J. Appl. Prob. **47,** 947–966 (2010)
Printed in England
© *Applied Probability Trust* 2010

MULTIOBJECTIVE STOPPING PROBLEM FOR DISCRETE-TIME MARKOV PROCESSES: CONVEX ANALYTIC APPROACH

F. DUFOUR,[*] *Université Bordeaux I*
A. B. PIUNOVSKIY,[**] *University of Liverpool*

Abstract

The purpose of this paper is to study an optimal stopping problem with constraints for a Markov chain with general state space by using the convex analytic approach. The costs are assumed to be nonnegative. Our model is not assumed to be transient or absorbing and the stopping time does not necessarily have a finite expectation. As a consequence, the occupation measure is not necessarily finite, which poses some difficulties in the analysis of the associated linear program. Under a very weak hypothesis, it is shown that the linear problem admits an optimal solution, guaranteeing the existence of an optimal stopping strategy for the optimal stopping problem with constraints.

Keywords: Optimal stopping; constrained control; Markov chain; linear program; convex analytic approach; Markov decision process; occupation measure

2010 Mathematics Subject Classification: Primary 60J10

1. Introduction

The purpose of this paper is to study an optimal stopping problem with constraints for a Markov chain with general state space by using the convex analytic approach. This method has proved to be very effective for solving constrained versions of different optimal control problems. Without attempting to present an exhaustive literature review, the interested reader may consult the surveys [15], [18], and the references therein to obtain a rather complete view of this research field. The key idea is to reformulate the control problem as a primal linear program (PLP) in a space of measures in which the main object of interest is the occupation measure of the controlled process. This approach is well developed for discounted Markov decision processes (MDPs) and for MDPs with long-run average rewards [1], [11], [12], [17]. Work on MDPs with total undiscounted rewards has received less attention, although investigations on transient and absorbing models with discrete state space have been treated in a recent book [1]. Of particular importance is the fact that, for absorbing models, occupation measure corresponding to each control policy is finite, while for transient models, it takes finite values on singletons. The convex analytic approach has also been used for investigating many other models, such as continuous-time Markov chains and more general Markov processes, including diffusions [5], [9], [19]. More particularly, optimal stopping problems for continuous-time Markov processes are shown in [5] to be equivalent to infinite-dimensional linear programs over a space of pairs

Received 24 January 2010; revision received 9 October 2010.
Research partially supported by the Royal Society, grant no. TG091905.

[*] Postal address: Université Bordeaux I and INRIA Bordeaux Sud Ouest, Institut de Mathématiques de Bordeaux, 351 cours de la Liberation, 33405 Talence Cedex, France. Email address: dufour@math.u-bordeaux1.fr
[**] Postal address: Department of Mathematical Sciences, University of Liverpool, L69 7ZL, Liverpool, UK. Email address: piunov@liv.ac.uk

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

of measures under the assumption that the expectation of the stopping time is finite. In [19], a simple standard continuous-time controlled Markov chain is studied using the notion of an occupation measure. A similar approach, but for a more general model, is presented in [9], where it is also proved that the convex analytic approach is dual to that of dynamic programming; namely these two approaches lead to a dual pair of linear programs.

On the other hand, there exist very few works devoted to the development of the convex analytic approach for the optimal stopping of a discrete-time Markov chain. To the best of the authors' knowledge, [13] and [14] are the only references addressing such a problem. In these works, the author investigated the case of a finite state space under conditions which guarantee that, formally speaking, the model is absorbing, and, for each control policy, the expected value of the stopping time is finite. We may also mention other references [3], [4], [6], [16], [21], in which single-objective optimal stopping problems are investigated using the dynamic programming approach. In particular, a linear programming formulation of the optimal stopping problem for MDPs is approximated using linear function approximation in [3]. In the book [4], the authors studied optimal stopping problems both in the discrete-time and continuous-time frameworks in the context of economics and finance applications. In [6], the authors investigated an optimal stopping problem for the class of piecewise-deterministic Markov processes and provided a numerical approximation scheme. In [16], the solution of the optimal stopping problem for processes with independent increments is analyzed. It is shown that the solution is related to the root of the Appell function associated with the maximum of the process. Finally, nonstationary stopped decision processes are studied in [21] using operator theory, rather than martingale theory.

Our work appears to be the first attempt to study the constrained version of optimal stopping of a discrete-time Markov chain with general state space by using the convex analytic approach. It can be shown that the optimal stopping problem is equivalent to an MDP with a total undiscounted cost. Therefore, this equivalence result provides a natural way to analyze the optimal stopping problem by using the convex analytic approach. In the current paper, all the costs are assumed to be nonnegative and there is at least one policy with a finite (vector) performance satisfying the required inequality constraints. It is important to point out that in our case we do not assume that the stopping time has a finite expectation. Moreover, the equivalent MDP is not necessarily transient or absorbing. As a result, the occupation measure is not necessarily finite, which renders the PLP very difficult to analyze. Moreover, admissible solutions to the PLP can be phantom solutions, i.e. they do not correspond to any control policy. This means that the linear equation on the space of measures, which usually completely characterizes the space of occupation measures, defines a wider space in our case.

In Section 2 we formulate the optimal stopping problem and show that it is equivalent to an MDP with total expected cost. In Section 3 we study the PLP related to the unconstrained optimal stopping problem through the MDP. We show how to construct an optimal policy from an optimal solution to the PLP. Moreover, an important property is derived showing that, for any admissible solution to the PLP, there exists a policy for which the associated occupation measure provides an admissible solution to the PLP with a better value of the cost. This result will be crucial for the analysis of the constrained problem provided in Section 5. An example is presented in Section 4 illustrating all the theoretical issues and, in particular, the existence of phantom solutions to the PLP. Finally, Section 5 is dedicated to the analysis of the constrained version of the optimal stopping problem. Contrary to the unconstrained case, it is far from trivial to show that the PLP with constraints admits an optimal solution. The main result of this section is to prove that this result holds under a very weak condition guaranteeing the existence

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

of an optimal stopping strategy for the constrained version of the optimal stopping problem. This is done by introducing a suitable topology on a space of occupation measures. For the sake of clarity, most of the proofs of our results are presented in Appendix A.

The following notation will be used in this paper: $\mathbb{N}$ denotes the set of natural numbers, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{R}_+$ denotes the set of nonnegative real numbers, and $\bar{\mathbb{R}}_+$ denotes $\mathbb{R}_+ \cup \{+\infty\}$. The term *measure* will always refer to a countably additive, $\bar{\mathbb{R}}_+$-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 $\lim_{n\to\infty} \gamma_n = \gamma_\infty$. Let f be a measurable function defined on E and $\eta \in \mathbb{M}(E)_+$. The integral of f with respect to η is denoted by $\eta(f) = \int_E f(y)\eta(\mathrm{d}y)$. Recall that if W_1 and W_2 are positive kernels on E given E, the *product* of W_1 and W_2 is defined by $W_1W_2(B \mid x) = \int_E W_2(B \mid y)W_1(\mathrm{d}y \mid x)$ for any $(x, B) \in E \times \mathcal{B}(E)$. For a kernel W on E given E, the iterates W^n for $n \in \mathbb{N} \cup \{0\}$ are defined by setting $W^0(x, B) = \delta_x(B)$ for any $(x, B) \in E \times \mathcal{B}(E)$, and, iteratively, $W^n = WW^{n-1}$. For any nonnegative measurable function f on E, Wf is the measurable function defined on E by $Wf(x) = \int_E f(y)W(\mathrm{d}y \mid x)$ for any $x \in E$. For a positive measure η on $(E, \mathcal{B}(E))$, ηW is the measure defined on $(E, \mathcal{B}(E))$ by $\eta W(B) = \int_E W(B \mid y)\eta(\mathrm{d}y)$ for any $B \in \mathcal{B}(E)$. The restriction on a set $B \in \mathcal{B}(E)$ of a measure η is denoted by $\eta^B(C) = \eta(B \cap C)$ for any $C \in \mathcal{B}(E)$.

2. Problem formulation

In this section we describe the optimal stopping problem using a *weak* formulation. For a weak formulation of the optimal stochastic control problem, we refer the reader to [7], [8], and [10]. We then introduce an auxiliary control problem defined in terms of an MDP and show that it is equivalent to the optimal stopping problem (see Theorem 2.1). Some basic definitions related to the occupation measures of the MDP are also presented.

2.1. Optimal stopping

Let E be a Borel space, let S be a stochastic kernel on E given E, and let ν be an arbitrary probability measure on E. In this subsection we define the optimal stopping problem for a Markov chain $\{x_t\}$ with state space E generated by the Markov kernel S. The objective is to stop the process at a random time τ in order to minimize a cost function in the presence of constraints.

Definition 2.1. The control is defined by

$$\lambda = (\Omega, \mathcal{F}, \mathrm{Q}, \{\mathcal{F}_t\}_{t\in\mathbb{N}}, \{x_t\}_{t\in\mathbb{N}}, \tau)$$

- $(\Omega, \mathcal{F}, \mathrm{Q}, \{\mathcal{F}_t\}_{t\in\mathbb{N}})$ is a filtered probability space;
- $\{x_t\}_{t\in\mathbb{N}}$ is an E-valued $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-Markov chain defined on $(\Omega, \mathcal{F}, \mathrm{Q})$, where S is its associated transition kernel and ν is its initial distribution;
- τ is an $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-stopping time.

The set of previous controls is denoted by Λ and E_Q denotes the expectation under the probability Q.

For the sake of clarity, different notation will be used for the cost functions of the unconstrained and constrained versions of the optimal stopping problem.

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

The unconstrained case. Let $r\colon E \to \mathbb{R}_+$ and $R\colon E \to \mathbb{R}_+$ be measurable functions. For a control $\lambda \in \Lambda$, the performance criterion is given by

$$J(\lambda) = \mathrm{E}_{\mathrm{Q}}\Bigg[\sum_{t=0}^{\tau-1} r(x_t) + \mathbf{1}_{\{\tau<\infty\}}\, R(x_\tau)\Bigg].$$

The unconstrained optimal stopping problem, labeled (P$_1$), we are interested in is to minimize $J(\lambda)$ over Λ.

The constrained case. Let $(r_n)_{n=0,\dots,N}$ and $(R_n)_{n=0,\dots,N}$ be $\mathbb{R}_+$-valued measurable functions defined on E, and let $(j_n)_{n=0,\dots,N}$ be numbers in $\mathbb{R}_+$. For a control $\lambda \in \Lambda$, the performance criterion is given by

$$J_0(\lambda) = \mathrm{E}_{\mathrm{Q}}\Bigg[\sum_{t=0}^{\tau-1} r_0(x_t) + \mathbf{1}_{\{\tau<\infty\}}\, R_0(x_\tau)\Bigg],$$

and the constraints are given by

$$J_n(\lambda) = \mathrm{E}_{\mathrm{Q}}\Bigg[\sum_{t=0}^{\tau-1} r_n(x_t) + \mathbf{1}_{\{\tau<\infty\}}\, R_n(x_\tau)\Bigg].$$

The constrained optimal stopping problem, labeled (P$_2$), we are interested in is to minimize $J_0(\lambda)$ over Λ subject to $J_n(\lambda) \le j_n$ for $n \in \{1, \dots, N\}$.

2.2. Auxiliary control problem

Following the notation of [11], we introduce the control model (E^Δ, A, L), where $E^\Delta = E \times \{0, \Delta\}$ is the state space, $A = \{0, 1\}$ is the control set, and L is the stochastic kernel on E^Δ given $E^\Delta \times A$ defined by

$$L(B \times C \mid y, z, a) = S(B \mid y)[\mathbf{1}_{\{z\in C\}}\,\mathbf{1}_{\{a=0\}} + \mathbf{1}_{\{\Delta\in C\}}\,\mathbf{1}_{\{a=1\}}] \tag{2.1}$$

for all $B \in \mathcal{B}(E)$, $C \subset \{0, \Delta\}$, $(y, z) \in E^\Delta$, and $a \in A$.

Let Π be the set of all randomized past-dependent control policies $\pi = \{\pi_t\}_{t\in\mathbb{N}}$, where π_t is a stochastic kernel on the control set A given $(E^\Delta \times A)^{t-1} \times E^\Delta$. A randomized control policy $\pi = \{\pi_t\}_{t\in\mathbb{N}} \in \Pi$ is said to be stationary if $\pi_t = \pi^s$ for any $t \in \mathbb{N}$, where π^s is a stochastic kernel on the control set A given E^Δ. By a slight abuse of notation, if π^s is a stochastic kernel on the control set A given E^Δ, then the corresponding stationary control policy will be denoted by π^s instead of $\{\pi_t\}_{t\in\mathbb{N}}$ with $\pi_t = \pi^s$ for any $t \in \mathbb{N}$. Define $G = (E \times \{0, \Delta\} \times A)^\infty$, and let $\mathcal{G}$ be its associated product σ-algebra. According to [11, Section 2.2], for an arbitrary policy $\pi \in \Pi$, there exists a probability measure P^π_ν on $(G, \mathcal{G})$ such that the coordinate projections y_t, z_t, and a_t from G to the sets E, $\{0, \Delta\}$, and A, respectively, satisfy

(i) $\mathrm{P}^\pi_\nu[(y_0, z_0) \in B \times C] = \nu(B)\,\mathbf{1}_{\{0\in C\}}$;

(ii) $\mathrm{P}^\pi_\nu[a_t \in D \mid \mathcal{G}_t] = \pi_t(D \mid g_t)$;

(iii) $\mathrm{P}^\pi_\nu[(y_{t+1}, z_{t+1}) \in B \times C \mid \mathcal{G}_t \vee \sigma\{a_t\}] = L(B \times C \mid y_t, z_t, a_t)$,

for any $B \in \mathcal{B}(E)$, $C \subset \{0, \Delta\}$, and $D \subset A$, where $\mathcal{G}_t = \sigma\{g_t\}$ with $g_0 = (y_0, z_0)$ and $g_t = (y_0, z_0, a_0, \dots, y_{t-1}, z_{t-1}, a_{t-1}, y_t, z_t)$ for $t \ge 1$.

A probability on $(G, \mathcal{G})$ is said to be induced by a control policy $\pi \in \Pi$ if it satisfies (i)–(iii), and it will be denoted by P^π_ν. In this case, E^π_ν denotes the expectation under the probability P^π_ν.

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Introduce the cost functions $c\colon E^{\Delta} \times A \to \mathbb{R}_+$ and $c_n\colon E^{\Delta} \times A \to \mathbb{R}_+$ for the auxiliary control problem

$$\begin{aligned} c(y, z, a) &= [r(y)\,\mathbf{1}_{\{a=0\}} + R(y)\,\mathbf{1}_{\{a=1\}}]\,\mathbf{1}_{\{z=0\}}, \\ c_n(y, z, a) &= [r_n(y)\,\mathbf{1}_{\{a=0\}} + R_n(y)\,\mathbf{1}_{\{a=1\}}]\,\mathbf{1}_{\{z=0\}}, \end{aligned} \tag{2.2}$$

for any $(y, z) \in E^{\Delta}$ and $a \in A$. As for the optimal stopping problem, two different MDPs are introduced. The unconstrained version where the performance criterion to minimize is defined by

$$V(\pi) = \mathrm{E}_{\nu}^{\pi}\Bigg[\sum_{t=0}^{\infty} c(y_t, z_t, a_t)\Bigg] \tag{2.3}$$

and the constrained version where the performance criterion to minimize is defined by

$$V_0(\pi) = \mathrm{E}_{\nu}^{\pi}\Bigg[\sum_{t=0}^{\infty} c_0(y_t, z_t, a_t)\Bigg],$$

subject to $V_n(\pi) \leq j_n$ for $n \in \{1, \dots, N\}$ with

$$V_n(\pi) = \mathrm{E}_{\nu}^{\pi}\Bigg[\sum_{t=0}^{\infty} c_n(y_t, z_t, a_t)\Bigg].$$

Remark 2.1. Note that our model is clearly not assumed to be transient or absorbing and the stopping time does not necessarily have a finite expectation. The definitions of the transient and absorbing Markov control models can be found, for example, in [12, pp. 104–105] and [1, p. 75].

For a policy π, let us introduce the following expected occupation measures:

$$\mu_o^{\pi}(\Gamma) = \sum_{t=0}^{\infty} \mathrm{P}_{\nu}^{\pi}[(y_t, z_t, a_t) \in \Gamma \times \{0\} \times \{0\}], \tag{2.4}$$

$$\mu_{\tau}^{\pi}(\Gamma) = \sum_{t=0}^{\infty} \mathrm{P}_{\nu}^{\pi}[(y_t, z_t, a_t) \in \Gamma \times \{0\} \times \{1\}], \tag{2.5}$$

for any $\Gamma \in \mathcal{B}(E)$. In [13] and [14], the measures μ_o^{π} and μ_{τ}^{π} are called the running and stopped occupation measures, respectively.

Since $c(y, \Delta, a) = 0$ for any $(y, a) \in E \times A$, the performance criterion for the auxiliary control problem defined by (2.3) can be rewritten as

$$V(\pi) = \int_E c(y, 0, 0)\mu_o^{\pi}(\mathrm{d}y) + \int_E c(y, 0, 1)\mu_{\tau}^{\pi}(\mathrm{d}y) = \mu_o^{\pi}(r) + \mu_{\tau}^{\pi}(R). \tag{2.6}$$

Similarly, we have $V_n(\pi) = \mu_o^{\pi}(r_n) + \mu_{\tau}^{\pi}(R_n)$ for $n \in \{0, \dots, N\}$.

For notational convenience, the measure $\mu_o^{\pi} + \mu_{\tau}^{\pi}$ will be denoted by μ^{π}. We note that the probabilistic interpretation of μ^{π} is

$$\mu^{\pi}(\Gamma) = \sum_{t=0}^{\infty} \mathrm{P}_{\nu}^{\pi}[(y_t, z_t, a_t) \in \Gamma \times \{0\} \times A], \qquad \Gamma \in \mathcal{B}(E).$$

The next result shows that the constrained and unconstrained optimal stopping problems are respectively equivalent to the constrained and unconstrained MDPs previously defined.

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Theorem 2.1. *For any* $\lambda = (\Omega, \mathcal{F}, \mathrm{Q}, \{\mathcal{F}_t\}_{t\in\mathbb{N}}, \{x_t\}_{t\in\mathbb{N}}, \tau) \in \Lambda$, *there exists a policy* $\pi \in \Pi$ *such that, for all* $n \in \{0, \dots, N\}$,

$$V_n(\pi) = J_n(\lambda), \qquad V(\pi) = J(\lambda), \quad \textit{and} \quad \mu_o^\pi(E) = \mathrm{E}_{\mathrm{Q}}[\tau], \qquad \mu_\tau^\pi(E) = \mathrm{Q}[\tau < \infty].$$

Conversely, for any $\pi \in \Pi$, *there exists a control* $\lambda = (\Omega, \mathcal{F}, \mathrm{Q}, \{\mathcal{F}_t\}_{t\in\mathbb{N}}, \{x_t\}_{t\in\mathbb{N}}, \tau) \in \Lambda$ *such that, for all* $n \in \{0, \dots, N\}$,

$$J_n(\lambda) = V_n(\pi), \qquad J(\lambda) = V(\pi), \quad \textit{and} \quad \mu_o^\pi(E) = \mathrm{E}_{\mathrm{Q}}[\tau], \qquad \mu_\tau^\pi(E) = \mathrm{Q}[\tau < \infty].$$

Proof. See Appendix A.

3. Convex analytic approach for the unconstrained problem

Having shown that the optimal stopping problem can be reformulated as an MDP, we now study in this section the PLP related to the unconstrained problem. We show how to construct an optimal policy from an optimal solution to the PLP; see Proposition 3.1 and Definition 3.2. An important property is obtained showing that, for any admissible solution to the PLP, there exists a policy for which the associated occupation measure provides an admissible solution to the PLP with better value of the cost; see Theorem 3.1. A crucial corollary of this result (see Corollary 3.2) is then derived for the analysis of the constrained problem provided in Section 5. By using a dynamic programming argument, it is proved in Theorem 3.2, under a very weak assumption, that the unconstrained PLP admits an optimal solution, guaranteeing the existence of an optimal stopping time for the unconstrained problem. Finally, we show that admissible solutions to the PLP can be phantom solutions, i.e. they do not correspond to any control policy. However, Theorem 3.3 provides a connection between the optimal phantom solutions of the PLP and the optimal solutions of the PLP associated to a policy.

The unconstrained PLP is defined as follows: minimize $\mu_o(r) + \mu_\tau(R)$ subject to

$$(\mu_o, \mu_\tau) \in \mathbb{M}(E)_+^2, \qquad \mu_o + \mu_\tau = \nu + \mu_o S. \tag{3.1}$$

A pair of measures (μ_o, μ_τ) on E is called an admissible solution to the unconstrained PLP if (μ_o, μ_τ) satisfies (3.1). Here (μ_o^*, μ_τ^*) is called an optimal solution to the unconstrained PLP if it is admissible and $\mu_o^*(r) + \mu_\tau^*(R)$ is equal to the infimum of $\mu_o(r) + \mu_\tau(R)$ over $(\mu_o, \mu_\tau) \in \mathbb{M}(E)_+^2$ satisfying (3.1).

Remark 3.1. Note that an admissible pair of measures (μ_o, μ_τ) can take values $+\infty$ and are not necessarily σ-finite.

Definition 3.1. For any stationary policy π^s, the operator $T^{\pi^s} : \mathbb{M}(E)_+ \to \mathbb{M}(E)_+$ is defined by

$$T^{\pi^s}\eta(\Gamma) = \nu(\Gamma) + \int_E \pi^s(0 \mid y, 0) S(\Gamma \mid y)\eta(\mathrm{d}y), \qquad \Gamma \in \mathcal{B}(E).$$

Lemma 3.1. *The following assertions hold.*

(a) *For any policy* $\pi \in \Pi$, *the pair of measures* $(\mu_o^\pi, \mu_\tau^\pi)$ *is admissible for the PLP.*

(b) *If* π^s *is a stationary policy then* μ^{π^s} *is the minimal solution to the equation* $T^{\pi^s}\eta = \eta$, $\eta \in \mathbb{M}(E)_+$. *Moreover,* $\mu^{\pi^s} = \lim_{t\to\infty} \nu_t$, *where* $\nu_{t+1} = T^{\pi^s}\nu_t$ *for* $t \ge 0$ *and* $\nu_0 = \nu$.

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

(c) *If π^s is a stationary policy then, for any $\Gamma \in \mathcal{B}(E)$,*

$$\mu_o^{\pi^s}(\Gamma) = \int_\Gamma \pi^s(0 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y), \qquad \mu_\tau^{\pi^s}(\Gamma) = \int_\Gamma \pi^s(1 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y).$$

Proof. See Appendix A.

Condition 3.1. *There exists an admissible pair of measures $(\mu_o, \mu_\tau) \in \mathbb{M}(E)_+^2$ satisfying*

$$\mu_o(r) + \mu_\tau(R) < \infty.$$

Introduce the kernels

$$\mathbb{I}_B(C \mid x) = \mathbf{1}_B(x)\delta_x(C) \quad \text{and} \quad U_B(C \mid x) = \sum_{k\geq 1}(S\mathbb{I}_{B^c})^{k-1}S(C \mid x),$$

for $x \in E$ and $(B, C) \in \mathcal{B}(E) \times \mathcal{B}(E)$. Here $U_B(C \mid x)$ is the average amount of time the Markov chain spends in the set C up to the time where the chain enters B for the first time. Define

$$\begin{aligned}
E^R &= \{x \in E \colon R(x) > 0\},\\
E_0^R &= \{x \in E \colon R(x) = 0\},\\
E^r &= \{x \in E \colon r(x) > 0\},\\
E_0^r &= \{x \in E \colon r(x) = 0\},\\
F &= \{x \in E_0^r \colon U_{E_0^R}(E^r \cap E^R \mid x) > 0\},\\
D &= E^R \cap [F \cup E^r],\\
F_0 &= \{x \in E_0^r \colon U_{E_0^R}(E^r \cap E^R \mid x) = 0\},\\
\hat{E} &= E^R \cap F_0.
\end{aligned}$$

Remark 3.2. Note that D, $\hat{E}$, and E_0^R are pairwise disjoint, and that $E = D \cup \hat{E} \cup E_0^R$.

Proposition 3.1. *Assume that Condition 3.1 is satisfied for the pair of measures (μ_o, μ_τ). Then measures μ_o and μ_τ are σ-finite on D.*

Proof. See Appendix A.

Remark 3.3. From Proposition 3.1, the measures μ_o^D and μ_τ^D are σ-finite. Consequently, the Radon–Nikodym derivative, $\mathrm{d}\mu_\tau^D/\mathrm{d}(\mu_\tau^D + \mu_o^D)$, exists. Clearly, there is no loss of generality to consider that $\mathrm{d}\mu_\tau^D/\mathrm{d}(\mu_\tau^D + \mu_o^D) \in [0, 1]$.

Definition 3.2. Assume that the measures (μ_o, μ_τ) satisfy Condition 3.1. Associated to (μ_o, μ_τ), introduce the stochastic kernel π^s on A given E^Δ defined by

$$\pi^s(1 \mid y, 0) = \begin{cases} \dfrac{\mathrm{d}\mu_\tau^D}{\mathrm{d}(\mu_\tau^D + \mu_o^D)}(y) & \text{if } y \in D,\\ 1 & \text{if } y \in E_0^R,\\ 0 & \text{if } y \in \hat{E},\end{cases}$$

and $\pi^s(0 \mid y, 0) = 1 - \pi^s(1 \mid y, 0)$. Moreover, $\pi^s(1 \mid y, \Delta) = 1$ for any $y \in E$.

The stationary control policy π^s will be called the stationary control policy induced by (μ_o, μ_τ).

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Theorem 3.1. *Assume that Condition 3.1 is satisfied for the pair of measures (μ_o, μ_τ). Then, the stationary control policy π^s induced by (μ_o, μ_τ) satisfies*

$$V(\pi^s) \le \mu_o(r) + \mu_\tau(R).$$

Proof. See Appendix A.

Corollary 3.1. *Assume that Condition 3.1 is satisfied, and let (μ_o^*, μ_τ^*) be an optimal solution to the PLP. Then there exists an optimal stationary control policy π^* for the auxiliary control problem such that*

$$\inf_{\pi\in\Pi} V(\pi) = V(\pi^*) = \mu_o^*(r) + \mu_\tau^*(R).$$

Proof. According to Lemma 3.1(a), for any control policy π, the pair of measures $(\mu_o^\pi, \mu_\tau^\pi)$ is admissible for the PLP. Consequently, according to (2.6), $V(\pi) = \mu_o^\pi(r) + \mu_\tau^\pi(R) \ge \mu_o^*(r) + \mu_\tau^*(R)$. However, combining Theorem 3.1 and the fact that (μ_o^*, μ_τ^*) is an optimal solution to the PLP, there exists a stationary policy π^* such that $V(\pi^*) = \mu_o^*(r) + \mu_\tau^*(R)$, completing the proof.

Remark 3.4. (a) For an optimal solution of the PLP, labelled (μ_o^*, μ_τ^*), it can happen that $\mu_\tau^*(E) > 1$ or $\mu_\tau^*(E) = \infty$. However, for the optimal policy π^* induced by (μ_o^*, μ_τ^*), $\mu_\tau^{\pi^*} = \mathrm{Q}[\tau < \infty] \le 1$ (see Theorem 2.1). Of course, in such cases, there also exists another optimal solution to the PLP given by $(\mu_o^{\pi^*}, \mu_\tau^{\pi^*})$.

(b) An admissible solution (μ_o, μ_τ) of the PLP is called a phantom solution if there does not exist any control policy $\pi \in \Pi$ for which $\mu_o = \mu_o^\pi$ and $\mu_\tau = \mu_\tau^\pi$. In the case where the state space is finite, the PLP has no phantom solution [13], [14].

(c) For an optimal solution (μ_o^*, μ_τ^*) to the PLP, an optimal control policy π^* can be constructed according to Definition 3.2.

Corollary 3.2. *Assume that Condition 3.1 is satisfied for the pair of measures (μ_o, μ_τ). Let π^s be the stationary control policy induced by (μ_o, μ_τ). Suppose that functions $\tilde{r}$ and $\tilde{R}$ are such that $0 \le \tilde{r}(y) \le r(y)$ and $0 \le \tilde{R}(y) \le R(y)$ for all $y \in E$. Let $\tilde{V}(\pi^s)$ be the performance criterion corresponding to the cost functions $\tilde{r}$, $\tilde{R}$, and associated to π. Then*

$$\tilde{V}(\pi^s) \le \mu_o(\tilde{r}) + \mu_\tau(\tilde{R}).$$

Proof. We can apply exactly the same argument as at the end of the proof of Theorem 3.1, because $\tilde{R}(y) = 0$ if $y \in E_0^R$ and $\tilde{r}(y) = 0$ if $y \in \hat{E}$.

Theorem 3.2. *Under Condition 3.1, the unconstrained PLP has an optimal solution leading to the existence of an optimal stopping time for the unconstrained optimal stopping problem (P_1).*

Proof. The optimization problem $\inf_{\pi\in\Pi} V(\pi)$ associated to the unconstrained MDP has a solution due to Corollary 9.17.1 of [2]. The dynamic programming approach to optimal stopping of discrete state space process is presented in [20, Section 7.2.8]. Let $\pi^*(a \mid y, z)$ be the corresponding optimal stationary (in fact, nonrandomized) policy. According to Lemma 3.1(a), the pair of measures $(\mu_o^{\pi^*}, \mu_\tau^{\pi^*})$ is admissible for the unconstrained PLP and $V(\pi^*) = \mu_o^{\pi^*}(r) + \mu_\tau^{\pi^*}(R)$. Let (μ_o, μ_τ) be any admissible pair for the PLP. Without loss of generality, we can assume that $\mu_o(r) + \mu_\tau(R) < \infty$. Now, according to Theorem 3.1, we have $\mu_o(r) + \mu_\tau(R) \ge V(\pi^*)$. Now, according to Theorem 2.1, the control $\lambda^* \in \Lambda$ associated to the optimal control policy $\pi^* \in \Pi$ is optimal for problem (P$_1$), completing the proof.

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Theorem 3.3. *Assume that Condition 3.1 is satisfied. Let* (μ_o^*, μ_τ^*) *be an optimal solution for the PLP, and let* π^* *be the stationary control policy induced by* (μ_o^*, μ_τ^*). *Then, for all* $\Theta \in \mathcal{B}(D)$,

$$\mu_o^*(\Theta) = \mu_o^{\pi^*}(\Theta), \qquad \mu_\tau^*(\Theta) = \mu_\tau^{\pi^*}(\Theta).$$

Proof. See Appendix A.

4. Example

In this section we present an example that illustrates the fact that (optimal) solutions (γ_o, γ_τ) can exist to the PLP for which there does not exist any policy π such that $\gamma_o = \mu_o^\pi$ and $\gamma_\tau = \mu_\tau^\pi$, namely there exist (optimal) phantom solutions to the PLP.

The state space is defined by $E = \{A_1, A_2, A_3, A_4, A_5, 1, 2, \dots, 1', 2', \dots, 1'', 2'', \dots\}$. The transition kernel S of the Markov chain is given by

$$\begin{aligned}
&S((i-1)' \mid i') = 1 \quad \text{for all } i' > 1', \\
&S((i-1)'' \mid i'') = 1 \quad \text{for all } i'' > 1, \\
&S(A_3 \mid 1') = S(1 \mid A_1) = S(A_1 \mid A_2) = S(A_2 \mid A_3) = S(A_5 \mid A_5) = 1, \\
&S(A_3 \mid A_4) = S(A_4 \mid A_4) = \tfrac{1}{2}, \\
&S(A_1 \mid 1'') = S(A_2 \mid 1'') = \tfrac{1}{2}, \\
&S(i+1 \mid i) = 1 \quad \text{for all } i \geq 1.
\end{aligned}$$

All other transition probabilities are 0. The initial distribution ν is defined by $\nu(A_1) = \nu(A_2) = \nu(A_3) = \nu(A_4) = \nu(A_5) = \frac{1}{5}$. The values of loss functions r and R are given in Table 1.

In the proof of Theorem 3.1, it was shown that the state space E of the Markov chain admits a decomposition into the subsets D, E_0^R, and $\hat{E}$: $E = D \cup E_0^R \cup \hat{E}$. Let us denote by E^R the set of states x for which $R(x) > 0$ and by E^r the set of states x for which $r(x) > 0$. Then the following assertions hold.

- D contains $E^r \cap E^R$ and such states y satisfying $R(y) > 0$ and $r(y) = 0$ and for which there exists a path in the set E^R of the Markov chain generated by the kernel S starting from y and reaching the set $E^r \cap E^R$. This gives $D = \{A_1, A_2, 1, 2, \dots, 1'', 2'', \dots\}$.
- E_0^R is the set of states x such that $R(x) = 0$, giving $E_0^R = \{A_3\}$.
- $\hat{E} = (D \cup E_0^R)^{\mathrm{c}}$ and so $\hat{E} = \{A_4, A_5, 1', 2' \dots\}$.

It is very easy to describe the optimal stopping policy given by

- $\pi^*(1 \mid A_3, 0) = 1$ because $R(A_3) = 0$;
- $\pi^*(0 \mid i', 0) = 1$ for all i' because $R(i') = 1$, but starting from i' the process will reach state A_3 and terminate, leading to the zero total loss;

Table 1.

	A_1	A_2	A_3	A_4	A_5	$i \geq 1$	$i' \geq 1'$	$i'' \geq 1''$
r	0	0	1	0	0	$\left(\frac{1}{2}\right)^{i-1}$	0	0
R	3	1	0	1	1	3	1	3

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

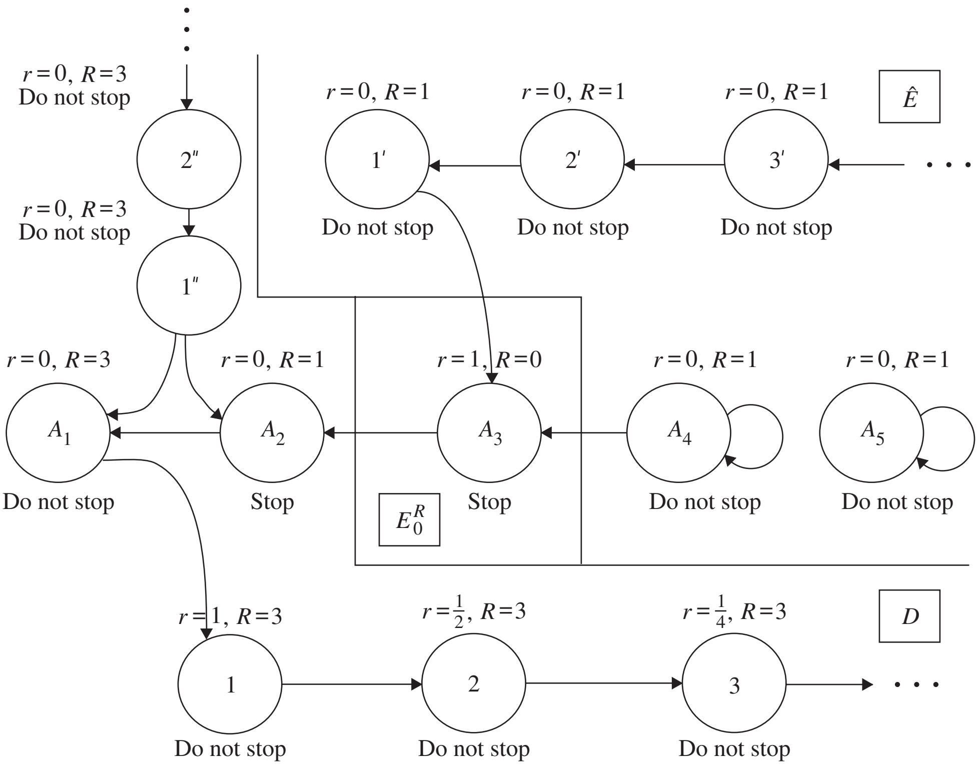

FIGURE 1: Example.

- $\pi^*(0 \mid A_4, 0) = 1$ because of the same reason;
- $\pi^*(0 \mid A_5, 0) = 1$ because $R(A_5) = 1 > 0$, but, never being stopped, the process, starting from the absorbing state A_5, provides no loss;
- $\pi^*(0 \mid A_1, 0) = \pi^*(0 \mid i, 0) = 1$ for all $i \geq 1$ because the total loss on the infinite horizon in the chain starting from A_1 or from $i \geq 1$ does not exceed 2, whereas the cost of stopping equals $R = 3$;
- $\pi^*(1 \mid A_2, 0) = 1$ because the stopping cost $R(A_2) = 1$ is smaller than 2, the total minimal possible future loss if the process starting from A_2 is not stopped;
- $\pi^*(0 \mid i'', 0) = 1$ for all i'' because $R(i'') = 3$, but starting from i'', the process will either terminate at A_2 or will never be stopped, leading to a total expected cost of $\frac{3}{2}$.

See Figure 1 for a pictorial representation of the above.

Combining parts (b) and (c) of Lemma 3.1, we can compute the occupation measures $\mu_o^{\pi^*}$ and $\mu_\tau^{\pi^*}$ given in Table 2, which provide an optimal solution to the PLP, according to Theorem 3.1.

It is interesting to note that the PLP has other optimal solutions (γ_o, γ_τ) which do not correspond to any stopping policy. Indeed, it is easy to check that, for any constant $c \in \mathbb{R} \cup \{\infty\}$, the measures (γ_o, γ_τ) defined by $\gamma_o(i') = c$ and $\gamma_o(x) = \mu_o^{\pi^*}(x)$ for $x \neq i'$, $i \in \mathbb{N}$, and $\gamma_\tau(A_3) = \frac{2}{5} + c$ and $\gamma_\tau(x) = \mu_\tau^{\pi^*}(x)$ for $x \neq A_3$ is admissible for the PLP and optimal since only the values of $\gamma_o(i')$ and $\gamma_\tau(A_3)$ have changed and $r(i') = R(A_3) = 0$. Hence, the objective

TABLE 2.

x	A_1	A_2	A_3	A_4	A_5	$i \geq 1$	$i' \geq 1'$	$i'' \geq 1''$
$\mu_o^{\pi^*}(x)$	$\frac{1}{5}$	0	0	$\frac{2}{5}$	∞	$\frac{1}{5}$	0	0
$\mu_\tau^{\pi^*}(x)$	0	$\frac{1}{5}$	$\frac{2}{5}$	0	0	0	0	0

value does not change and remains minimal. However, for any $c > 0$, the pair of measures so defined cannot be associated to any policy π because the states i' cannot be reached and so $\mu_o^\pi(i')$ must be 0.

Remark 4.1. (a) At this point, we would like to emphasize the following fact. Let (μ_o, μ_τ) be a phantom solution to the PLP which is not optimal, and let π be the stationary policy induced by (μ_o, μ_τ) according to Definition 3.2. Then it may happen that $\mu_o \neq \mu_o^\pi$ and $\mu_\tau \neq \mu_\tau^\pi$ on D, as illustrated in the previous example. Indeed, we can set $\mu_o(i'') = c$, where $c \geq 0$ is an arbitrary number, with the corresponding modification of measures μ_o and μ_π on D. Then we necessarily have $\mu_o(i'') \neq \mu_o^\pi(i'')$, since, by using the same argument as above, the states i'' cannot be reached and so $\mu_o^\pi(i'')$ must be 0. However, if (μ_o, μ_τ) is a phantom solution to the PLP which is optimal then, necessarily, $\mu_o = \mu_o^\pi$ and $\mu_\tau = \mu_\tau^\pi$ on D according to Theorem 3.3.

(b) Since, for all $y \in E$, $\max_{a \in A} c(y, 0, a) > 0$, we can consider only the absorption in the subset $\{(y, \Delta, a), y \in E, a \in A\}$, on which $c(y, \Delta, a) \equiv 0$. But the optimal policy π^* is not absorbing because, for $\tau = \min\{t \in \mathbb{N} : a_t = 1\}$, we have $\mathrm{E}_\nu^{\pi^*}[\tau] = \infty$; so the expected time to absorption $\mathrm{E}_\nu^{\pi^*}[\tau + 1] = \infty$. It is sufficient to look at the initial state A_1 with $\nu(A_1) = \frac{1}{5}$. The optimal policy π^* is also not transient because $\sum_{t=0}^{\infty} \mathrm{P}_\nu^{\pi^*}[y_t = A_5, \ \tau > t] = \infty$. Therefore, the auxiliary control problem under consideration is neither absorbing nor transient.

5. Constrained version of the optimal stopping problem

In this section, the constrained version of the optimal stopping problem is investigated through the related constrained version of the PLP. By introducing a suitable topology on a set of occupation measures, an existence result (see Theorem 5.1) for an optimal solution of the PLP with constraints is obtained, guaranteeing the existence of an optimal stopping time for the constrained optimal stopping problem. This result holds under a very weak assumption (see Condition 5.1). It must be pointed out that the dynamic programming argument used to prove the existence result for the unconstrained case cannot be used here in the presence of constraints.

The constrained version of the PLP (3.1) is defined as follows: minimize $\mu_o(r_0) + \mu_\tau(R_0)$ subject to

$$\begin{aligned} (\mu_o, \mu_\tau) \in \mathbb{M}(E)_+^2, &\qquad \mu_o + \mu_\tau = \nu + \mu_o S, \\ \mu_o(r_n) + \mu_\tau(R_n) \leq j_n &\qquad \text{for } n \in \{1, \ldots, N\}. \end{aligned} \tag{5.1}$$

For notational convenience, let $g_t : G \to E^\Delta \times A$ be defined by

$$g_t(\omega) = (y_t(\omega), z_t(\omega), a_t(\omega)),$$

and define

$$h_0(\omega) = (y_0(\omega), z_0(\omega)) \quad \text{and} \quad h_t(\omega) = (g_0(\omega), \ldots, g_{t-1}(\omega), y_t(\omega), z_t(\omega))$$

for $\omega \in G$ and $t \geq 1$. Denote by $\mathcal{P}$ the set of probability measures on $(G, \mathcal{G})$ and by $\mathcal{P}^\pi$ the set of probability measures on $(G, \mathcal{G})$ induced by control policies $\pi \in \Pi$. Now introduce $\mathcal{D}^\pi$ to be the set of occupation measures $(\mu_o^\pi, \mu_\tau^\pi) \in \mathbb{M}(E)_+^2$, where μ_o^π and μ_τ^π are respectively defined by (2.4) and (2.5) for a policy $\pi \in \Pi$. Let $\mathbb{B}$ be the mapping $\mathbb{B}\colon \mathcal{P}^\pi \to \mathcal{D}^\pi$ such that $\mathbb{B}(\mathrm{P}_\nu^\pi) = (\mu_o^\pi, \mu_\tau^\pi)$.

In order to define a topology on $\mathcal{P}$, we need to introduce the set of functions $\mathcal{W}_t$ as a subset of bounded measurable functions $f\colon (E^\Delta \times A)^t \times E^\Delta \to \mathbb{R}$ such that, for any $(e_0, \dots, e_t) \in (E^\Delta)^{t+1}$, the function $f(e_0, ., e_1, ., e_2, \dots, e_{t-1}, ., e_t)$ defined on A^t is continuous.

The ws$^\infty$-topology on $\mathcal{P}$ is defined as the coarsest topology rendering the mappings $\mathrm{P} \to \int_G f(h_t(\omega))\,\mathrm{dP}(\omega)$ continuous, where $f \in \mathcal{W}_t$ and $t \geq 0$. For more details on the ws$^\infty$-topology, we refer the reader to [22]. The set $\mathcal{P}^\pi$ is endowed by the induced topology. Note that items (1) and (2) of Conditions (S) of [22] are satisfied. Therefore, $\mathcal{P}^\pi$ is compact (see Theorem 6.6 of [22]).

The topology on $\mathcal{D}^\pi$ is defined as the strongest topology for which the mapping $\mathbb{B}$ is continuous (the final topology on $\mathcal{D}^\pi$ associated to the mapping $\mathbb{B}$).

Lemma 5.1. *For any nonnegative measurable functions r and R, and any $K \in \mathbb{R}_+$, define $\mathcal{D}_K^\pi$ by*

$$\mathcal{D}_K^\pi = \{(\mu_o^\pi, \mu_\tau^\pi) \in \mathcal{D}^\pi : \mu_o^\pi(r) + \mu_\tau^\pi(R) \leq K\}.$$

The set $\mathcal{D}_K^\pi$ is compact and the mapping $\mathbb{D}\colon \mathcal{D}^\pi \to \mathbb{R}_+$ defined by $\mathbb{D}(\mu_o^\pi, \mu_\tau^\pi) = \mu_o^\pi(r) + \mu^\pi\tau(R)$ is lower semi-continuous.

Proof. Let us show that the set $\mathcal{P}_K^\pi = \{\mathrm{P}_\nu^\pi \in \mathcal{P}^\pi : \sum_{t=0}^\infty \int_G c(g_t(\omega))\,\mathrm{dP}_\nu^\pi(\omega) \leq K\}$ is compact in the ws$^\infty$-topology (where c is given by (2.2)).

Clearly, by definition, the mappings $\mathbb{A}_n\colon \mathcal{P}^\pi \to \mathbb{R}_+$ defined by

$$\mathbb{A}_n(\mathrm{P}_\nu^\pi) = \sum_{t=0}^n \int_G c(g_t(\omega))\,\mathrm{dP}_\nu^\pi(\omega)$$

are continuous and so the mapping $\mathbb{A}\colon \mathcal{P}^\pi \to \mathbb{R}_+$ defined by

$$\mathbb{A}(\mathrm{P}_\nu^\pi) = \sum_{t=0}^\infty \int_G c(g_t(\omega))\,\mathrm{dP}_\nu^\pi(\omega)$$

is lower semi-continuous because $c \geq 0$. However, the set $\mathcal{P}_K^\pi = \mathbb{A}^{-1}([0, K])$ and so it is closed and compact. Since $\mathbb{B}(\mathcal{P}_K^\pi) = \mathcal{D}_K^\pi$, the set $\mathcal{D}_K^\pi$ is compact as a continuous image of a compact set, completing the first part of the proof.

Now, note that $\mathbb{D} \circ \mathbb{B} = \mathbb{A}$. Consequently, for any $M \in \mathbb{R}_+$, the set $\mathbb{B}^{-1}(\mathbb{D}^{-1}((M, \infty)))$ is an open set of $\mathcal{P}^\pi$ and so $\mathbb{D}^{-1}((M, \infty))$ is open in the topology of $\mathcal{D}^\pi$, showing that $\mathbb{D}$ is lower semi-continuous and completing the last part of the proof.

Condition 5.1. *There exist a control policy $\bar{\pi}$ and a constant $j_0 < \infty$ such that*

$$V_0(\bar{\pi}) = j_0 \quad \textit{and} \quad V_n(\bar{\pi}) \leq j_n \quad \textit{for } n \in \{1, \dots, N\}.$$

Theorem 5.1. *Under Condition 5.1, the constrained PLP has an optimal solution leading to the existence of an optimal control for the constrained optimal stopping problem (P_2).*

Proof. Let $\bar{\pi}$ be control policy, and let j_0 be the constant satisfying Condition 5.1. Introduce the sets

$$\mathcal{D}_{j_n}^\pi = \{(\mu_o^\pi, \mu_\tau^\pi) \in \mathcal{D}^\pi : \mu_o^\pi(r_n) + \mu_\tau^\pi(R_n) \leq j_n\}.$$

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Then $(\mu_o^{\bar{\pi}}, \mu_\tau^{\bar{\pi}}) \in \bigcap_{n=0}^{N} \mathcal{D}_{j_n}^{\pi}$, and so according to Lemma 5.1, $\bigcap_{n=0}^{N} \mathcal{D}_{j_n}^{\pi}$ is a nonempty compact set and the mapping $\mathbb{D}_0 \colon \mathcal{D}^{\pi} \to \mathbb{R}_+$ given by $\mathbb{D}_0(\mu_o^{\pi}, \mu_\tau^{\pi}) = \mu_o^{\pi}(r_0) + \mu_\tau^{\pi}(R_0)$ is lower semi-continuous. Consequently, there exists $\pi^* \in \Pi$ such that $(\mu_o^{\pi^*}, \mu_\tau^{\pi^*}) \in \bigcap_{n=1}^{N} \mathcal{D}_{j_n}^{\pi}$ and

$$\inf_{(\mu_o,\mu_\tau)\in\bigcap_{n=0}^{N} \mathcal{D}_{j_n}^{\pi}} \{\mu_o(r_0) + \mu_\tau(R_0)\} = \mu_o^{\pi^*}(r_0) + \mu_\tau^{\pi^*}(R_0).$$

However, according to Corollary 3.2 with $r = \sum_{n=0}^{N} r_n$ and $R = \sum_{n=0}^{N} R_n$, it follows that, for any $(\mu_o, \mu_\tau) \in \mathbb{M}(E)_+^2$ satisfying (5.1) and $\mu_o(r_0) + \mu_\tau(R_0) \leq j_0$, there exists a stationary policy $\pi^s \in \Pi$ such that $\mu_o^{\pi^s}(r_n) + \mu_\tau^{\pi^s}(R_n) \leq \mu_o(r_n) + \mu_\tau(R_n)$ for all $n \in \{0, \dots, N\}$. Therefore, the infimum of $\mu_o(r_0) + \mu_\tau(R_0)$ over $(\mu_o, \mu_\tau) \in \mathbb{M}(E)_+^2$ satisfying (5.1) is equal to

$$\inf_{(\mu_o,\mu_\tau)\in\bigcap_{n=0}^{N} \mathcal{D}_{j_n}^{\pi}} \{\mu_o(r_0) + \mu_\tau(R_0)\}.$$

Now, according to Theorem 2.1, the control $\lambda^* \in \Lambda$ associated to the optimal control policy $\pi^* \in \Pi$ is optimal for problem (P$_2$), completing the proof.

Appendix A

A.1. Proof of Theorem 2.1

Consider $\lambda \in \Lambda$, where $\lambda = (\Omega, \mathcal{F}, \mathrm{Q}, \{\mathcal{F}_t\}_{t\in\mathbb{N}}, \{x_t\}_{t\in\mathbb{N}}, \tau)$. On the probability space $(\Omega, \mathcal{F}, \mathrm{Q})$, let us introduce the random processes $\{u_t\}_{t\in\mathbb{N}}$ and $\{d_t\}_{t\in\mathbb{N}}$ defined by

$$u_t = \mathbf{1}_{\{\tau\leq t\}}, \qquad d_t = u_{t-1}\Delta \quad \text{for } t \geq 1 \text{ and } d_0 = 0.$$

Since τ is an $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-stopping time, then clearly $\{u_t\}_{t\in\mathbb{N}}$ is $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-adapted and $\{d_t\}_{t\in\mathbb{N}}$ is $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-predictable. Define $\mathcal{H}_t = \sigma\{x_0, d_0, u_0, \dots, x_{t-1}, d_{t-1}, u_{t-1}, x_t, d_t\}$ for $t \geq 1$ and $\mathcal{H}_0 = \sigma\{x_0, d_0\}$. Observe that $\mathcal{H}_t \vee \sigma\{u_t\} \subset \mathcal{F}_t$ for all $t \in \mathbb{N}$. For any $B \in \mathcal{B}(E)$ and $C \subset \{0, \Delta\}$, we have

$$\begin{aligned}\mathrm{Q}[(x_{t+1}, d_{t+1}) \in B \times C \mid \mathcal{H}_t \vee \sigma\{u_t\}] &= \mathrm{E}_\mathrm{Q}[\mathrm{E}_\mathrm{Q}[\mathbf{1}_{\{x_{t+1}\in B\}} \mid \mathcal{F}_t]\, \mathbf{1}_{\{d_{t+1}\in C\}} \mid \mathcal{H}_t \vee \sigma\{u_t\}] \\ &= S(B \mid x_t)[\mathbf{1}_{\{0\in C\}}\, \mathbf{1}_{\{u_t=0\}} + \mathbf{1}_{\{\Delta\in C\}}\, \mathbf{1}_{\{u_t=1\}}].\end{aligned}$$

Moreover, $\{u_t = 0\} \subset \{u_{t-1} = 0\} = \{d_t = 0\}$, and so

$$\mathrm{Q}[(x_{t+1}, d_{t+1}) \in B \times C \mid \mathcal{H}_t \vee \sigma\{u_t\}] = L(B \times C \mid x_t, d_t, u_t).$$

Now introduce the sequence $\pi = \{\pi_t\}_{t\in\mathbb{N}}$ of stochastic kernels defined on the control set A given $(E^\Delta \times A)^{t-1} \times E^\Delta$ by

$$\pi_t(D \mid h_t) = \mathrm{E}[\mathbf{1}_{\{1\in D\}}\, u_t + \mathbf{1}_{\{0\in D\}}(1 - u_t) \mid \mathcal{H}_t]$$

for any $D \subset A$, where $h_t = (x_0, d_0, u_0, \dots, x_{t-1}, d_{t-1}, u_{t-1}, x_t, d_t)$.

By the uniqueness property in the theorem of Ionescu-Tulcea (see Appendix C of [11, Section 2.2]), we obtain

$$\mathrm{Q}[(x_0, d_0, u_0, \dots, x_t, d_t, u_t, \dots) \in H] = \mathrm{P}_\nu^\pi[y_0, z_0, a_0, \dots, y_t, z_t, a_t, \dots) \in H]$$

for any $H \in \mathcal{G}$, where the processes $\{y_t\}_{t\in\mathbb{N}}$, $\{z_t\}_{t\in\mathbb{N}}$, and $\{a_t\}_{t\in\mathbb{N}}$ were introduced in Section 2.2.

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Note that the previous construction of the policy π does not depend on the cost functions. Observe that $\{t < \tau\} = \{u_t = 0\} = \{u_t = 0\} \cap \{d_t = 0\}$ and $\{t = \tau\} = \{u_t = 1\} \cap \{d_t = 0\}$, and so the cost is given by

$$J(\lambda) = \mathrm{E}_{\mathrm{Q}}\Bigg[\sum_{t=0}^{\infty} r(x_t)\,\mathbf{1}_{\{t<\tau\}} + R(x_t)\,\mathbf{1}_{\{t=\tau\}}\Bigg] = \mathrm{E}_{\mathrm{Q}}\Bigg[\sum_{t=0}^{\infty} c(x_t, d_t, u_t)\Bigg];$$

hence,

$$J(\lambda) = \mathrm{E}_{\mathrm{Q}}\Bigg[\sum_{t=0}^{\infty} c(x_t, d_t, u_t)\Bigg] = \mathrm{E}_{\nu}^{\pi}\Bigg[\sum_{t=0}^{\infty} c(y_t, z_t, a_t)\Bigg] = V(\pi).$$

Similarly, $J_n(\lambda) = V_n(\pi)$ for all $n \in \{0, \dots, N\}$, completing the first part of the proof.

Conversely, consider a policy π and the processes $\{y_t\}_{t\in\mathbb{N}}$, $\{z_t\}_{t\in\mathbb{N}}$, and $\{a_t\}_{t\in\mathbb{N}}$ as defined in Section 2.2 on the probability space $(G, \mathcal{G}, \mathrm{P}_{\nu}^{\pi})$. Define $\mathcal{F}_t = \sigma\{y_0, a_0, \dots, y_t, a_t\}$. Clearly, the process $\{y_t\}_{t\in\mathbb{N}}$ defined on $(G, \mathcal{G}, \mathrm{P}_{\nu}^{\pi})$ is an E-valued $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-Markov chain with transition kernel S and initial distribution ν. Now define $\tau = \inf\{t \in \mathbb{N} \colon a_t = 1\}$, and if this set is empty then set $\tau = \infty$. Then τ is an $\{\mathcal{F}_t\}_{t\in\mathbb{N}}$-stopping time. Observe that $\{t < \tau\} = \{a_t = 0\} \cap \{z_t = 0\}$ and $\{t = \tau\} = \{a_t = 1\} \cap \{z_t = 0\}$. Consequently, we have

$$\begin{aligned} V(\pi) &= \mathrm{E}_{\nu}^{\pi}\Bigg[\sum_{t=0}^{\infty} c(y_t, z_t, a_t)\Bigg] \\ &= \mathrm{E}_{\nu}^{\pi}\Bigg[\sum_{t=0}^{\infty} (r(y_t)\,\mathbf{1}_{\{a_t=0\}} + R(y_t)\,\mathbf{1}_{\{a_t=1\}})\,\mathbf{1}_{\{z_t=0\}}\Bigg] \\ &= \mathrm{E}_{\nu}^{\pi}\Bigg[\sum_{t=0}^{\tau-1} r(y_t) + \mathbf{1}_{\{\tau<\infty\}}\, R(y_\tau)\Bigg]. \end{aligned}$$

Using similar arguments, we have $V_n(\pi) = J_n(\lambda)$ for all $n \in \{0, \dots, N\}$. Therefore, the control λ defined by $(G, \mathcal{G}, \mathrm{P}_{\nu}^{\pi}, \{\mathcal{F}_t\}_{t\in\mathbb{N}}, \{y_t\}_{t\in\mathbb{N}}, \tau)$ belongs to Λ, and satisfies $J(\lambda) = V(\pi)$ and $J_n(\lambda) = V_n(\pi)$ for all $n \in \{0, \dots, N\}$.

In both cases, it is easy to check that $\tau = \inf\{t \in \mathbb{N} \colon a_t = 1\} = \sum_{t=0}^{\infty} \mathbf{1}_{E\times\{0\}\times\{0\}}(y_t, z_t, a_t)$ and $\mathbf{1}_{\{\tau<\infty\}} = \sum_{t=0}^{\infty} \mathbf{1}_{E\times\{0\}\times\{1\}}(y_t, z_t, a_t)$, implying that $\mu_o^{\pi}(E) = \mathrm{E}_{\mathrm{Q}}[\tau]$ and $\mu_{\tau}^{\pi}(E) = \mathrm{Q}[\tau < \infty]$. This completes the proof.

A.2. Proof of Lemma 3.1

(a) Clearly, $(\mu_o^{\pi}, \mu_{\tau}^{\pi}) \in \mathbb{M}(E)_+^2$. According to Lemma 9.4.3(c) of [12], we obtain, for $\Gamma \in \mathcal{B}(E)$,

$$\rho(\Gamma \times \{0\} \times A) = \nu(\Gamma) + \int_{E^{\Delta}\times A} L(\Gamma \times \{0\} \times A \mid y, z, a)\rho(\mathrm{d}(y, z, a)),$$

where the measure ρ is defined by $\rho(\Theta \times C \times D) := \sum_{t=0}^{\infty} \mathrm{P}_{\nu}^{\pi}[(y_t, z_t, a_t) \in \Theta \times C \times D]$ for $\Theta \in \mathcal{B}(E)$, $C \subset \{0, \Delta\}$, and $D \subset \{0, 1\}$. However, from the definitions of μ_o^{π} and μ_{τ}^{π} (see (2.4) and (2.5)), we have $\rho(\Gamma \times \{0\} \times A) = \mu_o^{\pi}(\Gamma) + \mu_{\tau}^{\pi}(\Gamma)$. Now, using the definition of the stochastic kernel L (see (2.1)), we have $\int_{E^{\Delta}\times A} L(\Gamma \times \{0\} \times A \mid y, z, a)\rho(\mathrm{d}(y, z, a)) = \mu_o^{\pi} S$, completing the proof of part (a).

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

(b) Let π^s be a stationary policy. Define

$$\nu_t(\Gamma) = \sum_{i=0}^{t} \hat{\nu}_i(\Gamma), \tag{A.1}$$

where $\hat{\nu}_i(\Gamma) = \mathrm{P}_\nu^{\pi^s}[y_i \in \Gamma,\ z_i = 0]$ for $\Gamma \in \mathcal{B}(E)$. Let us show by induction that $\nu_{t+1} = T^{\pi^s}\nu_t$. For $t = 0$, we have $\nu_1(\Gamma) = \nu_0(\Gamma) + \mathrm{P}_\nu^{\pi^s}[y_1 \in \Gamma,\ z_1 = 0]$. By definition, $\nu_0 = \nu$ and

$$\begin{aligned}\hat{\nu}_1(\Gamma) &= \mathrm{E}_\nu^{\pi^s}[L(\Gamma \times \{0\} \mid y_0, z_0)]\\ &= \mathrm{E}_\nu^{\pi^s}[S(\Gamma \mid y_0)\,\mathbf{1}_{\{z_0=0\}}\,\mathbf{1}_{\{a_0=0\}}]\\ &= \int_E S(\Gamma \mid y)\pi^s(0 \mid y, 0)\nu(\mathrm{d}y),\end{aligned}$$

showing that $\nu_1 = T^{\pi^s}\nu_0$. Now, assume that $\nu_t = T^{\pi^s}\nu_{t-1}$ for $t \geq 1$. Then, using similar arguments, it is easy to obtain

$$\hat{\nu}_{t+1}(\Gamma) = \mathrm{E}_\nu^{\pi^s}[S(\Gamma \mid y_t)\,\mathbf{1}_{\{z_t=0\}}\,\mathbf{1}_{\{a_t=0\}}] = \int_E S(\Gamma \mid y)\pi^s(0 \mid y, 0)\hat{\nu}_t(\mathrm{d}y).$$

However,

$$\nu_{t+1}(\Gamma) = \nu_t(\Gamma) + \hat{\nu}_t(\Gamma) = T^{\pi^s}\nu_{t-1} + \int_E S(\Gamma \mid y)\pi^s(0 \mid y, 0)\hat{\nu}_t(\mathrm{d}y),$$

which, upon using the definition of T^{π^s}, completes the induction step.

The operator T^{π^s} is monotone: if $\eta^1 \geq \eta^2$ then $T^{\pi^s}\eta^1 \geq T^{\pi^s}\eta^2$. Consequently, we have $\nu_{t+1} \geq \nu_t$ for $t \geq 0$, and so the limit of ν_t as t tends to ∞ exists, $\lim_{t\to\infty}\nu_t = \nu_\infty$, implying that $\nu_\infty = T^{\pi^s}\nu_\infty$. However, from the definition of ν_t, it follows that $\nu_\infty = \mu^{\pi^s}$, yielding $T^{\pi^s}\mu^{\pi^s} = \mu^{\pi^s}$.

If $\tilde{\mu} \geq 0$ is another measure on E satisfying $\tilde{\mu} = T^{\pi^s}\tilde{\mu}$ then, by the definition of T^{π^s}, we have $\nu \leq \tilde{\mu}$. Since the operator T^{π^s} is monotone, it is easy to show by induction that $\nu_t \leq \tilde{\mu}$, implying that $\nu_\infty = \mu^{\pi^s} \leq \tilde{\mu}$, completing the proof of part (b).

(c) Define $\nu_{o,t}(\Gamma) = \sum_{i=0}^{t}\mathrm{P}_\nu^{\pi^s}[y_i \in \Gamma,\ z_i = 0,\ a_i = 0]$ for $\Gamma \in \mathcal{B}(E)$. Similarly to the proof of part (b), it can be shown by induction that $\nu_{o,t}(\Gamma) = \int_\Gamma \pi^s(0 \mid y, 0)\nu_t(\mathrm{d}y)$, where ν_t is defined in (A.1). Moreover, note that $\lim_{t\to\infty}\nu_{o,t} = \mu_o^{\pi^s}$, and, from part (b), $\lim_{t\to\infty}\nu_t = \mu^{\pi^s}$. Consequently, $\mu_o^{\pi^s}(\Gamma) = \int_\Gamma \pi^s(0 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y)$. The second equality follows by similar arguments.

A.3. Proof of Proposition 3.1

The proof of this result is very involved and so it is divided into several preliminary results.

Lemma A.1. *For $y \in \hat{E}$, $S(D \mid y) = 0$.*

Proof. Let $y \in \hat{E}$. Note that $S(D \mid y) = S(E^R \cap F \mid y) + S(E^R \cap E^r \mid y)$. Clearly, we have $S(E^R \cap E^r \mid y) = 0$ since $S(E^R \cap E^r \mid y) \leq U_{E_0^R}(E^r \cap E^R \mid y)$. We will show by contradiction that $S(E^R \cap F \mid y) = 0$. Assume that $S(E^R \cap F \mid y) > 0$. Consequently, it follows from the definition of F that

$$\int_{E^R \cap E_0^r} U_{E_0^R}(E^r \cap E^R \mid z)S(\mathrm{d}z \mid y) = \int_{E^R \cap E_0^r} U_{E_0^R}(E^r \cap E^R \mid z)S\mathbb{I}_{E^R}(\mathrm{d}z \mid y) > 0,$$

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

implying that $\sum_{n\geq 1}(S\mathbb{I}_{E^R})^n S(E^r \cap E^R \mid y) > 0$. It follows that $U_{E_0^R}(x, E^r \cap E^R) > 0$, leading to a contradiction. This gives $S(E^R \cap F \mid y) = 0$, completing the proof.

Lemma A.2. *Let η be a positive measure on $(E, \mathcal{B}(E))$, and let W be a positive kernel on E given E. Assume that ηW is σ-finite on $B \in \mathcal{B}(E)$. Then η is σ-finite on*

$$\{y \in E \colon W(B \mid y) > 0\}.$$

Proof. There exists a sequence of pairwise disjoint sets $(B_j)_{j\in\mathbb{N}}$ such that $\eta W(B_j) < \infty$ and $B = \bigcup_{j\geq 1} B_j$. Define, for $j \in \mathbb{N}$ and $p \in \mathbb{N}$, $B_j^p = \{y \in E \colon W(B_j \mid y) > 1/p\}$. Clearly, we have

$$\{y \in E \colon W(B \mid y) > 0\} = \bigcup_{j\geq 1}\bigcup_{p\geq 1} B_j^p,$$

and

$$\eta(B_j^p) \leq \int_{B_j^p} pW(B_j \mid y)\eta(\mathrm{d}y) \leq p\eta W(B_j) < \infty,$$

completing the proof.

In what follows, (μ_o, μ_τ) is a fixed pair of measures satisfying Condition 3.1.

Proof of Proposition 3.1. Clearly, μ_τ is σ-finite on $D \subset E^R$. Moreover, μ_o is σ-finite on $E^r \cap E^R$. Let us show that μ_o is σ-finite on $E^R \cap F$ to obtain the result. Define $B_k = \{x \in E^R \colon S(\mathbb{I}_{E^R} S)^{k-1}(E^r \cap E^R \mid x) > 0\}$ for $k \in \mathbb{N}$. Note that $E^R \cap F \subset \bigcup_{k\geq 1} B_k$. Let us show by induction that the measure μ_o is σ-finite on B_k for $k \in \mathbb{N}$, and the result will follow.

The measure $\mu_o S$ is σ-finite on $E^r \cap E^R$ since $\mu_o S + \nu = \mu_o + \mu_\tau$ and $\mu_o + \mu_\tau$ is σ-finite on $E^r \cap E^R$. Consequently, by applying Lemma A.2 we find that μ_o is σ-finite on B_1. Assume that μ_o is σ-finite on B_k for $k \in \mathbb{N}$. Since $B_k \subset E^R$, μ_τ is σ-finite on B_k and we find that $\mu_o S$ is σ-finite on B_k. Again, applying Lemma A.2, we find that μ_o is σ-finite on $\{x \in E \colon S(B_k \mid x) > 0\}$. However, note that $B_{k+1} \subset \{x \in E \colon S(B_k \mid x) > 0\}$, which completes the induction.

A.4. Proof of Theorem 3.1

In order to prove this result, we first need the following technical lemma.

Lemma A.3. *For any $\Theta \in \mathcal{B}(D)$, the following inequalities hold:*

$$\int_\Theta \pi^s(0 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y) \leq \mu_o(\Theta), \tag{A.2}$$

$$\int_\Theta \pi^s(1 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y) \leq \mu_\tau(\Theta). \tag{A.3}$$

Proof. Note that, by the definition of π^s, for any $\Theta \in \mathcal{B}(D)$, we have

$$\int_\Theta \pi^s(0 \mid y, 0)(\mu_o + \mu_\tau)(\mathrm{d}y) = \mu_o(\Theta). \tag{A.4}$$

According to Lemma 3.1, $\mu^{\pi^s} = \lim_{t\to\infty} \nu_t$, where $\nu_{t+1} = T^{\pi^s}\nu_t$ for $t \geq 0$ and $\nu_0 = \nu$. Let us show by induction that, for all $t \geq 0$,

$$\int_\Theta \pi^s(0 \mid y, 0)\nu_t(\mathrm{d}y) \leq \mu_o(\Theta). \tag{A.5}$$

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Since (μ_o, μ_τ) satisfies (3.1), it follows that

$$\int_\Theta \pi^s(0 \mid y, 0)\nu_0(\mathrm{d}y) \le \int_\Theta \pi^s(0 \mid y, 0)(\mu_o + \mu_\tau)(\mathrm{d}y) = \mu_o(\Theta).$$

Now assume that, for any $\Theta \in \mathcal{B}(D)$, $\int_\Theta \pi^s(0 \mid y, 0)\nu_t(\mathrm{d}y) \le \mu_o(\Theta)$ for $t \ge 1$. From Lemma A.1 and bearing in mind that $\pi^s(0 \mid y, 0) = 0$ for any $y \in E_0^R$, we obtain

$$\nu_{t+1}(\Theta) = \nu(\Theta) + \int_D S(\Theta \mid y)\pi^s(0 \mid y, 0)\nu_t(\mathrm{d}y).$$

Therefore, $\nu_{t+1}(\Theta) \le \nu(\Theta) + \int_D S(\Theta \mid y)\mu_o(\mathrm{d}y) \le \mu_o(\Theta) + \mu_\tau(\Theta)$. Using (A.4), we complete the induction. Now by taking the limit in (A.5) we obtain (A.2). Similar arguments can be used to show (A.3).

Proof of Theorem 3.1. Combining Lemma 3.1(c) and (2.6), it follows that

$$V(\pi^s) = \int_E r(y)\pi^s(0 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y) + \int_E R(y)\pi^s(1 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y).$$

However, from the definition of π^s and the set D, we obtain

$$V(\pi^s) = \int_D r(y)\pi^s(0 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y) + \int_D R(y)\pi^s(1 \mid y, 0)\mu^{\pi^s}(\mathrm{d}y).$$

From the previous lemma, we obtain

$$V(\pi^s) \le \int_D r(y)\mu_o(\mathrm{d}y) + \int_D R(y)\mu_\tau(\mathrm{d}y).$$

A.5. Proof of Theorem 3.3

The proof of this result is divided into several steps.

Lemma A.4. *For all $\Theta \in \mathcal{B}(D)$, $\mu_\tau^*(\Theta) = \mu_\tau^{\pi^*}(\Theta)$ and $\mu_o^*(\Theta) \le \mu_o^{\pi^*}(\Theta)$. Moreover, $\mu_o^*(\Gamma) = \mu_o^{\pi^*}(\Gamma)$ for all $\Gamma \in \mathcal{B}(E^r \cap E^R)$.*

Proof. Clearly, combining Lemma A.3 and Lemma 3.1(c), we have, for all $\Theta \in \mathcal{B}(D)$,

$$\mu_\tau^{\pi^*}(\Theta) \le \mu_\tau^*(\Theta) \quad \text{and} \quad \mu_o^{\pi^*}(\Theta) \le \mu_o^*(\Theta).$$

Now assume that there exists $\Theta \in \mathcal{B}(D)$ such that $\mu_\tau^{\pi^*}(\Theta) < \mu_\tau^*(\Theta)$. Then $\mu_\tau^{\pi^*}(R) < \mu_\tau^*(R)$ since $R(x) > 0$ for all $x \in D$. According to Corollary 3.2 with $\widetilde{R}(x) = 0$ and $\widetilde{r}(x) = r(x)$, we have $\mu_o^{\pi^*}(r) \le \mu_o^*(r)$. This shows that $\mu_o^{\pi^*}(r) + \mu_\tau^{\pi^*}(R) < \mu_o^*(r) + \mu_\tau^*(R)$, in contradiction with the fact that (μ_o^*, μ_τ^*) is an optimal solution for the PLP. Consequently, $\mu_\tau^{\pi^*}(\Theta) = \mu_\tau^*(\Theta)$ for all $\Theta \in \mathcal{B}(D)$. Similarly, it can be shown that $\mu_o^*(\Gamma) = \mu_o^{\pi^*}(\Gamma)$ for all $\Gamma \in \mathcal{B}(E^r \cap E^R)$.

Unfortunately, it cannot be shown directly that $\mu_o^*(\Theta) = \mu_o^{\pi^*}(\Theta)$ for all $\Theta \in \mathcal{B}(D)$ as for μ_τ^* and $\mu_\tau^{\pi^*}$, mainly because it cannot be claimed that $r(x) > 0$ for all $x \in D$. The rest of the proof is devoted to showing that, for all $\Theta \in \mathcal{B}(F \cap E^R)$, $\mu_o^*(\Theta) = \mu_o^{\pi^*}(\Theta)$.

Lemma A.5. *For all $\Gamma \in \mathcal{B}(E^r \cap E^R)$ and all $k \ge 0$,*

$$\mu_o^{\pi^*} S(\mathbb{I}_D S)^k(\Gamma) = \mu_o^{\pi^*}(\mathbb{I}_D S)^{k+1}(\Gamma).$$

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Proof. By definition, $\pi^*(0 \mid y, 0) = 0$ for all $y \in E_0^R$. Therefore, using Lemma 3.1(c), we have $\mu_o^{\pi^*}(E_0^R) = 0$, implying that, for all $\Gamma \in \mathcal{B}(E)$ and $k \geq 0$,

$$\mu_o^{\pi^*} S(\mathbb{I}_D S)^k(\Gamma) = \int_{D \cup \hat{E}} S(\mathbb{I}_D S)^k(\Gamma \mid x)\mu_o^{\pi^*}(\mathrm{d}x). \tag{A.6}$$

Note that, since $D \subset E^R$, $S(\mathbb{I}_D S)^k(\Gamma \mid x) \leq S(\mathbb{I}_{E^R} S)^k(\Gamma \mid x)$ for all $x \in E$ and $\Gamma \in \mathcal{B}(E^r \cap E^R)$. Now by the definitions of $\hat{E}$ and F_0, $S(\mathbb{I}_{E^R} S)^k(\Gamma \mid x) = 0$ for all $x \in \hat{E}$ and $\Gamma \in \mathcal{B}(E^r \cap E^R)$, and so, using (A.6),

$$\mu_o^{\pi^*} S(\mathbb{I}_D S)^k(\Gamma) = \int_D S(\mathbb{I}_D S)^k(\Gamma \mid x)\mu_o^{\pi^*}(\mathrm{d}x)$$

for all $\Gamma \in \mathcal{B}(E^r \cap E^R)$, completing the proof.

Lemma A.6. *The measures $\mu_o^{\pi^*}(\mathbb{I}_D S)^k$ are σ-finite on $E^r \cap E^R$ and*

$$\mu_o^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) = \mu_o^*(\mathbb{I}_D S)^k(\Gamma)$$

for all $k \geq 0$ and $\Gamma \in \mathcal{B}(E^r \cap E^R)$.

Proof. Let us prove this result by induction. Clearly, the result holds for $k = 0$ because of Proposition 3.1 and Lemma A.4, and the fact that $E^r \cap E^R \subset D$. Now assume that the result holds for $k \geq 0$. Consider any $\Gamma \in \mathcal{B}(E^r \cap E^R)$. Then, using the fact that $(\mu_o^{\pi^*}, \mu_\tau^{\pi^*})$ is an admissible solution for the PLP, we obtain

$$\mu_o^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) + \mu_\tau^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) = \nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^{\pi^*} S(\mathbb{I}_D S)^k(\Gamma). \tag{A.7}$$

According to Theorem 2.1, the measure $\mu_\tau^{\pi^*}$ is finite. Recalling that the measure ν is finite and using the induction hypothesis, we find that the measure $\mu_o^{\pi^*} S(\mathbb{I}_D S)^k$ is σ-finite on $E^r \cap E^R$. However, $\mu_o^*\mathbb{I}_D \leq \mu_o^*$, and so we find that the measure $\mu_o^{\pi^*}(\mathbb{I}_D S)^{k+1}$ is σ-finite on $E^r \cap E^R$. Now, using (A.7) and Lemma A.5, we have

$$\mu_o^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) + \mu_\tau^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) \leq \nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^{\pi^*}(\mathbb{I}_D S)^{k+1}(\Gamma). \tag{A.8}$$

However, Lemma A.4 shows that $\mu_o^{\pi^*}\mathbb{I}_D \leq \mu_o^*\mathbb{I}_D$, implying that

$$\nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^{\pi^*}(\mathbb{I}_D S)^{k+1}(\Gamma) \leq \nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^*(\mathbb{I}_D S)^{k+1}(\Gamma). \tag{A.9}$$

Moreover, combining the facts that $\mu_o^*\mathbb{I}_D \leq \mu_o^*$ and the pair of measures (μ_o^*, μ_τ^*) is an admissible solution for the PLP, we have

$$\begin{aligned}\nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^*(\mathbb{I}_D S)^{k+1}(\Gamma) &\leq \nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^* S(\mathbb{I}_D S)^k(\Gamma)\\ &= \mu_o^*(\mathbb{I}_D S)^k(\Gamma) + \mu_\tau^*(\mathbb{I}_D S)^k(\Gamma).\end{aligned} \tag{A.10}$$

According to the induction hypothesis and Lemma A.4, it follows that

$$\mu_o^*(\mathbb{I}_D S)^k(\Gamma) + \mu_\tau^*(\mathbb{I}_D S)^k(\Gamma) = \mu_o^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) + \mu_\tau^{\pi^*}(\mathbb{I}_D S)^k(\Gamma). \tag{A.11}$$

Combining equations (A.8)– (A.11), we obtain

$$\begin{aligned}\mu_o^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) + \mu_\tau^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) &\leq \nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^{\pi^*}(\mathbb{I}_D S)^{k+1}(\Gamma)\\ &\leq \nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^*(\mathbb{I}_D S)^{k+1}(\Gamma)\\ &\leq \mu_o^{\pi^*}(\mathbb{I}_D S)^k(\Gamma) + \mu_\tau^{\pi^*}(\mathbb{I}_D S)^k(\Gamma).\end{aligned}$$

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

Therefore, since the left- and right-hand sides of the above inequalities are equal, we have

$$\nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^{\pi^*}(\mathbb{I}_D S)^{k+1}(\Gamma) = \nu(\mathbb{I}_D S)^k(\Gamma) + \mu_o^*(\mathbb{I}_D S)^{k+1}(\Gamma),$$

and since the measure ν is finite, the result follows.

Lemma A.7. *For all $x \in D$,*

$$U_{D^c}(E^r \cap E^R \mid x) = U_{E_0^R}(E^r \cap E^R \mid x).$$

Proof. In order to obtain the result, let us show that

$$S(\mathbb{I}_D S)^k(E^r \cap E^R \mid x) = S(\mathbb{I}_{E^R} S)^k(E^r \cap E^R \mid x)$$

for all $x \in D$ and $k \geq 0$ by induction. This is clearly true for $k = 0$. Now assume that the result holds for $k \geq 0$. Consider $x \in D$. Then,

$$\begin{aligned} S(\mathbb{I}_D S)^{k+1}(E^r \cap E^R \mid x) &= S\mathbb{I}_D S(\mathbb{I}_D S)^k(E^r \cap E^R \mid x) \\ &= S\mathbb{I}_D S(\mathbb{I}_{E^R} S)^k(E^r \cap E^R \mid x). \end{aligned}$$

Owing to Lemma A.1, we have $S(\mathbb{I}_{E^R} S)^k(E^r \cap E^R \mid y) = 0$ for all $y \in \hat{E}$. Therefore, $\mathbb{I}_{\hat{E}} S(\mathbb{I}_{E^R} S)^k(E^r \cap E^R \mid x) = 0$, and so

$$S\mathbb{I}_D S(\mathbb{I}_{E^R} S)^k(E^r \cap E^R \mid x) = S(\mathbb{I}_D + \mathbb{I}_{\hat{E}})S(\mathbb{I}_{E^R} S)^k(E^r \cap E^R \mid x).$$

However, $\mathbb{I}_{E^R} = \mathbb{I}_D + \mathbb{I}_{\hat{E}}$, completing the induction.

Lemma A.8. *For all $\Theta \in \mathcal{B}(F \cap E^R)$, $\mu_o^*(\Theta) = \mu_o^{\pi^*}(\Theta)$.*

Proof. According to Lemma A.6, the measure $\mu_o^{\pi^*}\mathbb{I}_D U_{D^c}$ is σ-finite on $E^r \cap E^R$. Consequently, there exists an increasing sequence of sets $(\Gamma_i)_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty} \Gamma_i = E^r \cap E^R$ with $\Gamma_i \subset E^r \cap E^R$ and $\mu_o^{\pi^*}\mathbb{I}_D U_{D^c}(\Gamma_i) < \infty$ for $i \geq 1$. Let us introduce the sequence of sets $(D_i)_{i\in\mathbb{N}}$ defined by $D_i = \{x \in D \colon U_{D^c}(\Gamma_i \mid x) > 0\}$. Using Lemma A.6, we have, for all $i \geq 1$,

$$\mu_o^{\pi^*}\mathbb{I}_D U_{D^c}(\Gamma_i) = \mu_o^*\mathbb{I}_D U_{D^c}(\Gamma_i). \tag{A.12}$$

Therefore, for all $i \geq 1$ and $\Gamma \in \mathcal{B}(D_i)$, we obtain

$$\mu_o^{\pi^*}(\Gamma) = \mu_o^*(\Gamma). \tag{A.13}$$

Indeed, assume the contrary, namely that there exists a set $\Gamma \in \mathcal{B}(D_i)$ such that $\mu_o^{\pi^*}(\Gamma) < \mu_o^*(\Gamma)$. Using Lemma A.4 and the fact that $\mathbb{I}_D U_{D^c}(\Gamma_i \mid x) > 0$ for all $x \in \Gamma$, we have $\mu_o^{\pi^*}\mathbb{I}_D U_{D^c}(\Gamma_i) < \mu_o^*\mathbb{I}_D U_{D^c}(\Gamma_i)$, leading to a contradiction with (A.12). This proves (A.13). According to Lemma A.7,

$$D_i \uparrow \{x \in D \colon U_{D^c}(E^r \cap E^R \mid x) > 0\} = \{x \in D \colon U_{E_0^R}(E^r \cap E^R \mid x) > 0\}.$$

However, note that $F \cap E^R \subset \{x \in D \colon U_{E_0^R}(E^r \cap E^R \mid x) > 0\}$, completing the proof.

Proof of Theorem 3.3. The result is a straightforward consequence of Lemmas A.4 and A.8, and the fact that $D = (E^r \cap E^R) \cup (F \cap E^R)$.

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press

References

[1] Altman, E (1999). *Constrained Markov Decision Processes*. Chapman and Hall, Boca Raton, FL.

[2] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control* (Math. Sci. Eng. **139**). Academic Press, New York.

[3] Borkar, V. S., Pinto, J. and Prabhu, T. (2009). A new learning algorithm for optimal stopping. *Discrete Event Dyn. Systems* **19,** 91–113.

[4] Boyarchenko, S. and Levendorskiĭ, S. (2007). *Irreversible Decisions Under Uncertainty* (Stud. Econom. Theory **27**). Springer, Berlin.

[5] Cho, M. J. and Stockbridge, R. H. (2002). Linear programming formulation for optimal stopping problems. *SIAM J. Control Optimization* **40,** 1965–1982.

[6] De Saporta, B., Dufour, F. and Gonzalez, K. (2010). Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Prob.* **20,** 1607–1637.

[7] Dufour, F. and Miller, B. (2004). Singular stochastic control problems. *SIAM J. Control Optim.* **43,** 708–730.

[8] El Karoui, N., Hủ̀u Nguyen, D. and Jeanblanc-Picqué, M. (1987). Compactification methods in the control of degenerate diffusions: existence of an optimal control. *Stochastics* **20,** 169–219.

[9] Guo, X. and Piunovskiy, A. (2010). Discounted continuous-time Markov decision processes with constraints: unbounded transition and loss rates. To appear in *Math. Operat. Res.*

[10] Haussmann, U. and Lepeltier, J.-P. (1990). On the existence of optimal controls. *SIAM J. Control Optimization* **28,** 851–902.

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

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

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

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

[15] Kallenberg, L. C. M. (1994). Survey of linear programming for standard and nonstandard Markovian control problems. I. Theory. *Z. Operat. Res.* **40,** 1–42.

[16] Novikov, A. and Shiryaev, A. (2007). On solution of the optimal stopping problem for processes with independent increments. *Stochastics* **79,** 393–406.

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

[18] Piunovskiy, A. B. (1998). Controlled random sequences: methods of convex analysis and problems with functional constraints. *Russian Math. Surveys* **53,** 1233–1293.

[19] Piunovskiy, A. B. (2005). Discounted continuous time Markov decision processes: the convex analytic approach. In *Proc. 16th Triennial IFAC World Congress* (Praha, Czech. Republic).

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

[21] Rieder, U. (1975). On stopped decision processes with discrete time parameter. *Stoch. Process. Appl.* **3,** 365–383.

[22] Schäl, M. (1975). On dynamic programming: compactness of the space of policies. *Stoch. Process. Appl.* **3,** 345–364.

https://doi.org/10.1239/jap/1294170511 Published online by Cambridge University Press