Adv. Appl. Prob. **45,** 837–859 (2013)
Printed in Northern Ireland
© *Applied Probability Trust* 2013

THE EXPECTED TOTAL COST CRITERION FOR MARKOV DECISION PROCESSES UNDER CONSTRAINTS

FRANÇOIS DUFOUR,* *Université Bordeaux, IMB and INRIA Bordeaux Sud-Ouest*
A. B. PIUNOVSKIY,** *University of Liverpool*

Abstract

In this work, we study discrete-time Markov decision processes (MDPs) with constraints when all the objectives have the same form of expected total cost over the infinite time horizon. Our objective is to analyze this problem by using the linear programming approach. Under some technical hypotheses, it is shown that if there exists an optimal solution for the associated linear program then there exists a randomized stationary policy which is optimal for the MDP, and that the optimal value of the linear program coincides with the optimal value of the constrained control problem. A second important result states that the set of randomized stationary policies provides a sufficient set for solving this MDP. It is important to note that, in contrast with the classical results of the literature, we do not assume the MDP to be transient or absorbing. More importantly, we do not impose the cost functions to be nonnegative or to be bounded below. Several examples are presented to illustrate our results.

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

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

1. Introduction

We consider a Markov decision process (MDP) with constraints under the total expected cost optimality criterion. MDPs are a general family of controlled stochastic processes, which are suitable for the modeling of several sequential decision-making problems under uncertainty. They arise in many applications, such as engineering, computer science, telecommunications, finance, among others. From a theoretical point of view, discrete-time MDPs have been extensively studied (see, e.g. [1], [3], [4], [8], [9], [10], [13], [14], and [15]) by using several techniques such as the dynamic programming and the linear programming approaches, among others. The latter technique has proved to be a very powerful tool for solving MDPs with constraints. A significant list of references on this research field can be found in [1], [5], [13], and the references therein. The key idea is to reformulate the original sequential decision-making problem as an infinite-dimensional static optimization problem over a space of measures where the admissible solutions are the so-called expected state-action frequencies or occupation measures of the controlled process. The significance and the attraction of this method lie mainly

Received 12 June 2012; revision received 26 October 2012.
* Postal address: Team CQFD, INRIA Bordeaux Sud-Ouest, 200 Avenue de la Vieille Tour, 33405 Talence cedex, France. Email address: dufour@math.u-bordeaux1.fr
** Postal address: Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 7ZL, UK. Email address: piunov@liverpool.ac.uk

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

in the following two points. Firstly, it gives an alternative approach to study theoretically and to solve numerically MDPs. Secondly, control problems with constraints can be easily analyzed with this method in contrast to the dynamic programming technique.

The linear programming approach for MDPs has been studied in great detail under a large variety of cost criteria and fairly general state and action spaces. These include the finite horizon cost, the infinite time horizon discounted cost, the cost up to exit time, the ergodic pathwise cost. However, it must be emphasized that the linear programming technique when applied to the expected total cost (ETC) criterion has received less attention in the literature, the reason being that the ETC criterion is very demanding from a technical point of view and yields some pathological behaviors as pointed out in [5, pp. 357–358] and [10, pp. 92–94]. These latter difficulties are twofold. The first important issue is that the expected state-action frequency may not be finite for some or all policies. The second point is related to the fact that an admissible solution of the linear program (LP) may not correspond to any expected state-action frequency of the process. The classical way used in the literature to overcome the first point is to impose suitable conditions on the model as the so-called transient and absorbing conditions. Roughly speaking, these conditions ensure that, for any stationary policy, the controlled process is absorbed at a particular state guaranteeing that the expected state-action frequency is finite for all other states and stationary policies. However, 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.

In the current paper, we investigate the ETC criterion with constraints by using the linear programming formulation. We work under the hypotheses that the state space is a general Borel space and the action space is a compact metric space. Moreover, it is assumed that the stochastic kernel and the cost functions are continuous in the action variable and that the value of the unconstrained LP associated to each cost function is finite. Finally, our last hypothesis consists of ensuring the existence of a reference probability measure with respect to which the stochastic kernel is absolutely continuous.

One of our main results is to show that if μ^* is an optimal solution of the constrained LP then there exists a randomized stationary policy φ^* which is optimal for the MDP under consideration and that the optimal value of the LP coincides with the optimal value of the constrained control problem (see Theorems 4.2 and 5.2). A second important result states that the set of randomized stationary policies provides a sufficient set for solving this optimization problem (see Corollaries 4.1 and 5.2). Similar results have been obtained in [7] in a considerably simpler framework where the cost functions are assumed to be nonnegative. Clearly, the approach developed in [7] strongly relies on the positiveness of the cost functions. In this work we deal with general cost functions that may take negative values. This general context induces important technical difficulties that cannot be dealt with using the technique presented in [7]. This issue imposes the development of a new and radically different approach to deal with our general framework.

It is important to emphasize that our assumptions are very weak and do not exclude the two issues described before. In particular, we do not impose the MDP to be transient or absorbing. The (optimal) occupation measures of the constrained control problem are not necessarily finite and an admissible solution of the LP may not correspond to any occupation measure of the controlled process. More importantly, we do not require the cost functions to be nonnegative or bounded below. Several examples presented in Section 6 will illustrate the necessity of the hypotheses that have been made. One example is dedicated to the presentation of a possible application of the results developed in the paper. To the best of our knowledge, the book [1]

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

is the only reference in the literature that analyses the ETC criterion with constraints by using the linear programming formulation under the hypotheses that the state and action spaces are discrete and the model is transient or absorbing. One can also mention the references [6], [7], [11], and [12], where MDPs under the assumption that the cost function is nonnegative are studied. Consequently, our results appear to be very general compared to those existing in the literature.

Actually, for the sake of clarity of exposition, we consider in two independent sections the case of a finite action space and the case of a compact action space. Although the results derived in the context of a finite action space can be obtained from the more general case where the action space is compact, we believe that the section dedicated to the finite action space is of interest by itself mainly because the proofs are simpler, more natural, and more intuitive.

The rest of the paper is organized as follows. In Section 2 we present the control problem that will be considered throughout this work. Preliminary results are derived in Section 3. Our main results and the analysis of the LP are presented in Section 4 in the case of a finite action space, and in Section 5 in the general case of a compact action space. Finally, Section 6 is dedicated to the presentation of several examples illustrating some theoretical issues if our set of assumptions is not satisfied and to the presentation of a possible application of the results developed in the paper.

2. Problem formulation

The main goal of this section is to introduce the notation, definitions, and problem formulation that will be used throughout the paper. The following notation will be used in this paper: $\mathbb{N}$ is 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 $\bar{\mathbb{R}}_+$ denotes $\mathbb{R}_+ \cup \{+\infty\}$. For any $q \in \mathbb{N}$, $\mathbb{N}_q$ is the set $\{1, \ldots, q\}$. If $(B_\lambda)_{\lambda \in \Lambda}$ is a family of sets, we write $\sum_{\lambda \in \Lambda} B_\lambda$ instead of $\bigcup_{\lambda \in \Lambda} B_\lambda$ to indicate that $B_\lambda \cap B_{\lambda'} = \varnothing$ for any $\lambda \neq \lambda'$. Let A be an arbitrary set, and let B and C be subsets of A. Then the difference of the sets B and C is denoted by $B \setminus C = B \cap C^c$ and the symmetric difference is denoted by $B \triangle C = (B \setminus C) \cup (C \setminus B)$. Let f be a real-valued mapping defined on a set E, the positive (respectively negative) part of f is given by $f^+(x) = \max(0, f(x))$ (respectively $f^-(x) = \max(0, -f(x))$). The term *measure* will always refer to a countably additive, $\bar{\mathbb{R}}_+$-valued set function. Let X be a Borel space, and denote by $\mathcal{B}(X)$ its associated Borel σ-algebra. For any set A, I_A denotes the indicator function of the set A. The set of measures defined on $(X, \mathcal{B}(X))$ is denoted by $\mathbb{M}(X)_+$. For any point $x \in X$, δ_x denotes the Dirac measure defined by $\delta_x(\Gamma) = I_\Gamma(x)$ for any $\Gamma \in \mathcal{B}(X)$. For two measures $(\gamma_1, \gamma_2) \in \mathbb{M}(X)_+^2$, $\gamma_1 \leq \gamma_2$ means that $\gamma_1(\Gamma) \leq \gamma_2(\Gamma)$ for any $\Gamma \in \mathcal{B}(X)$. 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 X is denoted by $\mathbb{C}(X)$. Let f be a real-valued measurable function defined on X, and let $\eta \in \mathbb{M}(X)_+$; the integral $\int_X f(y)\eta(\mathrm{d}y)$ 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 $\int_X f(x)W(\mathrm{d}x \mid y)$ for any $y \in Y$ is denoted by $Wf(y)$ provided it is well defined. For any measure η on $(Y, \mathcal{B}(Y))$, ηW is the measure defined on $(X, \mathcal{B}(X))$ by $\eta W(\Gamma) = \int_Y W(\Gamma \mid y)\eta(\mathrm{d}y)$ for any $\Gamma \in \mathcal{B}(X)$. Let μ be a measure in $\mathbb{M}(X \times Y)_+$ and $\Lambda \in \mathcal{B}(Y)$. Then μ_Λ denotes the measure defined on X by $\mu_\Lambda(\Gamma) = \mu(\Gamma \times \Lambda)$ for any $\Gamma \in \mathcal{B}(X)$. If $\Lambda = \{y\}$ then we will write μ_y instead of $\mu_{\{y\}}$.

Definition 2.1. Let X be a Borel space. If μ and ν are σ-finite measures in $\mathbb{M}(X)_+$ such that $\mu \leq \nu$ then $(\nu - \mu)$ is the σ-finite measure in $\mathbb{M}(X)_+$ defined for any $\Gamma \in \mathcal{B}(X)$ by

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

$(\nu - \mu)(\Gamma) = \sum_{k=1}^{\infty}[\nu(\Gamma \cap B_k) - \mu(\Gamma \cap B_k)]$, where $X = \sum_{k=1}^{\infty} B_k$ and $\nu(B_k) < \infty$ (this definition does not depend on the choice of $(B_k)_{k\in\mathbb{N}}$).

In order to define an MDP, we consider, as in Section 2 of [9], a five-tuple (X, A, Q, r) consisting of

(a) a Borel space X representing the state space,

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

(c) a stochastic kernel Q on X given $X \times A$,

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

(e) measurable functions $r_\theta \colon X \times A \to \mathbb{R}$ representing the constraints for $\theta \in \Theta$, where Θ is an arbitrary set (without loss of generality, assume that $0 \notin \Theta$).

Definition 2.2. 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 introduce different classes of control policies.

Definition 2.3. 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 $\pi = \{\pi_t\}$ 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 $\mathbb{F}$ (respectively Φ) with the class of all deterministic (respectively randomized) stationary policies. Therefore, $\mathbb{F} \subset \Phi \subset \Pi$. If π is a randomized stationary policy generated by $\varphi \in \Phi$ (according to Definition 2.3), 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, it can be defined a probability, labeled $\mathbb{P}_\nu^\pi$, 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 $t \in \mathbb{N}^0$, $\mathbb{P}_\nu^\pi(x_0 \in B) = \nu(B)$, $\mathbb{P}_\nu^\pi(a_t \in C \mid h_t) = \pi_t(C \mid h_t)$, and $\mathbb{P}_\nu^\pi(x_{t+1} \in B \mid h_t, a_t) = Q(B \mid x_t, a_t)$; see, for example, [9, Chapter 2] for such a construction. The expectation with respect to $\mathbb{P}_\nu^\pi$ is denoted by $\mathbb{E}_\nu^\pi$. If $\nu = \delta_x$ for $x \in X$, we write $\mathbb{P}_x^\pi$ for $\mathbb{P}_\nu^\pi$ and $\mathbb{E}_x^\pi$ for $\mathbb{E}_\nu^\pi$.

Next, we define our Markov control problem. Suppose that we are given an initial distribution ν on X, and constraint limits $(R_\theta)_{\theta\in\Theta}$ such that $R_\theta \in \mathbb{R}$. For any $\theta \in \Theta \cup \{0\}$, the integral $\mathbb{E}_\nu^\pi[\sum_{t=0}^{\infty} r_\theta(x_t, a_t)]$ is said to be well defined if either $\mathbb{E}_\nu^\pi[\sum_{t=0}^{\infty} r_\theta^+(x_t, a_t)]$ or

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

$\mathbb{E}_\nu^\pi[\sum_{t=0}^\infty r_\theta^-(x_t,a_t)]$ is finite. A policy $\pi \in \Pi$ is said to be feasible if the integrals $\mathbb{E}_\nu^\pi[\sum_{t=0}^\infty r_\theta(x_t,a_t)]$ are well defined for any $\theta \in \Theta \cup \{0\}$ and if

$$v_\theta(\nu,\pi) := \mathbb{E}_\nu^\pi\left[\sum_{t=0}^\infty r_\theta(x_t,a_t)\right] \le R_\theta \tag{2.1}$$

for any $\theta \in \Theta$. The set of feasible policies is denoted by Π_c.

The optimization problem we consider consists in minimizing the cost function

$$v(\nu,\pi) := \mathbb{E}_\nu^\pi\left[\sum_{t=0}^\infty r_0(x_t,a_t)\right], \tag{2.2}$$

over the set of feasible control policies Π_c. The optimal value function is denoted by

$$V^*(\nu) := \inf_{\pi\in\Pi_c} v(\nu,\pi). \tag{2.3}$$

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

$$\mu^\pi(\Gamma) = \sum_{t=0}^\infty \mathbb{P}_\nu^\pi((x_t,a_t)\in\Gamma)$$

for any $\Gamma \in \mathcal{B}(X\times A)$. According to Lemma 9.4.3 of [10], for any stationary policy $\varphi \in \Phi$, μ_A^φ is a solution of the equation

$$\gamma(\cdot) = \nu(\cdot) + \int_{X\times A} Q(\cdot \mid y,a)\varphi(\mathrm{d}a \mid y)\gamma(\mathrm{d}y), \tag{2.4}$$

with $\gamma \in \mathbb{M}(X)_+$. Moreover, a careful inspection of the proof of Lemma 9.4.3 of [10] shows that μ_A^φ is the minimal positive solution of (2.4) and μ^φ is given by

$$\mu^\varphi(\Gamma\times\Lambda) = \mu_\Lambda^\varphi(\Gamma) = \int_\Gamma \varphi(\Lambda \mid y)\mu_A^\varphi(\mathrm{d}y), \tag{2.5}$$

and so μ^φ is a solution of the equation

$$\mu_A = \nu + \mu Q, \tag{2.6}$$

with $\mu \in \mathbb{M}(X\times A)_+$.

We now introduce the LP related to the control problem defined by (2.1)–(2.3).

Definition 2.4. The constrained linear program $\mathbb{LP}$ associated with the constrained control problem (2.3) is defined as

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

where

$$\mathbb{L}_c = \{\mu \in \mathbb{L}\colon \mu(r_\theta) \le R_\theta \text{ for } \theta \in \Theta\}$$

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

with $\mathbb{L} = \bigcap_{\theta \in \Theta \cup \{0\}} \mathbb{L}_\theta$ and

$$\mathbb{L}_\theta = \{\mu \in \mathbb{M}(X \times A)_+ : \mu_A = \nu + \mu Q \text{ and for which } \mu(r_\theta) \text{ is well defined}\}.$$

Associated to the constrained linear program $\mathbb{LP}$, we introduce the family of auxiliary unconstrained linear programs $\mathbb{LP}_\theta$ for $\theta \in \Theta \cup \{0\}$, as

$$(\mathbb{LP}_\theta) \begin{cases} \text{minimize } \mu(r_\theta) \\ \text{subject to } \mu \in \mathbb{L}_\theta. \end{cases}$$

A measure μ is said to be *admissible* for $\mathbb{LP}$ if $\mu \in \mathbb{L}_c$. A measure μ is said to be *optimal* for $\mathbb{LP}$ if μ is admissible and if $\mu(r_0) \leq \gamma(r_0)$ for any admissible γ. The value of the linear program $\mathbb{LP}$ (respectively $\mathbb{LP}_\theta$ for $\theta \in \Theta \cup \{0\}$) is said to be finite if $\inf_{\mu \in \mathbb{L}_c} \mu(r_0)$ (respectively $\inf_{\mu \in \mathbb{L}_\theta} \mu(r_\theta)$) is finite.

3. Preliminary results

In this section we establish some preliminary results to obtain a decomposition of the state space in order to ensure that, for any $\mu \in \mathbb{M}(X \times A)_+$ satisfying the linear equation (2.6), the measure μ_A is σ-finite on some subset of the space X. In particular, we introduce the concept of (nearly) perfect sets (see Definitions 3.1 and 3.2). Roughly speaking, for two arbitrary measures λ and μ on $\mathcal{B}(X)$, a λ/μ-perfect set W is a *maximal* set for the inclusion having the property that either $\lambda(\Gamma) = 0$ or $\mu(\Gamma) = \infty$ for any subset Γ of W. Several technical properties of these sets are derived (see Lemma 3.1, Proposition 3.1, and Theorem 3.1). An important result which is a cornerstone of the paper is presented at the end of this section. It states that, for any measure $\mu \in \mathbb{M}(X)_+$ and under the assumption that λ is a probability measure, there exists a λ/μ-perfect set having the property that μ is σ-finite on the complement of this set (see Theorem 3.2).

Definition 3.1. Consider two measures μ and λ on $\mathcal{B}(X)$. A set W in $\mathcal{B}(X)$ is said to be λ/μ-perfect if

(a) for any $\Gamma \in \mathcal{B}(W)$, either $\lambda(\Gamma) = 0$ or $\mu(\Gamma) = \infty$,

(b) for any set $V \in \mathcal{B}(X)$ such that $W \subset V$ and $\lambda(V \setminus W) > 0$, (a) is violated, that is, there exists $\Gamma \in \mathcal{B}(X)$ with $\Gamma \subset V$ such that $\lambda(\Gamma) > 0$ and $\mu(\Gamma) < \infty$.

Remark 3.1. If W is λ/μ-perfect then Definition 3.1(b) can be replaced by the following.

(b′) For any set $U \in \mathcal{B}(X)$ such that $\lambda(U \setminus W) > 0$, there exists $\Lambda \in \mathcal{B}(X)$ with $\Lambda \subset U \setminus W$ such that $\lambda(\Lambda) > 0$ and $\mu(\Lambda) < \infty$.

Indeed, if $U \in \mathcal{B}(X)$ is such that $\lambda(U \setminus W) > 0$ then the set $V = U \cup W$ satisfies $W \subset V$ and $V \setminus W = U \setminus W$, and so $\lambda(V \setminus W) > 0$. So, from (b), there exists $\Gamma \in \mathcal{B}(X)$ with $\Gamma \subset V$ such that $\lambda(\Gamma) > 0$ and $\mu(\Gamma) < \infty$. Introduce the set $\Lambda = \Gamma \cap [U \setminus W]$. Then $\mu(\Lambda) < \infty$ and $\lambda(\Lambda) > 0$. The last inequality comes from the fact that $\Gamma = \Lambda \cup [\Gamma \cap W]$ and so if $\lambda(\Lambda) = 0$ then $\lambda(\Gamma \cap W) > 0$ and since $\mu(\Gamma \cap W) < \infty$ that would lead to a contradiction with (a).

Definition 3.2. Consider two measures μ and λ on $\mathcal{B}(X)$. A set W in $\mathcal{B}(X)$ is said to be λ/μ-nearly perfect if, for any $\Gamma \in \mathcal{B}(W)$, either $\lambda(\Gamma) = 0$ or $\mu(\Gamma) = \infty$.

Remark 3.2. It is important to observe that if a set W in $\mathcal{B}(X)$ is a subset of a λ/μ-perfect set then it is λ/μ-nearly perfect.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Definition 3.3. Consider W and $\widetilde{W}$ in $\mathcal{B}(X)$. The sets W and $\widetilde{W}$ are said to be λ-equivalent if $\lambda(W \triangle \widetilde{W}) = 0$. This is denoted by $W \overset{\lambda}{\simeq} \widetilde{W}$.

Clearly, $W \overset{\lambda}{\simeq} \widetilde{W}$ if and only if $I_W = I_{\widetilde{W}}$, λ-a.e. and so the relation '$\overset{\lambda}{\simeq}$' is reflexive, symmetric, and transitive. Moreover, if $W_1 \overset{\lambda}{\simeq} \widetilde{W}_1$ and $W_2 \overset{\lambda}{\simeq} \widetilde{W}_2$, we have $W_1 \cup W_2 \overset{\lambda}{\simeq} \widetilde{W}_1 \cup \widetilde{W}_2$ and $W_1^c \overset{\lambda}{\simeq} \widetilde{W}_1^c$. The next three technical results present some properties of perfect sets that will be used repeatedly in the next sections.

Lemma 3.1. *Consider two measures μ and λ on $\mathcal{B}(X)$. Assume that W and $\widetilde{W}$ are λ/μ-perfect. Then $W \overset{\lambda}{\simeq} \widetilde{W}$.*

Proof. Suppose that $\lambda(\widetilde{W} \setminus W) > 0$. Then introduce the set $\widehat{W} = \widetilde{W} \cup W$. Note that $W \subset \widehat{W}$ and $\lambda(\widehat{W} \setminus W) > 0$. Since W is perfect, it follows that there exists $\Gamma \in \widehat{W}$ such that $\lambda(\Gamma) > 0$ and $\mu(\Gamma) < \infty$. Consequently, $\mu(\Gamma \cap W) < \infty$ and so $\lambda(\Gamma \cap W) = 0$ since W is perfect. However, $\lambda(\Gamma \cap \widetilde{W}) = \lambda(\Gamma \cap \widehat{W}) = \lambda(\Gamma) > 0$. This implies that $\widetilde{W}$ cannot be perfect because $\mu(\Gamma \cap \widetilde{W}) \leq \mu(\Gamma) < \infty$, leading to a contradiction and so $\lambda(\widetilde{W} \setminus W) = 0$. By symmetry we obtain $\lambda(W \setminus \widetilde{W}) = 0$ and the result follows.

Proposition 3.1. *Suppose that the set W is λ/μ-perfect. Then $\widetilde{W}$ is λ/μ-perfect if and only if $W \overset{\lambda}{\simeq} \widetilde{W}$.*

Proof. The *only if* part was proved in Lemma 3.1. Suppose that $W \overset{\lambda}{\simeq} \widetilde{W}$. Let us show that items (a) and (b) of Definition 3.1 are satisfied. Regarding item (a), assume that there exists $\Gamma \in \mathcal{B}(X)$ such that $\Gamma \subset \widetilde{W}$ with $\lambda(\Gamma) > 0$ and $\mu(\Gamma) < \infty$. Then $\mu(\Gamma \cap W) < \infty$. Moreover, $\lambda(\Gamma \cap W) = \lambda(\Gamma \cap \widetilde{W})$ since $W \overset{\lambda}{\simeq} \widetilde{W}$. Consequently, $\lambda(\Gamma \cap W) = \lambda(\Gamma) > 0$ implies that W is not λ/μ-perfect, leading to a contradiction.

In order to show item (b), let us proceed by contradiction. Suppose that there exists a set $\widehat{W} \in \mathcal{B}(X)$ with $\widetilde{W} \subset \widehat{W}$ and $\lambda(\widehat{W} \setminus \widetilde{W}) > 0$ such that, for all $\Gamma \in \widehat{W}$, either $\lambda(\Gamma) = 0$ or $\mu(\Gamma) = \infty$. Define $U = \widehat{W} \cup (W \triangle \widetilde{W})$. Clearly, $W \subset U$. Since $W \overset{\lambda}{\simeq} \widetilde{W}$, we have $\lambda(U) = \lambda(\widehat{W})$. Recalling that $\lambda(\widehat{W} \setminus \widetilde{W}) > 0$, it follows that $\lambda(U) > \lambda(\widetilde{W})$ and so $\lambda(U) > \lambda(W)$. Consequently, $\lambda(U \setminus W) > 0$. However, $\lambda([U \setminus W] \cap \widehat{W}) = \lambda(U \setminus W) > 0$; hence, $\mu([U \setminus W] \cap \widehat{W}) = \infty$, implying that $\mu(U \setminus W) = \infty$. This shows that W is not λ/μ-perfect, leading to a contradiction. This gives the result.

Theorem 3.1. *Suppose that $\mu = \mu_1 + \mu_2$ and W (respectively W_1, W_2) is a λ/μ-perfect (respectively λ/μ_1-perfect, λ/μ_2-perfect) set. Then $W_1 \cup W_2 \overset{\lambda}{\simeq} W$ and $W_1 \cup W_2$ is λ/μ-perfect.*

Proof. The proof of the first statement consists in showing that $\lambda([W_1 \cup W_2] \setminus W) = 0$ and $\lambda(W \setminus [W_1 \cup W_2]) = 0$. Let us first show that $\lambda(W_1 \setminus W) = 0$, then by symmetry we have $\lambda(W_2 \setminus W) = 0$, giving $\lambda([W_1 \cup W_2] \setminus W) = 0$. Suppose that $\lambda(W_1 \setminus W) > 0$. Then, from Remark 3.1, there exists $\Gamma \in \mathcal{B}(X)$ such that $\Gamma \subset W_1 \setminus W$, $\lambda(\Gamma) > 0$, and $\mu(\Gamma) < \infty$ since W is λ/μ-perfect. However, $\mu_1(\Gamma) \leq \mu(\Gamma) < \infty$, implying that W_1 is not λ/μ_1-perfect. Consequently, $\lambda(W_1 \setminus W) = 0$, giving the first part of the proof. Now suppose that $\lambda(W \setminus [W_1 \cup W_2]) > 0$. Then there exists $\Gamma \in \mathcal{B}(X)$ such that $\Gamma \subset W \setminus [W_1 \cup W_2]$ with $\lambda(\Gamma) > 0$ and $\mu_1(\Gamma) < \infty$ by using the fact that W_1 is λ/μ_1-perfect. Define $\widehat{W} = \Gamma \cup W_2$. Since $W_2 \subset \widehat{W}$ and $\lambda(\widehat{W} \setminus W_2) = \lambda(\Gamma \setminus W_2) = \lambda(\Gamma) > 0$, there exists $\widetilde{\Gamma} \in \mathcal{B}(X)$ such that $\widetilde{\Gamma} \subset \Gamma$ with $\lambda(\widetilde{\Gamma}) > 0$ and $\mu_2(\widetilde{\Gamma}) < \infty$ since W_2 is λ/μ_2-perfect. Therefore, $\widetilde{\Gamma} \subset W$ and $\lambda(\widetilde{\Gamma}) > 0$ and

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

$\mu_1(\widetilde{\Gamma}) \leq \mu_1(\Gamma) < \infty$. Recalling that $\mu_2(\widetilde{\Gamma}) < \infty$, we have $\mu(\widetilde{\Gamma}) = \mu_1(\widetilde{\Gamma}) + \mu_2(\widetilde{\Gamma}) < \infty$, so that W is not λ/μ-perfect, leading to a contradiction. Consequently, $\lambda(W \setminus [W_1 \cup W_2]) = 0$, giving the second part of the proof and showing that $W_1 \cup W_2 \overset{\lambda}{\simeq} W$.

Now applying Proposition 3.1, we obtain the second statement and the result follows.

The following theorem shows how to construct perfect sets.

Theorem 3.2. *Let λ be a probability measure, and let μ be an arbitrary measure on X. Let us consider the construction of the following sets $(X_{i,0})_{i\in\mathbb{N}}$.*

0. *Put $i = 1$, $j = 1$, and define $X_{1,0} = X$.*
1. *If there exists $\Gamma \in \mathcal{B}(X)$ with $\Gamma \subset X_{i,j-1}$ such that $\lambda(\Gamma) \geq 1/2^i$ and $\mu(\Gamma) < \infty$ then define $X_{i,j} = X_{i,j-1} \setminus \Gamma$. Otherwise, put $X_{i,j} = X_{i,j-1}$ and go to step 3.*
2. *If $j < 2^i$, increase j by 1 and repeat step 1.*
3. *Increase i by 1 and define $X_{i,0} = X_{i-1,j}$. Put $j = 1$ and go to step 1.*

The set $\bigcap_{i\in\mathbb{N}} X_{i,0}$ is λ/μ-perfect. Moreover, the measure μ is σ-finite on $(\bigcap_{i\in\mathbb{N}} X_{i,0})^c$

Proof. Clearly, $X_{i+1,0} \subset X_{i,0}$. Note that if $\Gamma \in \mathcal{B}(X)$ with $\Gamma \subset X_{i,0}$ and $\lambda(\Gamma) \geq 1/2^{i-1}$ for some $i \geq 2$, then necessarily $\mu(\Gamma) = \infty$. Let us show that items (a) and (b) in Definition 3.1 are satisfied. Regarding item (a), let us proceed by contradiction and assume that there exists $\Gamma \in \mathcal{B}(X)$ such that $\Gamma \subset W$, $\lambda(\Gamma) > 0$, and $\mu(\Gamma) < \infty$. Let i^* be an integer satisfying $\lambda(\Gamma) \geq 1/2^{i^*-1}$. Since $\Gamma \subset X_{i^*,0}$, the previous comment implies that $\mu(\Gamma) = \infty$, leading to a contradiction. Consequently, Definition 3.1(a) holds. Now consider $V \in \mathcal{B}(X)$ such that $W \subset V$ and $\lambda(V \setminus W) > 0$. Since μ is σ-finite on W^c, it is σ-finite on $V \setminus W$. Consequently, there exists $\Gamma \subset V \setminus W$ such that $\lambda(\Gamma) > 0$ and $\mu(\Gamma) < \infty$, showing Definition 3.1(b). Since, the set $(\bigcap_{i\in\mathbb{N}} X_{i,0})^c$ consists of no more than a countable number of sets Γ with $\mu(\Gamma) < \infty$ as described in item 1 of the algorithm then the measure μ is σ-finite on $(\bigcap_{i\in\mathbb{N}} X_{i,0})^c$, giving the result.

4. Finite action space

In this section we restrict our attention to the case where the action set $A = \{a_1, \ldots, a_M\}$ is finite. In the next section, our results will be generalized to the case of a compact action space. The objective of the current section is to show that if the linear program $\mathbb{LP}$ admits an optimal solution μ^* then there exists a feasible randomized stationary policy $\varphi^* \in \Pi_c$ which is optimal for the control problem defined in (2.1)–(2.3) and that the optimal value of the $\mathbb{LP}$ coincides with the optimal value of the control problem (see Theorem 4.2). As a consequence, we show that the set of randomized stationary policies is a sufficient class of policies for the control problem under consideration (see Corollary 4.1). The proof of our main results strongly relies on the following key property: for any nearly perfect measure μ (see Definition 4.1 for a precise statement), one can construct a randomized stationary policy φ such that the occupation measure μ^φ associated with φ is *smaller* than μ (see Theorem 4.1). We introduce the following assumptions: there exists a probability measure q on $\mathcal{B}(X)$ with respect to which the stochastic kernel Q is absolutely continuous and the values of $\mathbb{LP}_\theta$ for any $\theta \in \Theta \cup \{0\}$ are finite (see Assumption 4). The former hypothesis is not very restrictive as explained in item (b) of Remark 4.1. Moreover, it appears necessary to impose the latter hypothesis. Indeed, if this assumption is violated then some examples presented in Section 6 show that the optimal solution of the linear program $\mathbb{LP}$ may have no meaning.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

We first introduce a special class of measures called nearly perfect.

Definition 4.1. A measure $\mu \in \mathbb{M}(X \times A)_+$ is said to be nearly perfect if it satisfies the linear equation (2.6) and if, for any $a \in A$, there exists a μ_A/μ_a-nearly perfect set V_a such that μ_A is σ-finite on V_A^c, where $V_A = \bigcup_{a\in A} V_a$.

In the above definition, there is no loss of generality to assume that $V_{a_i} \cap V_{a_j} = \varnothing$ for $i \neq j$.

Definition 4.2. If μ is nearly perfect then consider a stochastic kernel on A given X, labeled φ, such that, for any $\Gamma \in \mathcal{B}(V_A^c)$, $\mu_a(\Gamma) = \mu(\Gamma \times \{a\}) = \int_\Gamma \varphi(\{a\} \mid x)\mu_A(\mathrm{d}x)$ and, for any $x \in V_a$, $\varphi(\cdot \mid x) = \delta_a(\cdot)$. The randomized stationary policy generated by φ will be said to be induced by μ.

Sufficient conditions for the existence of nearly perfect measures is given in Lemma 4.4 below. In the following theorem we derive an important property of nearly perfect measures.

Theorem 4.1. *Let $\mu \in \mathbb{M}(X \times A)_+$ be nearly perfect, and let φ be its induced randomized stationary policy. Then $\mu^\varphi \leq \mu$.*

Proof. Introduce the mapping T defined on $\mathbb{M}(X)_+$ by

$$T\gamma(\Gamma) = \nu(\Gamma) + \sum_{a\in A} \int_X Q(\Gamma \mid y, a)\varphi(\{a\} \mid y)\gamma(\mathrm{d}y).$$

Define the sequence of measures $(\gamma_n)_{n\in\mathbb{N}}$ in $\mathbb{M}(X)_+$ by $\gamma_0 = 0$ and $\gamma_{n+1} = T\gamma_n$ for $n \in \mathbb{N}$. Since the operator T is monotone, it follows easily by induction that the sequence $(\gamma_n)_{n\in\mathbb{N}}$ is increasing and converges to μ_A^φ which is the minimal positive solution of $\gamma = T\gamma$ with $\gamma \in \mathbb{M}(X)_+$.

Now, let us show that $T\mu_A \leq \mu_A$. Indeed, for any $\Gamma \in \mathcal{B}(X)$,

$$\begin{aligned} T\mu_A(\Gamma) = \nu(\Gamma) &+ \sum_{a\in A} \int_{V_A^c} Q(\Gamma \mid y, a)\varphi(\{a\} \mid y)\mu_A(\mathrm{d}y) \\ &+ \sum_{a\in A} \int_{V_A} Q(\Gamma \mid y, a)\varphi(\{a\} \mid y)\mu_A(\mathrm{d}y). \end{aligned} \tag{4.1}$$

However, since $\mu(\Lambda \times \{a\}) = \int_\Lambda \varphi(\{a\} \mid x)\mu_A(\mathrm{d}x)$ for any $\Lambda \in \mathcal{B}(V_A^c)$, we have, for any $\Gamma \in \mathcal{B}(X)$,

$$\mu_A(\Gamma) = \nu(\Gamma) + \sum_{a\in A} \int_{V_A^c} Q(\Gamma \mid y, a)\varphi(\{a\} \mid y)\mu_A(\mathrm{d}y) + \sum_{a\in A} \int_{V_A} Q(\Gamma \mid y, a)\mu_a(\mathrm{d}y). \tag{4.2}$$

Consider a fixed set Γ in $\mathcal{B}(X)$. Suppose that $\mu_A(\{y \in V_{a_j} : Q(\Gamma \mid y, a_j) > 0\}) > 0$ for some $j \in \mathbb{N}_M$. Then, since V_{a_j} is μ_A/μ_{a_j}-nearly perfect, $\mu_{a_j}(\{y \in V_{a_j} : Q(\Gamma \mid y, a_j) > 0\}) = \infty$, implying that

$$\sum_{a\in A} \int_{V_A} Q(\Gamma \mid y, a)\mu_a(\mathrm{d}y) \geq \int_{V_{a_j}} Q(\Gamma \mid y, a_j)\mu_{a_j}(\mathrm{d}y) = \infty. \tag{4.3}$$

Combining (4.1)–(4.3), we have $T\mu_A(\Gamma) \leq \mu_A(\Gamma)$.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Now suppose that, for all $j \in \mathbb{N}_M$, $\mu_A(\{y \in V_{a_j} : Q(\Gamma \mid y, a_j) > 0\}) = 0$. Then

$$\sum_{a \in A} \int_{V_A} Q(\Gamma \mid y, a)\varphi(\{a\} \mid y)\mu_A(\mathrm{d}y) = \sum_{i=1}^{M} \int_{V_{a_i}} Q(\Gamma \mid y, a_i)\mu_A(\mathrm{d}y) = 0, \tag{4.4}$$

and so combining (4.1), (4.2), and (4.4), we have $T\mu_A(\Gamma) \le \mu_A(\Gamma)$.

In any case $T\mu_A \le \mu_A$. Consequently, it can be shown easily by induction that $\gamma_n \le \mu_A$, implying that $\mu_A^\varphi \le \mu_A$. Therefore, for any $\Lambda \in \mathcal{B}(V_A^c)$, $\mu(\Lambda \times \{a\}) = \int_\Lambda \varphi(\{a\} \mid x)\mu_A(\mathrm{d}x) \ge \int_\Lambda \varphi(\{a\} \mid x)\mu_A^\varphi(\mathrm{d}x) = \mu^\varphi(\Lambda \times \{a\})$. Now consider any $a \in A$ and $\Lambda \in \mathcal{B}(V_a)$. Then, if $\mu_a(\Lambda) < \infty$, we have $\mu_A(\Lambda) = 0$ since V_a is μ_A/μ_a-nearly perfect. Since $\mu_A^\varphi \le \mu_A$, we have $\mu^\varphi(\Lambda \times \{a\}) \le \mu^\varphi(\Lambda \times A) = 0$, showing that $\mu(\Lambda \times \{a\}) \ge \mu^\varphi(\Lambda \times \{a\})$. This concludes the proof.

Roughly speaking, we now introduce the *difference*, labeled Δ^μ, between the measures μ and μ^φ and derive some technical results related to it.

Definition 4.3. Let $\mu \in \mathbb{M}(X \times A)_+$ be nearly perfect, and let φ be its induced randomized stationary policy. Define the mapping Δ^μ on $\mathcal{B}(X \times A)$ by

$$\Delta^\mu(\Lambda) = \mu(\Lambda \cap [V_A \times A]) + (\mu - \mu^\varphi)(\Lambda \cap [V_A^c \times A]), \tag{4.5}$$

where $\Lambda \in \mathcal{B}(X \times A)$ (see Definition 2.1).

The dependence of Δ^μ on φ is not explicitly mentioned because φ itself depends on μ. Observe that, for any $\Lambda \in \mathcal{B}(X \times A)$, $(\mu - \mu^\varphi)(\Lambda)$ is well defined since, from Theorem 4.1, the measure μ is nearly perfect and $\mu^\varphi \le \mu$.

Lemma 4.1. *Let $\mu \in \mathbb{M}(X \times A)_+$ be nearly perfect, and let φ be its induced randomized stationary policy. The mapping Δ^μ is a measure in $\mathbb{M}(X \times A)_+$ that satisfies*

$$\mu = \mu^\varphi + \Delta^\mu. \tag{4.6}$$

Proof. It is easy to check that Δ^μ is a measure in $\mathbb{M}(X \times A)_+$. Consider $\Gamma \in \mathcal{B}(V_A^c)$ and $a \in A$. Then we clearly have $\mu(\Gamma \times \{a\}) = \mu^\varphi(\Gamma \times \{a\}) + \Delta^\mu(\Gamma \times \{a\})$ by the definition of Δ^μ, implying (4.6). According to Definition 4.1, $V_A = \sum_{b \in A} V_b$, where V_b is a μ_A/μ_b-nearly perfect set for any $b \in A$. Now consider $a \in A$ and $\Gamma \in \mathcal{B}(V_b)$ for $b \in A$. From the definition of Δ^μ, we need to check that $\mu(\Gamma \times \{a\}) = \mu^\varphi(\Gamma \times \{a\}) + \mu(\Gamma \times \{a\})$. Note that $\mu(\Gamma \times \{b\}) = 0$ or $\mu(\Gamma \times \{b\}) = \infty$ since V_b is a μ_A/μ_b-nearly perfect. Consequently, if $a = b$ then (4.6) follows immediately since $\mu^\varphi \le \mu$. If $b \ne a$ then $\mu^\varphi(\Gamma \times \{a\}) = \int_\Gamma \varphi(\{a\} \mid y)\mu_A^\varphi(\mathrm{d}y)$, but $\varphi(\{a\} \mid y) = 0$ for any $y \in \Gamma \subset V_b$ and so $\mu^\varphi(\Gamma \times \{a\}) = 0$, showing (4.6). This completes the proof.

Lemma 4.2. *Let $\mu \in \mathbb{M}(X \times A)_+$ be nearly perfect, and let φ be its induced randomized stationary policy. Consider $\Gamma \in \mathcal{B}(X)$. We have*

$$\Delta_A^\mu(\Gamma) \ge \Delta^\mu Q(\Gamma). \tag{4.7}$$

Moreover, if $\mu_A^\varphi(\Gamma) < \infty$ or $\Gamma \subset V_A^c$ then we have an equality in the above formula.

Proof. From (4.6), we have $\mu Q = \mu^\varphi Q + \Delta^\mu Q$ and so $\mu_A = \mu_A^\varphi + \Delta^\mu Q$. Let $\Gamma \in \mathcal{B}(X)$ be fixed. Let us consider two cases: $\mu_A^\varphi(\Gamma) < \infty$ and $\mu_A^\varphi(\Gamma) = \infty$. If $\mu_A^\varphi(\Gamma) < \infty$ then

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

$\Delta^{\mu} Q(\Gamma) = \mu_A(\Gamma) - \mu_A^{\varphi}(\Gamma)$. From (4.6) and since $\mu_A^{\varphi}(\Gamma) < \infty$, $\Delta_A^{\mu}(\Gamma) = \mu_A(\Gamma) - \mu_A^{\varphi}(\Gamma)$, showing the equality in (4.7). Now, if $\mu_A^{\varphi}(\Gamma) = \infty$ and $\Gamma \in \mathcal{B}(V_A^c)$, then, by using the fact that μ_A^{φ} is σ-finite on V_A^c, this case reduces to the previous one, and so the equality holds in (4.7). Finally, if $\mu_A^{\varphi}(\Gamma) = \infty$ and $\Gamma \in \mathcal{B}(V_a)$, then $\mu_a^{\varphi}(\Gamma) = \int_{\Gamma} \varphi(\{a\} \mid y) \mu_A^{\varphi}(\mathrm{d}y) = \infty$ by the definition of φ and (2.5). Consequently, $\Delta_A^{\mu}(\Gamma) \geq \Delta_a^{\mu}(\Gamma) = \mu_a(\Gamma) \geq \mu_a^{\varphi}(\Gamma) = \infty$ and so $\Delta_A^{\mu}(\Gamma) = \infty$, implying that inequality (4.7) is satisfied.

Lemma 4.3. *Let $\mu \in \mathbb{M}(X \times A)_+$ be nearly perfect, and let φ be its induced randomized stationary policy. Assume that $\gamma \in \mathbb{M}(X \times A)_+$ and satisfies $\gamma_A \geq \mu_A^{\varphi}$ and $\gamma_A = \nu + \gamma Q$. Then the measure $\lambda = \gamma + \Delta^{\mu}$ satisfies $\lambda_A = \nu + \lambda Q$.*

Proof. Let us consider $\Gamma \in \mathcal{B}(X)$ fixed. If $\mu_A^{\varphi}(\Gamma) = \infty$ then $\gamma_A(\Gamma) = \infty$. Consequently, we have $\lambda_A(\Gamma) = \gamma_A(\Gamma) + \Delta_A^{\mu}(\Gamma) = \nu(\Gamma) + \gamma Q(\Gamma) + \Delta^{\mu} Q(\Gamma) = \nu(\Gamma) + \lambda Q(\Gamma) = \infty$, showing the result. Now if $\mu_A^{\varphi}(\Gamma) < \infty$ then we have $\Delta_A^{\mu}(\Gamma) = \Delta^{\mu} Q(\Gamma)$ from Lemma 4.2. Therefore, $\lambda_A(\Gamma) = \nu(\Gamma) + \lambda Q(\Gamma)$, completing the proof.

We need the following assumptions to derive our main results.

Assumption A. *Suppose that*

(A1) *there exists a probability measure $q \in \mathbb{M}(X)_+$ such that $Q(\cdot \mid x, a) \ll q$ for any $(x, a) \in X \times A$,*

(A2) *the value of the linear program $\mathbb{LP}_{\theta}$ is finite for any $\theta \in \Theta \cup \{0\}$.*

Remark 4.1. (a) There is no loss of generality to assume that $\nu \ll q$ in assumption (A1). Indeed, if there exists a probability measure $\widetilde{q} \in \mathbb{M}(X)_+$ such that $Q(\cdot \mid x, a) \ll \widetilde{q}$ for any $(x, a) \in X \times A$ then the probability measure $q = \frac{1}{2}\nu + \frac{1}{2}\widetilde{q}$ satisfies the required property.

(b) Assumption (A1) is not very restrictive. In many applications, the evolution of an MDP is specified by a discrete-time equation of the form $x_{t+1} = F(x_t, a_t) + \xi_t$, where F is an $\mathbb{R}^n$-valued measurable mapping defined on $\mathbb{R}^n \times A$ and $(\xi_t)_{t \in \mathbb{N}^0}$ is an independent and identically distributed sequence of random variables with density α with respect to the Lebesgue measure on $\mathcal{B}(\mathbb{R}^n)$. By using the change of variable formula, we obtain $Q(A \mid x, a) = \int_A \alpha(y - F(x, a))\, \mathrm{d}y$. Consequently, assumption (A1) is satisfied for q defined by the standard normal distribution on $\mathcal{B}(\mathbb{R}^n)$.

Moreover, assumption (A1) is automatically satisfied if X is finite or countable; in these cases q can be defined as a geometric distribution.

The following lemma shows that hypothesis (A1) is a sufficient condition to ensure that any measure satisfying the linear equation (2.6) is nearly perfect.

Lemma 4.4. *Under assumption (A1), any measure $\mu \in \mathbb{M}(X)_+$ satisfying the linear equation (2.6) is nearly perfect.*

Proof. Since q is a probability measure, from Theorem 3.2, there exists a q/μ_a-perfect set, labeled V_a for any $a \in A$. Moreover, μ_a is σ-finite on V_a^c. Consequently, the measure $\mu_A = \sum_{a \in A} \mu_a$ is σ-finite on $(\bigcup_{a \in A} V_a)^c$. Furthermore, V_a is q/μ_a-nearly perfect set. From assumption (A1) and Remark 4.1(a), it can be easily shown that the measure μ_A is absolutely continuous with respect to q and so it is easy to show that V_a is μ_A/μ_a-nearly perfect, completing the proof.

We now present our two main results.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Theorem 4.2. *Suppose that Assumption A is satisfied. If μ^* is an optimal solution of the constrained linear program $\mathbb{LP}$ then the induced policy φ^* is a solution of the constrained control problem defined by (2.3) and the optimal value $V^*(\nu)$ coincides with the optimal value of the $\mathbb{LP}$, that is,*

$$v(\nu, \varphi^*) = \inf_{\pi \in \Pi_c} v(\nu, \pi) = \inf_{\mu \in \mathbb{L}_c} \mu(r_0) = \mu^*(r_0).$$

Proof. For any policy $\pi \in \Pi$, the corresponding occupation measure satisfies the linear equation (2.6) by Lemma 9.4.3 of [10]. The value of the constrained control problem defined in (2.3) cannot be smaller than the value of the $\mathbb{LP}$ given by $\mu^*(r_0)$. We will show that $\mu^*(r_\theta) \geq \mu^{\varphi^*}(r_\theta)$ for any $\theta \in \Theta \cup \{0\}$. Suppose that, for some $\theta \in \Theta$, $\mu^*(r_\theta) < \mu^{\varphi^*}(r_\theta)$. Then introduce Δ^{μ^*} according to (4.5). From Lemma 4.3 and Theorem 4.1, the measure λ^n defined by $\lambda^n = \mu^* + n\Delta^{\mu^*}$ for $n \in \mathbb{N}$ satisfies $\lambda^n_A = \nu + \lambda^n Q$ and so $\lambda^n(r_\theta) = \mu^*(r_\theta) + n\Delta^{\mu^*}(r_\theta)$. From assumption (A2), $-\infty < \mu^*(r_\theta)$ and $\mu^*(r_\theta) \leq R_\theta$, and so $\mu^*(r_\theta)$ is finite. Therefore, we obtain $\mu^{\varphi^*}(r^+_\theta) \leq \mu^*(r^+_\theta) < \infty$ and $\mu^{\varphi^*}(r^-_\theta) \leq \mu^*(r^-_\theta) < \infty$ by using Theorem 4.1. Consequently, $\mu^{\varphi^*}(r_\theta)$ is finite and we have $\Delta^{\mu^*}(r_\theta) < 0$ since $\mu^*(r_\theta) = \mu^{\varphi^*}(r_\theta) + \Delta^{\mu^*}(r_\theta) < \mu^{\varphi^*}(r_\theta)$. This implies that the value of the auxiliary unconstrained linear program $\mathbb{LP}_\theta$ is smaller than $\lambda^n(r_\theta) = \mu^*(r_\theta) + n\Delta^{\mu^*}(r_\theta)$ for any $n \in \mathbb{N}$ and, thus, equals $-\infty$, leading to a contradiction with assumption (A2). Therefore, for any $\theta \in \Theta$, $\mu^{\varphi^*}(r_\theta) \leq R_\theta$. Now, it is also obvious that $\mu^*(r_0) \leq \mu^{\varphi^*}(r_0)$, and it remains to prove that $\mu^*(r_0) = \mu^{\varphi^*}(r_0)$. If $\mu^*(r_0) = +\infty$ then necessarily $\mu^{\varphi^*}(r_0) = \mu^*(r_0)$ since $\infty = \mu^*(r_0) \leq \mu^{\varphi^*}(r_0)$. If $\mu^*(r_0) < +\infty$ then, from assumption (A2), $-\infty < \mu^*(r_0)$ and so $\mu^*(r_0)$ is finite. One can apply the same reasoning presented above for $\theta \in \Theta$ to show that $\mu^*(r_0) = \mu^{\varphi^*}(r_0)$. This completes the proof.

Corollary 4.1. *Suppose that Assumption A is satisfied. 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 4.2.

5. Compact action space

In this section, it is assumed that the action set A is compact. The main results we obtain in this section may appear similar to those of the previous section. However, there are several subtle differences that make the derivations more technical and complicated. The general idea of the proof relies on some kind of discretization of the action space (see Lemma 5.1 and Proposition 5.1). In this new context, we can still construct a randomized stationary policy φ associated to any measure μ satisfying the linear equation (2.6). However, we cannot guarantee that the occupation measure μ^φ associated to φ is *smaller* than μ as in Theorem 4.1. Roughly speaking, we can only prove that $\mu^\varphi(g) \leq \mu(g)$ for any $\mathbb{R}_+$-valued measurable function g defined on $X \times A$ such that $g(x, \cdot)$ is continuous on A for any $x \in X$. However, since the measures μ and μ^φ are not σ-finite, we cannot conclude that $\mu^\varphi \leq \mu$. These technical differences induce changes in the way we deal with the compact action space.

Assumption B. *Suppose that*

(B1) *there exists a probability measure $q \in \mathbb{M}(X)_+$ such that $Q(\cdot \mid x, a) \ll q$ for any $(x, a) \in X \times A$,*

(B2) *the set A is compact; the metric on A will be denoted by ρ,*

(B3) *for any $\Gamma \in \mathcal{B}(X)$ and $x \in X$, the mapping $Q(\Gamma \mid x, \cdot)$ is continuous on A,*

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

(B4) *the value of the linear program* $\mathbb{LP}_\theta$ *is finite for any* $\theta \in \Theta \cup \{0\}$,

(B5) *for any* $\theta \in \Theta \cup \{0\}$ *and any* $x \in X$, *the mapping* $r_\theta(x, \cdot)$ *is continuous on* A.

Observe that if assumption (B1) holds then any measure $\mu \in \mathbb{M}(X \times A)_+$ that satisfies the linear equation (2.6) can be considered as being absolutely continuous with respect to q according to Remark 4.1. Moreover, it must be noted that assumptions (B3) and (B5) are automatically satisfied in the case of a finite action space.

The next two technical results introduce some kind of discretization of the action space.

Lemma 5.1. *Suppose that assumptions (B1)–(B2) are satisfied. Consider a measure* $\mu \in \mathbb{M}(X \times A)_+$ *satisfying the linear equation (2.6). There exists a sequence* $(M_n)_{n\in\mathbb{N}}$ *in* $\mathbb{N}$, *a sequence* $(A_{i,j})_{i\in\mathbb{N},\, j\in\mathbb{N}_{M_i}}$ *in* $\mathcal{B}(A)$, *and a sequence* $(V_{i,j})_{i\in\mathbb{N},\, j\in\mathbb{N}_{M_i}}$ *in* $\mathcal{B}(X)$ *such that the following assertions hold.*

(a) $A = \sum_{j=1}^{M_1} A_{1,j}$ *with* $\rho(a,b) < 1$ *for any* $(a,b) \in A_{1,j}$. *The set* $V_{1,j}$ *is* $q/\mu_{A_{1,j}}$*-perfect and* $\mu_{A_{1,j}}$ *is* σ*-finite on* $V_{1,j}^c$ *for any* $j \in \mathbb{N}_{M_1}$.

(b) *For* $n > 1$, $A = \sum_{j=1}^{M_n} A_{n,j}$ *with* $\rho(a,b) < 1/2^{n-1}$ *for any* $(a,b) \in A_{n,j}$. *For any* $j \in \mathbb{N}_{M_n}$, $V_{n,j}$ *is a* $q/\mu_{A_{n,j}}$*-perfect set. Moreover, there exists a partition of* $\mathbb{N}_{M_n}$, *labeled* $(\mathcal{I}_{n,k})_{k\in\mathbb{N}_{M_{n-1}}}$, *such that* $V_{n-1,k} = \bigcup_{j\in\mathcal{I}_{n,k}} V_{n,j}$ *and* $A_{n-1,k} = \sum_{j\in\mathcal{I}_{n,k}} A_{n,j}$ *for any* $k \in \mathbb{N}_{M_{n-1}}$.

Proof. Part (a) is an easy consequence of Theorem 3.2 and the fact that A is compact. Let us show part (b) by induction. Consider $p \geq 1$, and assume that, for any $n \leq p$, (b) is satisfied. Since $A_{p,j}$ is relatively compact for any $j \in \mathbb{N}_{M_p}$, there exists a partition of $A_{p,j}$ into $N_{p,j}$ sets: $A_{p,j} = \sum_{k=1}^{N_{p,j}} B_{j,k}$ with $\rho(a,b) < 1/2^p$ for any $(a,b) \in B_{j,k}$. Then combining Theorems 3.1 and 3.2, there exists a $q/\mu_{B_{j,k}}$-perfect set, labeled $\widetilde{W}_{j,k}$ for any $k \in \mathbb{N}_{N_{p,j}}$ such that $V_{p,j} \overset{q}{\simeq} \bigcup_{k=1}^{N_{p,j}} \widetilde{W}_{j,k}$. Now, we have $V_{p,j} = \bigcup_{k=1}^{N_{p,j}} W_{j,k}$ with $W_{j,k} = [V_{p,j} \setminus \bigcup_{k=1}^{N_{p,j}} \widetilde{W}_{j,k}] \cup [V_{p,j} \cap \widetilde{W}_{j,k}]$. However, since $V_{p,j} \overset{q}{\simeq} \bigcup_{k=1}^{N_{p,j}} \widetilde{W}_{j,k}$, we have $W_{j,k} \overset{q}{\simeq} \widetilde{W}_{j,k}$. Consequently, $W_{j,k}$ is a $q/\mu_{B_{j,k}}$-perfect set according to Proposition 3.1. Defining $M_{p+1} = \sum_{j=1}^{M_p} N_{p,j}$ and reordering the sets $(B_{j,k})_{j\in\mathbb{N}_{M_p},\, k\in\mathbb{N}_{N_{p,j}}}$ and $(W_{j,k})_{j\in\mathbb{N}_{M_p},\, k\in\mathbb{N}_{N_{p,j}}}$, we can easily show the existence of $(A_{p+1,j})_{j\in\mathbb{N}_{M_{p+1}}}$, $(V_{p+1,j})_{j\in\mathbb{N}_{M_{p+1}}}$, and $(\mathcal{I}_{p+1,k})_{k\in\mathbb{N}_{M_p}}$, satisfying the desired properties.

Proposition 5.1. *Suppose that assumptions (B1)–(B2) are satisfied. Consider a measure* $\mu \in \mathbb{M}(X \times A)_+$ *satisfying the linear equation (2.6). There exists a sequence* $(M_n)_{n\in\mathbb{N}}$ *in* $\mathbb{N}$, *a sequence* $(A_{i,j})_{i\in\mathbb{N},\, j\in\mathbb{N}_{M_i}}$ *in* $\mathcal{B}(A)$, *and a sequence* $(\widehat{V}_{i,j})_{i\in\mathbb{N},\, j\in\mathbb{N}_{M_i}}$ *in* $\mathcal{B}(X)$ *such that the following assertions hold.*

(a) $A = \sum_{j=1}^{M_1} A_{1,j}$ *with* $\rho(a,b) < 1$ *for any* $(a,b) \in A_{1,j}$. *The set* $\widehat{V}_{1,j}$ *is* $q/\mu_{A_{1,j}}$*-nearly perfect and* $\mu_{A_{1,j}}$ *is* σ*-finite on* $\widehat{V}_{1,j}^c$ *for any* $j \in \mathbb{N}_{M_1}$. *The measure* μ_A *is* σ*-finite on* V_A^c, *where* $V_A = \sum_{i=1}^{M_1} \widehat{V}_{1,i}$.

(b) *For* $n > 1$, $A = \sum_{j=1}^{M_n} A_{n,j}$ *with* $\rho(a,b) < 1/2^{n-1}$ *for any* $(a,b) \in A_{n,j}$. *For any* $j \in \mathbb{N}_{M_n}$, $\widehat{V}_{n,j}$ *is a* $q/\mu_{A_{n,j}}$*-nearly perfect set and* $V_A = \sum_{j=1}^{M_n} \widehat{V}_{n,j}$. *Moreover, for any* $j \in \mathbb{N}_{M_n}$, *there exists* $k \in \mathbb{N}_{M_{n-1}}$ *such that* $\widehat{V}_{n,j} \subset \widehat{V}_{n-1,k}$ *and* $A_{n,j} \subset A_{n-1,k}$.

Proof. Part (a) follows from Lemma 5.1. Indeed, let us consider the sets $(A_{1,j})_{1\leq j\leq M_1}$ and $(V_{1,j})_{1\leq j\leq M_1}$ introduced in Lemma 5.1(a). Define $\widehat{V}_{1,1} = V_{1,1}$ and, for $j > 1$, $\widehat{V}_{1,j} = V_{1,j} \setminus \bigcup_{k=1}^{j-1} V_{1,k}$. Since $\widehat{V}_{1,j} \subset V_{1,j}$ for $j \in \mathbb{N}_{M_1}$ and according to Remark 3.2, $\widehat{V}_{1,j}$ is a

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

$q/\mu_{A_{1,j}}$-nearly perfect set. Clearly, the sets $(\widehat{V}_{1,j})_{j\in\mathbb{N}_{M_1}}$ are pairwise disjoint and $\bigcup_{i=1}^{M_1} V_{1,i} = \sum_{i=1}^{M_1} \widehat{V}_{1,i}$. Moreover, from Lemma 5.1(a), μ_A is σ-finite on $(\sum_{i=1}^{M_1} \widehat{V}_{1,i})^c$, showing part (a). Using similar arguments as before and from Lemma 5.1(b), there exits a $q/\mu_{A_{n,j}}$-nearly perfect set, labeled $\widetilde{V}_{n,j}$ for any $j \in M_n$, such that $V_{n-1,k} = \sum_{j\in\mathcal{I}_{n,k}} \widetilde{V}_{n,j}$. Now the set $\widehat{V}_{n,j}$ defined by $\widehat{V}_{n,j} = \widetilde{V}_{n,j} \cap \widehat{V}_{n-1,k}$ for $j \in \mathcal{I}_{n,k}$ satisfies the required properties, completing the proof.

Remark 5.1. It is worth mentioning that the proofs of Lemma 5.1 and Proposition 5.1 only required that q must be a probability measure. Actually, the second part of assumption (B1) is not needed, that is, $Q(\cdot \mid x, a) \ll q$ for any $(x, a) \in X \times A$. Moreover, observe that assumption (B2) is only needed to ensure that each element of the partition $(A_{n,j})_{j\in\mathbb{N}_{M_n}}$ of the action space A has a diameter less than $1/2^{n-1}$.

Let μ be a measure in $\mathbb{M}(X \times A)_+$ satisfying equation (2.6). According to Proposition 5.1, there exists a stochastic kernel on A given V_A^c, labeled ψ, such that, for any $\Gamma \in \mathcal{B}(V_A^c \times A)$, $\mu(\Lambda) = \int_\Lambda \psi(\mathrm{d}a \mid x)\mu_A(\mathrm{d}x)$. Moreover, introduce, for any $n \in \mathbb{N}$, the A-valued mapping f_n defined on V_A by $f_n(x) = a_{n,j}$ if $x \in \widehat{V}_{n,j}$, where $a_{n,j}$ is an arbitrary fixed element in $A_{n,j}$.

Lemma 5.2. *Suppose that assumptions (B1)–(B2) are satisfied. Then* $\lim_{n\to\infty} f_n(x)$ *exists for any* $x \in V_A$.

Proof. Let $x \in V_A$. For any $n > 1$, there exist $k \in \mathbb{N}_{M_{n-1}}$ and $j \in \mathbb{N}_{M_n}$ such that $f_{n-1}(x) \in A_{n-1,k}$ and $f_n(x) \in A_{n,j}$ with $A_{n,j} \subset A_{n-1,k}$ according to Proposition 5.1. Therefore, $\rho(f_n(x), f_{n-1}(x)) \le 1/2^{n-2}$, showing that the $(f_n(x))_{n\in\mathbb{N}}$ is a Cauchy sequence in the complete space A. This completes the proof.

Definition 5.1. Suppose that assumptions (B1)–(B2) are satisfied. Consider $\mu \in \mathbb{M}(X \times A)_+$ satisfying the linear equation (2.6). Let f be the A-valued measurable function defined on V_A by $f(x) = \lim_{n\to\infty} f_n(x)$, and let φ be the stochastic kernel defined on A given X by $\varphi(\cdot \mid x) = I_{V_A^c}(x)\psi(\cdot \mid x) + I_{V_A}(x)\delta_{f(x)}(\cdot)$. The randomized stationary policy generated by φ will be said to be induced by the measure μ.

Figure 1 illustrates the construction of $(A_{i,j})_{i\in\mathbb{N},\, j\in\mathbb{N}_{M_i}}$ and $(\widehat{V}_{i,j})_{i\in\mathbb{N},\, j\in\mathbb{N}_{M_i}}$ obtained in Lemma 5.1 and Proposition 5.1 and the associated mapping f introduced in Definition 5.1.

Lemma 5.3. *Suppose that assumptions (B1)–(B2) are satisfied. Consider* $\mu \in \mathbb{M}(X \times A)_+$ *satisfying the linear equation (2.6), and let* φ *be its induced randomized stationary policy.*

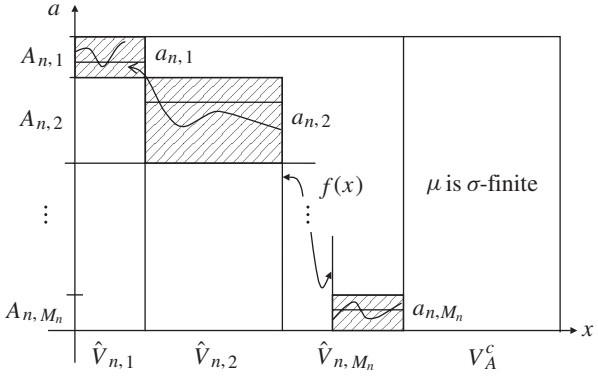

FIGURE 1: Construction of the sets $A_{i,j}$ and $\widehat{V}_{i,j}$.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Let g be an $\mathbb{R}_+$-valued measurable function defined on $X \times A$ such that $g(x,\cdot)$ is continuous on A for any $x \in X$. If $\int_{V_A \times A} g(y,a)\mu(\mathrm{d}y \times \mathrm{d}a) < \infty$ then $\int_{V_A} \times A g(y,a)\mu^{\varphi}(\mathrm{d}y \times \mathrm{d}a) = \int_{V_A} g(y, f(y))q(\mathrm{d}y) = 0$.

Proof. For $\varepsilon > 0$, $n \in \mathbb{N}$, and $j \in \mathbb{N}_{M_n}$, let us introduce $W^{\varepsilon} = \{x \in V_A : g(x, f(x)) \geq \varepsilon\}$, $W^{\varepsilon}_{n,j} = \{x \in \widehat{V}_{n,j} : \text{ for all } a \in A_{n,j},\ g(x,a) \geq \varepsilon\}$. According to Proposition 5.1, the sets $(W^{\varepsilon}_{n,j})_{j \in \mathbb{N}_{M_n}}$ are pairwise disjoint. Define $W^{\varepsilon}_n = \sum_{j=1}^{M_n} W^{\varepsilon}_{n,j}$. Since A is a compact metric space, any subset of A is separable. Therefore, we have $W^{\varepsilon}_{n,j} = \bigcap_{k \in \mathbb{N}} \{x \in X : G(x, a_k) \geq \varepsilon\} \cap \widehat{V}_{n,j}$, where $(a_k)_{k \in \mathbb{N}}$ is a dense subset of $A_{n,j}$ since $g(x,\cdot)$ is continuous for any $x \in X$. Consequently, $W^{\varepsilon}_{n,j} \in \mathcal{B}(X)$. Now let us show that $W^{\varepsilon} \subset \bigcup_{n \in \mathbb{N}} W^{\varepsilon/2}_n$. Indeed, consider $x \in W^{\varepsilon}$. Since $g(x,\cdot)$ is continuous on the compact set A, there exists $\beta_{x,\varepsilon} > 0$ such that $|g(x,a) - g(x, f(x))| \leq \varepsilon/2$ for any a satisfying $\rho(a, f(x)) \leq \beta_{x,\varepsilon}$. Now, for any $n \in \mathbb{N}$, we have $x \in \widehat{V}_{n,j_{n,x}}$ and $f_n(x) \in A_{n,j_{n,x}}$ for some $j_{n,x} \in \mathbb{N}_{M_n}$. According to Proposition 5.1 and Lemma 5.2, there exists $n_{x,\varepsilon} \in \mathbb{N}$ such that, for any $n \geq n_{x,\varepsilon}$ and $a \in A_{n,j_{n,x}}$, we have $\rho(f_n(x), a) \leq \beta_{x,\varepsilon}/2$ and $\rho(f_n(x), f(x)) \leq \beta_{x,\varepsilon}/2$. Consequently, for $n \geq n_{x,\varepsilon}$, we have $\rho(f(x), a) \leq \beta_{x,\varepsilon}$ for any $a \in A_{n,j_{n,x}}$ and so $x \in W^{\varepsilon/2}_{n,j_{n,x}} \subset W^{\varepsilon/2}_n$, showing that $W^{\varepsilon} \subset \bigcup_{n \in \mathbb{N}} W^{\varepsilon/2}_n$.

Now assume that $\int_{V_A \times A} g(y,a)\mu^{\varphi}(\mathrm{d}y \times \mathrm{d}a) > 0$. Then $\int_{V_A \times A} g(y, f(x))q(\mathrm{d}y) > 0$ according to the fact that $\mu^{\varphi}_A(\cdot) \ll q(\cdot)$ and so there exists $\varepsilon > 0$ such that $q(W^{\varepsilon}) > 0$. Therefore, $q(W^{\varepsilon/2}_{n,j}) > 0$ for some $n \in \mathbb{N}$ and $j \in \mathbb{N}_{M_n}$ since $W^{\varepsilon} \subset \bigcup_{n \in \mathbb{N}} \bigcup_{j \in \mathbb{N}_{M_n}} W^{\varepsilon/2}_{n,j}$. Observe that $W^{\varepsilon/2}_{n,j}$ is a $q/\mu_{A_{n,j}}$-nearly perfect set since it is a subset of $\widehat{V}_{n,j}$. This implies that $\mu_{A_{n,j}}(W^{\varepsilon/2}_{n,j}) = \infty$. Finally,

$$\int_{V_A \times A} g(y,a)\mu(\mathrm{d}y \times \mathrm{d}a) \geq \int_{W^{\varepsilon/2}_{n,j} \times A_{n,j}} g(y,a)\mu(\mathrm{d}y \times \mathrm{d}a) \geq \frac{\varepsilon}{2}\mu_{A_{n,j}}(W^{\varepsilon/2}_{n,j}) = \infty,$$

showing the contrapositive of the result.

Our next result is aimed towards comparing the measures μ and μ^{φ}. Clearly, in the context of a general compact action space, we cannot guarantee that the occupation measure μ^{φ} associated to φ is *smaller* than μ as in Theorem 4.1. This is an important difference with respect to the case of a finite action space. We can only prove that $\mu^{\varphi} \leq \mu$ on $\mathcal{B}(V^c_A \times A)$.

Theorem 5.1. *Suppose that assumptions (B1)–(B3) are satisfied. Consider $\mu \in \mathbb{M}(X \times A)_+$ satisfying the linear equation (2.6), and let φ be its induced randomized stationary policy. Then $\mu^{\varphi}_A \leq \mu_A$ and $\mu^{\varphi} \leq \mu$ on $\mathcal{B}(V^c_A \times A)$. Moreover, if g is an $\mathbb{R}_+$-valued measurable function defined on $X \times A$ such that $g(x,\cdot)$ is continuous on A for any $x \in X$ then*

$$\int_{V_A \times A} g(y,a)\mu^{\varphi}(\mathrm{d}y \times \mathrm{d}a) \leq \int_{V_A \times A} g(y,a)\mu(\mathrm{d}y \times \mathrm{d}a). \tag{5.1}$$

Proof. Introduce the mapping T defined on $\mathbb{M}(X)_+$ by

$$T\gamma(\Gamma) = \nu(\Gamma) + \int_{X \times A} Q(\Gamma \mid y,a)\varphi(\mathrm{d}a \mid y)\gamma(\mathrm{d}y).$$

Define the sequence of measures $(\gamma_n)_{n \in \mathbb{N}}$ in $\mathbb{M}(X)_+$ by $\gamma_0 = 0$ and $\gamma_{n+1} = T\gamma_n$ for $n \in \mathbb{N}$. Since the operator T is monotone, it follows easily by induction that the sequence $(\gamma_n)_{n \in \mathbb{N}}$ is increasing and converges to μ^{φ}_A which is the minimal positive solution of $\gamma = T\gamma$ with $\gamma \in \mathbb{M}(X)_+$.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Now, let us show that $T\mu_A \le \mu_A$. Indeed, by the definition of φ, for any $\Gamma \in \mathcal{B}(X)$,

$$T\mu_A(\Gamma) = \nu(\Gamma) + \int_{V_A^c \times A} Q(\Gamma \mid y, a)\varphi(\mathrm{d}a \mid y)\mu_A(\mathrm{d}y) + \int_{V_A} Q(\Gamma \mid y, f(y))\mu_A(\mathrm{d}y). \tag{5.2}$$

However, since $\mu(\Lambda) = \int_\Lambda \varphi(\mathrm{d}a \mid x)\mu_A(\mathrm{d}x)$ for any $\Lambda \in \mathcal{B}(V_A^c \times A)$, we have, for any $\Gamma \in \mathcal{B}(X)$,

$$\mu_A(\Gamma) = \nu(\Gamma) + \int_{V_A^c \times A} Q(\Gamma \mid y, a)\varphi(\mathrm{d}a \mid y)\mu_A(\mathrm{d}y) + \int_{V_A \times A} Q(\Gamma \mid y, a)\mu(\mathrm{d}y \times \mathrm{d}a). \tag{5.3}$$

We study two cases. Assume first that $\int_{V_A \times A} Q(\Gamma \mid y, a)\mu(\mathrm{d}y \times \mathrm{d}a) = \infty$. Then $\mu_A(\Gamma) = \infty$ and we clearly have $T\mu_A(\Gamma) \le \mu_A(\Gamma)$. Now suppose that $\int_{V_A \times A} Q(\Gamma \mid y, a)\mu(\mathrm{d}y \times \mathrm{d}a) < \infty$. From assumption (B3) and Lemma 5.3, taking into account the fact that $\mu_A \ll q$, it follows that

$$\int_{V_A} Q(\Gamma \mid y, f(y))\mu_A(\mathrm{d}y) = 0, \tag{5.4}$$

and so combining (5.2), (5.3), and (5.4), we have $T\mu_A \le \mu_A$.

Consequently, it can be shown easily by induction that $\gamma_n \le \mu_A$, implying that $\mu_A^\varphi \le \mu_A$. Therefore, according to the definition of φ (see Definition 5.1),

$$\mu(\Lambda) = \int_\Lambda \varphi(\mathrm{d}a \mid y)\mu_A(\mathrm{d}x) \ge \int_\Lambda \varphi(\mathrm{d}a \mid y)\mu_A^\varphi(\mathrm{d}x) = \mu^\varphi(\Lambda)$$

for any $\Lambda \in \mathcal{B}(V_A^c \times A)$. Moreover, (5.1) is a straightforward consequence of Lemma 5.3. This completes the proof.

Given the measures μ and μ^φ, we now give the definition of their difference.

Definition 5.2. Suppose that assumptions (B1)–(B3) are satisfied. Consider $\mu \in \mathbb{M}(X \times A)_+$ satisfying the linear equation (2.6), and let φ be its induced randomized stationary policy. Define the mapping Δ^μ on $\mathcal{B}(X \times A)$ by

$$\Delta^\mu(\Lambda) = \mu(\Lambda \cap [V_A \times A]) + (\mu - \mu^\varphi)(\Lambda \cap [V_A^c \times A]), \tag{5.5}$$

where $\Lambda \in \mathcal{B}(X \times A)$ (see Definition 2.1).

The dependence of Δ^μ on φ is not explicitly mentioned because φ itself depends on μ. Observe that, for any $\Lambda \in \mathcal{B}(V_A^c \times A)$, $(\mu - \mu^\varphi)(\Lambda)$ is well defined since, from Proposition 5.1, the measure μ is σ-finite on $\mathcal{B}(V_A^c \times A)$ and $\mu^\varphi \le \mu$ on $\mathcal{B}(V_A^c \times A)$. Our next result presents some technical properties of Δ^μ.

Lemma 5.4. *Suppose that assumptions (B1)–(B3) are satisfied. Consider $\mu \in \mathbb{M}(X \times A)_+$ satisfying the linear equation (2.6), and let φ be its induced randomized stationary policy. If g is an $\mathbb{R}_+$-valued measurable function defined on $X \times A$ such that $g(x, \cdot)$ is continuous on A for any $x \in X$ then*

$$\mu(g) = \mu^\varphi(g) + \Delta^\mu(g).$$

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Proof. Consider $\Lambda \in \mathcal{B}(V_A^c \times A)$. Then we have $\mu(\Lambda) = \mu^\varphi(\Lambda) + \Delta^\mu(\Lambda)$ by the definition of Δ^μ and Theorem 5.1, which implies that

$$\int_{V_A^c \times A} g(x,a)\mu(\mathrm{d}x,\ \mathrm{d}a) = \int_{V_A^c \times A} g(x,a)\mu^\varphi(\mathrm{d}x,\ \mathrm{d}a) + \int_{V_A^c \times A} g(x,a)\Delta^\mu(\mathrm{d}x,\ \mathrm{d}a). \quad (5.6)$$

Now, according to (5.5),

$$\int_{V_A \times A} g(x,a)\mu(\mathrm{d}x,\ \mathrm{d}a) = \int_{V_A \times A} g(x,a)\Delta^\mu(\mathrm{d}x,\ \mathrm{d}a).$$

Combining the above equality and Lemma 5.3, we obtain

$$\int_{V_A \times A} g(x,a)\mu(\mathrm{d}x,\ \mathrm{d}a) = \int_{V_A \times A} g(x,a)\mu^\varphi(\mathrm{d}x,\ \mathrm{d}a) + \int_{V_A \times A} g(x,a)\Delta^\mu(\mathrm{d}x,\ \mathrm{d}a). \quad (5.7)$$

From (5.6) and (5.7), the result follows.

Corollary 5.1. *Suppose that assumptions (B1)–(B3) are satisfied. Consider $\mu \in \mathbb{M}(X \times A)_+$ satisfying the linear equation (2.6), and let φ be its induced randomized stationary policy. For any $\Gamma \in \mathcal{B}(X)$, we have*

$$\Delta_A^\mu(\Gamma) \geq \Delta^\mu Q(\Gamma). \quad (5.8)$$

Moreover, if $\mu_A^\varphi(\Gamma) < \infty$ or $\Gamma \subset V_A^c$ then we have an equality in the above formula.

Proof. Let $\Gamma \in \mathcal{B}(X)$. According to assumption (B3), the mapping $Q(\Gamma \mid \cdot, \cdot)$ satisfies the hypothesis of Lemma 5.4 and so $\mu Q(\Gamma) = \mu^\varphi Q(\Gamma) + \Delta^\mu Q(\Gamma)$. Adding $\nu(\Gamma)$ to both sides of the above equality yields

$$\mu_A(\Gamma) = \mu_A^\varphi(\Gamma) + \Delta^\mu Q(\Gamma). \quad (5.9)$$

Applying Lemma 5.4 to $g = I_{\Gamma \times A}$ we have

$$\mu_A(\Gamma) = \mu_A^\varphi(\Gamma) + \Delta_A^\mu(\Gamma). \quad (5.10)$$

Consequently, if $\mu_A^\varphi(\Gamma) < \infty$ then combining (5.9) and (5.10) we get an equality in (5.8). Now, if $\mu_A^\varphi(\Gamma) = \infty$ and $\Gamma \in \mathcal{B}(V_A^c)$, then, by using the fact that μ_A^φ is σ-finite on V_A^c, this case reduces to the previous case, and so the equality holds in (5.8). Finally, if $\mu_A^\varphi(\Gamma) = \infty$ and $\Gamma \in \mathcal{B}(V_A)$, then, from Theorem 5.1 and by the definition of Δ^μ, we obtain $\infty = \mu_A^\varphi(\Gamma) \leq \mu_A(\Gamma) = \Delta_A^\mu(\Gamma)$, implying that inequality (5.8) is satisfied, showing the result.

Lemma 5.5. *Suppose that assumptions (B1)–(B3) are satisfied. Consider $\mu \in \mathbb{M}(X \times A)_+$ satisfying the linear equation (2.6), and let φ be its induced randomized stationary policy. Assume that $\gamma \in \mathbb{M}(X \times A)_+$ and satisfies $\gamma_A \geq \mu_A^\varphi$ and $\gamma_A = \nu + \gamma Q$. Then the measure $\lambda = \gamma + \Delta^\mu$ satisfies $\lambda_A = \nu + \lambda Q$.*

Proof. The proof of this result is similar to that of Lemma 4.3 by using Corollary 5.1 instead of Lemma 4.2.

In the next theorem and corollary, we present our main results.

Theorem 5.2. *Suppose that Assumption B holds. If μ^* is an optimal solution of the constrained linear program $\mathbb{LP}$ then the induced policy φ^* is a solution of the constrained control problem defined by (2.3) and the optimal value $V^*(\nu)$ coincides with the optimal value of the $\mathbb{LP}$.*

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Proof. For any policy $\pi \in \Pi$, the corresponding occupation measure satisfies the linear equation (2.6) by Lemma 9.4.3 of [10]. The value of the constrained control problem defined in (2.3) cannot be smaller than the value of the $\mathbb{LP}$ given by $\mu^*(r_0)$. We will show that $\mu^*(r_\theta) \geq \mu^{\varphi^*}(r_\theta)$ for any $\theta \in \Theta \cup \{0\}$. Suppose that, for some $\theta \in \Theta$, $\mu^*(r_\theta) < \mu^{\varphi^*}(r_\theta)$. Then introduce Δ^{μ^*} according to (5.5). From assumption (B3), $-\infty < \mu^*(r_\theta)$. Moreover, since μ^* is a solution of the constrained linear program $\mathbb{LP}$, $\mu^*(r_\theta) \leq R_\theta$ and so $\mu^*(r_\theta)$ is finite. According to assumption (B4) and Lemma 5.4, $\mu^*(r_\theta^+) = \mu^{\varphi^*}(r_\theta^+) + \Delta^{\mu^*}(r_\theta^+)$ and $\mu^*(r_\theta^-) = \mu^{\varphi^*}(r_\theta^-) + \Delta^{\mu^*}(r_\theta^-)$, implying that $\mu^{\varphi^*}(r_\theta)$ and $\Delta^{\mu^*}(r_\theta)$ are finite. Therefore, we obtain $\mu^*(r_\theta) = \mu^{\varphi^*}(r_\theta) + \Delta^{\mu^*}(r_\theta)$ and so $\Delta^{\mu^*}(r_\theta) < 0$ since it is assumed that $\mu^*(r_\theta) < \mu^{\varphi^*}(r_\theta)$. From Theorem 5.1 and Lemma 5.5, the measure λ^n defined by $\lambda^n = \mu^* + n\Delta^{\mu^*}$ for $n \in \mathbb{N}$ satisfies $\lambda^n_A = \nu + \lambda^n Q$. Therefore, $\lambda^1(r_\theta) = \mu^*(r_\theta) + \Delta^{\mu^*}(r_\theta)$ is well defined and finite and it can easily be shown by induction that $\lambda^n(r_\theta) = \mu^*(r_\theta) + n\Delta^{\mu^*}(r_\theta)$ is well defined and finite for any $n \in \mathbb{N}$. This implies that the value of the auxiliary unconstrained linear program $\mathbb{LP}_\theta$ is smaller than $\lambda^n(r_\theta) = \mu^*(r_\theta) + n\Delta^{\mu^*}(r_\theta)$ for any $n \in \mathbb{N}$ and, thus, equals $-\infty$, leading to a contradiction with assumption (B3). Therefore, for any $\theta \in \Theta$, $\mu^{\varphi^*}(r_\theta) \leq R_\theta$. Now, it is also obvious that $\mu^*(r_0) \leq \mu^{\varphi^*}(r_0)$, and it remains to prove that $\mu^*(r_0) = \mu^{\varphi^*}(r_0)$. If $\mu^*(r_0) = +\infty$ then necessarily $\mu^{\varphi^*}(r_0) = \mu^*(r_0)$. If $\mu^*(r_0) < +\infty$ then, since $\mu^*(r_0) > -\infty$, one can apply the same reasoning presented above for $\theta \in \Theta$ to show that $\mu^*(r_0) = \mu^{\varphi^*}(r_0)$. This completes the proof.

Corollary 5.2. *Suppose that Assumption B holds. 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 5.2.

6. Examples

6.1. Example 1

This example shows that assumption (A2) is important in Theorem 4.2 and, similarly, assumption (B4) is important in Theorem 5.2.

We consider an uncontrolled model ($A = \{a\}$, dummy action) with a single constraint: $\Theta = \{1\}$; $R_1 = 1$. The state space is $X = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$ and the stochastic kernel is defined by $Q(0 \mid 0, a) = 1$, $Q(1 \mid -1, a) = 1$, and $Q(i+1 \mid i, a) = 1$ for all $i \neq 0, -1$. Take $r_0(i, a) = r_1(i, a) = 0$ for all $i \leq 0$ and $r_0(i, a) = -r_1(i, a) = -(\frac{1}{2})^i$ for all $i > 0$ and choose $\nu(0) = 1$.

Obviously, $x_t \equiv 0$ and $v(\nu, \pi) = v_1(\nu, \pi) = 0$ for the (single) control policy π. The corresponding occupation measure is given by $\mu^\pi(0, a) = \infty$ and $\mu^\pi(i, a) = 0$ for $i \neq 0$. Observe that in this example $\mu_A(i) = \mu(i, a)$, which will be denoted by $\mu(i)$ for notational convenience. The linear program $\mathbb{LP}$ has the form

$$\text{minimize } -\sum_{j \geq 1} \mu(j) \left(\frac{1}{2}\right)^j$$

subject to

$$\sum_{j \geq 1} \mu(j) \left(\frac{1}{2}\right)^j \leq R_1 = 1,$$

$$\mu(i) = \mu(i-1) \quad \text{for all } i < 0 \text{ and } i > 1,$$

$$\mu(1) = \mu(-1), \qquad \mu(0) = 1 + \mu(0).$$

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Clearly, $\mu(0) = \infty$ and $\mu(1) = \mu(2) = \cdots = M$, where one can take any $M \geq 0$. The optimal solution to $\mathbb{LP}$ is given by $M = 1$, the minimal value is -1, but this solution does not correspond to any control policy. In this example, the value of $\mathbb{LP}_0$ equals $-\infty$, corresponding to $M = +\infty$.

6.2. Example 2

This example shows that condition (B2) and the continuity conditions (B3) and (B5) are crucial in Theorem 5.2. At the same time, in Theorem 4.2, it is critical that the control set A is finite.

We consider the unconstrained case: $\Theta = \varnothing$. The state space is $X = \{0, 1, 2, \ldots\}$, the action space is $A = \{1, 2, \ldots\}$, the stochastic kernel is defined by $Q(a \mid 0, a) = (\frac{1}{2})^{a-1}$, $Q(0 \mid 0, a) = 1 - (\frac{1}{2})^{a-1}$ for all $a \in A$, and $Q(0 \mid i, a) \equiv 1$ for all $i > 0$, $a \in A$. Take $r_0(0, a) = (\frac{1}{2})^{a-1}$ and $r_0(i, a) = -1$ for all $i > 0$, $a \in A$, and choose $\nu(0) = 1$. This model is presented in Figure 2. Observe that, for any $\mu \in \mathbb{M}(X \times A)_+$, $\mu(r_0^+) = \sum_{a \in A} \mu(0, a)(\frac{1}{2})^{a-1}$ and $\mu(r_0^-) = \sum_{i \geq 1} \mu_A(i)$. Consequently, $\mu(r_0)$ is well defined if $\sum_{a \in A} \mu(0, a)(\frac{1}{2})^{a-1} < \infty$ or $\sum_{i \geq 1} \mu_A(i) < \infty$. For such feasible measures, the linear program $\mathbb{LP}$ has the form

$$\text{minimize } \mu(r_0) = \sum_{a \in A} \mu(0, a)\left(\frac{1}{2}\right)^{a-1} - \sum_{i \geq 1} \mu_A(i)$$

subject to

$$\mu_A(0) = 1 + \sum_{a \in A} \mu(0, a)\left[1 - \left(\frac{1}{2}\right)^{a-1}\right] + \sum_{i \geq 1} \mu_A(i),$$

$$\mu_A(i) = \left(\frac{1}{2}\right)^{i-1} \mu(0, i), \qquad i \geq 1.$$

These equations show that $\mu(r_0)$ is well defined if and only if $\sum_{a \in A} \mu(0, a)(\frac{1}{2})^{a-1} < \infty$. One can easily see that in this case $\mu_A(0) = 1 + \mu_A(0)$ and so $\mu_A(0) = \infty$. The particular values $\mu(i, a)$ for $i \geq 1$ are of no importance; one can take $\mu(i, a) = \delta_i(\{a\})\mu_A(i) = \delta_i(\{a\})(\frac{1}{2})^{i-1}\mu(0, i)$. Moreover, it is easy to see that $\mu(r_0) = 0$ and so the optimal value of $\mathbb{LP}$ equals zero.

For this problem, any stationary policy is unfeasible. Indeed, if $\varphi(a \mid 0) > 0$ then $\mu(r_0^+) = \mu(r_0^-) = \infty$ and so $\mu(r_0)$ is not well defined since the running costs $+(\frac{1}{2})^{a-1}$ and -1 appear

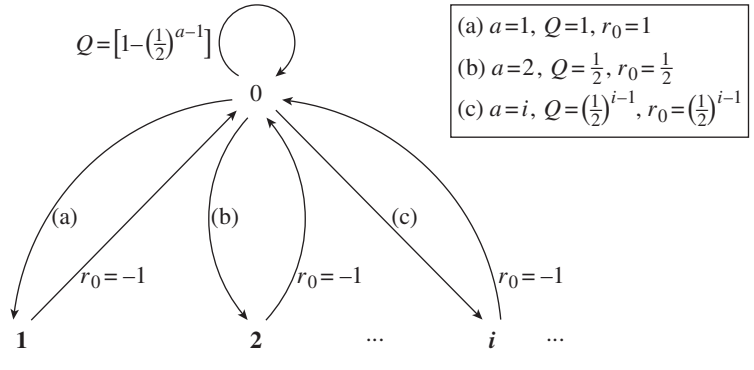

FIGURE 2: Illustration to Example 2.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

infinitely many times with positive probability. On the one hand, it is clear that $V^*(\nu) \geq 0$ since the optimal value of $\mathbb{LP}$ equals 0, and on the other hand, one can check that $v(\nu, \pi^*) = 0$, where

$$\pi^*(a \mid h_t) = \begin{cases} 1 & \text{if } x_t = 0 \text{ and } a \text{ equals the number of 0s in the history } h_t, \\ & \text{or if } a = x_t \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

For this policy, the occupation measure is given by $\mu^{\pi^*}(0, a) = 1$ and $\mu^{\pi^*}(i, a) = \delta_i(\{a\})(\frac{1}{2})^{i-1}$ for any $(i, a) \in \mathbb{N} \times A$.

In this example, one cannot construct the induced policy φ^* for the measure $\mu^* = \mu^{\pi^*}$ (and for any other feasible measure) because it is not nearly perfect. Assumption (A1) holds for instance for the geometric distribution $q(x) = (\frac{1}{2})^{x+1}$ on the state space X, but Lemma 4.4 does not hold because the action space A is not finite. Indeed, according to Definition 4.1, all sets V_a should be empty, because $\mu^*_A(i) > 0$ and $\mu^*_a(i) = \mu^*(i, a) < \infty$ for all $i \in X$, $a \in A$. However, the measure μ^*_A is not σ-finite on X since $\mu^*_A(0) = \infty$.

If we consider the discrete topology for the spaces X and A then assumptions (B3) and (B5) hold, but Theorem 5.2 cannot be applied because one cannot construct the induced policy since Lemma 5.2 requires the action space A to be compact. According to Remark 5.1, one can still construct a partition for the action space A as in Proposition 5.1, but we cannot ensure that each element of the partition $(A_{n,j})_{j \in \mathbb{N}_{M_n}}$ of the action space A will have a diameter less than $1/2^{n-1}$. Namely, for an arbitrary fixed $n \in N$, we define $M_n = n$. For $n > 1$, put $A_{n,j} = \{j\}$ for $j \in \mathbb{N}_{n-1}$ and $A_{n,n} = A_{n,M_n} = \{n, n+1, n+2, \ldots\}$. If $n = 1$, put $A = A_{1,1}$. The corresponding subsets $\widehat{V}_{n,j}$, which must be $q/\mu^*_{A_{n,j}}$-nearly perfect, are all empty except for $V_A = \widehat{V}_{n,M_n} = \{0\}$ because the geometric distribution takes positive values and $\mu^*_{A_{n,j}}(X) = 1 + (\frac{1}{2})^{j-1} < \infty$ for all $j < M_n$. On the other hand, $\mu^*_{A_{n,M_n}}(\{0\}) = \infty$. The measure $\mu^*_{A_{n,j}}$ is finite on X for all $j \in \mathbb{N}_{M_n - 1}$. The measure $\mu^*_{A_{n,M_n}}$ is σ-finite on $\widehat{V}^c_{n,M_n} = V^c_A = \{1, 2, \ldots\}$ and μ^*_A is also σ-finite on V^c_A. We can introduce the mapping f_n on $V_A = \{0\}$ as $f_n(0) = n \in \widehat{V}_{n,M_n}$, but the limit $\lim_{n\to\infty} f_n(0)$ does not exist.

A way to overcome these difficulties is to use a one-point compactification of the action space A. We add the point 0 to A and fix the following metric:

$$\rho(a, b) = \left| I_{\mathbb{N}}(a)\left(\tfrac{1}{2}\right)^a - I_{\mathbb{N}}(b)\left(\tfrac{1}{2}\right)^b \right|.$$

Assumptions (B3) and (B5) are satisfied if we put $Q(0 \mid x, 0) = 1$ for all $x \in X$. Now Lemma 5.2 and Theorem 5.2 can be applied. Starting from the measure μ^* (or from any other feasible measure), one can build the induced policy φ^*: $\lim_{n\to\infty} f_n(0) = 0$, so that $\varphi(\cdot \mid x) = \delta_x(\cdot)$ (see Definition 5.1). This policy is optimal for the problem (2.3).

6.3. Example 3

The following meaningful example describes the process of selling a property. Suppose that a landlord plans to sell the house and, once a month, receives the offers from the random market, denoted as $1, 2, \ldots, M$. The actual value of offer i is $f(i)$ (thousand pounds). We assume that the offers change according to an (uncontrolled) Markov chain with transition matrix $P = (p_{ij})$, $i, j = 1, 2, \ldots, M$. If a tenant currently rents the house, the landlord is not allowed to sell it, but the tenant may leave before the next offer with probability p_l. In case there is no tenant and the landlord does not accept the current offer, he/she can wait until the next month or can invite a new tenant who will appear by the next month with probability p_a.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Every month, the landlord pays the maintenance fee $c \geq 0$ (thousand pounds); the monthly revenue from a tenant is $d \geq 0$ (thousand pounds). Assume that the expected time for the whole selling period should not exceed a fixed constant R_1 (months). Moreover, the total maintenance-revenue cost should not exceed R_2 (thousand pounds). The goal is to elaborate a selling policy maximizing the expected value of the accepted offer under the imposed constraints. 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. A version of this example also appears in [7], where a pure positive MDP was studied.

To formulate the MDP, we introduce the state space $X = \{(i, N), (j, Y),\ i, j = 1, 2, \ldots, M\} \cup \{\Delta\}$, where the components $i,\ j$ represent the current offer and the letter N (respectively Y) means that there is no tenant currently (respectively there is a tenant). 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', w means 'wait'. The transition kernel is given by

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

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

The cost and constraint functions are defined by

$$\begin{aligned} r_0(\Delta, a) &= r_0((i, Y), a) = 0, \\ r_0((i, N), a) &= \begin{cases} f(i) & \text{if } a = s, \\ 0 & \text{otherwise,} \end{cases} \\ r_1(x, a) &= I_{\{x \neq \Delta\}}, \\ r_2(\Delta, a) &= r_2((i, N), s) = 0, \\ r_2((i, l), a) &= \begin{cases} c & \text{if } l = N,\ a \neq s, \\ c - d & \text{if } l = Y. \end{cases} \end{aligned}$$

Here $\Theta = \{1, 2\}$.

Below, we solve numerically the LP for the following data: $M = 6$; $f(i) = 100 + 20(i - 1)$:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \qquad p_l = 0.66, \qquad p_a = 0.2, \qquad c = 0.5, \qquad d = 1.$$

Suppose that $\nu((1, N)) = 1$, that is, there is no tenant and the first offer is 1 at the beginning.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

Remark 6.1. 1. One can easily reformulate the problem as a minimization problem by changing the sign of the function r_0.

2. Assumption A is satisfied: one can take q as the uniform distribution; the values of the linear programs $\mathbb{LP}_0$ and $\mathbb{LP}_1$ are obviously finite (the maximum of $\mathbb{LP}_0$ equals $f(6) = 200$, the minimum of $\mathbb{LP}_1$ equals 1). The value of $\mathbb{LP}_2$ equals 0 because the expected maintenance-revenue cost coming from r_2 increases with time.

3. In the case $p_l < p_a$ the expected maintenance-revenue cost decreases with time and the value of $\mathbb{LP}_2$ is $-\infty$. Therefore, Assumption A is not satisfied and the results of the present paper cannot be used.

Let $R_1 = 24$ and $R_2 = 10$. Then $\mu^*(r_0) = 167.1$ (thousand pounds), $\mu^*(r_1) = 24$ (months), and $\mu^*(r_2) = 10$ (thousand pounds), so that both constraints are active. The optimal policy solving the constrained control problem (see (2.3)) is as follows:

- accept offer numbers 5 and 6 if there is no tenant,
- offer number 4 should be accepted with probability 0.31, or the landlord should simply wait with the complementary probability 0.69,
- reject offer numbers 2 and 3 and just wait,
- if the landlord receives offer number 1 and there is no tenant then he/she should invite a tenant with probability 0.82 and wait with the complementary probability 0.18.

The optimal occupation measures are presented in Table 1. The values of $\mu^*((6, N), s)$ and $\mu^*((6, Y), w)$ are negligible but positive (approximately 0.0001).

If we combine together all the monetary objectives, i.e. replace r_0 with $r_0 - r_2$, then, under the same constraint $\mu(r_1) \le R_1$ and with the same values of the parameters, the optimal policy prescribes to invite a tenant in states $(1, N)$, $(2, N)$, and $(3, N)$ and accept the offers in states $(5, N)$ and $(6, N)$. In the state $(4, N)$, one should invite a tenant with probability 0.57 or accept the offer with probability 0.43. Under this policy, $\mu^*(r_0) = 160.7$; $\mu^*(r_1) = 24$.

We emphasize that this model coming from a real-life situation is not transient: for the policy 'never accept the offer', the expected time to the absorption at cemetery Δ equals $+\infty$. Of course, this policy 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. Moreover, since the cost function r_2 takes positive and negative values, and the running cost r_0, to be maximized, is positive, the results of [7] are also not applicable here.

TABLE 1.

a	x											
	$(1, N)$	$(2, N)$	$(3, N)$	$(4, N)$	$(5, N)$	$(6, N)$	$(1, Y)$	$(2, Y)$	$(3, Y)$	$(4, Y)$	$(5, Y)$	$(6, Y)$
s	0	0	0	0.643	0.357	0	0	0	0	0	0	0
t	4.950	0	0	0	0	0	0	0	0	0	0	0
w	1.092	8.686	5.352	1.420	0	0	0.672	0.742	0.077	0.008	0.001	0

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press

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. and Shreve, S. E. (1978). *Stochastic Optimal Control* (Math. Sci. Eng. **139**). Academic Press, New York.

[4] Borkar, V. S. (1991). *Topics in Controlled Markov Chains* (Pitman Res. Notes Math. Ser. **240**). Longman Scientific & Technical, Harlow.

[5] Borkar, V. S. (2002). Convex analytic methods in Markov decision processes. In *Handbook of Markov Decision Processes* (Internat. Ser. Operat. Res. Manag. Sci. **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] Dufour, F., Horiguchi, M. and Piunovskiy, A. B. (2012). The expected total cost criterion for Markov decision processes under constraints: a convex analytic approach. *Adv. Appl. Prob.* **44,** 774–793.

[8] Filar, J. and Vrieze, K. (1997). *Competitive Markov Decision Processes*. Springer, New York.

[9] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes* (Appl. Math. **30**). Springer, New York.

[10] Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes* (Appl. Math. **42**). Springer, New York.

[11] Horiguchi, M. (2001). Markov decision processes with a stopping time constraint. *Math. Meth. Operat. Res.* **53,** 279–295.

[12] Horiguchi, M. (2001). Stopped Markov decision processes with multiple constraints. *Math. Meth. Operat. Res.* **54,** 455–469.

[13] Piunovskiy, A. B. (1997). *Optimal Control of Random Sequences in Problems with Constraints* (Math. Appl. **410**). Kluwer Academic, Dordrecht.

[14] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley, New York.

[15] Sennott, L. I. (1999). *Stochastic Dynamic Programming and the Control of Queueing Systems*. John Wiley, New York.

https://doi.org/10.1239/aap/1377868541 Published online by Cambridge University Press