Morfismos, Vol. 3, No. 1, 1999, pp. 69–87

SAMPLE PATH AVERAGE COST OPTIMALITY FOR A CLASS OF PRODUCTION-INVENTORY SYSTEMS *

OSCAR VEGA-AMAYA [1]

Abstract

We show the existence of *sample-path average cost* (SPAC-) optimal policies for a class of production-inventory systems with *uncountable* state space and *strictly unbounded* one-step cost— that is, costs that growth without bound outside of compact subsets. In fact, for a specific case, we show that a K^*-threshold policy is both *expected and sample path average cost optimal*, where the constant K^* can be easily computed solving a static optimization problem, which, in turn, is obtained directly form the basic data of the production-inventory system.

1991 Mathematics Subject Classification: 93E20, 90C40.
Keywords and phrases: Production-inventory systems, Markov control processes, expected and sample path average cost criteria, strictly unbounded costs.

1 Introduction

We consider an inventory system with a single product, infinite storage and production capacities, for which the excess demand is not backlogged. Denote by x_t and a_t the inventory level and the amount of

*This paper is part of the author's Doctoral Thesis written at the Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa (March 1998)
[1]This work was partially supported by CONACYT under Grants 3115P-E9608 and 28309E

product ordered (and immediately supplied) at the beginning of each decision period $t = 0, 1, \cdots$, respectively. The product demand during period t is denoted by w_t, which is assumed to be a nonnegative random variable. The inventory level evolves in $\mathbf{X} = [0, +\infty)$ according to

$$x_{t+1} = \max(x_t + a_t - w_t, 0), \quad t = 1, 2, \cdots; \; x_0 = x \in \mathbf{X}, \tag{1}$$

and the performance of the system under an admissible control policy $\delta = \{a_t\} \subset \mathbf{A}$, given an initial state $x_0 = x \in \mathbf{X}$, is evaluated by means of the *sample-path average cost* (SPAC)

$$J_0(\delta, x) := \limsup_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t), \tag{2}$$

where C is a *lower semicontinuous* and *strictly unbounded* function on $\mathbf{K} := \mathbf{X} \times \mathbf{A}$; that is, it satisfies

$$\lim_{n\to\infty} \inf\{C(x, a) : (x, a) \notin K_n\} = \infty, \tag{3}$$

with $\{K_n\}$ being a sequence of compact subsets which converges increasingly to $\mathbf{K}$.

Thus, the optimal control problem we are concerned with is that of selecting a control policy $\delta^* = \{a_t^*\}$ that "minimizes" (2), that is,

$$J_0(\delta^*, x) \le J_0(\delta, x) \quad \textit{almost surely}, \tag{4}$$

for all admissible control policy $\delta = \{a_t\}$ and initial state $x_0 = x \in \mathbf{X}$. If such a policy $\delta^* = \{a_t^*\}$ exists, it is called *sample-path average cost* (SPAC) *optimal.*

We discuss the existence of SPAC optimal policies for several choices of one-step cost function satisfying (3) and, for a specific case, we prove that a K^*-threshold policy is both *expected and sample path average cost optimal*, where the constant K^* can be easily computed solving a static optimization problem, which, in turn, is obtained directly from the basic data of the production-inventory system [see Example A; Theorem 4.6 and Remark 4.5]. In order to prove these facts, in Section 3, we first discuss several results for general Markov control processes (MCPs) from [19]; then, in Section 4, we specialize these results to the inventory systems (1).

As it is well-known, the Markov Control Processes (MCPs) theory provides a very suitable framework for the control of production-inventory systems (and many other important dynamic optimization problems), but—to the best of our knowledge— the available results for MCPs do not include the production-inventory systems in which we are interested, since, to begin, most of the literature deals with the *expected average cost* (EAC) problem when the state space is *denumerable* or uncountable but assuming that the one-step cost is *bounded,* under strong recurrence/ergodicity assumptions (see [1], [9] and the references therein). On the other hand, the analysis of the *sample-path average cost* (SPAC) is seldom developed, and when it is done, it is also restricted to the denumerable state space or bounded one-step costs cases ([1], [3], [4], [14]).

Perhaps the only exceptions to this situation are the recent works by Hernández-Lerma et. al. [11], Lasserre [12] and Vega-Amaya [19] whose settings and approaches are not comparable at all. For example, in [11] it is used a V-uniform ergodicity assumptions to show the existence of a stationary policy SPAC-optimal (in the class of all policies) with minimum "variance" in the class of SPAC-optimal stationary polices, whereas [12] and [19] consider weaker recurrence conditions and suppose that the one-step cost is strictly unbounded, but equally, their results are weaker. Roughly speaking, in the former paper, it is shown that to find a SPAC-optimal policy is equivalent to solving a certain infinite-dimensional linear program, and in the latter work it is only guaranteed the existence of a relaxed (or randomized) stationary SPAC-optimal policy. However, from our point of view, the context provided in [19] is more suitable to solving the SPAC control problem for the system (1) with one-step costs satisfying (3), since it provides a direct approach and does not require strong ergodicity assumptions.

The remainder of the paper is organized as follow. Section 2 contains a brief description of the Markov control model of interest and assumptions. In Section 3 we introduce the optimality criteria and several results from Vega-Amaya [19] are stated without proof [see Theorems 3.6, 3.7 and 3.8]. Finally, in Section 4, we discuss several examples from inventory theory.

We shall use the following notation. Given a *Borel space* Y (i.e., a Borel subset of some separable complete metric space), $\mathcal{B}(Y)$ denotes its Borel $\sigma-$algebra and "measurable" will mean "Borel-measurable". $\mathbf{P}(Y)$ stands for the class of all probability measures on Y. Moreover, if

Y and Z are Borel spaces, then a *stochastic kernel* on Y given Z is a function $P(\cdot|\cdot)$ such that $P(\cdot|z)$ is a probability measure on Y, for each $z \in Z$, and $P(B|\cdot)$ is measurable function for each $B \in \mathcal{B}(Y)$. The family of all stochastic kernels on Y given Z is denoted by $\mathbf{P}(Y|Z)$. Finally, we denote by $\mathbf{N}$ (resp., $\mathbf{N}_0$) the set of positive integers (resp., nonnegative integers).

2 The Markov model

Since the *Markov control model* $(\mathbf{X}, \mathbf{A}, \{A(x) : x \in A(x)\}, Q, C)$ we are concerned with is quite standard, we only give a brief description. For details see, for instance, [9].

We assume that the state space $\mathbf{X}$ and the control space $\mathbf{A}$ are both Borel spaces. For each $x \in \mathbf{X}$, $A(x)$ is a nonempty Borel subset of $\mathbf{A}$ and, moreover, $\mathbf{K}:= \{(x,a) : a \in A(x), x \in \mathbf{X}\}$ is a Borel subset of the Cartesian product $\mathbf{X} \times \mathbf{A}$. Finally, the transition law Q is a stochastic kernel on $\mathbf{X}$ given $\mathbf{K}$ and the one step cost function C is a measurable function on $\mathbf{K}$.

Define

$$\mathbf{H}_0 := \mathbf{X} \ \text{ and } \quad \mathbf{H}_t := \mathbf{K}^t \times \mathbf{X} \ \text{ for } t \in \mathbf{N}.$$

An (admissible) *control policy* is a sequence $\delta = \{\delta_t\}$ such that, for each $t \in \mathbf{N}_0, \cdots$, $\delta_t \in \mathbf{P}(\mathbf{A}|\mathbf{H}_t)$ and it satisfies the constraint $\delta_t(A(x_t)|h_t) = 1 \ \forall h_t = (x_0, a_0, \cdots, x_{t-1}, a_t, x_t) \in \mathbf{H}_t$. A control policy $\delta = \{\delta_t\}$ is said to be: (i) *relaxed* (or randomized stationary) policy if there exists $\varphi \in \mathbf{P}(\mathbf{A}|\mathbf{X})$ such that, for each t, $\delta_t(\cdot|h_t) = \varphi(\cdot|x_t) \ \forall h_t \in \mathbf{H}_t$; (ii) (deterministic) *stationary* policy if there exists a measurable function $f : \mathbf{X} \mapsto \mathbf{A}$ such that, $f(x) \in A(x) \ \forall x \in \mathbf{X}$, and $\delta_t(\cdot|h_t)$ is concentrated at $f(x_t) \ \forall h_t \in \mathbf{H}_t$ and $t \in \mathbf{N}_0$.

The class of all control policies is denoted by Δ, while Φ and $\mathbf{F}$ stand for the subclasses formed by the relaxed and stationary policies, respectively.

For each policy $\delta \in \Delta$ and *initial distribution* $\nu \in \mathbf{P}(\mathbf{X})$, there exist a stochastic process $\{(x_t, a_t) : t = 0, 1, \cdots\}$ and a probability measure P_ν^δ —which governs the evolution of the process— both defined on the sample space $(\Omega, \mathcal{F})$, where $\Omega := (\mathbf{X} \times \mathbf{A})^\infty$ and $\mathcal{F}$ is the corresponding product σ-algebra. The expectation operator with respect to P_ν^δ is denoted by E_ν^δ. We will refer to x_t and a_t as the state and control at time t, respectively. If the initial probability measure ν is concentrated at an initial state $x_0 = x \in \mathbf{X}$, we write P_x^δ and E_x^δ instead of P_ν^δ and E_ν^δ, respectively.

When using a relaxed policy $\varphi \in \Phi$, the state process $\{x_t\}$ is a Markov chain on $\mathbf{X}$ with time-homogeneous transition kernel

$$Q(\cdot|x,\varphi) := \int_{\mathbf{X}} Q(\cdot|x,a)\varphi(d\,a|x), \quad x \in \mathbf{X}. \tag{5}$$

We also write

$$C(x,\varphi) := \int_{\mathbf{X}} C(x,a)\varphi(d\,a|x). \tag{6}$$

For a deterministic stationary policy $f \in \mathbf{F}$, (5)-(6) become

$$Q(\cdot|x,f) := Q(\cdot|x,f(x)) \text{ and } C(x,f) := C(x,f(x)). \tag{7}$$

We also suppose that the Markov control model satisfies the following properties:

Assumption 2.1.(a) *C is nonnegative and lower semicontinuous on $\mathbf{K}$;*
(b) *C is strictly unbounded on $\mathbf{K}$, i.e., there exists an increasing sequence of compact sets $\mathbf{K}_n \uparrow \mathbf{K}$ such that*

$$\lim_{n\to\infty} \inf\{C(x,a) : (x,a) \notin \mathbf{K}_n\} = \infty;$$

(c) *$Q(\cdot|x,a)$ is weakly continuous in $(x,a) \in \mathbf{K}$, that is, $\int_{\mathbf{X}} u(y)Q(dy|x,a)$ is continuous in $(x,a) \in \mathbf{K}$ for every bounded continuous function u on $\mathbf{X}$.*

The property in Assumption 2.1(b) is also referred to saying that C is a *moment* or that C is a *norm-like function on* $\mathbf{K}$. Its main appealing stems from that it provides an easy way to prove tightness of measures, which has been exploited in several contexts . See, for instance, [5], [7], [8], [10], [15], [16], and references therein.

3 Sample path and expected average cost

Our main interest is to evaluate the stochastic control system when a policy $\delta \in \Delta$ is used, given an initial distribution $\nu \in \mathbf{P}(\mathbf{X})$, by means of the *sample path average cost* (SPAC) defined as

$$J_0(\delta,\nu) := \limsup_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t,a_t), \tag{8}$$

but, we also consider the *expected average cost* (EAC) given by

$$J(\delta,\nu) := \limsup_{n\to\infty} \frac{1}{n} E_\nu^\delta \sum_{t=0}^{n-1} C(x_t,a_t). \tag{9}$$

Moreover, we define the *optimal* (minimum) *expected average cost* as

$$j^* := \inf_\nu \ \inf_\delta J(\delta,\nu) \tag{10}$$

To avoid a trivial problem we shall use the following assumption.

Assumption 3.1. *There exists a policy δ_* and initial distribution ν_* such that $J(\delta_*,\nu_*)$ is finite.*

The optimality criteria we are concerned with are the following.

Definition 3.2. *Let δ^* be a policy and ν^* an initial distribution.*

(a) *δ^* is said to be expected average cost (EAC-)optimal if*

$$J(\delta,x) \geq J(\delta^*,x) \quad \forall x \in \mathbf{X}, \delta \in \Delta;$$

(b) *δ^* is said to be strong expected average cost (strong EAC-) optimal if*

$$\liminf_{n\to\infty} \frac{1}{n} E_x^\delta \sum_{t=0}^{n-1} C(x_t,a_t) \geq J(\delta^*,x) \quad \forall x \in \mathbf{X}, \delta \in \Delta;$$

(c) *(δ^*,ν^*) is said to be a minimum pair if $J(\delta^*,\nu^*) = j^*$;*

(d) *δ^* is said to be sample path average cost* (SPAC-) *optimal if for every $\delta \in \Delta$ and $\nu \in \mathbf{P}(\mathbf{X})$:*

$$J_0(\delta,\nu) \geq j^* \quad P_\nu^\delta - \textit{almost surely} \tag{11}$$

and, moreover,

$$J_0(\delta^*,\nu) = j^* \quad P_\nu^{\delta^*} - \textit{almost surely}, \tag{12}$$

We now introduce some notation and several classes of polices

Definition 3.3. *A relaxed policy φ, or theMarkov chain induced by it, is said to be irreducible if there exists a non-trivial σ-finite measure λ on* **X**, *such that $\lambda(B) > 0$ implies that*

$$E_x^\varphi \sum_{t=1}^{\infty} \mathbf{I}_B(x_t) > 0$$

where $\mathbf{I}_B(\cdot)$ *denotes the indicator function of the subset* B.

If the policy φ is irreducible, from [17, Proposition 4.2.2], there exists a *maximal irreducibility measure* ψ, which means that any other iredubility measure ψ' is absolutely continuous with respect to ψ, i.e., $\psi' \prec \psi$. Thus, the class of subsets of "positive measure"

$$\mathcal{B}_+^\varphi(\mathbf{X}) := \{B \in \mathcal{B}(\mathbf{X}) : \psi(B) > 0\}$$

is uniquely defined.

Definition 3.4. *A relaxed policy* $\varphi \in \Phi$ *is said to be:*

(a) *stable if there exists an invariant probability measure* $\mu_\varphi \in \mathbf{P}(\mathbf{X})$ *for the transition law* $Q(\cdot|x, \varphi)$, *i.e.,*

$$\mu_\varphi(\cdot) = \int_{\mathbf{X}} Q(\cdot|y, \varphi)\mu_\varphi(dy),$$

which satisfies

$$J(\varphi, \mu_\varphi) = \int_{\mathbf{X}} C(y, \varphi)\mu_\varphi(dy) < \infty;$$

(b) *recurrent if it is irreducible and*

$$E_x^\varphi \sum_{t=1}^{\infty} \mathbf{I}_B(x_t) = \infty \quad \forall x \in \mathbf{X}, B \in \mathcal{B}_+^\varphi(\mathbf{X});$$

(c) *Harris recurrent if it is irreducible and*

$$P_x^\varphi \left[\sum_{t=1}^{\infty} \mathbf{I}_B(x_t) = \infty \right] = 1 \quad \forall x \in \mathbf{X}, B \in \mathcal{B}_+^\varphi(\mathbf{X}).$$

We denote by Φ_S the class of relaxed stable policies and by Φ_R (Φ_{HR}, resp.) the class of relaxed policies which are recurrent (Harris recurrent, resp.).

Remark 3.5. Let φ be a policy with invariant probabiltiy measure μ_φ. If it is also irreducible, then:

(a) φ is recurrent [17, Proposition 10.1.1.];

(b) Then, from [17, Propositions 9.1.5 and 4.2.3.], there exist subsets H_φ and N_φ such that:

(i) $\mathbf{X} = H_\varphi \cup N_\varphi$;

(ii) H_φ is absorbing and full, i.e., $Q(H_\varphi|x,\varphi) = 1$ $\forall x \in H_\varphi$ and $\mu_\varphi(H_\varphi) = 1$;

(iii) the policy φ restricted to H_φ is Harris recurrent, i.e.,

$$P_x^\varphi\left[\sum_{t=1}^{\infty} \mathbf{I}_B(x_t) = \infty\right] = 1 \quad \forall x \in H_\varphi,$$

for each $B \in \mathcal{B}_+(\mathbf{X})$ contained in H_φ;

(c) Thus, φ is Harris recurrent if and only if $N_\varphi = \emptyset$.

We suppose throughout the following that Assumptions 2.1 and 3.1 hold.

We now state one of the main results. The proof of this and the other results in this section are given in [19].

Theorem 3.6. *For each policy $\delta \in \Delta$ and measure $\nu \in \mathbf{P}(\mathbf{X})$,*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t) \geq j^* \quad P_\nu^\delta - \textit{almost surely.} \tag{13}$$

The next theorem contains some interesting relations among the concepts of minimum pair, sample path and expected average costs, which are direct consequences of Theorem 3.6. The result in Theorem 3.7(c) was previously proved in [7].

Theorem 3.7.(a) *A policy $\delta^* \in \Delta$ is EAC-optimal if and only if it is strong EAC-optimal;*

(b) *If (δ, ν) is a minimum pair, with $\delta \in \Delta$ and $\nu \in \mathbf{P}(\mathbf{X})$, then*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} C(x_t, a_t) = j^* \quad P_\nu^\delta - \textit{almost surely.}$$

(c) *Let $\varphi \in \Phi_S$ and μ_φ an associated invariant probability measure. Then, (φ, μ_φ) is a minimum pair if and only if $J(\varphi, x) = j^*$ for (μ_φ-) almost all $x \in \mathbf{X}$.*

The first part of the next theorem states the existence of a minimum pair (φ^*, μ^*) with φ^* being a stable policy and μ^* an associated invariant probability measure. This results was already proved in [7], but the approach used in that paper differs from the used in the present one in that his analysis relies on the well behavior of the expected average cost whereas our analysis is on the discounted cost. Roughly speaking, our proof of the existence of a minimum pair yields, at the same time, that the optimal average cost may be approximated by discounted programs, which exhibits other nice property of the control problem with strictly unbounded cost. The second part of the theorem states that if the policy φ^* is positive Harris recurrent then it is SPAC-optimal. To state precisely these facts, we introduce the following notation.

For each $\alpha \in (0,1)$, the (expected) *α-discounted cost* when it is used a policy $\delta \in \Delta$, given the initial distribution measure $\nu \in \mathbf{P}(\mathbf{X})$, is defined by

$$V_\alpha(\delta,\nu) := E_\nu^\delta \sum_{t=0}^{\infty} \alpha^t C(x_t,a_t), \tag{14}$$

and the *α-discounted optimal value* is given by

$$m_\alpha := \inf_\nu \inf_\delta V_\alpha(\delta,\nu). \tag{15}$$

Theorem 3.8.(a) *There exists a stable policy $\varphi^* \in \Phi_S$ [with invariant probability measure μ^*] such that (φ^*,μ^*) is a minimum pair. Hence, from Theorem 3.7(c) and Remark 3.5(b), for μ^*-*almost all $x \in \mathbf{X}$:

$$J(\varphi^*,x) = j^*, \tag{16}$$

and

$$J_0(\varphi^*,x) = \lim_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n-1} C(x_t,a_t) = j^* \quad P_x^{\varphi^*} - \textit{almost surely}.$$

Moreover,

$$j^* = \lim_{\alpha\to 1^-}(1-\alpha)m_\alpha; \tag{17}$$

(b) *If the policy φ^* is positive Harris recurrent, then it is SPAC-optimal.*

4 Examples

We now discuss a number of examples from inventory theory to illustrate the potential of the approach used in the previous section; indeed, in Example A we show that a K^*-threshold policy is both strong EAC and SPAC optimal, where the nonnegative constant K^* solves a static optimization problem coming from the basic data of the inventory system. In [7], [10] and [16] are given other interesting examples, including the LQ control problem, which satisfy the assumptions in Theorems 3.4, 3.5 and 3.6.

Recall that the stock level $\{x_t\}$ evolves in $\mathbf{X} := [0, \infty)$ according to

$$x_{t+1} = \min(x_t + a_t - w_t, 0), \quad t = 1, 2, \cdots; \ x_0 = x \in \mathbf{X}, \qquad (18)$$

where $\{a_t\}$ and $\{w_t\}$ denote the control and the demand processes, respectively.

EXAMPLE A. We first consider the case in which the one step cost function has the following structured cost form

$$C(x, a) = F_1(x + a) + ba \quad (x, a) \in \mathbf{K}, \qquad (19)$$

where $b \geq 0$ is a constant and $F_1(\cdot)$ is a function on $[0, \infty)$ satisfying the following properties.

Assumption 4.1.(a) *$F_1(\cdot)$ is a continuous convex function bounded from below;*
(b) $F_1(y) \to \infty$ as $y \to \infty$.

We also suppose that the next conditions hold.

Assumption 4.2(a) $\mathbf{A} = A(x) = [0, \infty)\ \forall x \in \mathbf{X}$;
(b) *The process $\{w_t\}$ is formed by i.i.d. nonnegative random variables. The common cumulative distribution function is denoted by $G(\cdot)$;*
(c) $G(y) < 1\ \forall y \geq 0$.

Throughtout this section we denote by E the expectation with respect to the joint distribution of the random variables $w_0, w_1, \cdots$.

Theorem 4.3. *Suppose that Assumptions 4.1 and 4.2 hold. Then, there exists an stable policy φ^* with invariant probability measure μ_{φ^*} such that $(\varphi^*, \mu_{\varphi^*})$ is a minimum pair.*

Proof of Theorem 4.3. We shall verify that Assumptions 2.1 and 3.1 hold; hence, the existence of the minimum pair $(\varphi^*, \mu_{\varphi^*})$ is ensured by Theorem 3.8(a). To do this, first note that Assumption 4.1 implies Assumption 2.1 (a) and (b), whereas Assumption 4.2(b) implies Assumption 2.1(c), that is, that the transition law of the inventory system (18)

$$Q(B|x, a) = E\mathbf{I}_B[\min(x + a - w_0, 0)] \quad B \in \mathcal{B}(\mathbf{X}),$$

is weakly continuos in $(x, a) \in \mathbf{K}$. Finally, to verify that Assumption 3.1 holds, observe that for the policy $f(x) = 0\ \forall x \in \mathbf{X}$, the stock level evolves according to

$$x_{t+1} = \max(x_t - w_t, 0), \quad t = 1, 2, \cdots;\ x_0 = x \in \mathbf{X},$$

and also that they form a decreasing sequence bounded above by $x_0 = x$. Moreover, since

$$C(x_t, f) = F_1(x_t), \quad t = 0, 1, 2, \cdots,$$

and $F_1(\cdot)$ is bounded on $[0, x]$ by a constant, say M_x, we have that

$$J(f, x) < \infty \quad \forall x \in \mathbf{X}. \Box$$

We shall prove the existence of a threshold policy which is both strong expected and sample path average cost optimal. We now introduce this class of polices: for a nonnegative constant K, a stationary policy f_K is said to be K-*threshold policy* if $f_K(x) = K - x$ for $x \in [0, x]$, and $f_K(x) = 0$ otherwise. We show in the next lemma that any threshold policy is stable and Harris recurrent.

Lemma 4.4. *Suppose that Assumptions 4.1 and 4.2 hold. Then, for each $K \geq 0$, the policy f_K is stable Harris recurrent and*

$$J(f_K, x) = F_1(K) + bE[\min(K, w_0)] \quad \forall x \in \mathbf{X}. \tag{20}$$

Proof of Lemma 4.4. Let K be a nonnegative fixed constant. Note that, from Assumption 4.2(c), f_K is irreducible respect to the measure

$$\lambda(B) := \mathbf{I}_B(0) \quad B \in \mathcal{B}(\mathbf{X}).$$

Moreover, direct computations yield that

$$\mu_K(B) := \int_B \min(K - w, 0) G(dw) \quad B \in \mathcal{B}(\mathbf{X}),$$

is the invariant probability measure for f_K and also that

$$\mu_K(B) = Q(B|x, f_K) \quad \forall B \in \mathcal{B}(\mathbf{X}), x \in [0, K].$$

Thus, $[0, K]$ is a *petite* subset of $\mathbf{X}$, since

$$Q(B|x, f_K) \geq \mathbf{I}_{[0,K]}(x)\, \mu_K(B) \quad \forall B \in \mathcal{B}(\mathbf{X}), x \in \mathbf{X}.$$

Then, from [17, Theorem 10.4.10(ii), p. 246], to prove that f_K is Harris recurrent, it suffices to show that

$$E_x^{f_K}\tau < \infty \quad \forall x \in \mathbf{X}, \tag{21}$$

where

$$\tau := \min\{n \geq 1 : x_n \leq K\}. \tag{22}$$

It is obvious that (21) holds for $x_0 = x \in [0, K]$. Now consider the case $x_0 = x > K$ and observe that

$$\begin{aligned} P_x^{f_K}\tau &= \textstyle\sum_{n=0}^{\infty} P_x^{f_K}[\tau > n] \\ &= 1 + \textstyle\sum_{n=1}^{\infty} G^{(n)}(K - x) \\ &\leq 1 + \textstyle\sum_{n=1}^{\infty}[G(K - x)]^n = [1 - G(K - x))]^{-1} < \infty \end{aligned}$$

where $G^{(n)}(\cdot)$ denotes the *n-fold convolution* of $G(\cdot)$ and the last equality follows from Assumption 4.2(c). Hence, (21) holds.

Now, It is easy to check that (20) holds for $x_0 = x \in [0, K]$. Thus, consider $x_0 = x > K$ and define $\tau(n) := \min(\tau, n), n \in \mathbf{N}$. Then, from (21) and the strong Markov propety, we see that

$$\begin{aligned} E_x^{f_K}\textstyle\sum_{t=0}^{n} C(x_t, f_K) &= E_x^{f_K}\textstyle\sum_{t=0}^{\tau(n)-1} C(x_t, f_K) \\ &\qquad + E_x^{f_K}\textstyle\sum_{t=\tau(n)}^{n} C(x_t, f_K) \\ &= E_x^{f_K}\textstyle\sum_{t=0}^{\tau(n)-1} F_1(x_t) \\ &\qquad + E_x^{f_K}\textstyle\sum_{t=\tau(n)}^{n}[F_1(K) + b(K - x_t)] \\ &= E_x^{f_K}\textstyle\sum_{t=0}^{\tau(n)-1} F_1(x_t) \\ &\qquad + [F_1(K) + bE\min(K, w_0)]E_x^{f_K}[n + 1 - \tau(n)]. \end{aligned}$$

Thus, using the fact that $F_1(\cdot)$ is bounded on $[0, x]$ and (21), we see that (20) holds.

Finally, note that

$$\int_{\mathbf{X}} C(y, f_K)\mu_K(dy) = F_1(K) + bE\min(K, w_0) < \infty;$$

that is, f_K is a stable Harris recurrent policy.□

Remark 4.5(a) Define

$$L(y) := F_1(y) + bE\min(y, w_0) \quad y \geq 0, \quad \rho^* := \inf_{y\geq 0} L(y), \tag{23}$$

and observe that

$$J(f_K, x) = L(K) \quad \forall x \in \mathbf{X}. \tag{24}$$

(b) Moreover, there exists a constant $K^* \geq 0$ such that

$$L(K^*) = \rho^* = \inf_{y\geq 0} L(y). \tag{25}$$

Indeed, this follows from the continuity of $L(\cdot)$ and the fact that $L(y) \to \infty$ as $y \to \infty$. Hence, from (24), $J(f_{K^*}, \cdot) = \rho^*$.

Theorem 4.6 *Suppose that Assumptions 4.1 and 4.2 hold. Then, the K^*-threshold policy is strong expected and sample path average cost optimal, where K^* is as in (25).*

Proof of Theorem 4.6. For the proof we require some results on discounted-cost control problems. For each $\alpha \in (0,1)$, recall from (14) that

$$V_\alpha(\delta, x) = E_x^\delta \sum_{t=0}^{\infty} \alpha^t C(x_t, a_t), \quad x \in \mathbf{X} \text{ and } \delta \in \Delta,$$

and define

$$V_\alpha(x) := \inf_{\delta\in\Delta} V_\alpha(\delta, x), \quad x \in \mathbf{X}. \tag{26}$$

Now, from Theorem 4.3, there exists a stable policy φ^* with invariant probability measure μ^* such that

$$J(\varphi^*, x) = j^* \quad \mu^*\text{-almost all } x \in \mathbf{X};$$

thus, from a well-known Abelian Theorem (see [9], Lemma 5.3.1, p. 84),

$$j^* = \lim_{\alpha\to 1^-} (1-\alpha)V_\alpha(\varphi, x) \geq \limsup_{\alpha\to 1^-}(1-\alpha)V_\alpha(x) \quad \mu^*\text{-almost all } x \in \mathbf{X}.$$

Then, since $V_\alpha(\cdot) \geq m_\alpha$ $\forall\alpha \in (0,1)$, we see from this and (17) that

$$j^* = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(x) \quad \mu^*\text{-almost all } x \in \mathbf{X}. \tag{27}$$

Then, since $\mu^*(\{0\}) > 0$, to conclude that the K^*-threshold policy is strong EAC and SPAC optimal, it suffices to prove that

$$\rho^* = \lim_{\alpha \to 1^-} (1-\alpha) V_\alpha(0). \tag{28}$$

In order to do this, first note that

$$V_\alpha(x) \leq V_\alpha(f_K, x) < \infty \quad 0 \leq x \leq K,$$

where f_K is the K-threshold policy; then, taking K large enough we see that $V_\alpha(\cdot) < \infty$ $\forall \alpha \in (0,1)$. Now, using Assumption 4.1, it is easy to prove that $V_\alpha(\cdot)$ is a convex function; thus, the function

$$T_\alpha(y) := F_1(y) + by + \alpha E V_\alpha[(y - w_0)^+], \quad y \geq 0,$$

is convex and $\lim_{y \to +\infty} T(y) = +\infty$, which imply that there exists a constant $K_\alpha \geq 0$ such that $T_\alpha(K_\alpha) = \inf_{y \geq 0} T_\alpha(y)$. Hence, for each $\alpha \in (0,1)$, $V_\alpha(\cdot)$ satisfies the *α-Discounted Cost Optimality Equation* ([6])

$$V_\alpha(x) = \min_{a \in \mathbf{A}} \left[F_1(x+a) + ba + \alpha E V_\alpha[(x + a - w_0)^+] \right] \quad \forall x \in \mathbf{X}, \tag{29}$$

and the K_α-threshold policy attains the minimum at the right-hand side of (29), that is, for all $x \in \mathbf{X}$

$$V_\alpha(x) = F_1(x + f_\alpha(x)) + b f_\alpha(x) + \alpha E V_\alpha[(x + f_\alpha(x) - w_0)^+], \tag{30}$$

where, for each $\alpha \in (0,1)$, f_α denotes the $K_\alpha-$threshold policy.

Then, standard arguments yield

$$V_\alpha(x) = V_\alpha(f_\alpha, x) \;\; \forall x \in \mathbf{X}, \alpha \in (0,1). \tag{31}$$

Moreover, simple computations show that $\forall \alpha \in (0,1)$

$$(1-\alpha) V_\alpha(f_\alpha, 0) = F_1(K_\alpha) + \alpha E \min(K_\alpha, w_0) + b(1-\alpha) K_\alpha. \tag{32}$$

Now define

$$L_\alpha(y) := F_1(y) + \alpha E\min(y, w_0) + b(1-\alpha)y, \quad y \geq 0 \text{ and } \alpha \in (0,1),$$

and note that, from (31)-(32), $L_\alpha(K_\alpha) = \inf_{y\geq 0} L_\alpha(y)$ for each $\alpha \in (0,1)$, and also that $L_\alpha(\cdot) \downarrow L(\cdot)$ as $\alpha \uparrow 1$, where $L(\cdot)$ is the function in (23). From these facts, we see that

$$L_\alpha(K^*) \geq L_\alpha(K_\alpha) \geq L(K_\alpha) \geq L(K^*) \quad \forall \alpha \in (0,1),$$

where K^* is as in (25). Thus, we also obtain

$$\rho^* = L(K^*) = \lim_{\alpha\to 1^-} L_\alpha(K_\alpha) = \lim_{\alpha\to 1^-} (1-\alpha)V_\alpha(0),$$

Therefore, the K^*-threshold policy is strong EAC and SPAC optimal. In fact,

$$j^* = \rho^* = L(K^*) = J(f_{K^*}, x) \quad \forall x \in \mathbf{X}. \ \Box$$

Example B. We now consider a one step cost function more general than (19). Specifically, we assume that

$$C(x,a) = F_1(x+a) + F_2(a), \tag{33}$$

where $F_1(\cdot)$ and $F_2(\cdot)$ are functions from $[0,+\infty)$ into itself satisfying the following:

Assumption 4.7.(a) *$F_1(\cdot)$ and $F_2(\cdot)$ are lower semicontinuous functions bounded from below;*
(b) $\lim_{y\to\infty} F_1(y) = \infty$;
(c) $EF_2(\min(y, w_0)) < \infty \ \forall y \geq 0$.

Note that Assumption 4.3 is general enough to include problems with a set-up cost, that is, a fixed cost for placing orders ([2], [13]).

In order to guarante the existence of SPAC optimal policy for the one step cost function (33) we require suitable strengthening of Assumption 4.2.

Assumption 4.8.(a) *The process $\{w_t\}$ is formed by nonnegative i.i.d random variables;*
(b) $w^* := \int wG(dw) < \infty$;
(c) $\mathbf{A} = A(x) = [0,\theta] \ \forall x \in \mathbf{X}$, where $\theta < w^*$.
(d) $G(y) < 1 \ \forall y \geq 0$.

The main consequence of Assuption 4.8 is given in the next lemma.

Lemma 4.9. *Under Assuption 4.8, any stable policy $\varphi \in \Phi_0$ is Harris recurrent.*

Proof of Lemma 4.9. Let $\varphi \in \Phi_S$ be an arbitrary but fixed stable policy. We shall prove that the Markov chain $\{x_t\}$, induced by φ, is an irreducible *T-chain* and *non-evanescent* (see definitons in [17, pp. 127 and 207]). Thus, from [17, Theorem 9.22, p. 208], the policy φ is Harris recurrent.

Now, that $\{x_t\}$ is an irreducible T-chain follows from the inequality

$$Q(B|x,\varphi) \geq \mathbf{I}_B(0)[1 - G(x+\theta)] \quad B \in \mathcal{B}(\mathbf{X}), x \in \mathbf{X},$$

which can be verified by direct computations.

On the other hand, one can to verify that the function $V(x) := x,\ x \in \mathbf{X}$, satisfies the "drift condition"

$$\int_{\mathbf{X}} V(y)Q(dy|x,\varphi) - V(x) \leq \int_0^{x+\theta} (\theta - w)G(dw) \quad \forall x \in \mathbf{X}.$$

Next, from Assumption 4.8(c), there exists a nonnegative real number $y_0 > \theta$ such that

$$\int_0^{x+\theta} (\theta - w)G(dw) < 0 \quad \forall x > y_0 - \theta.$$

Then, since $V(\cdot)$ is strictly unbounded on $\mathbf{X}$, we have that $\{x_t\}$ is Harris recurrent (see [17, Theorem 9.4.1, p. 208]).

Theorem 4.10. *If Assumptions 4.1 and 4.3 hold, then there exist a relaxed policy $\varphi^* \in \Phi_{HR}$ which is SPAC-optimal and, moreover, $J(\varphi^*, x) = j^*$ μ^*-almost all $x \in \mathbf{X}$.*

Proof of Theorem 4.10. It is easy to check that Assuptions 2.1 and 3.1 hold; then, from Theorem 3.8(a), there exists policy $\varphi^* \in \Phi_S$ with invariant probability measure μ^* which form a minimum pair. Hence, from Lemma 4.9 and Theorem 3.8(b), φ^* is SPAC-optimal.□

Example C. An alternative to measure the inventory system performance is to consider quadratic holding and production costs, that is,

$$C(x,a) = R(x - \overline{x})^2 + S(a - \overline{a})^2, \quad (x,a) \in \mathbf{K}, \tag{34}$$

where R and S are positive constants, and $\overline{x} \in \mathbf{X}$ and $\overline{a} \in \mathbf{A}$ denote the target inventory and production levels, respectively. In addition to Assumptions 4.8, we also suppose the following:

Assumption 4.11. *The second moment of the demand variables is finite, that is,* $\int_0^{+\infty} y^2 G(dy) < \infty$.

For the cost function (34), the Assumption 2.1(a)-(b) trivially holds, while Assumption 4.10 ensures that j^* is finite. Indeed, consider the stationary policy $f(x) = 0$, $\forall x \in \mathbf{X}$, and compute its average cost to obtain

$$J(f, x) = \overline{x}^2 + \overline{a}^2, \quad \forall x \in \mathbf{X}.$$

This facts yield the following:

Theorem 4.12. *Suppose that Assumptions 4.8 and 4.11 hold. Then, there exists a relaxed policy* $\varphi^* \in \Phi_{HR}$ *which is SPAC-optimal.*

Example D. The paper [18] studies a finite horizon control problem for an inventory system considering a variant of (34), in which there is a "cost free interval" containing the target stock level. More precisely, they take as the holding cost the function

$$\overline{C}(y) := \begin{cases} R_1(y-\alpha)^2 & \text{if} \quad 0 \le y < \alpha \\ 0 & \text{if} \quad \alpha \le y \le \beta \\ R_2(y-\beta)^2 & \text{if} \quad y > \beta \end{cases}$$

with $0 < \alpha < \beta$ and R_1, R_2 are positive constants, and the one-step cost function is given as

$$C(x, a) = E\overline{C}(x + a - w_0) + S(a - \overline{a})^2, \quad (x, a) \in \mathbf{K}. \tag{35}$$

As in Example C, it is easy to establish the following results.

Theorem 4.13. *Suppose that Assumptions 4.8 and 4.11 hold. Then, there exists* $\varphi^* \in \Phi_{HR}$ *which is SPAC-optimal.*

Acknowledgement

The author wishes to thank Professor Onésimo Hernández-Lerma for many useful comments on this paper. Acknowledgments are also due to the referees for their criticisms, which improved in several ways this work.

Oscar Vega-Amaya
Departmento de Matemáticas,
Universidad de Sonora,
Blvd. Luis Encinas y Rosales,
Hermosillo, Sonora, México,
ovega@gauss.mat.uson.mx

References

[1] Arapostathis, A. et. al.(1993), *Discrete time controlled Markov processes with an average cost criterion: A survey*, SIAM Journal on Control and Optimization, 31, 282-344.

[2] Bertsekas, D.P. (1987), *Dynamic Programming : Deterministic and Stochastic Models*, Prentice Hall, Englewood Cliffs, New Jersey.

[3] Borkar, V.S. (1991), *Topics in Controlled Markov Chains*, Pitman Research Notes Series 240, Longman Scientific & Technical, United Kingdom,

[4] Cavazos-Cadena, R. and Fernández-Gaucherand, E. (1995), *Denumerable controlled Markov chains with average reward criterion: sample path optimality*, ZOR-Mathematical Methods of Operations Research 41, 89-108.

[5] González-Hernández, J. and Hernández-Lerma, O. (1996), *Envelopes of sets of measures, tightness, and Markov Control processes,* to appear in Appl. Math. Optim.

[6] Hernández-Lerma, O. and Muñoz-de-Osak, M.(1992), *Discrete-time Markov control processes with discounted unbounded cost: optimality criteria*, Kybernetika 28, 191-212.

[7] Hernández-Lerma, O. (1993), *Existence of average optimal policies in Markov control processes with strictly unbounded costs,* Kybernetika 29**,** 1-17.

[8] Hernández-Lerma, O. and Lasserre, J.B.(1995), *Invariant probabilities for Feller-Markov chains*, Journal of Applied Mathematics and Stochastic Analysis 8, 341-345.

[9] Hernández-Lerma, O. and Lasserre, J.B.(1996), *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer-Verlag, New York.

[10] Hernández-Lerma, O. and Lasserre, J.B.(1997), *Policy iteration for average cost Markov control processes on Borel spaces,* Acta Applicandae Mathematicae 47, 125-154.

[11] Hernández-Lerma, O. Vega-Amaya, O. and Carrasco, G. (1998), *Sample-path optimality and variance-minimization of average cost Markov control processes*, Reporte Interno #236, Departamento de Matemáticas, CINVESTAV-IPN, México City, MÉXICO.(To appear in SIAM J. Control Optim.)

[12] Lasserre, J.B. (1997), *Sample-path average optimality for Markov control processes*, Report No. 97102, LASS-CNRS, Toulouse, France.(To appear IEEE Trans. Autom. Control)

[13] Lee, H.L. and Nahmias, S. (1993), *Single-product, single-location models*, in Logistic of Production and Inventory (Eds. S. C. Graves, A.H.G. Rinnooy Kan and P.H. Zipkin), pp. 3-51, Handbooks in Operations Research and Management Science, Vol. 4, North Holland.

[14] Mandl, P. and Lausmanová, M. (1991), *Two extensions of asymptotic methods in controlled Markov chains*, Annals of Operations Research 28, 67-80.

[15] Meyn, S.P. (1989), *Ergodic theorems for discrete time stochastic systems using a stochastic Lyapunov function*, SIAM Journal on Control and Optimization 27, 1409-1439.

[16] Meyn, S.P. (1997), *The policy iteration algorithm for average reward Markov decision processes with general state space*, IEEE Trans. Autom. Control 42, 1663-1679.

[17] Meyn, S.P. and Tweedie, R.L. (1993), *Markov Chains and Stochastic Stability*, Springer-Verlag, London.

[18] Parlar, M. and Rempala, R. (1992), *Stochastic inventory problem with piecewise quadratic holding cost function containing a cost-free interval*, Journal of Optimization Theory and Applications, 75, 133-153.

[19] Vega-Amaya, O. (1998), *Markov Control Processes in Borel Spaces: Undiscouted Cost Criteria*, Doctoral Thesis, Division de Ciencias e Ingeniería, UAM-Iztapalapa, México.(In Spanish).