Morfismos, Vol. 11, No. 1, 2007, pp. 15–36

Average optimality for semi-Markov control processes *

Anna Jaśkiewicz Andrzej S. Nowak

Abstract

This paper is a survey of some recent results on the average cost optimality equation for semi-Markov control processes. We assume that the embedded Markov chain is V-geometric ergodic and show that there exist a solution to the average cost optimality equation as well as an (ε-)optimal stationary policy. Moreover, we also prove the equivalence of two optimality cost criteria: ratio-average and time-average, in the sense that they lead to the same optimal costs and (ε-)optimal stationary policies.

2000 Mathematics Subject Classification: 90C40, 93E20, 60K15.
Keywords and phrases: Borel state space, the average cost optimality equation.

1 Introduction

In this paper we deal with the ratio-average and time-average cost optimality criteria for semi-Markov control processes on a Borel space. First, we only assume that the one-step cost function is lower semianalytic, and the transition probability function satisfies certain ergodicity conditions. For such a model, a lower semianalytic solution to the optimality equation and a universally measurable ε-optimal stationary policy are shown to exist with respect to the ratio-average cost criterion. Next we indicate that additional regularity assumptions (either (**B1**) or (**B2**)) allow us to obtain either a Borel measurable or continuous solution to the optimality equation. Moreover, we show that in these cases there exists a Borel measurable optimal stationary policy.

*__Invited article__

In order to establish the optimality equation we apply a fixed point theorem. The idea of using this method for semi-Markov control processes satisfying general ergodicity assumptions belongs to Vega-Amaya [26]. He solved the optimality equation in the case when regularity assumptions (**B1**) are satisfied. Then, his concept was applied by Jaśkiewicz [11] to semi-Markov models with lower semicontinuous cost functions and weakly continuous transition laws (assumptions (**B2**)). However, one can imagine examples of (semi-)Markov control processes with weakly continuous transition probabilities and one-step payoff functions, which are not neccessarily lower semicontinuous [12]. Moreover, there are examples of such models that meet neither conditions (**B1**) nor (**B2**). Nevertheless, the optimality equation can be still derived using a fixed point argument. This fact, in turn, allows us to consider the time-average cost criterion within such a general framework. Starting with the optimality equation we are able to obtain (ε-)optimal stationary policies and optimal costs with respect to the time-average cost criterion.

The paper is organized as follows. First we recall certain terminology and facts concerning lower semianalytic functions and Borel as well as universally measurable selectors. Then we present the model and introduce our assumptions. In Section 4 we discuss a solution to the average cost optimality equation, whilst Section 5 is devoted to a study of the time-average cost criterion. We end this section with an example illustrating that the average cost optimality criteria may lead to different optimal policies and optimal costs.

2 Preliminaries

At the beginning we give the definitions of Borel, analytic and universally measurable sets and functions. For further and more complete terminology the reader is referred to [1].

Definition 1. We call X a *Borel space*, if X is a non-empty Borel subset of some Polish space, i.e., complete separable metric space, and it is endowed with σ-algebra $\mathcal{B}(X)$ of all its Borel subsets.

Let N^N be the set of sequences of positive integers, equipped with the product topology. So N^N is a Polish space. Let A be a separable metric space.

Definition 2. A is called an *analytic set* or *analytic space* provided there is a continuous function g on N^N whose range $g(N^N)$ is A.

There are other equivalent definitions of analytic sets in a Borel space X. One posiibility is to define them as the projection on X of the Borel subsets of $X \times Y$, where Y is some uncountable Borel space.

Now let E be an analytic subset of an analytic space X and let p be any probability measure on the Borel subsets of X.

Definition 3. E is *universally measurable*, if E is in the completion of the Borel σ-algebra with respect to every probability measure p.

From now on let X and Y be Borel spaces. Let $\mathcal{A}(X)$ be the analytic σ-algebra and $\mathcal{U}(X)$ be the σ-algebra of all universally measurable subsets of X.

Definition 4. We say that a function $f : X \mapsto Y$ is *analytically measurable* [*universally measurable*] if $f^{-1}(B) \in \mathcal{A}(X)$ $[f^{-1}(B) \in \mathcal{U}(X)]$ for every $B \in \mathcal{B}(Y)$.

We have $\mathcal{B}(X) \subset \mathcal{A}(X) \subset \mathcal{U}(X)$. Hence, every Borel measurable function is analytically measurable, and every analytically measurable function is universally measurable.

Definition 5. Let $B \subset X$ and $f : B \mapsto R$. If B is analytic and the set $\{x \in B : \ f(x) < c\}$ is analytic for each $c \in R$, then f is said to be *lower semianalytic* (l.s.a.).

Now we are in a position to recall some basic results on l.s.a. functions and universally measurable selectors.

Lemma A. (Proposition 7.48 in [1]) Let $f : X \times Y \mapsto R$ be l.s.a., and $q(dy|x)$ a Borel measurable stochastic kernel on Y given X. Then, the function $\bar{f} : X \mapsto R$ defined by

$$\bar{f}(x) = \int_Y f(x,y)q(dy|x)$$

is l.s.a.

Lemma B. (Jankov-von Neumann theorem) If $K \subset X \times Y$ is analytic, then there exists an analytically measurable function $\phi : \text{proj}_X(K) \mapsto Y$ such that

$$\text{Gr}(\phi) := \{(x,y) : \ y = \phi(x), \ x \in \text{proj}_X(K)\} \subset K.$$

For the proof the reader is referred to [1], p. 182. This lemma brings us to the following selection theorem for l.s.a. functions.

Lemma C. Let $K \subset X \times Y$ be analytic and $f : K \mapsto R$ be l.s.a. Define $f^* : \mathrm{proj}_X(K) \mapsto R$ by

$$f^*(x) = \inf_{y \in Y(x)} f(x, y),$$

with $Y(x) := \{y \in Y : (x, y) \in K\}$. Then, the following holds

(a)f^* is l.s.a. function,

(b) the set

$$I = \{x \in \mathrm{proj}_X(K) | \text{for some } y_x \in Y(x),\ f(x, y_x) = f^*(x)\}$$

is universally measurable, and for every $\varepsilon > 0$ there exists a universally measurable function $\phi : \mathrm{proj}_X(K) \mapsto Y$ such that $\mathrm{Gr}(\phi) \subset K$ and for all $x \in \mathrm{proj}_X(K)$

$$f(x, \phi(x)) = f^*(x), \text{ if } x \in I, \quad \text{and} \quad f(x, \phi(x)) \le f^*(x) + \varepsilon, \text{ if } x \notin I.$$

Part (a) follows from the proof of Proposition 7.47 in [1], whilst part (b) is a consequence of Proposition 7.50 in [1].

Definition 6. Let K be a Borel set and $\mathrm{proj}_X(K) = X$. It is said that K admits a *graph* (Borel measurable selection or uniformization), if there exists a Borel measurable function $\phi : X \mapsto Y$ such that

$$\phi(x) \in Y(x).$$

It is worth mentioning that a Borel set K need not have a graph [2]. However, it is well-known that if $Y(x)$ is σ-compact for each $x \in X$, then K contains a graph [3].

3 The model

A *semi-Markov control process* is described by the following objects:

(i) The *state space* X is a standard Borel space.

(ii) A is a Borel *action space*.

(iii) K is a non-empty analytic subset of $X \times A$. We assume that for each $x \in X$, the non-empty x-section

$$A(x) = \{a \in A : (x, a) \in K\}$$

of K represents the *set of actions available* in state x.

(iv) $Q(\cdot|x,a)$ is a regular transition measure from $X \times A$ into $R_+ \times X$, where $R_+ = [0,\infty)$. It is assumed that $Q(D|x,a)$ is a Borel function on $X \times A$ for any Borel subset $D \subset R_+ \times X$ and $Q(\cdot|x,a)$ is a probability measure on $R_+ \times X$ for any $x \in X$, $a \in A(x)$. Denote

$$Q(t,\hat{X}|x,a) := Q([0,t] \times \hat{X}|x,a)$$

for any Borel set $\hat{X} \subset X$. If $a \in A(x)$ is selected in state x, then $Q(t,\hat{X}|x,a)$ is the joint probability that the sojourn time is not greater than $t \in R_+$ and the next state $y \in \hat{X}$. By $H(\cdot|x,a)$ denote a distribution of the sojourn time when the process is in state x and action $a \in A(x)$ is selected, that is, $H(t|x,a) = Q(t,X|x,a)$. Let $\tau(x,a)$ be the mean holding time, i.e.,

$$\tau(x,a) = \int_0^\infty tH(dt|x,a).$$

Put $q(\cdot|x,a) := Q(R_+,\cdot|x,a)$. Then, q is called *the transition law of the embedded Markov process.* Moreover, the distribution of the sojourn time and the next state are independent conditional on (x,a), i.e.,

$$Q(t,\hat{X}|x,a) = q(\hat{X}|x,a)H(t|x,a).$$

(v) Let $c_i : K \mapsto R$, $i = 1,2$. Then, the expected one-step *cost function* $c : K \mapsto R$ equals

$$c(x,a) = c_1(x,a) + \tau(x,a)c_2(x,a).$$

Here c_1 is an immediate cost paid by the decision maker at the transition time and the cost c_2 is incurred until the next transition occurs.

Put $T_0 := 0$. Let $\{T_n\}$ denote a sequence of random decision (jump) epochs. If the initial state is $x = x_0$ and the action $a_0 \in A(x)$ is selected, then the immediate cost $c_1(x,a_0)$ is incurred for the decision maker and the process remains in state x up to the time $T = T_1 - T_0 = T_1$. The cost $c_2(x,a_0)$ per unit time is incurred until the next transition occurs. Afterwards the system jumps to the state x_1 according to the probability measure $q(\cdot|x,a_0)$. The decision maker chooses again an action $a_1 \in A(x_1)$ and the process remains in state x_1 for a random time $T_2 - T_1$. The cost $c_1(x_1,a_1) + (T_2 - T_1)c_2(x_1,a_1)$ is incurred and a new state x_2 is generated according to the distribution $q(\cdot|x_1,a_1)$. This situation repeats itself yielding a trajectory $(x_0,a_0,t_1,x_1,a_1,t_2,\ldots)$ of some stochastic process, where x_n and a_n describe the state and the chosen action, respectively, on the nth stage of the process. Obviously, t_n is a realization of the random variable T_n, and a distribution function of the random holding time $T_{n+1} - T_n$ is $H(\cdot|x_n, a_n)$.

A *policy* is a sequence $\pi = \{\pi_n\}$ where π_n $(n \geq 0)$ is a universally measurable stochastic kernel on A given $(X \times A \times R_+)^n \times X$ satisfying $\pi_n(A(x_n)|h_n) = 1$ for any history $h_n = (x_0, a_0, t_1, \ldots, x_n)$ of the process (Clearly, $h_0 = x_0$.) By Π^0 we denote the class of all policies. Let F^0 be the set of all universally measurable transition probabilities f from X to A such that $f(x) \in A(x)$ for each $x \in X$. A *stationary policy* π is of the form $\pi = \{f, f, \ldots\}$, where $f \in F^0$. Thus, every stationary policy $\pi = \{f, f, \ldots\}$ can be identified with the mapping $f \in F^0$. Since K is analytic, the Jankov-von Neumann theorem guarantees that there exists at least one $f \in F^0$. Therefore, F^0 and Π^0 are non-empty.

Let $\Omega = (K \times R_+)^\infty$ be the space of all infinite histories of the process endowed with $\mathcal{U}$ (σ-algebra of universally measurable sets in Ω). According to Proposition 7.45 in [1], for any $\pi \in \Pi$ and an initial state $x_0 = x \in X$ there exists a unique probability measure P_x^π defined on Ω. By E_x^π we denote the expectation operator with respect to P_x^π.

Let $\pi \in \Pi^0$, $x \in X$ and $t \geq 0$ be fixed. Put

$$N(t) := \max\{n \geq 0 : T_n \leq t\}$$

as the counting process. By our assumptions, which are presented below, $P_x^\pi(N(t) < \infty) = 1$ (see Remark 2 in [10]).

We shall consider the two average expected costs: - the *ratio-average cost*

$$J(x, \pi) := \limsup_{n \to \infty} \frac{E_x^\pi \left(\sum_{k=0}^{n-1} c(x_k, a_k) \right)}{E_x^\pi \sum_{k=0}^{n-1} (\tau(x_k, a_k))},$$

- the *time-average cost*

$$j(x, \pi) := \limsup_{t \to \infty} \frac{E_x^\pi \left(\sum_{k=0}^{N(t)} c(x_k, a_k) \right)}{t}.$$

For functions $J(x, \pi)$ and $j(x, \pi)$ we define the optimal costs as

$$J(x) := \inf_{\pi \in \Pi^0} J(x, \pi), \quad j(x) := \inf_{\pi \in \Pi^0} j(x, \pi).$$

A policy π^ε is called ε-optimal with respect to the ratio-average cost criterion if

$$J(x, \pi^\varepsilon) - \varepsilon \leq J(x)$$

for all $x \in X$. In a similar way we define the ε-optimality with respect to the time-average cost criterion. Now we are in a position to introduce our assumptions.

(**B0**) *Basic assumptions*: (i) the set K is analytic;

(ii) there exist a constant $B > 0$ and a Borel measurable function $V : X \mapsto [1, +\infty)$ such that

$$|c(x, a)| \leq BV(x) \quad \text{and} \quad |\tau(x, a)| \leq BV(x)$$

for every $(x, a) \in K$;
(iii) the function τ is Borel measurable, whilst c is l.s.a on K.

(**GE**) *V-geometric ergodicity assumptions*: (i) there exists a Borel set $C \subset X$ such that for some $\lambda \in (0, 1)$ and $\eta > 0$, we have

$$\int_X V(y)q(dy|x, a) \leq \lambda V(x) + \eta 1_C(x)$$

for each $(x, a) \in K$; V is the function introduced in (**B0**);
(ii) the function V is bounded on C, i.e.,

$$v_C := \sup_{x \in C} V(x) < \infty;$$

(iii) there exist some $\delta \in (0, 1)$ and a probability measure μ concentrated on the Borel set C with the property that

$$q(D|x, a) \geq \delta\mu(D)$$

for each Borel set $D \subset C$, $x \in C$ and $a \in A(x)$.

For any function $u : X \mapsto R$ define the V-norm

$$\|u\|_V := \sup_{x \in X} \frac{|u(x)|}{V(x)}.$$

Under (**GE**) the embedded state process $\{x_n\}$ governed by a stationary policy $f \in F^0$ is a positive recurrent aperiodic Markov chain and there exists a unique invariant probability measure π_f (consult Theorem 11.3.4 and page 116 in [16]). Moreover, by Theorem 2.3 in [17], $\{x_n\}$ is V-ergodic, that is, there exist $\theta > 0$ and $\alpha \in (0, 1)$ such that

$$\text{(1)} \qquad \Big| \int_X u(y)q^n(dy|x, f(x)) - \int_X u(y)\pi_f(dy) \Big| \leq V(x)\|u\|_V \theta\alpha^n$$

for every u with $\|u\|_V < \infty$, and $x \in X$, $n \geq 1$. Here $q^n(\cdot|x, f(x))$ denotes the n-stage transition probability induced by q and a stationary policy f. As an immediate consequence of (1), one can easily get

$$J(f) := J(x, f) = \frac{\int_X c(x, f(x))\pi_f(dx)}{\int_X \tau(x, f(x))\pi_f(dx)}, \tag{2}$$

for every $f \in F^0$.

Lemma 1. Let (**GE**) hold. Then

(a) $\inf_{f \in F^0} \pi_f(C) \geq \frac{1-\lambda}{\eta}$;

(b) $\sup_{f \in F^0} \int_X V(y)\pi_f(dy) \leq \frac{\eta}{1-\lambda}$;

Proof. Let the process be governed by a stationary policy $f \in F^0$. Integrating both sides of (**GE**, i) with respect to the invariant probability measure π_f we get

$$\int_X V(y)\pi_f(dy) \leq \lambda \int_X V(y)\pi_f(dy) + \eta\pi_f(C).$$

Now part (a) easily follows from the fact that $V \geq 1$, whilst part (b) is a consequence of $\pi_f(C) \leq 1$. □

We also make two additional assumptions on the sojourn time T.

(**R**) *Regularity condition*: there exist $\kappa > 0$ and $\beta < 1$ such that

$$H(\kappa|x, a) \leq \beta$$

for all $x \in C$ and $a \in A(x)$.

(**I**) *Uniform integrability condition*:

$$\lim_{t\to\infty} \sup_{x \in C} \sup_{a \in A(x)} [1 - H(t|x, a)] = 0.$$

Assumption (**R**) ensures that an infinite number of transitions does not occur in a finite time interval. Ross [20], Sennott [21] and Yushkevich [25] assume that assumption (**R**) holds for the whole state space. However, we require (**R**) to hold only for the states $x \in C$. This is because condition (**GE**) implies that the embedded Markov process governed by any policy returns to the set C within the finite number of transitions with probability 1. Therefore, we have to control its behaviour only on the set C. From (**R**) one can easily deduce that

$$\tau(x, a) \geq \kappa(1 - \beta) \quad \text{for } x \in C \text{ and } a \in A(x). \tag{3}$$

For broader discussion of the assumptions the reader is referred to [7, 9, 13, 16, 17, 22].

4 The average cost optimality equation

We begin with an auxiliary result that enables us to replace the function V used in (**GE**) by a new function W. The W-norm defined below will play an essential role in the proofs of our main results.

Lemma 2. Let assumption (**GE**) hold. Then, there exist a measurable function $W > V$ and a constant $\lambda' \in (0,1)$ such that

$$\int_X W(y)q(dy|x,a) \le \lambda' W(x) + \delta 1_C(x) \int_C W(y)\mu(dy).$$

Proof. Define $W(x) := V(x) + \frac{\eta}{\delta}$. Then, simple calculations give

$$\begin{aligned}\int_X W(y)q(dy|x,a) &= \int_X V(y)q(dy|x,a) + \frac{\eta}{\delta} \le \left[\lambda V(x) + \frac{\eta}{\delta}\right] + \eta 1_C(x) \\ &\le \frac{\lambda + \frac{\eta}{\delta}}{1 + \frac{\eta}{\delta}} W(x) + 1_C(x)\delta \left[\frac{\eta}{\delta} + \int_C V(y)\mu(dy)\right].\end{aligned}$$

Hence, the result holds with $\lambda' = \frac{\lambda + \frac{\eta}{\delta}}{1 + \frac{\eta}{\delta}}$. □

For any function $u : X \mapsto R$ we define the W-norm as

$$\|u\|_W := \sup_{x \in X} \frac{|u(x)|}{W(x)}.$$

From now on, we shall take into consideration the functions, which have finite W-norm. Moreover, note that

$$\|u\|_V < \infty \quad \text{iff} \quad \|u\|_W < \infty.$$

Let L^0_W denote the set of all l.s.a. functions whose W-norm is finite. Note that L^0_W is a Banach space.

For any $(x,a) \in K$ set

$$p(\cdot|x,a) := q(\cdot|x,a) - 1_C(x)\delta\mu(\cdot).$$

Observe that from Lemma 1 we have

$$\int_X W(y)p(dy|k) = \int_X W(y)p(dy|x,a) \le \lambda' W(x). \tag{4}$$

Put

$$g := \inf_{f \in F^0} J(f).$$

From (**B0**), (**GE**, i) and (**R**) we conclude that $g < \infty$. Indeed, by (2) and (3)

$$|J(f)| \leq \frac{\int_X |c(x,f(x))|\pi_f(dx)}{\int_X \tau(x,f(x))\pi_f(dx)} \leq \frac{B\int_X V(x)\pi_f(dx)}{\kappa(1-\beta)\pi_f(C)}.$$

Now Lemma 1 yields

$$g \leq \frac{B\frac{\eta}{1-\lambda}}{\frac{1-\lambda}{\eta}\kappa(1-\beta)}.$$

For any function $u \in L^0_W$ define the operator $\mathbf{T}$ in the following way

$$(5) \qquad (\mathbf{T}u)(x) := \inf_{a\in A(x)}\left(c(x,a) - g\tau(x,a) + \int_X u(y)p(dy|x,a)\right)$$

for all $x \in X$.

Theorem 1. Assume (**B0**, **GE**, **R**). (a) There exist a constant g^* and a function $h \in L^0_W$ such that

$$(6) \qquad h(x) = \inf_{a\in A(x)}\left(c(x,a) - g^*\tau(x,a) + \int_X h(y)q(dy|x,a)\right)$$

for all $x \in X$.(b) For any $\varepsilon > 0$ there exists a universally measurable function $f^\varepsilon \in F$ such that

$$(7) \quad h(x) \geq c(x,f^\varepsilon(x)) - g^*\tau(x,f^\varepsilon(x)) + \int_X h(y)q(dy|x,f^\varepsilon(x)) - \varepsilon$$

for all $x \in X$.(c) Moreover, $g^* = g = \inf_{\pi\in\Pi^0} J(x,\pi)$ and $g^* \geq J(f^\varepsilon) - \varepsilon$.

The proof of Theorem 1 is similar to that of Theorem 1 in [12] and makes use of an idea presented by Vega-Amaya [26]. We first notice that by (4) the operator $\mathbf{T}$ is contractive and maps the Banach space L^0_W into itself. This follows from our assumption (**B0**) and (4). Hence, from the Banach fixed point theorem there exists $h \in L^0_W$ such that

$$(8) \qquad h(x) = \inf_{a\in A(x)}\Big(c(x,a) - g\tau(x,a) + \int_X h(y)q(dy|x,a) - 1_C(x)\delta\int_X h(y)\mu(dy)\Big).$$

Clearly, if $x \notin C$ then (8) becomes (6). If, on the other hand, $x \in C$ we define

$$d := -\delta\int_C h(x)\mu(dx)$$

and are going to show that $d = 0$. On the contrary, we assume that $d \neq 0$ and proceed along the same lines as in [12]. Therefore, optimality equation (6) is satisfied with $g^* := g$ and the function h. Then, part (b) follows directly from Lemma C(b), whilst part (c) is an immediate consequence of a standard dynamic programming argument (see [7, 9]).

Now we shall describe more specific results, when certain assumptions of regularity are imposed. Assumptions (**B1**) correspond to the model with a strongly continuous transition probability function, whilst conditions (**B2**) are in agreement with a semi-Markov control process, whose transition law is weakly continuous.

(**B1**) *Basic assumptions*:

(i) the set K is Borel and $A(x)$ is compact for any $x \in X$;

(ii) for each $x \in X$, $c(x, \cdot)$ is lower semicontinuous and $\tau(x, \cdot)$ is continuous on $A(x)$;

(iii) for each $x \in X$ and Borel set $D \subset X$, the function $q(D|x, \cdot)$ is continuous on $A(x)$;

(iv) there exists a constant $B > 0$ and a Borel measurable function $V : X \mapsto [1, +\infty)$ such that

$$|c(x, a)| \leq BV(x) \quad \text{and} \quad |\tau(x, a)| \leq BV(x)$$

for every $(x, a) \in K$;

(v) for each $x \in X$, the function

$$\int_X V(y)q(dy|x, \cdot)$$

is continuous on $A(x)$.

(**B2**) *Basic assumptions*: (i) the set K is Borel, $A(x)$ is compact for any $x \in X$, and moreover, the set-valued mapping $x \mapsto A(x)$ is upper semicontinuous, that is, $\{x \in X : A(x) \cap D \neq \emptyset\}$ is closed for every closed set D in A;

(ii) c is lower semicontinuous and τ is continuous on K;

(iii) the transition law q is weakly continuous on K, that is,

$$\int_X u(y)q(dy|x, a)$$

is a continuous function of $(x, a) \in K$ for each bounded and continuous function u;

(iv) there exists a constant $B > 0$ and a continuous function $V : X \mapsto [1, +\infty)$ such that

$$|c(x,a)| \le BV(x) \quad \text{and} \quad |\tau(x,a)| \le BV(x)$$

for every $(x, a) \in K$;

(v) the function

$$\int_X V(y)q(dy|\cdot,\cdot)$$

is continuous in $(x, a) \in K$;

(vi) there exists an open set $\widetilde{C} \subset C$ such that $\mu(\widetilde{C}) > 0$ (recall that μ is a probability measure on the set C).

Let Π $[F]$ be the set of all Borel measurable [stationary Borel measurable] policies. Note that since either (**B1**, i) or (**B2**, i) hold, then from Corollary 1 in [3] F is non-empty. Thus, let $k \in F$. Define

$$g = \inf_{f \in F} J(f).$$

Remark 1. It is worth pointing out that under (**B1**, i) or (**B2**, i) the optimal costs within the classes F to F^0 are same, that is,

$$g = \inf_{f \in F} J(f) = \inf_{f \in F^0} J(f). \tag{9}$$

Indeed, let $f \in F^0$ and D be any Borel subset of X. Then, by the definition of π_f we have

$$\pi_f(D) = \int_X q(D|x, f(x))\pi_f(dx) = \int_X \int_A q(D|x,a)f(da|x)\pi_f(dx).$$

From Lemma 7.28(c) in [1], there exist a Borel set $\hat{X} \subset X$ and a Borel measurable function $\hat{f} : \hat{X} \mapsto A$ such that $\pi_f(\hat{X}) = 1$ and $\hat{f}(x) = f(x)$ for each $x \in \hat{X}$. Now define

$$f^*(x) = \begin{cases} \hat{f}(x), & x \in \hat{X}, \\ k(x), & \text{otherwise.} \end{cases}$$

Hence, $f^* \in F$. Further, we observe that

$$\nu(D) = \int_X q(D|x, f^*(x))\nu(dx), \quad \text{with } \pi_f = \nu.$$

Since our assumptions (**GE**) imply the uniqueness of an invariant probability measure, we conclude that $\nu = \pi_{f^*}$. Therefore, (9) holds.

An analogous conclusion may be drawn in Lemma 1. It allows to re-formulate the lemma with the set F instead of F^0.

Observe that from our discussion in Section 2 it follows that we cannot only assume K is a Borel set. This is because it may occur that $F = \emptyset$, and consequently $\inf_{f \in F} J(f) = \infty$. Therefore, we add the additional assumption: compactness of $A(x)$, in order to apply Corollary 1 in [3].

By B_W [L_W] we denote the space of all Borel measurable [lower semicontinuous] functions that have finite W-norm.

Now we are ready to present our next results.

Theorem 2. Assume (**B1**) [(**B2**)] and (**GE**, **R**).
(a) There exist a constant g^* and a function $h \in B_W$ [$h \in L_W$] such that

$$h(x) = \inf_{a \in A(x)} \left(c(x,a) - g^* \tau(x,a) + \int_X h(y) q(dy|x,a) \right) \tag{10}$$

for all $x \in X$.
(b) There exists a Borel measurable function $\widetilde{f} \in F$ such that

$$h(x) = c(x, \widetilde{f}(x)) - g^* \tau(x, \widetilde{f}(x)) + \int_X h(y) q(dy|x, \widetilde{f}(x))$$

for all $x \in X$.
(c) Moreover, $g^* = g = \inf_{\pi \in \Pi} J(x, \pi)$ and $g^* = J(\widetilde{f})$.

The average cost optimality equation for semi-Markov control processes with strongly continuous transition probability functions satisfying quite general ergodicity assumptions has been established in a few papers [8, 9, 26]. For instance, Hernández-Lerma and Luque-Vásquez [8] apply the so-called Schweitzer's data transformation, Jaśkiewicz [9] examines auxiliary perturbed models, whilst Vega-Amaya [26] makes use of a fixed point theorem, which directly leads to the solution of the optimality equation. In particular, (under slightly different ergodicity assumptions) he defines the operator $\mathbf{T}$ as in (5). Since c and τ are Borel measurable functions, it is easy to observe that $\mathbf{T}$ maps the space B_W into itself and is a contractive operator. Therefore, by the Banach fixed point theorem it follows that there exists a function $h \in B_W$ such that

$$h(x) = \inf_{a \in A(x)} \left(c(x,a) - g\tau(x,a) + \int_X h(y) q(dy|x,a) - 1_C(x) \int_X h(y) \mu(dy) \right).$$

Now it suffices to prove that $\int_X h(y)\mu(dy) = 0$. But this fact can be shown in much the same way as in the proofs of Theorems 3.5 and 3.6 in [26] and by applying Lemma 1. Hence, the optimality equation holds with the function h and constant $g^* := g$, and moreover, by our semi-continuity/compactness assumptions (**B1**) we may replace inf in (10) by min . The existence of a Borel measurable selector of the minima on the right-hand side of (10) follows from a measurable selection theorem [3]. The part (c) is a consequence of a dynamic programming argument. Here, we would like to emphasise that the idea of making use of a fixed point theorem to solve the optimality equation in this set-up belongs to Vega-Amaya [26].

As far as semi-Markov control processes with weakly continuous and V-geometric ergodic transition probabilities are concerned, they were only examined in [11, 14]. A solution to the optimality equation has been obtained in [11] by a fixed point argument. However, in contrast to the previous case the Banach theorem cannot be applied directly. This is because the operator $\mathbf{T}$ need not map the space L_W into itself. To see this peculiarity observe that the function $k(x) := 1_C(x)\int_X u(y)\mu(dy)$ is only Borel measurable. Even if the set C was closed (or open), we would not know the sign of the integral $\int_X u(y)\mu(dy)$ for different functions $u \in L_W$. Consequently, $k(x)$ is not necessarily lower semicontinuous. Therefore, we first have to regularise/smooth an appropriate function in the following way

$$\begin{aligned}\Phi^u(x,a) \quad := \quad & \liminf_{x'\to x,\, a'\to a}\Big[c(x',a') - g\tau(x',a') + \int_X u(y)q(dy|x',a') \\ & -1_C(x')\int_X u(y)\mu(dy)\Big]\end{aligned}$$

Then, the function Φ^u is lower semicontinuous on K and the operator $\widetilde{\mathbf{T}}$ defined as

$$(\widetilde{\mathbf{T}}u)(x) := \inf_{a\in A(x)} \Phi^u(x,a)$$

maps the space L_W into itself and is contractive. For these properties the reader is referred to Lemma 3.3 in [11]. Consequently, there exists a fixed point of $\widetilde{\mathbf{T}}$, a function $h \in L_W$ such that

$$h(x) \quad = \quad \inf_{a\in A(x)} \Phi^h(x,a) \tag{11}$$

$$= \inf_{a\in A(x)} \left[\liminf_{x'\to x,\, a'\to a} \left(c(x',a') - g\tau(x',a') + \int_X h(y)q(dy|x',a') - 1_C(x') \int_X h(y)\mu(dy) \right) \right].$$

Now it suffices to prove that $\int_X h(y)\mu(dy) = 0$ and at the same time dispose of the liminf in (11). This fact has been shown in [11]. Thus, (10) holds with the function h and constant $g^* = g$. Parts (b) and (c) are obvious.

Generally, semi-Markov control processes with Feller transition probabilities require more delicate handling. Indeed, even for V-geometric ergodic Markov decision models, for which the jump times occur at integer points, the issue is not a simple matter, see [6, 13, 15, 22]. For instance, Jaśkiewicz and Nowak [13] and Schäl [22] apply the Fatou lemma for weakly convergent measures, which only yields the optimality inequality. Küenle [15], on the other hand, introduces certain contraction operators that lead to a parametrized family of functional equations. Making use of some continuity and monotonicity properties of the solutions to these equations (with respect to the parameter) he obtains a lower semicontinuous solution of the optimality equation. In contrast to his approach, González-Trejo et al. [6] apply directly the Banach fixed point theorem. Nevertheless, their method has some disadvatages, namely, it requires stronger assumptions and excludes many interesting examples (see Remark 4(b) in [13]).

For further interesting examples of (semi-)Markov control processes the reader is referred to [5, 6, 7, 8, 11, 18, 23] and reference therein.

5 The time-average cost criterion

The following result is concerned with the equivalence of the ratio-average and time-average cost criteria. Generally, these two criteria may have nothing to do with each other, and may lead to different optimal policies and costs (see Example).

Theorem 3. Assume (**GE**, **R**, **I**).

(a)If (**B0**) is satisfied and K admits a graph, then $g^* = \inf_{\pi\in\Pi^0} j(x,\pi)$, and for any $\varepsilon_0 > 0$, the policy $f^{\varepsilon_0} \in F^0$ is an ε_0-optimal.

(b)If either (**B1**) or (**B2**) holds, then $g^* = \inf_{\pi\in\Pi} j(x,\pi) = j(x,\widetilde{f})$.

The proof of Theorem 3 is based on the optional sampling theorem applied to the appropriate sub- and supermartingales that are uniformly integrable. The property of uniform integrability is the main difficulty in the proof and in order to overcome it one needs to employ some basic facts from renewal theory and certain consequences of V-geometric ergodicity. Let us mention that this issue was thoroughly studied in [10] under assumptions (**B1**) and in [14] under conditions (**B2**). Therefore, part (b) follows immediately from Theorem in [10] and Theorem 2 in [14].

As far as part (a) is concerned, its proof is similar to that of part (b). There are only two matters in question that require some explanation. Firstly, we claim that any universally measurable policy can be replaced by a Borel measurable one. This fact is formulated in Lemma 3 below. Secondly, we note that the proof of Theorem 2 in [14] is also valid for any Borel measurable q (not only weakly continuous). Hence, all lemmas in [10, 14] hold true within our general framework. We provide a rough idea of the proof.

Lemma 3. Assume that $\pi \in \Pi$ and $x \in X$ is fixed and let $d : X \mapsto R$ be a Borel measurable function such that $\|d\|_W \leq +\infty$. If K contains a graph,
(a) there exists a Borel measurable semi-Markov policy $\widetilde{\pi}_x$, for which

$$E_x^{\pi} d(x_n) = E_x^{\widetilde{\pi}_x} d(x_n), \quad n = 0, 1, ...$$

(b) the function

$$D(x) := \sup_{\pi \in \Pi^0} E_x^{\pi} d(x_n) = \sup_{\widetilde{\pi} \in \Pi} E_x^{\widetilde{\pi}} d(x_n), \quad n = 0, 1, ...$$

is universally measurable in x.

The proof of part (a) consists of two steps. We first follow Proposition 8.1 in [1] replacing π by an universally measurable semi-Markov policy. Next this policy is superseded by a Borel measurable one (note that this can be done since K contains a graph). Part (b) is a consequence of part (a) and Lemma 7.1 in [24].

Proof of Theorem 3. Let $\mathcal{F}_n$ be the σ-algebra generated by all events up to the nth state. By (6) we infer that

$$S_n = \sum_{k=0}^{n-1} (c(x_k, a_k) - g^* \tau(x_k, a_k)) + h(x_n)$$

is a submartingale with respect to $\mathcal{F}_n$, and by (7)

$$\widetilde{S_n} = \sum_{k=0}^{n-1} (c(x_k, f^\varepsilon(x_k)) - \varepsilon - g^* \tau(x_k, f^\varepsilon(x_k))) + h(x_n)$$

is a supermartingale with respect to $\mathcal{F}_n$. From [10, 14] it follows that $\{S_n\}$ and $\{\widetilde{S_n}\}$ are uniformly integrable, that is,
(I) $E_x^\pi |S_{N(t)+1}|$ and $E_x^{f^\varepsilon} |\widetilde{S_{N(t)+1}}|$ are well defined;
(II) $E_x^\pi [|S_n|; N(t) \geq n]$ and $E_x^{f^\varepsilon} [|\widetilde{S_n}|; N(t) \geq n]$ tend to 0, when $n \to \infty$.
Now applying the optional sampling theorem to $\{S_n\}$ with two stopping times 0 and $N(t)+1$, we get

$$h(x) \leq E_x^\pi \left(\sum_{k=0}^{N(t)} (c(x_k, a_k) - g^* \tau(x_k, a_k)) \right) + E_x^\pi h(x_{N(t)+1}).$$

Simple rearrangments and the fact that

$$E_x^\pi \left(\sum_{k=0}^{N(t)} \tau(x_k, a_k) \right) = E_x^\pi T_{N(t)+1}$$

yield

$$g^* \frac{1}{t} E_x^\pi T_{N(t)+1} \leq \frac{1}{t} E_x^\pi \left(\sum_{k=0}^{N(t)} c(x_k, a_k) \right) + E_x^\pi \left(\frac{h(x_{N(t)+1})}{t} \right) - \frac{h(x)}{t}. \tag{12}$$

From the proofs of Lemma 8 in [10] and Theorem 2 in [14], it follows that

$$\lim_{t \to \infty} \frac{1}{t} E_x^\pi T_{N(t)+1} = 1.$$

In addition, since $h \in L_W$, then $\|h\|_V < +\infty$ (because $\|h\|_W < +\infty$). Hence, making use of Lemma 8 in [10] and Theorem 2 in [14] we conclude

$$\lim_{t \to \infty} \frac{1}{t} E_x^\pi h(x_{N(t)+1}) = 0.$$

Letting $t \to \infty$ in (12) we infer

$$g^* \leq \limsup_{t \to \infty} \frac{E_x^\pi \left(\sum_{k=0}^{N(t)} c(x_k, a_k) \right)}{t} = j(x, \pi).$$

Since $\pi \in \Pi$ is arbitrary, we have

$$g^* \le \inf_{\pi\in\Pi^0} j(x,\pi). \tag{13}$$

Now consider $\widetilde{S_n}$. Applying the optional sampling theorem to $\{\widetilde{S_n}\}$ with two stopping times 0 and $N(t)+1$ we obtain

$$\frac{\varepsilon E_x^{f^\varepsilon} N(t)}{t} + \frac{h(x)}{t} + g^* \frac{E_x^{f^\varepsilon} T_{N(t)+1}}{t} \ge \frac{E_x^{f^\varepsilon}\left(\sum_{k=0}^{N(t)} \left(c(x_k, f^\varepsilon(x_k))\right)\right)}{t} + \frac{E_x^{f^\varepsilon} h(x_N(t))}{t}. \tag{14}$$

Letting $t \to \infty$ in (14) and arguing in the same way as above, we deduce

$$\limsup_{t\to\infty} \frac{\varepsilon E_x^{f^\varepsilon} N(t)}{t} + g^* \ge j(x, f^\varepsilon).$$

Let $M(t)$ be a renewal function that corresponds to an i.i.d. sequence of random variables, each with the following distribution

$$\widetilde{H}(t) := \begin{cases} \beta, & t \in [0,\kappa) \\ 1, & t \ge \kappa, \end{cases} \qquad \widetilde{H}(t) = 0, \quad t < 0.$$

The constants κ and β were introduced in (**R**). From Lemma 6(b) in [10], it follows that

$$E_x^{f^\varepsilon} N(t) \le \theta_C M(t) + \theta(x),$$

with

$$\theta(x) := \frac{1}{\ln(1/\lambda)} \left(\ln V(x) + \frac{\eta}{\lambda} 1_C(x) \right), \quad \theta_C := \sup_{x\in C} \theta(x) \quad (\text{see } (\mathbf{GE})).$$

Moreover,

$$\lim_{t\to\infty} \frac{M(t)}{t} = \frac{1}{\kappa(1-\beta)} \qquad \text{(by Theorem 3.3.2(a) in [19])}.$$

Thus,

$$\limsup_{t\to\infty} \frac{\varepsilon E_x^{f^\varepsilon} N(t)}{t} \le \varepsilon \limsup_{t\to\infty} \frac{\theta_C M(t) + \theta(x)}{t} = \frac{\varepsilon}{\kappa(1-\beta)} =: \varepsilon_0$$

Let $f^{\varepsilon_0} := f^{\varepsilon}$. We infer that

$$g^* \geq j(x, f^{\varepsilon_0}) - \varepsilon_0.$$

From the fact that $\varepsilon > 0$ is arbitrary (so is ε_0), we obtain

$$g^* \geq \inf_{\pi \in \Pi^0} j(x, \pi). \tag{15}$$

Combining (13) and (15) together yields part (a). □

Remark 2. The proof of Theorem 3 discovers a suprising feature of an ε-optimal policy f^{ε} obtained in Theorem 1. Namely, it turns out that f^{ε} is ε_0-optimal within the time-average cost criterion, and ε_0 does not have to be ε. However, ε_0 can be expressed by ε and other constants used in the assumptions.

We present Example 3.2 in [4]. It shows that the two average optimal costs and corresponding optimal policies may be different.

Example. Let $X = \{1, 2, 3\}$, $A(1) = \{c, s\}$ and $A(2) = A(3) = \{s\}$. The mean holding time equals $\tau(1, c) = \tau(1, s) = \tau(2, s) = 1$, $\tau(3, s) = 2$. The transition probabilities are given by $q(2|1, s) = q(3|1, s) = \frac{1}{2}$, $q(2|2, s) = q(3|3, s) = q(1|1, c)$, whilst the one-step cost function is $r(2, s) = 1$, $r(1, c) = \frac{2}{5}$, and 0 otherwise.

In this model, there are two stationary policies $f(1) = c$ and $d(1) = s$. Let $x_0 = 1$, then

$$j(1, d) = \frac{1}{2}, \quad J(1, d) = \limsup_{n \to \infty} \frac{\frac{1}{2} n}{1 + \frac{3}{2} n} = \frac{1}{3}$$

and

$$j(1, f) = \frac{2}{5}, \quad J(1, f) = \frac{2}{5}.$$

Therefore, for the time-average cost criterion, policy f is better than d, whereas for the other one d is better.

Anna Jaśkiewicz
Institute of Mathematics and Computer Science,
Wrocław University of Technology,
Wyspiańskiego 27,
50-370 Wrocław, Poland.
ajaskiew@im.pwr.wroc.pl

Andrzej S. Nowak
Faculty of Mathematics, Computer Science, and Econometrics,
University of Zielona Góra,
Podgórna 50,
65-246 Zielona Góra, Poland,
a.nowak@wmie.uz.zgora.pl

References

[1] Bertsekas D.P.; Shreve, S.E., *Stochastic Optimal Control: The Discrete Time Case,* Academic Press, New York, 1978.

[2] Blackwell D., *A Borel set not containing a graph*, Ann. Math. Stat. **39** (1968), 1345–1347.

[3] Brown L.D.; Purves R., *Measurable selections of extrema*, Ann. Stat. **1** (1973), 902–912.

[4] Feinberg E.A., *Constrained semi-Markov decision processes with average rewards*, Math. Methods Oper. Res. **39** (1994), 257–288.

[5] Feinberg E.A.; Lewis M.E., *Optimality of four-threshold policies in inventory systems with customer returns and borrowing/storage options*, Prob. Eng. Info. Sci. **19** (2005), 45–71.

[6] González-Trejo J.I.; Hernández-Lerma O.; Hoyos-Reyes L.F., *Minimax control of discrete-time stochastic systems*, SIAM J. Control Optim. **41** (2003), 1626–1659.

[7] Hernández-Lerma O.; Lasserre J.B., *Further Topics on Discrete-Time Markov Control Process*, New York, Springer-Verlag, 1999.

[8] Hernández-Lerma O.; Luque-Vásquez F., *Semi-Markov control models with average costs*, Appl. Math. **26** (1999), 315–331.

[9] Jaśkiewicz A., *An approximation approach to ergodic semi-Markov control processes*, Math. Methods Oper. Res. **54** (2001), 1–19.

[10] Jaśkiewicz A., *On the equivalence of two expected average cost criteria for semi-Markov control processes*, Math. Oper. Res. **29** (2004), 326–338.

[11] Jaśkiewicz A., *A fixed point approach to solve the average cost optimality equation for semi-Markov decision processes with Feller transition probabilities*, Comm. Stat. -Theory and Methods **36** (2007), 2559-2575.

[12] Jaśkiewicz A., *Semi-Markov control processes with noncompact action spaces and discontinuous costs*, in preparation (2007).

[13] Jaśkiewicz A.; Nowak A.S., *On the optimality equation for average cost Markov control processes with Feller transition probabilities*, J. Math. Anal. Appl. **316** (2006), 495–509.

[14] Jaśkiewicz A.; Nowak A.S., *Optimality in Feller semi-Markov control processes*, Oper. Res. Lett. **34** (2006), 713–718.

[15] Küenle H.-U., *On Markov games with average reward criterion and weakly continuous transition probabilities*, SIAM J. Control Optim. **45** (2007), 2156–2168.

[16] Meyn S.P.; Tweedie R.L., *Markov Chains and Stochastic Stability*, New York, Springer-Verlag, 1993.

[17] Meyn S.P.; Tweedie R.L., *Computable bounds for geometric convergence rates of Markov chains*, Ann. Appl. Probab. **4** (1994), 981–1011.

[18] Nishimura K.; Stachurski J., *Stochastic optimal policies when the discount rate vanishes,* J. Econ. Dynamics Control **31** (2007), 1416–1430.

[19] Resnick S.I., *Adventures in Stochastic Processes*, Birkhäuser, Boston, 1992.

[20] Ross S.M., *Applied Probability Models with Optimization Applications*, San Francisco, Holden-Day, 1970.

[21] Sennott L.I., *Average cost semi-Markov decision processes and the control of queueing systems*, Prob. Eng. Info. Sci. **3** (1989), 247–272.

[22] Schäl M., *Average optimality in dynamic programming with general state space*, Math. Oper. Res. **18** (1993), 163–172.

[23] Stockey N.L.; Lucas R.E.; Prescott E.C., *Recursive Methods in Economic Dynamics*, Harvard University Press, Massachusetts, 1989.

[24] Strauch R.E., *Negative dynamic programming*, Ann. Stat. **37** (1966), 871–890.

[25] Yushkevich A., *On semi-Markov controlled models with an average reward criterion*, Theory Probab. Appl. **26** (1981), 796–803.

[26] Vega-Amaya O., *The average cost optimality equation: a fixed point approach*, Bol. Soc. Mat. Mexicana **9** (2003), 185–195.