{N}_q$ is the set $\{1, \ldots, q\}$. The term *measure* will always refer to a countably additive, $\overline{\mathbb{R}}_+$-valued set function. Let E be a Borel space, and denote by $\mathcal{B}(E)$ its associated Borel σ-algebra. The set of measures defined on $(E, \mathcal{B}(E))$ is denoted by $\mathbb{M}(E)_+$. For two measures $(\gamma_1, \gamma_2) \in \mathbb{M}(E)_+^2$, $\gamma_1 \leq \gamma_2$ means that $\gamma_1(\Gamma) \leq \gamma_2(\Gamma)$ for any $\Gamma \in \mathcal{B}(E)$. The setwise convergence of a sequence of measures $(\gamma_n)_{n\in\mathbb{N}}$ to a measure γ_∞ is denoted by $\lim_{n\to\infty} \gamma_n = \gamma_\infty$. The set of bounded real-valued continuous functions defined on E is denoted by $\mathbb{C}(E)$. Let f be a real-valued measurable function, and let $\eta \in \mathbb{M}(E)_+$. The integral $\int_E f(y)\eta(\mathrm{d}y)$ is denoted by $\eta(f)$ provided it is well defined. Let X and Y be Borel spaces. If W is a stochastic kernel on X given Y then, for any real-valued measurable function f, the integral $\int_E f(x)W(\mathrm{d}x \mid y)$ for any $y \in Y$ is denoted by $Wf(y)$ provided it is well defined. For any positive measure η on $(Y, \mathcal{B}(Y))$, ηW is the measure defined on $(X, \mathcal{B}(X))$ by

$$\eta W(\Gamma) = \int_Y W(\Gamma \mid y)\eta(\mathrm{d}y) \quad \text{for any } \Gamma \in \mathcal{B}(X).$$

Let μ be a measure in $\mathbb{M}(X \times Y)_+$; the marginal of μ on X is denoted by μ_Y: $\mu_Y(\Gamma) = \mu(\Gamma \times Y)$ for any $\Gamma \in \mathcal{B}(X)$.

In order to define an MDP, we consider, as in Section 2 of [8], a four-tuple (X, A, Q, r) consisting of

(a) a Borel space X which is the state space,

(b) a Borel space A, representing the control or action set,

(c) a stochastic kernel Q on X given $X \times A$ which stands for the transition law of the controlled process,

(d) a measurable function $r_0 \colon X \times A \to \mathbb{R}$ representing the running cost,

(e) measurable functions $r_i \colon X \times A \to \mathbb{R}$ for $i \in \mathbb{N}_q$ representing the constraints.

Definition 2.1. The set of all stochastic kernels φ on A given X is denoted by Φ, and $\mathbb{F}$ stands for the set of all measurable functions $f \colon X \to A$.

To introduce the optimal control problem we are concerned with, it is necessary to define different classes of control policies.

Definition 2.2. Define $H_0 = X$ and $H_t = X \times A \times H_{t-1}$ for $t \geq 1$. A control policy is a sequence $\pi = (\pi_t)_{t\in\mathbb{N}^0}$ of stochastic kernels π_t on A given H_t. Let Π be the class of all policies. A policy $\pi = (\pi_t)_{t\in\mathbb{N}^0}$ is said to be

- a randomized stationary policy if there exists $\varphi \in \Phi$ such that $\pi_t(\cdot \mid h_t) = \varphi(\cdot \mid x_t)$,

- a deterministic Markov policy if there exists a sequence $(f_t)_{t\in\mathbb{N}^0} \subset \mathbb{F}$ such that $\pi_t(\cdot \mid h_t) = \delta_{f_t(x_t)}(\cdot)$,
- a deterministic stationary policy if there exists $f \in \mathbb{F}$ such that $\pi_t(\cdot \mid h_t) = \delta_{f(x_t)}(\cdot)$,

where $h_t = (x_0, a_0, \ldots, x_{n-1}, a_{t-1}, x_t)$.

According to the standard convention, we identify $\mathbb{F}$ and Φ with the classes of all deterministic and randomized, respectively, stationary policies. Therefore, $\mathbb{F} \subset \Phi \subset \Pi$. If π is a randomized stationary policy generated by $\varphi \in \Phi$ (according to Definition 2.2), we will write φ instead of π; similarly if π is a deterministic stationary policy generated by $f \in \mathbb{F}$, we will write f instead of π.

Let $(\Omega, \mathcal{F})$ be the canonical space consisting of the sample path $\Omega = (X \times A)^\infty$ and the associated σ-algebra $\mathcal{F}$. For any policy $\pi \in \Pi$ and any initial distribution ν on X, we can define a probability, labeled P_ν^π, and a stochastic process $((x_t, a_t))_{t\in\mathbb{N}^0}$, where $(x_t)_{t\in\mathbb{N}^0}$ is the state process and $(a_t)_{t\in\mathbb{N}^0}$ is the control process satisfying, for any $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$, and $h_t \in H_t$ with $t \in \mathbb{N}^0$, $\mathrm{P}_\nu^\pi(x_0 \in B) = \nu(B)$, $\mathrm{P}_\nu^\pi(a_t \in C \mid h_t) = \pi_t(C \mid h_t)$, and $\mathrm{P}_\nu^\pi(x_{t+1} \in B \mid h_t, a_t) = Q(B \mid x_t, a_t)$; see, for example, [8, Chapter 2] for such a construction. The expectation with respect to P_ν^π is denoted by E_ν^π. If $\nu = \delta_x$ for $x \in X$, we write P_x^π for P_ν^π and E_x^π for E_ν^π.

Next, we define our Markov control problem. Suppose that we are given an initial distribution ν on X, and constraint limits $(R_1, \ldots, R_q) \in \mathbb{R}^q$. The optimization problem we consider consists in minimizing the cost function

$$v(\nu, \pi) = \mathrm{E}_\nu^\pi\left[\sum_{t=0}^{\infty} r_0(x_t, a_t)\right] \tag{2.1}$$

over the set of feasible control policies, labeled Π_{c}, defined by the set of policies $\pi \in \Pi$ such that

$$v_i(\nu, \pi) = \mathrm{E}_\nu^\pi\left[\sum_{t=0}^{\infty} r_i(x_t, a_t)\right] \le R_i \tag{2.2}$$

for $i \in \mathbb{N}_q$. The optimal value function is denoted by

$$V^*(\nu) = \inf_{\pi\in\Pi_{\mathrm{c}}} v(\nu, \pi). \tag{2.3}$$

For a policy $\pi \in \Pi$, let us introduce the following expected state-action frequency or occupation measure induced by $\pi \in \Pi$:

$$\mu^\pi(\Gamma) = \sum_{t=0}^{\infty} \mathrm{P}_\nu^\pi((x_t, a_t) \in \Gamma) \quad \text{for any } \Gamma \in \mathcal{B}(X \times A). \tag{2.4}$$

In this paper, we make the following assumptions on the parameters of the MDP.

Assumption 2.1. *Assume that the following hypotheses hold.*

(i) *The control space A is compact.*

(ii) *The mappings r_0 and r_i for all $i \in \mathbb{N}_q$ are nonnegative and lower semicontinuous.*

(iii) *The kernel Q is weakly continuous, that is, Qf is continuous on $X \times A$ for any $f \in \mathbb{C}(X)$.*

Remark 2.1. Note that Assumption 2.1 is a standard and weak hypothesis in the literature on MDPs. Moreover, from (iii) of Assumption 2.1, it is easy to see that $Q(\Gamma \mid \cdot, \cdot)$ is lower semicontinuous for any open set $\Gamma \in \mathcal{B}(X)$.

3. Preliminary results

The main goal of this section is to derive preliminary results that will be used in Section 4 to obtain the main results of our paper. In particular, a special set, labeled V_r, is constructed with remarkable properties; see Proposition 3.2 for its definition. It is proved in Theorem 3.2 that on the complement of V_r all the occupation measures of interest are σ-finite. Another key characteristic of the set V_r is that, roughly speaking, its complement is *stochastically closed*. More precisely, this means, see Proposition 3.2, that V_r is topologically closed and that if the process starts from a point in V_r then there exists an action that keeps the process in V_r with an associated cost being equal to 0. These properties will play a crucial role in showing in Theorem 3.3 and Corollary 3.1 that, for any admissible solution for the LP, we can construct a randomized stationary policy, giving a better value for the cost. We conclude this section by deriving two technical results (Lemmas 3.4 and 3.5) that will be used in Section 4 to show that any optimal solution of the LP coincides with an optimal occupation measure generated by a randomized stationary policy on V_r^c.

In this section we will denote by r a lower-semicontinuous nonnegative function defined on $X \times A$ and we will suppose that the following assumption holds.

Assumption 3.1. *There exists a measure* $\eta \in \mathbb{M}(X \times A)_+$ *satisfying* $\eta_A = \nu + \eta Q$ *and* $\eta(r) < \infty$.

Remark 3.1. Note that we do not require the measure η to be finite. Actually, η can take the value $+\infty$ and it is not necessarily σ-finite.

Remark 3.2. Since the function r is lower semicontinuous and the whole space $X \times A$ is closed, parts (i) and (iii) of Assumption 2.1 imply that we deal with a semicontinuous model according to Definition 8.7 of [4]. Therefore, according to Corollary 9.17.2 of [4], the cost function associated to r defined by

$$r^*(x) = \inf_{\pi \in \Pi} \mathrm{E}_\nu^\pi \left[\sum_{t=0}^{\infty} r(x_t, a_t) \right]$$

is a lower-semicontinuous function on X satisfying the optimality equation

$$r^*(x) = \inf_{a \in A} [r(x, a) + Qr^*(x, a)].$$

Moreover, there exists a measurable mapping $f^* \in \mathbb{F}$ such that the deterministic stationary policy f^* is optimal, that is,

$$r^*(x) = \mathrm{E}_\nu^{f^*} \left[\sum_{t=0}^{\infty} r(x_t, a_t) \right] = r(x, f^*(x)) + Qr^*(x, f^*(x)) \tag{3.1}$$

for any $x \in X$.

Let us first show a technical result that will be used repeatedly without reference.

Lemma 3.1. *Let g be a lower-semicontinuous function on $X \times A$. For $\alpha \in \mathbb{R}$, the set*

$$\{x \in X\colon \text{ for all } a \in A,\ g(x,a) > \alpha\}$$

is an open set and so belongs to $\mathcal{B}(X)$.

Proof. From item (b) of Proposition D.5 of [8], there exists a measurable function $\phi\colon X \to A$ such that $\inf_{a\in A} g(x,a) = g(x,\phi(x))$ and $g(x,\phi(x))$ is lower semicontinuous on X. Consequently, $\{x \in X\colon \inf_{a\in A} g(x,a) > \alpha\} = \{x \in X\colon g(x,\phi(x)) > \alpha\}$ is an open set in X. For any $x \in X$, $g(x,\cdot)$ is lower semicontinuous and so $\{x \in X\colon \text{ for all } a \in A,\ g(x,a) > \alpha\} = \{x \in X\colon \inf_{a\in A} g(x,a) > \alpha\}$, proving the result.

Lemma 3.2. *Let g be a lower-semicontinuous function on $X \times A$. Then,*

$$\bigcup_{p\in\mathbb{N}} \left\{x \in X\colon \text{ for all } a \in A,\ g(x,a) > \frac{1}{p}\right\} = \{x \in X\colon \text{ for all } a \in A,\ g(x,a) > 0\}.$$

Proof. It is easy to see that

$$\bigcup_{p\in\mathbb{N}} \left\{x \in X\colon \text{ for all } a \in A,\ g(x,a) > \frac{1}{p}\right\} \subset \{x \in X\colon \text{ for all } a \in A,\ g(x,a) > 0\}.$$

Now take an arbitrary $y \in \{x \in X\colon \text{ for all } a \in A,\ g(x,a) > 0\}$. Since $g(y,\cdot)$ is lower semicontinuous on the compact set A, there exists $u \in A$ such that, for all $a \in A$, $g(y,a) \geq \inf_{a\in A} g(y,a) = g(y,u) > 0$, implying that there exists $p \in \mathbb{N}$ such that $y \in \{x \in X\colon \text{ for all } a \in A,\ g(x,a) > 1/p\}$.

Proposition 3.1. *Suppose that Assumption 3.1 holds, and assume that there exists an increasing sequence of open sets $(B_j)_{j\in\mathbb{N}} \subset \mathcal{B}(X)$ such that η_A is finite on B_j. Then there exists a sequence of open sets $(E_i)_{i\in\mathbb{N}} \subset \mathcal{B}(X)$ such that $\{x \in X\colon \text{for all } a \in A,\ Q(\bigcup_{j\in\mathbb{N}} B_j \mid x,a) + r(x,a) > 0\} = \bigcup_{i\in\mathbb{N}} E_i$ and η_A is finite on E_i for any $i \in \mathbb{N}$.*

Proof. Introduce for $j \in \mathbb{N}$ and $p \in \mathbb{N}$ the following sets:

$$B_j^p = \left\{(y,a) \in X \times A\colon Q(B_j \mid y,a) + r(y,a) > \frac{1}{p}\right\},$$
$$C_j^p = \left\{y \in X\colon \text{ for all } a \in A,\ Q(B_j \mid y,a) + r(y,a) > \frac{1}{p}\right\}.$$

Since ν is a probability measure, we have $\eta Q(B_j) < \infty$ and so

$$\eta(B_j^p) \leq \int_{B_j^p} p[Q(B_j \mid y,a) + r(y,a)]\eta(\mathrm{d}y \times \mathrm{d}a) \leq p[\eta Q(B_j) + \eta(r)] < \infty.$$

Consequently, η_A is finite on C_j^p. Since, for any $j \in \mathbb{N}$, B_j is open, $Q(B_j \mid \cdot,\cdot) + r(\cdot,\cdot)$ is lower semicontinuous on $X \times A$. From Lemma 3.1, it follows that the set C_j^p is open. Therefore, we will obtain the result if we show that

$$\bigcup_{j\in\mathbb{N}} \bigcup_{p\in\mathbb{N}} C_j^p = \left\{x \in X\colon \text{ for all } a \in A,\ Q\left(\bigcup_{j\in\mathbb{N}} B_j \,\middle|\, x,a\right) + r(x,a) > 0\right\}. \tag{3.2}$$

However, from Lemma 3.2 we have

$$\bigcup_{p\in\mathbb{N}} C_j^p = \{x \in X\colon \text{ for all } a \in A,\ Q(B_j \mid x, a) + r(x, a) > 0\}, \tag{3.3}$$

and so

$$\bigcup_{j\in\mathbb{N}}\bigcup_{p\in\mathbb{N}} C_j^p \subset \{x \in X\colon \text{ for all } a \in A,\ Q(B \mid x, a) + r(x, a) > 0\}, \tag{3.4}$$

where $B = \bigcup_{j\in\mathbb{N}} B_j$. Let us show the reverse inclusion. Consider $y \in X$ such that $Q(B \mid y, a) + r(y, a) > 0$ for all $a \in A$. Since B_i is open, $\{Q(B_i \mid y, \cdot) + r(y, \cdot)\}_{i\in\mathbb{N}}$ is an increasing sequence of lower-semicontinuous functions on A. Consequently, Lemma 2.1 of [16] gives

$$\begin{aligned}\lim_i \inf_{a\in A}[Q(B_i \mid y, a) + r(y, a)] &= \inf_{a\in A} \lim_i [Q(B_i \mid y, a) + r(y, a)] \\ &= \inf_{a\in A}[Q(B \mid y, a) + r(y, a)].\end{aligned}$$

The set B being open, $Q(B \mid y, \cdot) + r(y, \cdot)$ is lower semicontinuous on the compact set A and so $\inf_{a\in A}[Q(B \mid y, a)+r(y, a)] > 0$. Therefore, $y \in \{x \in X\colon \text{ for all } a \in A,\ Q(B_j \mid x, a)+ r(x, a) > 0\}$ for some $j \in \mathbb{N}$ and so by using (3.3) we obtain the reverse inclusion:

$$\{x \in X\colon \text{ for all } a \in A,\ Q(B \mid x, a) + r(x, a) > 0\} \subset \bigcup_{j\in\mathbb{N}}\bigcup_{p\in\mathbb{N}} C_j^p. \tag{3.5}$$

Combining (3.4) and (3.5), we obtain (3.2) and the result follows.

Theorem 3.1. *Suppose that Assumption 3.1 holds. Define the set*

$$W = \bigcup_{j\in\mathbb{N}} W_j,$$

where $W_1 = \{y \in X\colon \text{ for all } a \in A,\ r(y, a) > 0\}$ *and, for any* $j \in \mathbb{N}$,

$$W_{j+1} = \left\{x \in X\colon \text{ for all } a \in A,\ Q\left(\bigcup_{i=1}^{j} W_i \,\middle|\, x, a\right) + r(x, a) > 0\right\}.$$

Then the set W_j *is open for any* $j \in \mathbb{N}$ *and the measure* η_A *is* σ*-finite on* W.

Proof. From Lemma 3.2, $W_1 = \bigcup_{j\in\mathbb{N}} D_j$, where $D_j = \{y \in X\colon \text{ for all } a \in A,\ r(y, a) > 1/j\}$. By using Lemma 3.1, D_j is open as well as W_1. Moreover, the measure η_A is finite on the set D_j. Indeed, $\eta(D_j \times A) \le \int_{D_j} jr(y, a)\eta(\mathrm{d}y \times \mathrm{d}a) \le j\eta(r) < \infty$. Now, using Proposition 3.1, it can be easily shown by induction that the set W_j is open for any $j \in \mathbb{N}$ and η_A is σ-finite on W_j for all $j \in \mathbb{N}$, implying that η_A is σ-finite on W.

Proposition 3.2. *Define the set*

$$V_r = \{x \in X\colon r^*(x) = 0\}.$$

The set V_r *is closed. Moreover, a state* x *belongs to* V_r *if and only if there exists* $a \in A$ *such that* $r(x, a) = 0$ *and* $Q(V_r \mid x, a) = 1$.

Proof. From Remark 3.2, r^* is lower semicontinuous on X and so V_r is closed, showing the first statement of the proposition. Let $x \in V_r$. From (3.1), $r^*(x) = r(x, f^*(x)) + Qr^*(x, f^*(x)) = 0$. Consequently, for $a = f^*(x)$, we have $r(x, a) = 0$ and $Qr^*(x, a) = 0$. However,

$$Qr^*(x,a) = \int_{V_r} r^*(y)Q(\mathrm{d}y \mid x,a) + \int_{V_r^c} r^*(y)Q(\mathrm{d}y \mid x,a) = \int_{V_r^c} r^*(y)Q(\mathrm{d}y \mid x,a).$$

Since on V_r^c the function r^* is strictly positive, it follows that $Q(V_r^c \mid x, a) = 0$, showing the first part of the result.

Now assume that $y \in X$ satisfies $r(y, a) = 0$ and $Q(V_r \mid y, a) = 1$ for $a \in A$. Define the deterministic Markov policy $\pi = \{\psi_t\}$ by $\psi_0(x) = a$ for any $x \in X$ and $\psi_t = f^*$ for $t \geq 1$. It is easy to see that $v^\pi(y) = 0$, implying that $r^*(y) = 0$ and so $y \in V_r$.

Lemma 3.3. *Let G be a closed set in $\mathcal{B}(X)$. Assume that, for any $x \in G$, there exists $a_x \in A$ such that $r(x, a_x) = 0$ and $Q(G \mid x, a_x) = 1$. Then $G \subset V_r$.*

Proof. Assume without loss of generality that $G \neq \varnothing$. We have

$$\inf_{b\in A}[r(x,b) + Q(G^c \mid x,b)] = 0 \quad \text{for } x \in G.$$

Since G^c is open, the function $r(x, \cdot) + Q(G^c \mid x, \cdot)$ is lower semicontinuous on the compact set A. Therefore, it follows from Proposition D.5 of [8] that there exists a measurable mapping $\psi_G : G \to A$ such that, for any $x \in G$, $r(x, \psi_G(x)) + Q(G^c \mid x, \psi_G(x)) = 0$. Fix an arbitrary $u \in A$, and define the measurable mapping $\psi : X \to A$ by $\psi(x) = \psi_G(x)$ if $x \in G$ and $\psi(x) = u$ otherwise. Then, for $x \in G$ and the deterministic stationary policy ψ, $v^\psi(x) = 0$ and so $x \in V_r$, proving the result.

Proposition 3.3. *The set W^c is included in V_r.*

Proof. Consider $x \in W^c$. Then, by the definition of W, it follows that, for any $j \in \mathbb{N}$, there exists $a_j \in A$ satisfying $Q(\bigcup_{i=1}^j W_i \mid x, a_j) + r(x, a_j) = 0$. This implies that

$$\lim_j \inf_{a\in A}\left[Q\left(\bigcup_{i=1}^j W_i \;\middle|\; x,a\right) + r(x,a)\right] = 0.$$

Clearly, $\{Q(\bigcup_{i=1}^j W_i \mid x, \cdot) + r(x, \cdot)\}_{j\in\mathbb{N}}$ is an increasing sequence of lower-semicontinuous functions defined on the compact set A since $\bigcup_{i=1}^j W_i$ is open from Theorem 3.1. Consequently,

$$\lim_j \inf_{a\in A}\left[Q\left(\bigcup_{i=1}^j W_i \;\middle|\; x,a\right) + r(x,a)\right] = \inf_{a\in A}\lim_j\left[Q\left(\bigcup_{i=1}^j W_i \;\middle|\; x,a\right) + r(x,a)\right],$$

and so

$$\inf_{a\in A}[Q(W \mid x,a) + r(x,a)] = 0.$$

Again, from Theorem 3.1, W is open and so $Q(W \mid x, \cdot) + r(x, \cdot)$ is a lower-semicontinuous function defined on the compact set A. Consequently, for $x \in W^c$, there exists $a_x \in A$ such that $Q(W \mid x, a_x) + r(x, a_x) = 0$. The set W^c being closed, we obtain the result by using Lemma 3.3.

Theorem 3.2. *The measure η_A is σ-finite on V_r^c.*

Proof. This result is a straightforward consequence of Theorem 3.1 and Proposition 3.3.

According to Theorem 3.2, there exists a partition $(U_k)_{k\in\mathbb{N}} \subset \mathcal{B}(X)$ of the set V_r^c such that, for all $k \in \mathbb{N}$, $\eta_A(U_k) < \infty$. From Proposition D.8 of [8], for any $k \in \mathbb{N}$ such that $\eta_A(U_k) > 0$, there exists a stochastic kernel, labeled φ_k, on A given U_k satisfying $\eta(\Gamma_k \times \Gamma_A) = \int_{\Gamma_k} \varphi_k(\Gamma_A \mid x)\eta_A(\mathrm{d}x)$ with $\Gamma_k \in \mathcal{B}(U_k)$ and $\Gamma_A \in \mathcal{B}(A)$. Consider $\widetilde{\varphi}$, an arbitrary kernel on A given X. Define $\widehat{\varphi}^\eta$, the kernel on A given V_r^c, by

$$\widehat{\varphi}^\eta(\Gamma_A \mid x) = \begin{cases} \varphi_k(\Gamma_A \mid x) & \text{if } x \in U_k \text{ with } \eta(U_k) > 0, \\ \widetilde{\varphi}(\Gamma_A \mid x) & \text{otherwise,} \end{cases}$$

for any $\Gamma_A \in \mathcal{B}(A)$. Clearly, it is a stochastic kernel satisfying

$$\eta(\Gamma \times \Gamma_A) = \int_\Gamma \widehat{\varphi}^\eta(\Gamma_A \mid x)\eta_A(\mathrm{d}x)$$

for any $\Gamma \in \mathcal{B}(V_r^c)$ and $\Gamma_A \in \mathcal{B}(A)$.

Definition 3.1. Suppose that Assumption 3.1 holds. Associated to η and r, we introduce the stochastic kernel $\varphi^{\eta,r}$ on A given X defined by

$$\varphi^{\eta,r}(\Gamma \mid x) = \widehat{\varphi}^\eta(\Gamma \mid x)\delta_x(V_r^c) + \delta_{f^*(x)}(\Gamma)\delta_x(V_r).$$

We say that the randomized stationary policy $\varphi^{\eta,r}$ is induced by (η, r).

Remark 3.3. (i) The subscripts η and r in $\varphi^{\eta,r}$ indicate the possible dependence of the policy on the measure η and the function r. For notational ease, however, 'r' will be omitted if there is no possibility of confusion, that is, it will be written φ^η instead of $\varphi^{\eta,r}$.

(ii) Note that, for any $x \in V_r$, $r(x, f^*(x)) = 0$.

Theorem 3.3. *Suppose that Assumption 3.1 holds. Then the randomized stationary policy φ^η induced by η and r satisfies $\mu_A^{\varphi^\eta}(\Gamma) \le \eta_A(\Gamma)$ for any $\Gamma \in \mathcal{B}(V_r^c)$, and, if the measures $\mu_A^{\varphi^\eta}$ and η_A coincide on V_r^c, then the measures μ^{φ^η} and η coincide on $V_r^c \times A$. Moreover,*

$$\mathrm{E}_\nu^{\varphi^\eta}\left[\sum_{t=0}^{\infty} r(x_t, a_t)\right] \le \eta(r).$$

Proof. Introduce the sequence of measures $(\eta_A^n)_{n\in\mathbb{N}} \subset \mathbb{M}(V_r^c)_+$ by η_A^1 equal to the restriction of η_A to V_r^c and $\eta_A^{n+1} = S^\eta \eta_A^n$, where $S^\eta \colon \mathbb{M}(V_r^c)_+ \to \mathbb{M}(V_r^c)_+$ is given by

$$S^\eta\gamma(\Gamma) = \nu(\Gamma) + \int_{V_r^c}\int_A Q(\Gamma \mid x, a)\varphi^\eta(\mathrm{d}a \mid x)\gamma(\mathrm{d}x)$$

for $\gamma \in \mathbb{M}(V_r^c)_+$ and $\Gamma \in \mathcal{B}(V_r^c)$. Let us show that

$$\eta_A^2 = S^\eta\eta_A^1 \le \eta_A^1. \tag{3.6}$$

Indeed, for any $\Gamma \in \mathcal{B}(V_r^c)$,

$$S^\eta\eta_A^1(\Gamma) = \nu(\Gamma) + \int_{V_r^c}\int_A Q(\Gamma \mid x, a)\varphi^\eta(\mathrm{d}a \mid x)\eta_A^1(\mathrm{d}x). \tag{3.7}$$

Observe that, for any $\Gamma \in \mathcal{B}(V_r^c)$,

$$\int_{V_r^c}\int_A Q(\Gamma \mid x,a)\varphi^\eta(\mathrm{d}a \mid x)\eta_A(\mathrm{d}x) = \int_{V_r^c}\int_A Q(\Gamma \mid x,a)\eta(\mathrm{d}x\times \mathrm{d}a) \le \eta Q(\Gamma), \quad (3.8)$$

by the definition of φ^η. However, we have $\eta_A = \nu + \eta Q$, and so combining (3.7) and (3.8), we obtain $S^\eta \eta_A^1(\Gamma) \le \nu(\Gamma) + \eta Q(\Gamma) = \eta_A^1(\Gamma)$ for any $\Gamma \in \mathcal{B}(V_r^c)$.

Using (3.6) and the fact that S^η is monotone, it is easy to show by induction that $\eta_A^{n+1} \le \eta_A^n \le \eta_A^1$. Therefore, the limit, labeled η_A^∞, of the decreasing sequence $(\eta_A^n)_{n\in\mathbb{N}}$ as n tends to ∞ exists and satisfies, for any $\Gamma \in \mathcal{B}(V_r^c)$,

$$\eta_A^\infty(\Gamma) \le \eta_A(\Gamma). \quad (3.9)$$

Define the mapping $T^\eta\colon \mathbb{M}(X)_+ \to \mathbb{M}(X)_+$ by

$$T^\eta\mu(\cdot) = \nu(\cdot) + \int_X\int_A Q(\cdot \mid x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu(\mathrm{d}x).$$

The occupation measure $\mu_A^{\varphi^\eta}$ is the minimal positive solution of $\mu = T^\eta\mu$ for $\mu \in \mathbb{M}(X)_+$. The mapping T^η is monotone and so it can be easily shown that $\mu_A^{\varphi^\eta}$ is the limit of the increasing sequence of measures $(\mu^n)_{n\in\mathbb{N}} \subset \mathbb{M}(X)_+$ defined by $\mu^1 = \nu$ and $\mu^{n+1} = T^\eta\mu^n$ for $n \in \mathbb{N}$. Now consider the sequence of measures $(\nu^n)_{n\in\mathbb{N}} \subset \mathbb{M}(V_r^c)_+$ defined by $\nu_n(\Gamma) = \mu_n(\Gamma)$ for $\Gamma \in \mathcal{B}(V_r^c)$ and $n \in \mathbb{N}$. Therefore, $(\nu^n)_{n\in\mathbb{N}}$ is an increasing sequence of measures that converges to ν^∞ defined by $\nu^\infty(\Gamma) = \mu_A^{\varphi^\eta}(\Gamma)$ for any $\Gamma \in \mathcal{B}(V_r^c)$. However, note that, for any $\Gamma \in \mathcal{B}(V_r^c)$ and any measure $\mu \in \mathbb{M}(X)_+$, we have

$$T^\eta\mu(\Gamma) = \nu(\Gamma) + \int_{V_r^c}\int_A Q(\Gamma \mid x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu(\mathrm{d}x). \quad (3.10)$$

Indeed, for any $\Gamma \in \mathcal{B}(V_r^c)$,

$$\begin{aligned} T^\eta\mu(\Gamma) = \nu(\Gamma) &+ \int_{V_r^c}\int_A Q(\Gamma \mid x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu(\mathrm{d}x) \\ &+ \int_{V_r}\int_A Q(\Gamma \mid x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu(\mathrm{d}x). \end{aligned}$$

However, by the definition of φ^η, $\int_A Q(\Gamma \mid x,a)\varphi^\eta(\mathrm{d}a \mid x) = Q(\Gamma \mid x, f^*(x))$ for any $x \in V_r$. Since $\Gamma \in \mathcal{B}(V_r^c)$ and $x \in V_r$, we have, from Proposition 3.2, $Q(\Gamma \mid x, f^*(x)) = 0$ and so $\int_{V_r}\int_A Q(\Gamma \mid x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu(\mathrm{d}x) = 0$, showing (3.10).

Using (3.10), we can therefore equivalently define the sequence $(\nu^n)_{n\in\mathbb{N}} \subset \mathbb{M}(V_r^c)_+$ as ν^1 equal to the restriction of ν on V_r^c and $\nu^{n+1} = S^\eta\eta^n$ for $n \in \mathbb{N}$. However, the mapping S^η is monotone, and since $\nu^1 \le \eta_A^1$, we have $\nu^n \le \eta_A^n$, so $\mu_A^{\varphi^\eta}(\Gamma) = \lim_{n\to\infty}\nu^n(\Gamma) \le \lim_{n\to\infty}\eta_A^n(\Gamma) = \eta_A^\infty(\Gamma)$ for any $\Gamma \in \mathcal{B}(V_r^c)$. From (3.9) we obtain $\mu_A^{\varphi^\eta}(\Gamma) \le \eta_A(\Gamma)$ for any $\Gamma \in \mathcal{B}(V_r^c)$. When the measures $\mu_A^{\varphi^\eta}$ and η_A coincide on V_r^c, the measures μ^{φ^η} and η coincide on $V_r^c \times A$ according to the definition of the policy φ^η induced by η and r. This shows the first statement of the theorem.

Now, let us show the last statement of the theorem. We have

$$\int_{V_r^c}\int_A r(x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu_A^{\varphi^\eta}(\mathrm{d}x) \le \int_{V_r^c}\int_A r(x,a)\varphi^\eta(\mathrm{d}a \mid x)\eta_A(\mathrm{d}x),$$

and so

$$\eta(r) \geq \int_{V_r^c} \int_A r(x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu_A^{\varphi^\eta}(\mathrm{d}x).$$

However, by the definition of φ^η we have

$$\int_{V_r} \int_A r(x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu_A^{\varphi^\eta}(\mathrm{d}x) = \int_{V_r} r(x, f^*(x))\mu_A^{\varphi^\eta}(\mathrm{d}x),$$

implying that

$$\int_{V_r} \int_A r(x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu_A^{\varphi^\eta}(\mathrm{d}x) = 0$$

since $r(x, f^*(x)) = 0$ for $x \in V_r$. Combining the two previous inequalities, we obtain

$$\eta(r) \geq \int_X \int_A r(x,a)\varphi^\eta(\mathrm{d}a \mid x)\mu_A^{\varphi^\eta}(\mathrm{d}x) = \mathrm{E}_\nu^{\varphi^\eta}\left[\sum_{t=0}^{\infty} r(x_t, a_t)\right],$$

proving the result.

Corollary 3.1. *Under the conditions of Theorem 3.3, let $\widetilde{r} \leq r$ be a nonnegative real-valued function defined on $X \times A$. Then*

$$\mathrm{E}_\nu^{\varphi^\eta}\left[\sum_{t=0}^{\infty} \widetilde{r}(x_t, a_t)\right] \leq \eta(\widetilde{r}).$$

Proof. Observe that $\widetilde{r}(x,a) = 0$ if $(x,a) \in V_r$. Consequently, similar arguments as those used at the end of the proof of Theorem 3.3 can be applied to get the result.

Lemma 3.4. *Define the set*

$$\widetilde{W} = \bigcup_{j \in \mathbb{N}} \widetilde{W}_j,$$

where $\widetilde{W}_1 = \{x \in V_r^c\colon \text{for all } a \in A,\ r(x,a) > 0\}$ and, for $j \in \mathbb{N}$,

$$\widetilde{W}_{j+1} = \left\{x \in V_r^c\colon \text{for all } a \in A,\ Q\left(\bigcup_{i=1}^{j} \widetilde{W}_i \;\middle|\; x, a\right) + r(x,a) > 0\right\}.$$

Then $\widetilde{W} = V_r^c$.

Proof. Combining Lemma 3.1 and Proposition 3.2, it can be easily shown by induction that the set $\widetilde{W}_j$ is open for any $j \in \mathbb{N}$. Clearly, by the definition of $\widetilde{W}$ we have $\widetilde{W} \subset V_r^c$. Now let us take $x \in \widetilde{W}^c$. Then, for any $j \in \mathbb{N}$, there exists $a_j \in A$ such that $Q(\bigcup_{i=1}^{j} \widetilde{W}_i \mid x, a_j) + r(x, a_j) = 0$, implying that, for any $j \in \mathbb{N}$, $\inf_{a \in A}[Q(\bigcup_{i=1}^{j} \widetilde{W}_i \mid x, a) + r(x,a)] = 0$. Note that $(Q(\bigcup_{i=1}^{j} \widetilde{W}_i \mid x, \cdot) + r(x, \cdot))_{n \in \mathbb{N}}$ is an increasing sequence of lower-semicontinuous functions on the compact set A since $\bigcup_{i=1}^{j} \widetilde{W}_i$ is open. Therefore, it follows from Lemma 2.1 of [16] that

$$\begin{aligned}
\inf_{a \in A}[Q(\widetilde{W} \mid x, a) + r(x,a)] &= \inf_{a \in A} \lim_{n \to \infty}\left[Q\left(\bigcup_{i=1}^{j} \widetilde{W}_i \;\middle|\; x, a\right) + r(x,a)\right] \\
&= \lim_{n \to \infty} \inf_{a \in A}\left[Q\left(\bigcup_{i=1}^{j} \widetilde{W}_i \;\middle|\; x, a\right) + r(x,a)\right] \\
&= 0.
\end{aligned}$$

However, $\widetilde{W}$ being open, $Q(\widetilde{W} \mid x, \cdot) + r(x, \cdot)$ is a lower-semicontinuous function on the compact set A and so reaches its infimum on A, showing that there exists $a \in A$ satisfying $Q(\widetilde{W} \mid x, a) + r(x, a) = 0$. Lemma 3.3 implies that $x \in V_r$, showing the reverse inclusion: $\widetilde{W}^c \subset V_r$.

Lemma 3.5. *Suppose that $\gamma \in \mathbb{M}(V_r^c \times A)_+$ satisfies $\gamma_A \geq \gamma Q$ with $\gamma_A(V_r^c) > 0$. Then*

$$\int_{V_r^c} \int_A r(x, a)\gamma(\mathrm{d}x \times \mathrm{d}a) > 0.$$

Proof. Let us show by induction that if $\gamma_A(\bigcup_{i=1}^j \widetilde{W}_i) > 0$ then

$$\int_{\bigcup_{i=1}^j \widetilde{W}_i} \int_A r(x, a)\gamma(\mathrm{d}x \times \mathrm{d}a) > 0.$$

This is clearly true at step $j = 1$ because of the definition of $\widetilde{W}_1$. Now assume that if $\gamma_A(\bigcup_{i=1}^j \widetilde{W}_i) > 0$ then $\int_{\bigcup_{i=1}^j \widetilde{W}_i} \int_A r(x, a)\gamma(\mathrm{d}x \times \mathrm{d}a) > 0$. Consider $\gamma_A(\bigcup_{i=1}^{j+1} \widetilde{W}_i) > 0$. Then either $\gamma_A(\bigcup_{i=1}^j \widetilde{W}_i) > 0$ or $\gamma_A(\widetilde{W}_{j+1}) > 0$. In the first case, we obtain the claim by using the induction hypothesis. In the latter case, we have $\gamma_A(\widetilde{W}_{j+1}) > 0$, implying by the definition of $\widetilde{W}_{j+1}$ that either

$$\int_{\widetilde{W}_{j+1}} \int_A r(x, a)\gamma(\mathrm{d}x \times \mathrm{d}a) > 0 \quad \text{or} \quad \int_{\widetilde{W}_{j+1}} \int_A Q\Bigg(\bigcup_{i=1}^j \widetilde{W}_i \;\Bigg|\; x, a\Bigg)\gamma(\mathrm{d}x \times \mathrm{d}a) > 0.$$

In the latter case, using the fact that $\gamma_A \geq \gamma Q$, we have $\gamma_A(\bigcup_{i=1}^j \widetilde{W}_i) > 0$ and so, by the induction hypothesis, $\int_{\bigcup_{i=1}^j \widetilde{W}_i} \int_A r(x, a)\gamma(\mathrm{d}x \times \mathrm{d}a) > 0$. Consequently, in either case we obtain $\int_{\bigcup_{i=1}^{j+1} \widetilde{W}_i} \int_A r(x, a)\gamma(\mathrm{d}x \times \mathrm{d}a) > 0$. Now, by using Lemma 3.4, it follows that there exists $k \in \mathbb{N}$ such that $\gamma_A(\bigcup_{i=1}^k \widetilde{W}_i) > 0$ since $\gamma_A(V_r^c) > 0$, implying that

$$\int_{V_r^c} \int_A r(x, a)\gamma(\mathrm{d}x \times \mathrm{d}a) \geq \int_{\bigcup_{i=1}^k \widetilde{W}_i} \int_A r(x, a)\gamma(\mathrm{d}x \times \mathrm{d}a) > 0.$$

This proves the lemma.

4. Linear program

In this section we present the main results of the paper, supposing that the assumption on the parameters of the MDP presented in Section 2 is satisfied and assuming that there exists an admissible solution for the LP with finite cost. Our first main result, Theorem 4.1, consists of showing that there exists a feasible randomized stationary policy $\varphi^* \in \Pi_{\mathrm{c}}$ which generates an optimal solution μ^{φ^*} to the LP and which is optimal in the optimization problem (2.1)–(2.3). As a consequence, we show that the set of randomized stationary policies is a sufficient class of policies for the control problem under consideration. Our second main result states that any optimal solution of the LP on the complement of a special subset V_r coincides with an optimal occupation measure generated by the induced randomized stationary policy.

The constrained LP is defined as

$$\text{(LP)} \begin{cases} \text{minimize } \mu(r_0) \text{ subject to } \mu \in \mathbb{L}, \end{cases}$$

where

$$\mathbb{L} = \{\mu \in \mathbb{M}(X \times A)_+ \colon \mu_A = \nu + \mu Q \text{ and } \mu(r_n) \leq R_n \text{ for } n \in \mathbb{N}_q\}.$$

A measure μ is said to be *admissible* for the LP if $\mu \in \mathbb{L}$, and a measure μ is said to be *optimal* for the LP if μ is admissible and if $\mu(r_0) \leq \gamma(r_0)$ for any $\gamma \in \mathbb{L}$.

For notational convenience, let $g_t \colon \Omega \to X \times A$ be defined by $g_t(\omega) = (x_t(\omega), a_t(\omega))$, and denote by $h_0(\omega) = x_0(\omega)$ and $h_t(\omega) = (g_0(\omega), \ldots, g_{t-1}(\omega), x_t(\omega))$ for $\omega \in \Omega$ and $t \geq 1$. Denote by $\mathcal{P}$ the set of probability measures on $(\Omega, \mathcal{F})$ and by $\mathcal{P}^\pi$ the set of probability measures on $(\Omega, \mathcal{F})$ induced by the control policies $\pi \in \Pi$. Now introduce $\mathcal{O}^\pi$ as the set of occupation measures $\mu^\pi \in \mathbb{M}(X \times A)_+$ defined by (2.4). Let the mapping $\mathbb{O} \colon \mathcal{P}^\pi \to \mathcal{O}^\pi$ be defined by $\mathbb{O}(P_\nu^\pi) = \mu^\pi$.

The w-topology on $\mathcal{P}$ is defined as the coarsest topology rendering the mappings

$$P \to \int_\Omega f(h_t(\omega)) \, \mathrm{dP}(\omega)$$

continuous, where $f \in \mathbb{C}((X \times A)^t \times X)$. The set $\mathcal{P}^\pi$ is endowed by the induced topology. From items (i) and (iii) of Assumption 2.1, it is easy to see that Conditions (1) and (2) of (W) in [16, Section 5] are satisfied. Therefore, according to Theorem 5.6 of [16], $\mathcal{P}^\pi$ is compact. The topology on $\mathcal{O}^\pi$ is defined as the finest topology for which the mapping $\mathbb{O}$ is continuous (the final topology on $\mathcal{O}^\pi$ associated to the mapping $\mathbb{O}$).

Before stating our first main result, we need the following technical lemma.

Lemma 4.1. *For any nonnegative lower-semicontinuous function r defined on $X \times A$ and any $R \in \mathbb{R}_+$, define the set $\mathcal{O}_{r,R}^\pi$ by*

$$\mathcal{O}_{r,R}^\pi = \{\mu^\pi \in \mathcal{O}^\pi \colon \mu^\pi(r) \leq R\}.$$

The set $\mathcal{O}_{r,R}^\pi$ is compact and the mapping $\mathbb{J} \colon \mathcal{O}^\pi \to \overline{\mathbb{R}}_+$ defined by $\mathbb{J}(\mu^\pi) = \mu^\pi(r)$ is lower semicontinuous.

Proof. Let us show that the set $\mathcal{P}_R^\pi = \{\mathrm{P}_\nu^\pi \in \mathcal{P}^\pi \colon \sum_{t=0}^\infty \int_\Omega r(g_t(\omega)) \, \mathrm{dP}_\nu^\pi(\omega) \leq R\}$ is compact in the w-topology. Since r is lower semicontinuous and nonnegative, by the definition of the w-topology, the mappings $\mathbb{H}_n \colon \mathcal{P}^\pi \to \overline{\mathbb{R}}_+$ defined by

$$\mathbb{H}_n(P_\nu^\pi) = \sum_{t=0}^n \int_\Omega r(g_t(\omega)) \, \mathrm{dP}_\nu^\pi(\omega)$$

are lower semicontinuous and so the mapping $\mathbb{H} \colon \mathcal{P}^\pi \to \overline{\mathbb{R}}_+$ defined by

$$\mathbb{H}(P_\nu^\pi) = \sum_{t=0}^\infty \int_\Omega r(g_t(\omega)) \, \mathrm{dP}_\nu^\pi(\omega)$$

is lower semicontinuous because $r \geq 0$. However, $\mathcal{P}_R^\pi = \mathbb{H}^{-1}([0, R])$, and so it is closed and compact. Since $\mathbb{O}(\mathcal{P}_R^\pi) = \mathcal{O}_{r,R}^\pi$, the set $\mathcal{O}_{r,R}^\pi$ is compact as a continuous image of a compact set, showing the first part of the result.

Now, observe that $\mathbb{J} \circ \mathbb{O} = \mathbb{H}$. Consequently, for any $M \in \mathbb{R}_+$, the set $\mathbb{O}^{-1}(\mathbb{J}^{-1}((M, \infty]))$ is an open set of $\mathcal{P}^\pi$ and so $\mathbb{J}^{-1}((M, \infty])$ is open in the topology of $\mathcal{O}^\pi$, showing that $\mathbb{J}$ is lower semicontinuous, giving the last part of the result.

The following hypothesis states that there exists an admissible solution for the LP with finite cost.

Assumption 4.1. *There exists a measure* $\overline{\mu} \in \mathbb{L}$ *such that* $\overline{\mu}(r_0) < \infty$.

In the next theorem we show that the LP is solvable, leading to the existence of an optimal randomized stationary policy for the constrained control problem (2.1)–(2.3).

Theorem 4.1. *Under Assumptions 2.1 and 4.1, there exists a randomized stationary policy* $\varphi^* \in \Pi_{\mathrm{c}}$ *such that*

$$\inf_{\gamma \in \mathbb{L}} \gamma(r_0) = \mu^{\varphi^*}(r_0) = \inf_{\pi \in \Pi_{\mathrm{c}}} v(\nu, \pi) = v(\nu, \varphi^*).$$

Proof. According to Assumption 4.1, the constant $R_0 = \overline{\mu}(r_0)$ is finite. Therefore, we clearly have

$$\inf_{\gamma \in \mathbb{L}} \gamma(r_0) \le \inf_{\gamma \in \bigcap_{n=0}^{q} \mathcal{O}^{\pi}_{r_n, R_n}} \gamma(r_0), \tag{4.1}$$

since $\bigcap_{n=0}^{q} \mathcal{O}^{\pi}_{r_n, R_n} \subset \mathbb{L}$. Let γ be any measure in $\mathbb{L}$ satisfying $\gamma(r_0) \le R_0$. Applying Corollary 3.1 with $r = \sum_{n=0}^{q} r_n$ and $R = \sum_{n=0}^{q} R_n$, it follows that there exists a randomized stationary policy φ^γ such that

$$\mathrm{E}_{\nu}^{\varphi^\gamma}\left[\sum_{t=0}^{\infty} r_0(x_t, a_t)\right] = \mu^{\varphi^\gamma}(r_0) \le \gamma(r_0) \quad \text{and} \quad \mathrm{E}_{\nu}^{\varphi^\gamma}\left[\sum_{t=0}^{\infty} r_n(x_t, a_t)\right] = \mu^{\varphi^\gamma}(r_n) \le \gamma(r_n)$$

for all $n \in \mathbb{N}_q$. Therefore, it follows that $\bigcap_{n=0}^{q} \mathcal{O}^{\pi}_{r_n, R_n}$ is a nonempty set and

$$\inf_{\gamma \in \mathbb{L}} \gamma(r_0) \ge \inf_{\gamma \in \bigcap_{n=0}^{q} \mathcal{O}^{\pi}_{r_n, R_n}} \gamma(r_0). \tag{4.2}$$

Now, by using the same arguments as in the proof of Lemma 4.1, $\bigcap_{n=0}^{q} \mathcal{O}^{\pi}_{r_n, R_n}$ is nonempty and compact, and the mapping $\mathbb{J}_0 \colon \mathcal{O}^{\pi} \to \mathbb{R}_+$ given by $\mathbb{J}_0(\mu^\pi) = \mu^\pi(r_0)$ is lower semicontinuous. Consequently, there exists $\pi^* \in \Pi$ such that $\mu^{\pi^*} \in \bigcap_{n=0}^{q} \mathcal{O}^{\pi}_{r_n, R_n}$ and

$$\inf_{\mu \in \bigcap_{n=0}^{q} \mathcal{O}^{\pi}_{r_n, R_n}} \mu(r_0) = \mu^{\pi^*}(r_0). \tag{4.3}$$

We have $\inf_{\pi \in \Pi_{\mathrm{c}}} v(\nu, \pi) = \inf_{\mu \in \bigcap_{n=0}^{q} \mathcal{O}^{\pi}_{r_n, R_n}} \mu(r_0)$ and so, combining (4.1)–(4.3), we obtain

$$\inf_{\gamma \in \mathbb{L}} \gamma(r_0) = \inf_{\pi \in \Pi_{\mathrm{c}}} v(\nu, \pi) = \mu^{\pi^*}(r_0). \tag{4.4}$$

Again, by using Corollary 3.1 with $r = \sum_{n=0}^{q} r_n$ and $R = \sum_{n=0}^{q} R_n$, there exists a randomized stationary policy $\varphi^* \in \Pi$ such that

$$\mathrm{E}_{\nu}^{\varphi^*}\left[\sum_{t=0}^{\infty} r_0(x_t, a_t)\right] = \mu^{\varphi^*}(r_0) \le \mu^{\pi^*}(r_0) \quad \text{and} \quad \mathrm{E}_{\nu}^{\varphi^*}\left[\sum_{t=0}^{\infty} r_n(x_t, a_t)\right] = \mu^{\varphi^*}(r_n) \le \mu^{\pi^*}(r_n)$$

for all $n \in \mathbb{N}_q$, implying that $\varphi^* \in \Pi_{\mathrm{c}}$ and, by using (4.4),

$$\inf_{\gamma \in \mathbb{L}} \gamma(r_0) = \inf_{\pi \in \Pi_{\mathrm{c}}} v(\nu, \pi) = \mu^{\varphi^*}(r_0) = v(\nu, \varphi^*),$$

proving the result.

Corollary 4.1. *The set of randomized stationary policies is a sufficient set of policies for the optimization problem (2.1)–(2.3).*

Proof. This result is a straightforward consequence of Theorem 4.1.

Condition 4.1. (Slater condition.) *There exists a control policy $\overline{\pi}$ such that, for all $n \in \mathbb{N}_q$, $v(\nu, \overline{\pi}) < \infty$ and $v_n(\nu, \overline{\pi}) < R_n$.*

Clearly, this condition implies Assumptions 3.1 and 4.1. Let us introduce the concept of a strictly active constraint.

Definition 4.1. For $n \in \mathbb{N}_q$, the nth constraint, that is, $\mu(r_n) \leq R_n$, is called strictly active if, upon excluding this constraint, the minimal value of $\mu(r_0)$ in the LP becomes strictly smaller.

Before presenting our second main result, we need the following technical result.

Lemma 4.2. *Suppose that the Slater condition is satisfied. Then the measure $\mu^* \in \mathbb{L}$ solves the LP if and only if there exists a vector of (optimal) Lagrange multipliers $\lambda^* \in \mathbb{R}^q_+$ such that*

$$\sum_{n=1}^{q} \lambda_n^* (\mu^*(r_n) - R_n) = 0$$

and

$$\mu^*(r_0) + \sum_{n=1}^{q} \lambda_n^* (\mu^*(r_n) - R_n) = \min_{\mu \in \{\gamma \in \mathbb{M}(X \times A)_+ \,:\, \gamma_A = \nu + \gamma Q\}} \left\{ \mu(r_0) + \sum_{n=1}^{q} \lambda_n^* (\mu(r_n) - R_n) \right\}.$$

Proof. Introduce the set

$$D = \{(\mu(r_0), \ldots, \mu(r_q)),\ \mu \in \mathbb{M}(X \times A)_+,\ \mu_A = \nu + \mu Q\} \cap \mathbb{R}^{q+1}_+.$$

Clearly, this set is convex and the LP can be rewritten as

$$\text{minimize } d_0 \text{ subject to } (d_0, \ldots, d_q) \in D \text{ and } d_n - R_n \leq 0 \text{ for all } n \in \mathbb{N}_q. \tag{4.5}$$

According to the Kuhn–Tucker theorem (see [12, Proposition 4, Chapter 11] or [15, Section 28]), a vector $d^* \in D$ solves the convex program (4.5) if and only if $d_n^* \leq R_n$ for all $n \in \mathbb{N}_q$ and there is a vector $\lambda^* \in \mathbb{R}^q_+$ such that

$$\sum_{n=1}^{q} \lambda_n^* (d_n^* - R_n) = 0$$

and

$$d_0^* + \sum_{n=1}^{q} \lambda_n^* (d_n^* - R_n) = \min_{d \in D} \left\{ d_0 + \sum_{n=1}^{q} \lambda_n^* (d_n - R_n) \right\},$$

proving the result.

The following theorem states that any optimal solution of the LP on the complement of a special subset V_r coincides with an optimal occupation measure generated by the induced randomized stationary policy.

Theorem 4.2. *Suppose that the Slater condition is satisfied and that all the constraints are strictly active. Let μ^* be an optimal solution to the LP, and let $\varphi^* \in \Phi$ be the stationary policy induced by μ^* and $r = \sum_{n=0}^{q} r_n$. Then μ^* and μ^{φ^*} coincide on $V_r^c \times A$.*

Proof. Note that $\lambda_n^* > 0$ for any $n \in \mathbb{N}_q$. Indeed, if $\lambda_n^* = 0$ for some n then, according to Lemma 4.2, μ^* is a solution to the LP excluding the nth constraint, so the minimal value of $\mu(r_0)$ is equal to $\mu^*(r_0)$, in contradiction to the fact that all the constraints are active. Now, define $\widetilde{r} = r_0 + \sum_{n=1}^{q} \lambda_n^* r_n$ and $\widetilde{R} = R_0 + \sum_{n=1}^{q} \lambda_n^* R_n$. Note that $V_r = V_{\widetilde{r}}$. Indeed, there exist constants $0 < e_1 < e_2 < \infty$ such that $e_1\widetilde{r} < r < e_2\widetilde{r}$. Consequently, it is easy to show that, by definition (see Proposition 3.2), $V_r \subset V_{e_1\widetilde{r}}$ and $V_{e_2\widetilde{r}} \subset V_r$. Moreover, we have $V_{\widetilde{r}} = V_{e_1\widetilde{r}} = V_{e_2\widetilde{r}}$ and so $V_r = V_{\widetilde{r}}$. Since $V_r = V_{\widetilde{r}}$, the stationary policy induced by μ^* and $\widetilde{r}$ is given by φ^*; see Definition 3.1 and the associated construction.

According to Theorem 3.3, $\mu_A^* \geq \mu_A^{\varphi^*}$ on $V_{\widetilde{r}}^c$. Suppose by contradiction that $\mu_A^*(V_{\widetilde{r}}^c) > \mu_A^{\varphi^*}(V_{\widetilde{r}}^c)$. Since φ^* is the stationary control policy induced by μ^* and $\widetilde{r}$, $\mu^*(\mathrm{d}x \times \mathrm{d}a) = \varphi^*(\mathrm{d}a \mid x)\mu_A^*(\mathrm{d}x)$ on $V_{\widetilde{r}}^c \times A$ and so the measure μ^* satisfies, for any $\Gamma \in \mathcal{B}(V_{\widetilde{r}}^c)$,

$$\begin{aligned}\mu_A^*(\Gamma) &= \nu(\Gamma) + \mu^* Q(\Gamma) \\ &= \nu(\Gamma) + \int_{V_{\widetilde{r}}^c} \int_A Q(\Gamma \mid x, a)\varphi^*(\mathrm{d}a \mid x)\mu_A^*(\mathrm{d}x) + \int_{V_{\widetilde{r}} \times A} Q(\Gamma \mid x, a)\mu^*(\mathrm{d}x \times \mathrm{d}a).\end{aligned}$$

By the definition of φ^*, we have $\int_A Q(\Gamma \mid x, a)\varphi^*(\mathrm{d}a \mid x) = Q(\Gamma \mid x, f^*(x))$ for any $x \in V_{\widetilde{r}}$. For any $\Gamma \in \mathcal{B}(V_{\widetilde{r}}^c)$ and $x \in V_{\widetilde{r}}$, we have, from the proof of Proposition 3.2, $Q(\Gamma \mid x, f^*(x)) = 0$ and so

$$\int_{V_{\widetilde{r}}} \int_A Q(\Gamma \mid x, a)\varphi^*(\mathrm{d}a \mid x)\mu_A^{\varphi^*}(\mathrm{d}x) = 0.$$

Consequently, the measure $\mu_A^{\varphi^*}$ can be written as

$$\mu_A^{\varphi^*}(\Gamma) = \nu(\Gamma) + \int_{V_{\widetilde{r}}^c} \int_A Q(\Gamma \mid x, a)\varphi^*(\mathrm{d}a \mid x)\mu_A^{\varphi^*}(\mathrm{d}x)$$

for any $\Gamma \in \mathcal{B}(V_{\widetilde{r}}^c)$. Thus, the measure γ defined by $\gamma(\Lambda) = \mu^*(\Lambda) - \mu^{\varphi^*}(\Lambda)$ for any $\Lambda \in \mathcal{B}(V_{\widetilde{r}}^c \times A)$ belongs to $\mathbb{M}(V_{\widetilde{r}}^c \times A)_+$ and satisfies $\gamma_A \geq \gamma Q$ with $\gamma_A(V_{\widetilde{r}}^c) > 0$. Therefore, applying Lemma 3.5, we obtain

$$\int_{V_{\widetilde{r}}^c} \int_A \widetilde{r}(x, a)\mu^*(\mathrm{d}x \times \mathrm{d}a) > \int_{V_{\widetilde{r}}^c} \int_A \widetilde{r}(x, a)\mu^{\varphi^*}(\mathrm{d}x \times \mathrm{d}a),$$

but recalling that $\widetilde{r}(x, f^*(x)) = 0$ for any $x \in V_{\widetilde{r}}$, we have

$$\int_{V_{\widetilde{r}}} \int_A \widetilde{r}(x, a)\mu^{\varphi^*}(\mathrm{d}x \times \mathrm{d}a) = \int_{V_{\widetilde{r}}} \int_A \widetilde{r}(x, f^*(x))\mu_A^{\varphi^*}(\mathrm{d}x) = 0,$$

implying that $\mu^*(\widetilde{r}) > \mu^{\varphi^*}(\widetilde{r})$. However, for any $\mu \in \mathbb{M}(X \times A)_+$, $\mu(\widetilde{r}) = \mu(r_0) + \sum_{n=1}^{q} \lambda_n^* \mu(r_n)$ and, according to Lemma 4.2,

$$\mu^*(\widetilde{r}) = \min_{\mu \in \{\gamma \in \mathbb{M}(X \times A)_+ :\ \gamma_A = \nu + \gamma Q\}} \mu(\widetilde{r}),$$

leading to a contradiction. This shows that $\mu_A^* = \mu_A^{\varphi^*}$ on $V_{\widetilde{r}}^c$, and the measures μ^* and μ^{φ^*} coincide on $V_{\widetilde{r}}^c \times A$ owing to Theorem 3.3.

Remark 4.1. If all the constraints are strictly active then $\lambda_n^* > 0$ for all $n \in \mathbb{N}_q$; however, the opposite does not hold. Theorem 4.2 remains valid if $\lambda_n^* > 0$ for all $n \in \mathbb{N}_q$ and all the constraints are not strictly active.

5. Examples

In this section we provide examples to illustrate some technical issues and pathologies. In the first example we show that if the running cost r_0 can take negative values then the convex analytic approach to nonfinite models becomes problematic and the solutions to the LP can have no meaning.

Example 5.1. We consider an unconstrained ($q = 0$) and uncontrolled model ($A = \{a\}$, dummy action). The state space is given by $X = \{0, 1, 2, \ldots\}$, and the transition kernel is defined by $Q(0 \mid 0, a) = 1$ and $Q(i - 1 \mid i, a) = 1$ for $i \geq 1$, and

$$r_0(x, a) = \begin{cases} -1 & \text{if } x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\nu(1) = 1$. Obviously, $v(\nu, \pi) = -1$ and the occupation measure (for the unique control policy) is given by $\mu^\pi(1, a) = 1$, $\mu^\pi(0, a) = +\infty$, and $\mu^\pi(x, a) = 0$ for $x \geq 2$. Thus, μ^π is obviously admissible for the LP. However, for any $d \geq 0$, the measure

$$\mu(1, a) = 1 + d, \qquad \mu(0, a) = +\infty, \qquad \mu(x, a) = d \quad \text{for } x > 1$$

is also admissible, resulting in the value $\mu(r_0) = -(1 + d)$. Thus, the solution to the LP corresponds to $d = +\infty$, and the minimal value of $\mu(r_0)$ equals $-\infty$. Observe that the measures μ with $d > 0$ do not correspond to any control policy. In the case $r_0 \geq 0$, such measures are ignored as they do not solve the LP.

In the next example we confirm that the requirement in Theorem 4.2, i.e. that all the constraints are strictly active, is important.

Example 5.2. Let $X = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, and let $A = \{a\}$, a dummy action, meaning that the process is in fact not controlled. The transition kernel is given by $Q(i + 1 \mid i, a) = 1$ for $i < 0$, $Q(j - 1 \mid j, a) = 1$ for $j > 0$, and $Q(0 \mid 0, a) = 1$. We consider the case with one constraint ($q = 1$): $r_0(j, a) = \mathbf{1}_{\{j=1\}}$, $r_1(i, a) = \mathbf{1}_{\{i=-1\}}$, $R_1 = 2$. Finally, the initial distribution is defined by $\nu(-1) = \nu(1) = \frac{1}{2}$.

By defining $r = r_0 + r_1$, we have $V_r = \{0\}$ since $r^*(0) = 0$ and $r^*(x) = 1$ for all $x \neq 0$. The (single) control policy π is admissible: $v_1(\nu, \pi) = \frac{1}{2} < 2$; $V^*(\nu) = \frac{1}{2}$. The corresponding occupation measure μ^π is as follows:

$$\mu^\pi(i, a) = 0 \quad \text{for all } i \geq 2 \text{ and } i \leq -2,$$
$$\mu^\pi(1, a) = \mu^\pi(-1, a) = \tfrac{1}{2}, \qquad \mu^\pi(0, a) = \infty.$$

On $X \times A$, this measure obviously solves the LP. Since $\mu^\pi(r_1) = \frac{1}{2} < R_1 = 2$, the Lagrange multiplier $\lambda_1^* = 0$ and the conditions of Theorem 4.2 are violated. As a result, there exist other solutions to the LP which do not correspond to any control policies. Indeed, let

$$\mu(i, a) = d > 0 \quad \text{for } i \leq -2, \qquad \mu(i, a) = 0 \quad \text{for } i \geq 2,$$
$$\mu(1, a) = \tfrac{1}{2}, \qquad \mu(-1, a) = d + \tfrac{1}{2}, \qquad \mu(0, a) = \infty,$$

where $d > 0$ is an arbitrary number. This measure satisfies $\mu(r_0) = \frac{1}{2}$ and $\mu(r_1) = \frac{1}{2} + d$, so the constraint $\mu(r_1) \le R_1$ is satisfied if $d \le \frac{3}{2}$. However, the induced policy φ^* coincides with the unique control policy π in this model, and $\mu \ne \mu^{\varphi^*}$ on $V_r^c \times A$ if $d > 0$ because $\mu^{\varphi^*}(i, a) = 0$ for all $i \le -2$ and $\mu^{\varphi^*}(-1, a) = \frac{1}{2}$.

The following meaningful example describes the process of selling a property.

Example 5.3. Suppose that a landlord plans to sell the house and, once a month, receives offers from the random market, taking values in $\{1, 2, \dots, M\}$. Accepting offer i results in a loss of $f(i)$ units (e.g. a thousand pounds). Such losses are the result of not accepting a perfect offer. We assume that the offers change according to an (uncontrolled) Markov chain with transition matrix $P = (p_{ij})$, $i, j = 1, 2, \dots, M$. If a tenant is currently renting the house, the landlord is not allowed to sell it, but the tenant can leave before the next month with probability p_l. If there is no tenant and the landlord does not accept the current offer, the landlord can wait until the next month or invite a new tenant to rent the house by the next month with probability p_a. In either case, the landlord must pay a maintenance cost of $c \ge 0$, which is not applicable if a tenant is present. Finally, assume that the expected time for the whole selling period does not exceed a fixed constant R. The goal is to devise a selling policy that minimizes the total expected cost under the imposed time constraint. A similar example was solved in [2, Example 10.3.1] using the dynamic programming approach, but the problem was unconstrained and the authors considered the finite horizon case.

To formulate the MDP, we introduce the state space

$$X = \{(i, N), (j, Y),\ i, j = 1, 2, \dots, M\} \cup \{\Delta\},$$

where component i represents the current offer, and the letters N and Y respectively correspond to an untenanted property and a tenanted property. The state Δ means the house is sold. The action space is given by $A = \{s, t, w\}$, where s means 'accept the offer (sell the house)', t means 'invite a tenant', and w means 'wait'. The transition kernel is given by

$$Q(\Delta \mid \Delta, a) \equiv 1, \qquad Q(\Delta \mid (i, l), a) = \begin{cases} 1 & \text{if } l = N, a = s, \\ 0 & \text{otherwise,} \end{cases}$$

$$Q((j, k) \mid (i, l), a) = p_{ij} \cdot \begin{cases} p_l & \text{if } l = Y, k = N, \\ 1 - p_l & \text{if } l = Y, k = Y, \\ p_a & \text{if } l = N, a = t, k = Y, \\ 1 - p_a & \text{if } l = N, a = t, k = N, \\ 1 & \text{if } l = N, a = w, k = N, \\ 0 & \text{otherwise.} \end{cases}$$

The cost and constraint functions are defined by

$$r_0(\Delta, a) = r_0((i, Y), a) = 0,$$
$$r_0((i, l), a) = \begin{cases} c & \text{if } l = N, a \ne s, \\ f(i) & \text{if } l = N, a = s. \end{cases}$$
$$r_1(x, a) = \mathbf{1}_{\{x \ne \Delta\}}.$$

TABLE 1.

a	x									
	$(1, N)$	$(2, N)$	$(3, N)$	$(4, N)$	$(5, N)$	$(1, Y)$	$(2, Y)$	$(3, Y)$	$(4, Y)$	$(5, Y)$
s	0	0.219	0.622	0.131	0.028	0	0	0	0	0
t	2.350	1.650	0	0	0	0	0	0	0	0
w	0	0	0	0	0	1.628	2.304	0.830	0.196	0.042

Numerically solving the LP for $M = 5$ and $f(i) = 5 - i$, we obtain

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad p_l = \tfrac{2}{5}, \qquad p_a = \tfrac{1}{2}, \qquad c = \tfrac{1}{10}.$$

Suppose that $\nu((1, N)) = 1$, i.e. initially there is no tenant and the first offer is 1. In any case, for the optimal solution, $\mu_A^*(\Delta) = \infty$ and $\mu_A^*(x) < \infty$ for $x \neq \Delta$. If $R = 10$ then $\mu^*(r_1) = 10$ and $\mu^*(r_0) = 2.432$. The optimal policy is

- accept the offer if its value is 3, 4, or 5 and there is no tenant,
- if the value of the offer is 2 and there is no tenant, accept the offer with probability 0.12 or invite a tenant with the complementary probability 0.88,
- invite a tenant if the house is untenanted and the current offer is 1,
- wait otherwise.

The optimal occupation measure is presented in Table 1.

We note that this model, arising from a real-world situation, is not transient: for the policy 'never accept the offer', the expected time to the absorbtion at cemetery state Δ equals $+\infty$. Of course, this solution is far from optimal, the corresponding occupation measure equals $+\infty$ at most of the state-action pairs, and all the performance functionals equal $+\infty$. The theory developed in [1] is not applicable here.

Acknowledgements

This research was partially supported by the Royal Society (grant number TG091905), the EPSRC (grant number EP/I001238/1), and the RCMM, University of Liverpool (grant number 2510-02PIU).

References

[1] ALTMAN, E. (1999). *Constrained Markov Decision Processes*. Chapman & Hall/CRC, Boca Raton, FL.
[2] BÄUERLE, N. AND RIEDER, U. (2011). *Markov Decision Processes with Applications to Finance*. Springer, Heidelberg.
[3] BERTSEKAS, D. P. (1987). *Dynamic Programming*. Prentice Hall, Englewood Cliffs, NJ.
[4] BERTSEKAS, D. P. AND SHREVE, S. E. (1978). *Stochastic Optimal Control* (Math. Sci. Eng. **139**). Academic Press, New York.
[5] BORKAR, V. S. (2002). Convex analytic methods in Markov decision processes. In *Handbook of Markov Decision Processes* (Internat. Ser. Operat. Res. Manag. **40**), Kluwer, Boston, MA, pp. 347–375.

[6] Dufour, F. and Piunovskiy, A. B. (2010). Multiobjective stopping problem for discrete-time Markov processes: convex analytic approach. *J. Appl. Prob.* **47,** 947–966.
[7] Feinberg, E. A. (2002). Total reward criteria. In *Handbook of Markov Decision Processes* (Internat. Ser. Operat. Res. Manag. **40**), Kluwer, Boston, MA, pp. 173–207.
[8] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes* (Appl. Math. **30**). Springer, New York.
[9] Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes* (Appl. Math. **42**). Springer, New York.
[10] Horiguchi, M. (2001). Markov decision processes with a stopping time constraint. *Math. Meth. Operat. Res.* **53,** 279–295.
[11] Horiguchi, M. (2001). Stopped Markov decision processes with multiple constraints. *Math. Meth. Operat. Res.* **54,** 455–469.
[12] Luenberger, D. G. and Ye, Y. (2010). *Linear and Nonlinear Programming* (Internat. Ser. Operat. Res. Manag. Sci. **116**), 3rd edn. Springer, New York.
[13] Piunovskiy, A. B. (1997). *Optimal Control of Random Sequences in Problems with Constraints* (Math. Appl. **410**). Kluwer, Dordrecht.
[14] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.
[15] Rockafellar, R. T. (1970). *Convex Analysis* (Princeton Math. Ser. **28**). Princeton University Press.
[16] Schäl, M. (1975). On dynamic programming: compactness of the space of policies. *Stoch. Process. Appl.* **3,** 345–364.