OPTIMAL SOLUTIONS OF CONSTRAINED DISCOUNTED SEMI-MARKOV CONTROL PROBLEMS

JUAN GONZÁLEZ-HERNÁNDEZ AND CÉSAR EMILIO VILLARREAL-RODRÍGUEZ

ABSTRACT. We give conditions for the existence of optimal solutions to the constrained semi-Markov decision problem on Borel spaces, with possibly unbounded costs and discounted performance index. We also demonstrate the existence of optimal solutions which are given by a convex combinations of $N + 1$ measurable selectors, where $N$ is the number of constraints.

1. Introduction


---

2010 Mathematics Subject Classification: 60K15, 47N10, 47N30.

Keywords and phrases: semi-Markov decision process, performance index, stochastic kernel, Polish space.

For a deeper study on Markov Decision Processes see [11] or [25]. The cases with constraints can be seen in [7, 6, 8].

So far there are no known works that address to the constrained semi-Markov control problem whose state space and control space are Borel spaces. The main contributions of this article is the focus by occupation measures.

Section 2 contains preliminary notions of semi-Markov control model with constraints. Section 3 works with a generalization of the occupation measures for semi-Markov decision processes with discounted cost. This allows to give an equivalent problem in terms of a family of measures and to use the direct approach. In Section 4 we prove the existence of solutions to the control problem. We conclude the article with Section 5, where the existence of stationary optimal solutions which are convex combinations of at most \( N + 1 \) measurable selectors is demonstrated.

\textit{Notation} (1.1). A Polish space \( Z \) is a complete separable metric space and a Borel space is a measurable subset of a Polish space. We denote by \( \mathcal{B}(Z) \) its Borel \( \sigma \)-algebra. Measurable always means Borel-measurable. If \( Z \) and \( W \) are Borel spaces, a stochastic kernel \( P \) on \( Z \) given \( W \) is a function \( (w, B) \mapsto P(B|w) \) such that \( P(\cdot|w) \) is a probability measure on \( \mathcal{B}(Z) \) for each \( w \in W \), and \( P(B|\cdot) \) is a measurable function on \( W \) for each \( B \in \mathcal{B}(Z) \). We also shall denote by \( \mathcal{P}(Z) \) the set of all probability measures on \( (Z, \mathcal{B}(Z)) \).

2. The semi-Markov control model

\textit{Definition} (2.1). A constrained semi-Markov decision process (CSMDP)

\((X, \mathcal{A}, A, Q, F, c_j, d_j; j \in \{0, 1, \ldots, N\})\)

consists of:

(a) A Borel space space \( X \), called the state space.
(b) A Borel space \( \mathcal{A} \), the control (or action) space.
(c) A function \( A : X \to (B : B \text{ is a measurable subset of } \mathcal{A}) \). For each \( x \) we have that \( A(x) \neq \emptyset \) and it is the set of admissible controls (or actions) at the state \( x \). Moreover we assume that the set

\[ \mathcal{K} := \{(x, a) : x \in X, \ a \in A(x)\} \]

is a Borel subset of \( X \times \mathcal{A} \) and contains the graph of a measurable map from \( X \) to \( \mathcal{A} \).
(d) A stochastic kernel \( Q \) on \( X \) given \( \mathcal{K} \) called the transition law.
(e) A continuous function \( t \mapsto F(t|x, a, y) \) which is a probability distribution function, for each \((x, a, y) \in \mathcal{K} \times X\), and we assume \( F(t|\cdot) \) is jointly measurable for each real number \( t \).
(f) The nonnegative measurable real functions on \( X \times \mathcal{A} \), \( c_j \) and \( d_j \) for \( j \in \{0, 1, \ldots, N\} \) are the so called cost functions.
The CSMDP represents a stochastic system that evolves in the next way: At stage $i$ the system is in the state $x_i \in X$ and a control $a_i \in A(x_i)$ is applied, then the following things happen: The immediate costs $c_j(x_i,a_i)$ for $j \in \{0,1,\ldots,N\}$ are incurred. The system moves to the next state $x_{i+1} \in X$ according to the probability measure $Q(\cdot|x_i,a_i)$. Conditional to $(x_i,a_i,x_{i+1})$ the time $t_{i+1}$ from the transition $i$ occurs until the transition $i+1$ occurs has the distribution function $F(\cdot|x_i,a_i,x_{i+1})$. The costs $d_j(x_i,a_i)$ for $j \in \{0,1,\ldots,N\}$ are imposed until the transition $i+1$ occurs. After the transition $i+1$ occurs, a control $a_{i+1} \in A(x_{i+1})$ is chosen and the process continues in this way.

For each $i \in \mathbb{N} \cup \{0\}$, define the space of admissible histories up to stage $i$ by $H_0 := X$ and $H_i := X^{i-1} \times X = X \times H_{i-1}$. A generic element $h_i \in H_i$ is a vector, or history, of the form $h_i = (x_0,a_0,\ldots,x_{i-1},a_{i-1},x_i)$, where $(x_j,a_j) \in X$ for $j \in \{0,1,2,\ldots,i-1\}$ and $x_i \in X$.

**Definition (2.2).**
(a) A control policy is a sequence $\pi = (\pi_i)$ of stochastic kernels $\pi_i$ on $\mathcal{A}$ given $H_i$, satisfying the constraint $\pi_i(A(x_i)|h_i) = 1$ for all $h_i \in H_i$ and $i \in \mathbb{N} \cup \{0\}$. We denote by $\Pi$ the class of all policies.
(b) A control policy is said to be randomized stationary, if there exists a stochastic kernel $\varphi$ on $\mathcal{A}$ given $X$, satisfying the constraint $\varphi(A(x)|x) = 1$ such that $\pi_i(\cdot|h_i) = \varphi(\cdot|x_i)$ for all $h_i \in H_i$ and $i \in \mathbb{N} \cup \{0\}$. We identify the policy $\pi$ with $\varphi$ and denote by $\Phi$ the set of all such policies.
(c) A randomized stationary policy $\varphi$ is said to be stationary deterministic if there exists a function $f$ from $X$ to $\mathcal{A}$, satisfying the constraint $f(x) \in A(x)$ such that $\varphi(\cdot|x) = f(x)$ is concentrated at $f(x)$. These functions are called measurable selectors. We identify the policy $\varphi$ with $f$ and we denote by $\mathcal{F}$ the set of all such policies.
(d) A randomized stationary policy is said to be $N$-randomized policy if it is a convex combination of at most $N$ stationary deterministic policies. We denote the set of all such policies by $\Phi_N$.

Given an initial distribution $\nu \in \mathcal{P}(X)$ and $\pi = (\pi_i) \in \Pi$, by Ionescu-Tulcea Theorem [13, Th. 2.7.2], [13, Sec. 11] or [19, pp. 137-139]), there exists a probability space $(\Omega,\mathcal{A},\mathbb{P})$ such that

1. $\mathbb{P}^v(\pi_0 \in B) = \nu(B)$ for $B \in \mathcal{B}(X)$;
2. $\mathbb{P}^v(\pi_i \in B|h_i,a_i) = Q(B|x_i,a_i)$ for all $B \in \mathcal{B}(X)$, $h_i \in H_i$ and $a_i \in A(x_i)$, $i \in \mathbb{N} \cup \{0\}$;
3. $\mathbb{P}^v(a_i \in C|h_i) = \pi_i(C|h_i)$ for all $C \in \mathcal{B}(\mathcal{A})$ and $h_i \in H_i$, $i \in \mathbb{N} \cup \{0\}$;
4. $\mathbb{P}^v(t_i \leq r|h_{i+1}) = F(r|x_i,a_i,x_{i+1})$ for all $r \in \mathbb{R}$, $h_{i+1} \in H_{i+1}$ and $a_i \in A(x_i)$, $i \in \mathbb{N} \cup \{0\}$;

From the dynamic of the process given by Definitions (2.1) (d) and (e) and the nature of a policy, Definition (2.2) (a), we can see that:

**Remark (2.3).** The random variables $t_1,t_2,\ldots$ are conditionally independent given the process $(x_0,a_0,x_1,a_1,\ldots)$.

We denote by $\mathbb{E}^v_\pi$ the expectation with respect to $\mathbb{P}^v_\pi$, and we denote by $\mathbb{P}^x_\pi$ and $\mathbb{E}^x_\pi$ respectively when $\nu$ is concentrated at $x$.

In order to assure that an infinite number of transitions $t_1,t_2,\ldots$ does not occur in a finite interval, we need to impose a condition: (see Vega-Amaya [31]). To
do this, we introduce the following notations: for the distribution function of the holding time $t_i$ conditional to $(x,a) \in \mathcal{K}$ we put

$$G(t|x,a) := \int_{\mathcal{X}} F(t|x,a,y)Q(dy|x,a) \quad \text{for } t \geq 0;$$

for the conditional expected value of $t_i$ given $(x,a)$ we put

$$\tau(x,a) := \int_0^{+\infty} (1-G(t|x,a))dt$$

and we introduce the auxiliary functions $\tau_a$ such that $\tau_a(x,a)$ is the conditional expected value of $\int_0^{t_i} e^{-at}ds$ given $(x,a)$ putting

$$\chi_a(x,a) := \int_0^{+\infty} e^{-at}G(dt|x,a) \quad \text{and}$$

$$\tau_a(x,a) := (1-\chi_a(x,a))/a,$$

for $a \in (0,1)$ (the discount rate) and $(x,a) \in \mathcal{K}$. 

**Condition (2.7).** There exist $\epsilon > 0$ and $\bar{t} > 0$ such that $P(t > \bar{t}|x,a) = 1-G(\bar{t}|x,a) \geq \epsilon$ for all $(x,a) \in \mathcal{K}$. 

So we have the next lemma (Proposition 2.4 of [31]):

**LEMMA (2.8).** If Condition (2.7) holds, then

1. $\inf_{\mathcal{K}} \tau(x,a) \geq \epsilon \bar{t}$;
2. $\tau_a := \sup_{\mathcal{K}} \tau_a(x,a) < 1$;
3. $P_x^\pi \left( \sum_{i=1}^{\infty} t_i = +\infty \right) = 1$ for every $x \in \mathcal{X}$ and $\pi \in \Pi$.

**Performance index.** Let us define the sum of the transition times as $T_0 := t_0 := 0$, and $T_i := T_{i-1} + t_i$ for $i \in \mathbb{N}$. If we consider $V^\pi_j(x,v)$ for $\pi \in \Pi$ and $v$ a fixed initial distribution on $\mathcal{X}$ as:

$$V^\pi_j(x,v) := E^\pi_v \left( \sum_{i=1}^{\infty} e^{-aT_i} \left( c_j(x,a) + d_j(x,a) \int_0^{t_{i+1}} e^{-at}dt \right) \right),$$

for $j \in \{0,1,\ldots,N\}$. Now we define new current costs when the process is in state $x$ and an action $a$ is chosen as:

$$C^a_j(x,a) := c_j(x,a) + d_j(x,a)\tau_a(x,a),$$

where $\tau_a$ is given in (2.6). So, by properties of conditional probability, we can express $V^\pi_j(x,v)$ as:

$$V^\pi_j(x,v) := E^\pi_v \left( \sum_{i=0}^{\infty} e^{-aT_i} C^a_j(x_i,a_i) \right),$$

for $j \in \{0,1,\ldots,N\}$. 

Definition (2.11). For $\alpha \in (0,1)$, an initial distribution $\nu$ on $\mathcal{X}$ and a policy $\pi \in \Pi$, the values of $V^\alpha_j(\pi, \nu)$ given in (2.10) are called the $\alpha$-discounted expected costs. When $\nu$ is concentrated at some $x \in \mathcal{X}$ we write $V^\alpha_j(\pi, x)$ instead.

Let $k_j > 0$ for $j \in \{1, \ldots, N\}$ given. The discounted control problems is:

\[
\text{DCP} \quad \begin{align*}
\min & \quad V^\alpha_0(\pi, \nu), \\
\text{subject to} & \quad V^\alpha_j(\pi, \nu) \leq k_j \text{ for } j \in \{1, \ldots, N\}, \\
& \pi \in \Pi.
\end{align*}
\]

If a policy $\pi^*$ reaches this minimum, it is said that it is an optimal solution for DCP and it is called an alpha-discounted optimal policy.

3. Discounted occupation measures

Notation (3.1). For an arbitrary Borel space $Z$ we shall denote by $\mathcal{M}(Z)$ the family of finite (signed) measures on $(Z, \mathcal{B}(Z))$. We shall denote by $\mathcal{M}(Z|Z')$ the family of all conditional finite measures on $Z$ given $Z'$. That is, an element of $\mathcal{M}(Z|Z')$ is a function $(z', B) \mapsto m(B|z')$, such that $m(B|\cdot)$ is a measurable function on $Z'$ for each $B \in \mathcal{B}(Z)$ and $m(\cdot|z') \in \mathcal{M}(Z)$ for each $z' \in Z'$. Also when $C \subset Z$, $\mathcal{M}(C)$ shall denote the family of all finite measures with support on $C$. Similarly we denote by $\mathcal{M}_+(Z)$, $\mathcal{M}_+(Z|Z')$ and $\mathcal{M}_+(C)$ the corresponding families for nonnegative finite measures.

Let us define the conditional measure $H_\alpha \in \mathcal{M}(\mathcal{X}|\mathcal{X} \times A)$ as

\[
H_\alpha(C|x,a) := \int_C e^{-at}G(dt|x,a)
\]

\[
= E_\pi^\nu (e^{-at_i} \mathbb{I}_C(x_i)|x_{i-1} = x, a_{i-1} = a),
\]

for all $i \in \mathbb{N}$, where $\mathbb{I}_C$ is the indicator function of the set $C$, that is $\mathbb{I}_C(x) = 1$, if $x \in C$, and $\mathbb{I}_C(x) = 0$ otherwise.

Definition (3.2). Given $\alpha \in (0,1)$, $\nu$ an initial distribution on $\mathcal{P}(\mathcal{X})$ and a policy $\pi \in \Pi$, the $\alpha$-discounted occupation measure $m(\cdot: \pi, \nu, \alpha)$ is defined as

\[
m(B: \pi, \nu, \alpha) := E_\pi^\nu \left( \sum_{i=0}^{\infty} e^{-aT_i} \mathbb{I}_B(x_i, a_i) \right),
\]

for $B \in \mathcal{B}(\mathcal{X} \times A)$.

Remark (3.4). Observe that $m(\cdot: \pi, \nu, \alpha) \in \mathcal{M}_+(\mathcal{X} \times A)$. In fact

\[
m(B: \pi, \nu, \alpha) := E_\pi^\nu \left( \sum_{i=0}^{\infty} e^{-aT_i} \mathbb{I}_B(x_i, a_i) \right) \leq \sum_{i=0}^{\infty} (\bar{\tau}_{\alpha})^i,
\]

Remark (3.5). (a) If $\mu \in \mathcal{M}(\mathcal{X} \times A)$, there is a randomized control $\varphi \in \Phi$ and a signed measure $\hat{\mu} \in \mathcal{M}(\mathcal{X})$ such that

\[
\mu(B \times C) = (\hat{\mu} \circ \varphi)(B \times C) := \int_B \varphi(C|x)\hat{\mu}(dx)
\]

for $B \in \mathcal{B}(\mathcal{X})$ and $C \in \mathcal{B}(A)$. The measure $\hat{\mu}$ is called the marginal measure of $\mu$ on $\mathcal{X}$ and it is obtained by mean of $\hat{\mu} := \mu(\cdot \times A)$. Observe that for each $C \in \mathcal{B}(A)$, $\hat{\mu}(B) = 0 \implies \mu(B \times C) = 0$ and the function $\varphi(C|\cdot)$ is the Radon-Nikodým derivative of $\mu(\cdot \times C)$ with respect to $\hat{\mu}$.
(b) Conversely, if \( \varphi \) is a randomized control and \( \mu \in \mathcal{M}(\mathcal{X}) \), there is an unique signed measure \( \mu \in \mathcal{M}(\mathcal{X} \times \mathcal{A}) \) such that \([3.6]\) is satisfied.

**Lemma (3.7).** Let \( f \) be a nonnegative measurable function on \( \mathcal{B}(\mathcal{X} \times \mathcal{A}) \), and let \( a \in (0, 1) \), \( \nu \) an initial distribution on \( \mathcal{P}(\mathcal{X}) \) and a policy \( \pi \in \Pi \). Set \( m(B) = m(B : \pi, \nu, a) \), thus

\[
\int_{\mathcal{X} \times \mathcal{A}} f(x, a) m(d(x, a)) = E^\nu_0 \left( \sum_{i=0}^{\infty} e^{-aT_i} f(x_i, a_i) \right).
\]

**Proof.** This property can be proved by following the classic way of supposing first the case when the function \( f \) is an indicator function, then simple function, then increasing limit of simple functions. \( \square \)

**Notation (3.8).** Given a measurable function \( f \) on \( \mathcal{X} \), a stochastic kernel \( P \) on \( \mathcal{X} \), a stochastic kernel \( \varphi \) on \( \mathcal{A} \) given \( \mathcal{X} \) and a Borel subset of \( \mathcal{X} \), we denote by

\[
f(x, \varphi) := \int_\mathcal{A} f(x, a) \varphi(da|x)
\]

and

\[
P(B|x, \varphi) = \int_\mathcal{A} P(B|x, a) \varphi(da|x).
\]

**Theorem (3.9).** A measure \( m \in \mathcal{M}_+(\mathcal{X}) \) is an \( \alpha \)-discounted occupation measure if and only if it satisfies

\[
\dot{m}(B) = \nu(B) + \int_{\mathcal{X} \times \mathcal{A}} H_\alpha(B|x, a) m(d(x, a)),
\]

for every Borel subset \( B \) of \( \mathcal{X} \), where \( \dot{m} \) is the marginal measure of \( m \) on \( \mathcal{X} \), that is \( \dot{m}(B) := m(B \times \mathcal{A}) \) for \( B \in \mathcal{B}(\mathcal{X}) \).

**Proof.** Let us take \( m(B) = m(B : \pi, \nu, a) \). We have

\[
\dot{m}(B) = m(B \times \mathcal{A}) = E^\nu_0 \left( \sum_{i=0}^{\infty} e^{-aT_i} \mathbb{1}_B \times \mathcal{A}(x_i, a_i) \right)
\]

\[
= \nu(B) + E^\nu_0 \left( \sum_{i=1}^{\infty} e^{-aT_i} \mathbb{1}_B \times \mathcal{A}(x_i, a_i) \right)
\]

\[
= \nu(B) + E^\nu_0 \left( \sum_{i=1}^{\infty} E^\nu_0(e^{-aT_i} \mathbb{1}_B \times \mathcal{A}(x_i, a_i)|h_{i-1}, a_{i-1}) \right)
\]

\[
= \nu(B) + E^\nu_0 \left( \sum_{i=1}^{\infty} e^{-aT_{i-1}} E^\nu_0(e^{-aT_{i-1}} \mathbb{1}_B \times \mathcal{A}(x_i, a_i)|x_{i-1}, a_{i-1}) \right)
\]

\[
= \nu(B) + \int_{\mathcal{X} \times \mathcal{A}} H_\alpha(B|x_{i-1}, a_{i-1}) m(d(x, a)).
\]
Conversely, let us consider a measure \( m \in \mathcal{M}_{+}(\mathbb{K}) \) such that satisfies (3.11). Let us disintegrate \( m = \hat{m} \otimes \varphi \), then iterations of this equation produce

\[
\hat{m}(B) = \nu(B) + \int_{x} H_{a}(B|x_{0}, \varphi) \hat{m}(dx_{0}) = \nu(B) + \int_{x} H_{a}(B|x_{0}, \varphi)\nu(dx_{0}) \\
+ \int_{x} \int_{x} H_{a}(B|x_{1}, \varphi) H_{a}(dx_{1}|x_{0}, \varphi) \hat{m}(dx_{0}) = \cdots = \nu(B) \\
+ \sum_{i=1}^{M-1} \int_{x} \cdots \int_{x} H_{a}(B|x_{i}, \varphi) \prod_{k=1}^{i} H_{a}(dx_{k}|x_{k-1}, \varphi) \nu(dx_{0}) \\
+ \int_{x} \cdots \int_{x} H_{a}(B|x_{M}, \varphi) \prod_{k=1}^{M} H_{a}(dx_{k}|x_{k-1}, \varphi) \hat{m}(dx_{0}),
\]

for all \( M \in \mathbb{N} \). In this last expression we consider an empty product as equal to 1. The last sumand tends to zero, in fact

\[
\int_{x} \cdots \int_{x} H_{a}(B|x_{M}, \varphi) \prod_{k=1}^{M} H_{a}(dx_{k}|x_{k-1}, \varphi) \hat{m}(dx_{0}) \leq (\tau_{a})^{M} \rightarrow 0,
\]
as \( M \rightarrow \infty \). Now, for \( i \in \mathbb{N} \) we have

\[
\int_{x} \cdots \int_{x} H_{a}(B|x_{i}, \varphi) \prod_{k=1}^{i} H_{a}(dx_{k}|x_{k-1}, \varphi) \nu(dx_{0}) = \mathbb{E}_{\nu}^{\varphi}(e^{-aT_{i}} \mathbb{I}_{B}(x_{i}))
\]

and

\[
\nu(B) = \mathbb{E}_{\nu}^{\varphi}(e^{-aT_{0}} \mathbb{I}_{B}(x_{0})).
\]

Hence

\[
\hat{m}(B) = \sum_{i=0}^{\infty} \mathbb{E}_{\nu}^{\varphi}(e^{-aT_{i}} \mathbb{I}_{B}(x_{i})) = m(B \times \mathbb{A} : \pi, \nu, \alpha) = \hat{m}(B : \pi, \nu, \alpha),
\]

therefore

\[
m = m \otimes \varphi = \hat{m} \otimes \varphi(\cdot : \pi, \nu, \alpha) = m(\cdot : \pi, \nu, \alpha).
\]

\( \square \)

**Corollary (3.11).** The family of stationary policies is sufficient for the control problems.

Let us denote

\[
\langle m, f \rangle := \int f(x, \alpha) m(dx, \alpha)
\]

Let \( k_{j} \geq 0 \) for \( j \in \{0, 1, \ldots, N\} \) given.

The control problems now has the form

\[
\text{MDCP} \quad \min \langle m, C_{0}^{a} \rangle \\
\text{subject to} \quad \langle m, C_{j}^{a} \rangle \leq k_{j} \quad \text{for} \quad j \in \{1, 2, \ldots, N\} \\
\text{and} \quad \hat{m} = \nu + (m \otimes H_{a}) \\
m \in \mathcal{M}_{+}(\mathbb{K}).
\]

Actually.
Corollary (3.12). The problems DCP and MDCP are equivalent.

4. Existence of solution

Remember that $X$ and $A$ are Borel spaces. The functions $c_j$ and $d_j$ are non-negative for $j \in \{1, 2, \ldots, N\}$ and $F$ is a continuous function. By using the former corollary we shall prove that MDCP is solvable in Theorem (4.4) below.

Condition (4.1). Let us suppose:

(a) There is a policy $\pi \in \Pi$ such that $V^\alpha_{0j}(\pi, \nu) \leq k_j$ for $j \in \{1, 2, \ldots, N\}$ and $V^\alpha_{00}(\pi, \nu) < +\infty$.

(b) The function $c_0$ is inf-compact or is a moment and lower semicontinuous.

(c) The functions $c_k$ and $d_j$ are lower semicontinuous functions for $k \in \{1, 2, \ldots, N\}$ and $j \in \{0, 1, \ldots, N\}$.

(d) The stochastic kernel $Q$ is weakly continuous.

(e) There is a density function $f(t|x, a, y)$ for $F(t|x,a,y)$ such that $f$ is an uniformly continuous function on all its variables.

Remark (4.2). Claus (a) is in order the problem DCP (or MDCP) makes sense. Claus (b) assure we have a tight family, so we can apply Prohorov’s Theorem. Claus (c) implies that with limits we fulfill the constraints. Clauses (d) and (e) force the function $\tau_\alpha$ to be continuous.

Let us define the set $M_f$ of all feasible occupation measures, that is, the set of all measures $m \in M_+(K)$ such that

$$\langle m, C^0 \rangle < +\infty,$$
$$\langle m, C^j \rangle \leq k_j \text{ for } j \in \{1, 2, \ldots, N\},$$

and $\hat{m} = \nu + (m \otimes H_\alpha)$

and let us define the value of the program MDCP by

$$\bar{V} = \inf \{ \langle m, C^0 \rangle : m \in M_f \}.$$

Lemma (4.3). Under Condition (4.1) the function $\tau_\alpha$ given in (2.6) is a continuous function.

Proof. As $f(t|x,a,y)$ is an uniformly continuous function on $(t,y)$, given $\epsilon > 0$ and $(x,a) \in X \times A$, there is $\delta > 0$, such that

$$|f(t|x,a,y) - f(t|x,a,y')| < \epsilon,$$

for all $(t,y), (t,y') \in (0, +\infty) \times X$ such that $d_1(y,y') < \delta$, where $d_1$ is the distance in the space $X$. Hence

$$\left| \int_0^{+\infty} e^{-at} f(dt|x,a,y) - \int_0^{+\infty} e^{-at} f(dt|x,a,y') \right| = \epsilon.$$

That is, for each $(x_0,a_0) \in X \times A$, the function

$$y \mapsto \int_0^{+\infty} e^{-at} f(dt|x_0,a_0,y)$$
is continuous on $\mathcal{X}$. Hence, by Condition (4.1) (d), the function
\[
(x, a) \mapsto \int_{\mathcal{X}} \int_{0}^{+\infty} e^{-at} f(dt|x_0, a_0, y) Q(dy|(x, a))
\]
is a continuous function on $\mathcal{X} \times \mathcal{A}$ for each $(x_0, a_0) \in \mathcal{X} \times \mathcal{A}$. Now from the inequality
\[
\left| \int_{\mathcal{X}} \int_{0}^{+\infty} e^{-at} f(dt|x', a', y) Q(dy|x, a') \right|
- \left| \int_{\mathcal{X}} \int_{0}^{+\infty} e^{-at} f(dt|x', a, y) Q(dy|x, a) \right|
\leq \left| \int_{\mathcal{X}} \int_{0}^{+\infty} e^{-at} f(dt|x', a', y) Q(dy|x', a') \right|
- \left| \int_{\mathcal{X}} \int_{0}^{+\infty} e^{-at} f(dt|x', a', y) Q(dy|x, a) \right|
+ \int_{\mathcal{X}} \int_{0}^{+\infty} |f(dt|x', a', y) - f(dt|x, a, y)| Q(dy|x, a),
\]
we can see that the function $\tau_a$ is a continuous function. □

**Theorem (4.4).** If conditions (2.7) and (4.1) holds then the MDCP is solvable.

**Proof.** Let $(m_i)_{i=1}^{\infty}$ be a sequence of occupation measures in $\mathcal{M}_f$ such that $\langle m_i, C_0^0 \rangle \leq V$. Let $m_i(\cdot) = m_i(\cdot : \pi, \nu, a)$. By Condition (2.7) the family of occupation measures is uniformly bounded by $\frac{1}{1-a}$.

Now, from Condition (4.1), (a) and (b) the family of occupation measures $\mathcal{M}_f$ is tight. Hence by Prohorov’s Theorem, there is a measure $m_0$ and a subsequence $(m_{n_i})_{i=1}^{\infty}$ which is weakly convergent to $m_0$.

From this we obtain that its marginals $\hat{m}_{n_i} \rightarrow \hat{m}_0$ and by Theorem (3.9) we have $\hat{m}_0 = \nu + (m_0 \otimes \varphi_0)$, then $m_0(\cdot) = m_0(\cdot : \varphi, \nu, a)$.

Now by (4.1), (b) and (c) $c_j$ and $d_j$ are lower semicontinuous functions for $j \in \{0, 1, \ldots, N\}$, then by Lemma (2.6) $C_j^a$ is a semicontinuous function for $j \in \{0, 1, \ldots, N\}$. By Fatou Lemma and since $\hat{m}_{n_i} \rightarrow \hat{m}_0$, $\langle m_0, C_j^a \rangle \leq \liminf_k \langle m_k, C_j^a \rangle \leq k_j$ for $j \in \{1, \ldots, N\}$. Finally $V \leq \langle m_0, C_0^0 \rangle \leq \liminf_k \langle m_k, C_0^0 \rangle \leq V$. □

### 5. Characterization of the solutions

In this section we shall prove that if the stochastic kernel is nonatomic, then there exists an $N+1$-randomized optimal policy. For this we shall need some preliminaries definitions and lemmas.

**Definition (5.1).** Let $\mu$ a finite (nonnegative) measure on $\mathcal{B}(\mathcal{A})$. Then $\mu$ is said to be:

(a) **regular** if $\mu(D) = \sup\{\mu(C) : C \subset D \text{ and } C \text{ is closed}\}$ for each Borel set $D \in \mathcal{B}(\mathcal{A})$;
(b) \(\tau\)-smooth if for every decreasing net \((F_\eta)\eta\) of closed subsets of \(S\) we have \(\mu(\cap_\eta F_\eta) = \inf_\eta \mu(F_\eta)\).

Remark (5.2). (a) If \(\mathcal{A}\) is a Hausdorff (or \(T_2\)) space, then every Radon measure on \(\mathcal{A}\) is \(\tau\)-smooth, and if \(\mathcal{A}\) is regular (or \(T_3\)), then every \(\tau\)-smooth measure is regular (see e.g., [30 Proposition I.3.1]).

(b) If \(\mathcal{A}\) is strongly Lindelöf (which is the case e.g., if \(\mathcal{A}\) is a Suslin space, see [26 p. 104]), then every finite measure on \(B(\mathcal{A})\) is \(\tau\)-smooth. The latter fact and (a) yield, in particular the following.

(c) In particular, if \(\mathcal{A}\) is a locally compact and separable metric space, the parts (a) and (b) imply that each p.m. on \(B(\mathcal{A})\) is both \(\tau\)-smooth and regular.

By Remark (3.5) and [6 Th. 2.6] we get immediately the next theorem.

Theorem (5.3). Let \(\mathcal{X}\) be an arbitrary topological space, and \(\mathcal{A}\) a topological space such that every p.m. on \(B(\mathcal{A})\) is \(\tau\)-smooth and regular. Fix an arbitrary finite measure \(\hat{\mu}\) on \(B(\mathcal{X})\), nonnegative real-valued measurable functions \(C_1^\alpha, C_2^\alpha, \ldots, C_N^\alpha\) on \(\mathcal{X}\), and real numbers \(k_1, \ldots, k_N\). Consider the set \(\Lambda \subset \Phi\) that consists of all the randomized strategies \(\varphi \in \Phi\) for which

\[
\int C_j^\alpha(x, \varphi(x)) \hat{\mu}(dx, a)) \leq k_j \quad \text{for all} \quad j \in \{1, \ldots, N\},
\]

and let \(\text{ex}(\Lambda)\) be the set of extreme points of \(\Lambda\). Then:

1. \(\Lambda\) is convex and

\[
\text{ex}(\Lambda) \subset \Phi_{N+1}^0
\]

where \(\Phi_{N+1}^0\) is the set of all the \((N + 1)\)-randomizations of the form \(\varphi(|x) = \sum_{j=1}^{N+1} \lambda_j \mathbb{1}_{J(f_j(x))} \in \Phi_{N+1}\) for some \(f_j \in \mathcal{F}\) and nonnegative numbers \(\lambda_j\) such that

\[
\sum_{j=1}^{N+1} \lambda_j = 1 \quad \text{and the vectors}
\]

\[
\left(\int C_1^\alpha(x, f_j(x)) \hat{\mu}(dx), \ldots, \int C_N^\alpha(x, f_j(x)) \hat{\mu}(dx)\right) \in \mathbb{R}^{N+1},
\]

for \(j \in \{1, \ldots, N\}\), are linearly independent.

2. If equality holds in (5.4), then we have equality of the sets in (5.5).

Remark (5.7). (a) Theorem 5.3 requires the action set \(\mathcal{A}\) is a topological space such that

\[
\text{every p.m. on } B(\mathcal{A}) \text{ is } \tau\text{-smooth and regular.}
\]

This condition ensures that the set of extreme points of the space \(\mathcal{P}(\mathcal{A})\) of p.m.’s on \(\mathcal{A}\) coincides with the set of Dirac measures \(\delta_a\) for all \(a \in \mathcal{A}\) (see, for instance, [29 Th. 11.1]).

(b) If a p.m. is tight, then it is \(\tau\)-smooth and regular (see [29 p. xiii]). It follows that to obtain (5.8) it suffices to give conditions on \(\mathcal{A}\) so that every p.m. on \(B(\mathcal{A})\) is tight. This is the case if, for instance, \(\mathcal{A}\) is: (i) a \(\sigma\)-compact Hausdorff space; (ii) a Polish space; or (iii) a locally compact separable metric space.

(See [26, 29].)

In the remainder we also consider the following sets:
• The convex cone $D_+:={\mathbb R}_+ \cdot \Phi$ of the so-called transition measures restricted to $\mathcal K$;
• the linear space $D:=D_+-D_+$ of signed transition measures with the obvious definitions of sum and scalar multiplication; and
• $\mathcal M(\mathcal K)$ the linear space of finite signed measures on $\mathcal X \times \mathcal A$ concentrated on $\mathcal K$.

Let $l:D \to \mathcal M(\mathcal K)$ be the linear mapping defined by $l(\varphi):=\nu \otimes \varphi$, where $\nu$ is finite measure on $\mathcal X$ and $\nu \otimes \varphi$ is as in Remark (3.5). We define the quotient space $\overline D:=D/\ker(l)$, where $\ker(l):=\{\varphi \in D: (\varphi)=0\}$ is the kernel of $l$. For each $\varphi \in D$, let $\overline \varphi:=(\varphi' \in D: \nu \otimes \varphi' = \nu \otimes \varphi)$ be the corresponding equivalence class in $\overline D$, and the quotient sets $\overline{\Phi}, \overline{\Lambda}, \overline{\Phi}^0_{N+1}$ are defined similarly. For instance, $\overline{\Lambda}:=(\overline{\varphi}: \varphi \in \Lambda)$.

In the next lemma we use the following notation. If $\nu$ is a finite measure on $\mathcal X$ and $\Phi'$ is a subfamily of randomized strategies in $\Phi$, then $\nu \otimes \Phi':=\{\nu \otimes \varphi: \varphi \in \Phi'\}$.

Next lemma can be proved as in [6, Th. 2.6] and [8, Th. 5.6].

**Lemma (5.9).** 1. If $\nu \otimes \varphi$ is an extreme point of $\nu \cdot \Phi$ (resp., $\nu \cdot \Lambda$), then $\overline{\varphi}$ is an extreme point of $\overline{\Phi}$ (resp., $\overline{\Lambda}$).
2. If $\overline{\varphi}$ is an extreme point of $\overline{\Phi}$, then $\overline{\varphi}$ has an extreme point $f$ in $\overline{\Phi} \cap \overline{\varphi}$.
3. If $\overline{\varphi}$ is an extreme point of $\overline{\Lambda}$, with $\Lambda$ as in Theorem (5.3), then $\overline{\varphi}$ has an extreme point $\varphi^*$ of $\Lambda$ with $\varphi^*$ in $\overline{\Phi} \cap \overline{\varphi}$.

**Theorem (5.10).** (Bauer’s extremum principal (see [2])) If $S$ is a compact convex subset of a locally convex Hausdorff topological vector space, then every l.s.c. concave function on $S$ achieves its minimum at an extreme point.

Finally we get as in [6, Th. 6.2] and [8, Th. 5.8].

**Theorem (5.11).** Suppose that $\mathcal A$ is locally compact separable metric space and that $Q$ and $v$ are nonatomic, then there exists an $N+1$-randomized optimal policy.

### 6. Example

Let $\mathcal X = \mathcal A = [0, +\infty)$, $A(x) = [0, x]$ and $v(B) = \int_B e^{-t}dt$, for $B \in \mathcal B(\mathcal X)$.

Let us consider a device such that the probability of passing from a state $x \in \mathcal X$ with an action $a \in A(x)$ to a state in $B \in \mathcal B(\mathcal X)$ is given by

$$Q(B|x,a) = \int_B \lambda_2(x+1-a) \exp(-\lambda_2(x+1-a)y)dy$$

and the transition occurs in a random time whose distribution function is given by

$$F(t|x,a,y) = 1 - \exp(-\lambda_1(x+1-a)yt^2),$$

and so

$$f(t|x,a,y) = (-2\lambda_1(x+1-a)yt) \exp(-\lambda_1(x+1-a)yt^2).$$

The current operation cost functions to be minimized are $c_0(x,a) = \gamma_0 a^2 + \gamma_1 a^2$ and $d_0(x,a) = \eta_0 x^2$. The cost function for which is important to keep under some bounds are $c_1(x,a) = \gamma_2(x-a)^2$ and $d_1(x,a) = \eta_1 x$ which represent some measure of risk associated with big values. Then the distribution $G$ and its density $g$ are independent of $(x,a)$. $G_0(t) = G(t|x,a) = 1 - \frac{\lambda_2}{\lambda_1 t^2 + \lambda_2}, \quad g_0(t) = g(t|x,a) = \frac{2\lambda_1 \lambda_2}{(\lambda_1 t^2 + \lambda_2)^2}$. 

The expected value of the time $T$ is $\tau_\alpha = \tau_\alpha(x,a) = \left(\frac{\lambda_2}{\lambda_1}\right)^{\frac{1}{2}} \frac{\pi}{2} = \bar{\tau}_\alpha$. The other costs are $C^a_0(x,a) = \gamma_0 a^2 + \eta_0 \tau_\alpha x^2$, $C^a_1(x,a) = \gamma_1 (x - a)^2 + \eta_1 \tau_\alpha x$. The kernel $H_\alpha$ of finite measures is $H_\alpha(C|x,a) = \int e^{-at}2\lambda_1 2\lambda_2 \frac{t}{(\lambda_1 t^2 + \lambda_2)^2} dt$. Now we shall see that this example satisfies all the conditions. First, Condition (2.7) is fulfilled if

$$\frac{\lambda_2}{\lambda_1 t^2 + \lambda_2} \geq \bar{\varepsilon} > 0.$$  

Condition (4.1): (a) To see that there is a policy $\pi \in \Pi$ such that $V^a_j(\pi,\nu) \leq k_1$ and $V^a_0(\pi,\nu) < +\infty$ let us consider the deterministic stationary policy given by the function $f(x) = x$. Hence

$$V^a_0(\pi,\nu) \leq 2(\gamma_0 + \gamma_1 + \eta_0 \tau_\alpha) \sum_{i=1}^{\infty} (\tau_\alpha)^i$$

and

$$V^a_1 \leq 1 + \frac{1}{\lambda_1} \sum_{i=1}^{\infty} (\tau_\alpha)^i,$$

that is, this condition is hold if $1 + \frac{1}{\lambda_1} \sum_{i=1}^{\infty} (\tau_\alpha)^i \leq k_1$.

(b), (c) and (d) hold.

(e) The fact that the density function $f$ is uniformly continuous is a consequence of its properties. It is nonnegative, bounded, analytic, the function itself and all its derivatives tend to zero when its argument tend to infinity and is such that the maximum of the absolute value of all its second partial derivatives are reached, hence it is uniformly continuous.

Finally the space $A$ is locally compact separable metric space and $Q$ and $\nu$ are nonatomic.

7. Conclusions and open problems

In this article, for discounted constrained semi-Markov decision processes in Borel spaces, we transform the original control problem in an optimization problem in the space of finite measures. This allowed to demonstrate the existence of solutions to the control problem, to characterize the extreme points of this family and to show there are solutions which are extreme points.

A work in process that we can mention is to find the analogous family of occupation measures for average constrained semi-Markov decision processes in Borel spaces. The target is to demonstrate existence of solutions to the control problem and to characterize the solutions.

In the model worked in this article it was considered that the actions are taken just in base of the previous states and actions, independent of sojourn times. A variant is to allow the actions may depend on the sojourn times also and that the dynamic of the system were described by a stochastic kernel on $\mathbb{X} \times [0,\infty) \times A$. The problem is to find the family of occupation measures and to follow the scheme of this article.

Other open problem is to pose the control problem as an infinite linear programming. To do this, the first step is to characterize the family of occupation measures. Moreover this family allows to use also convex programming.
Acknowledgment

The present work was partially supported by the Programa de Apoyo a la Investigación Científica y Tecnológica of the Universidad Autónoma de Nuevo León under the grant CE396-10. The authors wish to thank the anonymous reviewers for their valuable suggestions.

Received May 23, 2012

Final version received April 2, 2013

JUAN GONZÁLEZ-HERNÁNDEZ
DEPARTAMENTO DE PROBABILIDAD Y ESTADÍSTICA
I. I. M. A. S.
UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
A.P. 20-726, 01000 MÉXICO, D.F.
MÉXICO
juan@sigma.iimas.unam.mx

CESAR EMILIO VILLARREAL-RODRÍGUEZ
FACULTAD DE INGENIERÍA MECÁNICA Y ELÉCTRICA
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
A.P. 66-F, CIUDAD UNIVERSITARIA
66455 SAN NICOLÁS DE LOS GARZA, N. L.
MÉXICO
cesar.villarrealrd@uanl.edu.mx

REFERENCES


