11

Estimation and Control of Stochastic Systems under Discounted Criterion

Hilgert Nadine[1] and Minjárez-Sosa J. Adolfo[2]
[1]*UMR 729 ASB, INRA SUPAGRO, Montpellier,*
[2]*Departamento de Matemáticas, Universidad de Sonora, Hermosillo*
[1]*France,* [2]*Mexico*

1. Introduction

We consider a class of discrete-time Markov control processes evolving according to the equation

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, \ldots, \tag{1}$$

where x_t, a_t and ξ_t are the state, action and random disturbance at time t respectively, taking values on Borel spaces. F is a known continuous function. Moreover, $\{\xi_t\}$ is an observable sequence of independent and identically distributed (i.i.d.) random vectors with distribution θ_ξ. This class of control systems has been widely studied assuming that all the components of the corresponding control model are known by the controller. In this context, the evolution of the system is as follows. At each stage t, on the knowledge of the state $x_t = x$ as well as the history of the system, the controller has to select a control or action $a_t = a$. Then a cost c, depending on x and a, is incurred, and the system moves to a new state $x_{t+1} = x'$ according to the transition probability determined by the equation (1). Once the transition to state x' occurs, the process is repeated. Moreover, the costs are accumulated throughout the evolution of the system in an infinite horizon using a discounted criterion. The actions applied at any given time are selected according to rules known as control policies, and therefore the standard optimal control problem is to determine a control policy that minimizes a discounted cost criterion.

However, assuming the knowledge of all components of the control model might be non realistic from the point of view of the applications. In this sense we consider control models that may depend on an unknown component.

Two cases are discussed in the present chapter. In the first one we assume that the disturbance distribution θ_ξ is unknown, whereas in the second one we consider a cost function depending on an exogenous random variable η_t at time t, whose distribution θ_η is unknown. First situation is well documented in the literature and will be briefly described, while the second is less known (even if it is of great interest for application problems) and will be largely developed.

Thus, in contrast with the evolution of a standard system as described above, in both cases, before choosing the control a_t, the controller has to implement a statistical estimation procedure of θ_ξ (or θ_η) to get an estimate $\theta_{\xi,t}$ (or $\theta_{\eta,t}$), and combines this with the history of

the system to select a control $a = a_t^{\theta_\xi, t}$ (or $a_t^{\theta_\eta, t}$). The resulting policy in this estimation and control process is called *adaptive*. Therefore, the optimal control problem we are dealing with in this chapter is to construct adaptive policies that minimize a discounted cost criterion. Furthermore, we study the optimality of such policies in an asymptotic sense.

The chapter is organized as follows. The models and definitions are introduced in the Section 2, as well as an overview of adaptive Markov control processes under discounted criteria. In particular, the required sets of assumptions are introduced and commented. The Section 3 is dedicated to the adaptive control of stochastic systems in the case where the cost function depends on an exogenous random variable with unknown distribution. Here we present two approaches to construct optimal adaptive policies. Finally we conclude in Section 4 with some remarks.

Remark 1.1 *Given a Borel space X (that is, a Borel subset of a complete and separable metric space) its Borel sigma-algebra is denoted by $\mathcal{B}(X)$, and "measurable", for either sets or functions, means "Borel measurable". The space of probability measures on X is denoted by $\mathbb{P}(X)$. Let X and Y be Borel spaces. Then a stochastic kernel $Q(dx \mid y)$ on X given Y is a function such that $Q(\cdot \mid y)$ is a probability measure on X for each fixed $y \in Y$, and $Q(B \mid \cdot)$ is a measurable function on Y for each fixed $B \in \mathcal{B}(X)$.*

2. Adaptive stochastic optimal control problems

2.1 Markov control models

We consider a class of discrete-time Markov control models

$$\mathcal{M} := (X, A, \{A(x) \subset A | x \in X\}, c, Q) \tag{2}$$

satisfying the following conditions. The state space *X* and action space *A* are Borel spaces endowed with their Borel *σ*-algebras (See Remark 1.1). For each state $x \in X$, $A(x)$ is a nonempty Borel subset of *A* denoting the set of admissible controls when the system is in state *x*. The set

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

of admissible state-action pairs is assumed to be a Borel subset of the Cartesian product of *X* and *A*. In addition, the cost-per-stage *c*(*x*, *a*) is a nonnegative measurable real-valued function, possibly unbounded, and depends on the pair $(x, a) \in \mathbb{K}$. Finally, the transition law of the system $Q(\cdot|\cdot)$ is a stochastic kernel on *X* given $\mathbb{K}$. That is, for all $t \in \mathbb{N}$, $(x, a) \in \mathbb{K}$ and $B \in \mathcal{B}(X)$,

$$Q(B|x, a) := P\left[x_{t+1} \in B | x_t = x, a_t = a\right]. \tag{3}$$

We will consider independently the two following cases:

- 1st case: the stochastic kernel *Q* is unknown, as depending on the system disturbance distribution θ_ξ, which is unknown. We have, for all (*x*, *a*) ∈ $\mathbb{K}$ and $B \in \mathcal{B}(X)$,

$$\begin{aligned} Q(B|x, a) &= \int_S 1_B\left(F(x, a, s)\right) \theta_\xi(ds) \\ &= \theta_\xi\left(\{s \in S \ : \ F(x, a, s) \in B\}\right), \end{aligned}$$

where S is the Borel space of the disturbance ξ in (1). The Markov control model under consideration can also be noted $\mathcal{M}_\xi := (X, A, \{A(x) \subset A | x \in X\}, S, F, c, \theta_\xi)$.

- 2nd case: the cost function c is poorly known, as depending on the unknown distribution θ_η, of a stochastic variable η through the relation:

$$c(x,a) := E_{\theta_\eta}\, \bar{c}\,(x,a,\eta)\,, \tag{4}$$

where η is an exogenous variable belonging to a Borel space S and E_{θ_η} denotes the expectation operator with respect to the probability distribution θ_η. Thus

$$\theta_\eta(B) = P\,(\eta_t \in B)\,,\ t \in \mathbb{N},\ B \in \mathcal{B}(S). \tag{5}$$

The function $\bar{c}$ is in fact the true cost function, whose mean c is unknown, which yields the following Markov control model $\mathcal{M}_\eta := (X, A, \{A(x) \subset A | x \in X\}, Q, S, \theta_\eta, \bar{c})$.

Throughout the paper we suppose that the random variables ξ and η are defined on an underlying probability space $(\Omega, \mathcal{F}, P)$, and *a.s.* means *almost surely with respect to P*. In addition, we assume the complete observability of the states x_0, x_1, ..., and also of the realizations $\eta_0, \eta_1, ...$ or $\xi_0, \xi_1, ...$ when their distribution is unknown.

2.2 Set of admissible policies

We define the spaces of admissible histories up to time t by $\mathbb{H}_0 := X$ and $\mathbb{H}_t := (\mathbb{K}\times S)^t \times X, t \geq 1$. A generic element of $\mathbb{H}_t$ is written as $h_t = (x_0, a_0, \eta_0, ..., x_{t-1}, a_{t-1}, \eta_{t-1}, x_t)$. A control policy $\pi = \{\pi_t\}$ is a sequence of measurable functions $\pi_t : \mathbb{H}_t \to A$ such that $\pi_t(h_t) \in A(x_t)$, $h_t \in \mathbb{H}_t$, $t \geq 0$. Let Π be the set of all control policies and $\mathbb{F} \subset \Pi$ the subset of stationary policies. If necessary, see for example (Dynkin & Yushkevich, 1979); (Hernández-Lerma & Lasserre, 1996 and 1999); (Hernández-Lerma, 1989) or (Gordienko & Minjárez-Sosa, 1998) for further information on those policies. As usual, each stationary policy $\pi \in \mathbb{F}$ is identified with a measurable function $f : X \to A$ such that $f(x) \in A(x)$ for every $x \in X$, so that π is of the form $\pi = \{f, f, f, ...\}$. In this case we denote π by f, and we write

$$c(x,f) := c(x, f(x)) \quad \text{and} \quad Q(B|x,f) := Q(B|x, f(x)),$$

for all $x \in X$.

2.3 Discounted criterion

Once we are given a Markov control model $\mathcal{M}$ and a set Π of admissible policies, to complete the description of an optimal control problem we need to specify a performance index, that is, a function measuring the system's performance when a given policy $\pi \in \Pi$ is used and the initial state of the system is $x_0 = x$. This study concerns the α-discounted cost, whose definition is as follows:

$$V(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right], \quad x \in X,\ \pi \in \Pi. \tag{6}$$

where $\alpha \in (0,1)$ is the so-called discount factor, and E_x^π denotes the expectation operator with respect to the probability measure P_x^π induced by the policy π, given the initial state $x_0 = x$.

The α-discounted criterion is one of the most famous long run criteria. Among the main motivations to study this optimality criterion are to analyze an economic or financial model as an optimal control problem (for instance optimal growth of capital model, see Stockey & Lucas (1989)), and the mathematical convenience (the discounted criterion is the best understood of all performance index). In fact, it is often studied before other more complicated criteria, like for example the expected average cost, which can be seen as the limit of $V(\pi, x)$ when α tends to 1.
The optimal control problem is then defined as follows: determine a policy $\pi^* \in \Pi$ such that:

$$V(\pi^*, x) = \inf_{\pi \in \Pi} V(\pi, x) \text{ for all } x \in X.$$

The function V^* defined by

$$V^*(x) := \inf_{\pi \in \Pi} V(\pi, x), \ x \in X \tag{7}$$

is called the value (or optimal cost) function. A policy $\pi^* \in \Pi$ is said to be α-discount optimal (or simply α-optimal) for the control model $\mathcal{M}$ if

$$V^*(x) = V(\pi^*, x) \text{ for all } x \in X. \tag{8}$$

Note that, in the case of model $\mathcal{M}_\eta$ we are in fact interested by looking for optimal policies with respect to the general α-discounted cost

$$V_\eta(\pi, x) := E_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t \bar{c}(x_t, a_t, \eta_t) \right], \quad x \in X, \ \pi \in \Pi.$$

But, as c is the mean cost of function $\bar{c}$, see (4), and using properties of conditional expectation, we have that $V_\eta(\pi, x) = V(\pi, x)$. So, looking for optimal policies for V_η, is equivalent to looking for optimal policies for V.
Since θ_ξ and θ_η are unknown, we combine suitable statistical estimation methods and control procedures in order to construct the adaptive policy. That is, we use the observed history of the system to estimate θ_ξ or θ_η and then adapt the decision or control to the available estimate. On the other hand, as the discounted cost depends heavily on the controls selected at the first stages (precisely when the information about the unknown distribution is poor or deficient), we can't ensure the existence of an α-optimal adaptive policy (see Hernández-Lerma, 1989). Thus the α-optimality of an adaptive policy will be understood in the following asymptotic sense:
Definition 2.1 *(Schäl, 1987). A policy $\pi \in \Pi$ is said to be asymptotically discounted optimal for the control model $\mathcal{M}$ if*

$$\left| V^{(k)}(\pi, x) - E_x^\pi \left[V^*(x_k) \right] \right| \to 0 \ \ as \ \ k \to \infty, \ for \ all \ x \in X,$$

where

$$V^{(k)}(\pi, x) := E_x^\pi \left[\sum_{t=k}^{\infty} \alpha^{t-k} c(x_t, a_t) \right]$$

is the expected total discounted cost from stage k onward and $a_t = \pi_t(h_t)$.
In the above definition, the model $\mathcal{M}$ stands either for $\mathcal{M}_\xi$ or for $\mathcal{M}_\eta$

Remark 2.2 *Let* $\pi \in \Pi$ *be a policy such that* $V(\pi, x) < \infty$ *for each* $x \in X$, *and* $\{(x_t, a_t)\}$ *be a sequence of state-actions pairs corresponding to application of* π. *In (Hernández-Lerma & Lasserre, 1996), it has been proved that* π *is an asymptotically discounted optimal policy if, and only if,* $E_x^\pi \Phi(x_t, a_t) \to 0$, *as* $t \to \infty$, *where*

$$\Phi(x,a) := c(x,a) + \alpha \int_X V^*(y) Q(dy|x,a) - V^*(x), \quad (x,a) \in \mathbb{K}, \tag{9}$$

is the well-known discrepancy function, which is nonnegative from (15).
In the remainder of the paper, we fix an arbitrary discount factor $\alpha \in (0,1)$.

2.4 Overview of adaptive Markov control processes with Borel state and action spaces, and possibly unbounded costs

Even in the non adaptive case, handling Markov control processes with Borel state and action spaces, and possibly unbounded costs, requires much attention in the work space setting towards specific assumptions. Three types of hypotheses are usually imposed, see (Hernández-Lerma & Lasserre, 1999). The first one is about compactness-continuity conditions for Markov control models. The second one introduces a weight function *W* to impose a growth condition on the cost function, which will yield that the dynamic programming operator *T:*

$$Tu(x) := \min_{A(x)} \left(c(x,a) + \alpha \int_X u(y) Q(dy|x,a) \right) \tag{10}$$

is a contraction (on some space that will be specified later, see §3.1). The third type of assumptions is a further continuity condition, which combined with the previous ones, will ensure the existence of measurable minimizers for *T*. We don't detail these assumptions for the general non adaptive case. They are extended to model $\mathcal{M}_\eta$ in the adaptive case as follows:

Assumption 2.3 *a) For each* $x \in X$, *the set A(x) is σ-compact.*
b) For each $x \in X$ *the function* $a \to \bar{c}(x,a,s)$ *is l.s.c. on A*(*x*) *for all s. Moreover, there exists a measurable function* $W : X \to [1, \infty)$ *such that* $\sup_{a \in A(x)} \sup_{s \in S} \bar{c}(x,a,s) \le W(x)$ *for all* $x \in X$. *(Recall that* $\bar{c}$ *is assumed to be nonnegative.)*
c)There exist three constants $p > 1$, $\beta_0 < 1$ *and* $b_0 < +\infty$ *such that for all* $x \in X$, $a \in A(x)$,

$$\int_X W^p(y) Q(dy|x,a) \le \beta_0 W^p(x) + b_0. \tag{11}$$

d) The function $(x,a) \to \int_X v(y) Q(dy|x,a)$ *is continuous and bounded on* $\mathbb{K}$ *for every bounded and continuous function v on X.*
e) For each $x \in X$, *the function* $a \to \int_X W(y) Q(dy|x,a)$ *is continuous on A(x).*
Remark 2.4 *Note that from Jensen's inequality, (11) implies*

$$\int_X W(y) Q(dy|x,a) \le \beta W(x) + b, \quad \text{for all } (x,a) \in \mathbb{K}, \tag{12}$$

where $\beta = \beta_0^{1/p}$ *and* $b = b_0^{1/p}$. *Moreover, a consequence for both inequalities (11) and (12), is (see (Gordienko & Minjdrez-Sosa, 1998) or (Hernández-Lerma & Lasserre, 1999))*

$$\sup_{n\geq 0} E_x^{\pi}\left(W^p(x_n)\right) < \infty \text{ and } \sup_{n\geq 0} E_x^{\pi}\left(W(x_n)\right) < \infty, \tag{13}$$

for each $\pi \in \Pi$ *and* $x \in X$.
We denote by L_W^{∞} the normed linear space of all measurable functions $u : X \to \Re$ with a finite norm $\|u\|_W$ defined as

$$\|u\|_W := \sup_{x\in X} \frac{|u(x)|}{W(x)}. \tag{14}$$

A first consequence of Assumption 2.3 is the following proposition, which states the existence of a stationary α-discount optimal policy in the general case:
Proposition 2.5 *(Hernández-Lerma & Lasserre, 1999) Suppose that Assumption 2.3 holds. Then: a) The function V* belongs to* L_W^{∞} *and satisfies the* α*-discounted optimality equation*

$$V^*(x) = \min_{a\in A(x)} \left(c(x,a) + \alpha \int_X V^*(y) Q\left(dy|x,a\right) \right), \; x \in X \tag{15}$$

Moreover, we have $0 \leq V^*(x) \leq W(x)/(1-\alpha)$, $x \in X$.
b) There exists $f \in \mathbb{F}$ *such that* $f(x) \in A(x)$ *attains the minimum in (15), i.e.*

$$V^*(x) = c(x,f) + \alpha \int_X V^*(y) Q\left(dy|x,f\right), \; x \in X, \tag{16}$$

and the stationary policy f is optimal.
As we already mentioned, our main concern is in the two cases of adaptive control we introduced in §2.1 where the distribution θ_ξ or θ_η is unknown. Thus, the solution given in the Proposition 2.5 is not accessible to the controller. In fact, an estimation process has to be chosen, which depends on the knowledge we have of this distribution, for example: absolutely continuous with respect to the Lebesgue measure (and so with an unknown density). With the estimator on hand we can apply the "principle of estimation and control" proposed by Kurano (1972) and Mandl (1974). That is, we obtain an estimated optimality equation with which we can construct the adaptive policies.
The case of the model $\mathcal{M}_\xi$, and assuming that θ_ξ has a density, is described in (Gordienko & Minjárez-Sosa, 1998), and also in (Minjárez-Sosa, 1999) for the expected average cost. The estimation of θ_ξ is obtained by means of an estimator of its density function. However the unboundedness assumption on the cost c makes difficult the implementation of the density estimation process. The estimator is defined by the projection (of an auxiliary estimator) on some special set of density functions to ensure good properties of the estimated model. Beyond the complexity of the estimation procedure, the assumption of absolutely continuity excludes the case of discrete distributions, which appears in some inventory-production and queuing systems. On the other hand, the case of an arbitrary distribution θ_ξ (without a priori assumption) has been treated in (Hilgert & Minjárez-Sosa, 2006) and relies on the empirical distribution. It may seem an obvious choice, but this was a great improvement on what was done previously. The assumptions used are even weaker than in the non adaptive case and wouldn't be sufficient to prove the existence of a stationary optimal policy with a known distribution θ_ξ. The extension to the expected average cost is the subject of (Minjárez-Sosa, 2008).
The case of model $\mathcal{M}_\eta$, is less known in the literature and is treated in detail in the following section.

3. Adaptive control of stochastic systems with poorly-known cost function

The construction of the adaptive policies is based mainly on the cost estimation process which, in turns, is obtained by implementing suitable estimation methods of the probability distribution θ_η. In general our approach consists in getting an estimator c_n of the cost such that

- it converges to c (in a sense that will be given later);
- it leads up to the convergence of the following sequence: $V_0^* = 0$,

$$V_n^*(x) := \min_{a \in A(x)} \left(c_n(x,a) + \alpha \int_X V_{n-1}^*(y) Q(dy|x,a) \right), n \geq 1, x \in X \tag{17}$$

to the unknown value function V^* given in (7).
In particular, we take

$$c_n(x,a) = E_{\theta_{\eta,n}} \bar{c}(x,a,s), \quad (x,a) \in \mathbb{K}, \tag{18}$$

where $\{\theta_{\eta,n}\} \subset \mathbb{P}(S)$ is a sequence of "consistent" estimators of θ_η.
Now, applying standard arguments on the existence of minimizers, under Assumption 2.3, we have that for each $n \in \mathbb{N}$ there exists $f_n^{\theta_{\eta,n}} \in \mathbb{F}$ such that,

$$V_n^*(x) = c_n(x, f_n^{\theta_{\eta,n}}) + \alpha \int_X V_{n-1}^*(y) Q(dy|x, f_n^{\theta_{\eta,n}}), \; x \in X, \tag{19}$$

where the minimization is done for every $\omega \in \Omega$. Moreover, by a result of (Schäl, 1975), there is a stationary policy $f_\infty^{\theta_\eta}$ such that for each $x \in X$, $f_\infty^{\theta_\eta}(x) \in A(x)$ is an accumulation point of $\{f_n^{\theta_{\eta,n}}(x)\}$.
We state our main result as follows:
Theorem 3.1 *a) Let $\hat{\pi} = \{\hat{\pi}_n\}$ be the policy defined by $\hat{\pi}_n(h_n) = \hat{\pi}_n(h_n, \theta_{\eta,n}) := f_n^{\theta_{\eta,n}}(x_n)$, $n \in \mathbb{N}$ and $\hat{\pi}_0$ any fixed action.*
Then, under Assumption 2.3 and if $\{\theta_{\eta,n}\}$ is an appropriate sequence of "consistent" estimators of θ_η, $\hat{\pi}$ is asymptotically discount optimal.
b) In addition, the stationary policy $f_\infty^{\theta_\eta}$ is optimal for the control model $\mathcal{M}_\eta$.
The remainder of this section is devoted to the proof of Theorem 3.1 for two estimators of the cost function that correspond to two different assumptions on the unknown distribution θ_η. In the first one, Subsection 3.2, we suppose that θ_η is absolutely continuous with respect to the Lebesgue measure and has an unknown density function g. The estimator c_n of the cost function is then based on a nonparametric estimator of g. Next, in Subsection 3.4, we don't make any a priori assumption on θ_η. The estimator c_n is based on the empirical distribution of θ_η. We first give some preliminary definitions and developments that are useful for both situations.

3.1 Preliminaries

We present some preliminary facts that will be useful in the proof of our main result.
Let us define the operator T_n in the same way as T in (10):

$$\Gamma_n u(x) := \min_{A(x)} \left(c_n(x,a) + \alpha \int_X u(y) Q(dy|x,a) \right) \tag{20}$$

for all $x \in X$ and $u \in L_W^\infty$. Observe that from (15) and (17)

$$TV^* = V^* \quad \text{and} \quad T_n V_{n-1}^* = V_n^*, \quad n \in \mathbb{N}. \tag{21}$$

In addition, from Assumption 2.3(a), $\sup_{A(x)} c_n(x,a) \leq W(x)$, and applying the inequality (12), a straightforward calculation shows that, for some constant C,

$$V_n^*(x) \leq C\, W(x),\ n \in \mathbb{N}, x \in X. \tag{22}$$

Thus, from Assumption 2.3, T and T_n maps L_W^∞ into itself.
We fix an arbitrary number $\epsilon \in (\alpha, 1)$ and define the function $\bar{W}(x) := W(x) + d$ for $x \in X$, where $d := b\,(\epsilon/\alpha - 1)^{-1}$. Let $L_{\bar{W}}^\infty$ be the space of measurable functions $u : X \to \mathbb{R}$ with norm

$$\|u\|_{\bar{W}} := \sup_{x \in X} \frac{|u(x)|}{\bar{W}(x)} < \infty.$$

Observe that the norms $\|\cdot\|_W$ and $\|\cdot\|_{\bar{W}}$ are equivalent because

$$\|u\|_{\bar{W}} \leq \|u\|_W \leq (1+d)\ \|u\|_{\bar{W}}\,. \tag{23}$$

A consequence of Lemma 2 in (Van Nunen & Wessels, 1978) is that the inequality (12) implies respectively that the operators T_n and T, $n \in \mathbb{N}$, are contractions with modulus ϵ, with respect to the norm $\|\cdot\|_{\bar{W}}$, i.e. for all $u, v \in L_{\bar{W}}^\infty$:

$$\|Tv - Tu\|_{\bar{W}} \leq \epsilon\, \|v - u\|_{\bar{W}}\,, \tag{24}$$

$$\|T_n v - T_n u\|_{\bar{W}} \leq \epsilon\, \|v - u\|_{\bar{W}} \ \text{a.s.} \tag{25}$$

Hence, from (21), for each $n \in \mathbb{N}$, we have

$$\begin{aligned} \|V^* - V_{n+1}^*\|_{\bar{W}} &\leq \|TV^* - T_{n+1}V^*\|_{\bar{W}} + \|T_{n+1}V^* - T_{n+1}V_n^*\|_{\bar{W}} \\ &\leq \|TV^* - T_{n+1}V^*\|_{\bar{W}} + \epsilon \|V^* - V_n^*\|_{\bar{W}}. \end{aligned} \tag{26}$$

Now, let $\pi \in \Pi$ and $x \in X$ be arbitrary, and define

$$l := \limsup_{n\to\infty} E_x^\pi \|V^* - V_{n+1}^*\|_{\bar{W}}$$

and

$$l' := \limsup_{n\to\infty} \|V^* - V_{n+1}^*\|_{\bar{W}}.$$

Observe that $l < \infty$ and $l' < \infty$ (see Proposition 2.5, (22) and (23)). Then from (26)

$$l \leq \frac{1}{1-\epsilon} \limsup_{n\to\infty} E_x^\pi \|TV^* - T_{n+1}V^*\|_{\bar{W}} \tag{27}$$

and

$$l' \leq \frac{1}{1-\epsilon} \limsup_{n\to\infty} \|TV^* - T_{n+1}V^*\|_{\bar{W}} \quad a.s. \tag{28}$$

3.2 Cost estimation when θ_η has a density

In this part, we suppose the existence of a density of θ_η as stated below. We will then start step by step the proof of Theorem 3.1.

Assumption 3.2 *a)* $S = \Re^k$.
b) The distribution θ_η *is absolutely continuous with respect to the Lebesgue measure on* $\Re^k$ *and has a density function* g. *That is,*

$$\theta_\eta(B) = \int_B g(s)ds, \quad B \in \mathcal{B}(\Re^k). \tag{29}$$

Under this context, from (4) we have

$$c(x,a) = \int_{\Re^k} \bar{c}(x,a,s)g(s)ds, \quad (x,a) \in \mathbb{K}; \tag{30}$$

Let $\eta_1, ..., \eta_t$ be independent realizations (observed up to time *t*), of r.v.'s with the unknown density *g*, and $g_n(s) := g_n(s; \eta_1, .., \eta_n)$, $s \in \Re^k$, be an arbitrary estimator of g such that

$$E \int_{\Re^k} |g(s) - g_n(s)| \to 0 \text{ as } n \to \infty. \tag{31}$$

Defining, for each $n \in \mathbb{N}$,

$$\theta_{\eta,n}(B) = \int_B g_n(s)ds, \quad B \in \mathcal{B}(\Re^k), \tag{32}$$

the relation (18) becomes

$$c_n(x,a) := \int_{\Re^k} \bar{c}(x,a,s)g_n(s)ds, \quad (x,a) \in \mathbb{K}. \tag{33}$$

Now, let us define the approximate discrepancy function Φ_n, for each $n \in \mathbb{N}$ as (see (9))

$$\Phi_n(x,a) := c_n(x,a) + \alpha \int_X V^*_{n-1}(y)Q(dy|x,a) - V^*_n(x), \tag{34}$$

for all $(x,a) \in \mathbb{K}$, where $\{V^*_n\}$ is the sequence defined in (17) corresponding to the cost (33), and denote

$$\Psi_n := \sup_{x \in X} \left\{ (W(x))^{-1} \sup_{a \in A(x)} |\Phi(x,a) - \Phi_n(x,a)|, \quad n \in \mathbb{N}. \right\} \tag{35}$$

The following Lemma will be useful to prove Theorem 3.1.
Lemma 3.3 *Assumptions 2.3, 3.2 and (31) imply that:*
a) For each $\pi \in \Pi$ *and* $x \in X$,

$$\lim_{n \to \infty} E^\pi_x \|V^*_n - V^*\|_W = 0. \tag{36}$$

b)For each $x \in X$ *and* $\pi \in \Pi$,

$$\lim_n E^\pi_x \Psi_n = 0. \tag{37}$$

c)For each $x \in X$ *and* $\pi \in \Pi$,

$$\lim_n E_x^\pi \left(W(x_n)\Psi_n\right) = 0. \tag{38}$$

d) For each $(x,a) \in \mathbb{K}$,

$$\lim_n E_x^\pi \left| \int_X V_{n-1}^*(y)Q(dy|x,a) - \int_X V^*(y)Q(dy|x,a) \right| = 0. \tag{39}$$

Proof:
a) From (23) and (27), to prove the part a), it is sufficient to show that

$$E_x^\pi \|TV^* - T_{n+1}V^*\|_{\bar{W}} \to 0, \text{ as } n \to \infty. \tag{40}$$

To this end, from (10), (20), (30), and (33), and Assumption 2.3(a), for each $x \in X$ and $n \in \mathbb{N}$,

$$\begin{aligned} |TV^*(x) - T_{n+1}V^*(x)| &\le \sup_{a\in A(x)} |c(x,a) - c_n(x,a)| \\ &\le \sup_{a\in A(x)} \int_{\Re^k} \bar{c}(x,a,s)\,|g(s) - g_n(s)|\,ds \\ &\le W(x) \int_{\Re^k} |g(s) - g_n(s)|\,ds. \end{aligned}$$

Hence, as $\bar{W}(\cdot) > W(\cdot)$,

$$\|TV^* - T_nV^*\|_{\bar{W}} \le \int_{\Re^k} |g(s) - g_n(s)|ds, \;\; n \in \mathbb{N}. \tag{41}$$

Taking expectation E_x^π on both sides of (41) and observing that (since g_n does not depend on $x \in X$ and $\pi \in \Pi$)

$$E_x^\pi \int_{\Re^k} |g(s) - g_n(s)|ds = E \int_{\Re^k} |g(s) - g_n(s)|ds, \;\; n \in \mathbb{N},$$

relation (40) follows thanks to (31).
b) From definitions of the function Φ and Φ_n, the norm $\|\cdot\|_{\bar{W}}$ in (14), and (12), for each $(x,a) \in \mathbb{K}$ and $n \in \mathbb{N}$,

$$\begin{aligned} |\Phi(x,a) - \Phi_n(x,a)| \le& |c(x,a) - c_n(x,a)| + |V^*(x) - V_n^*(x)| \\ &+ \alpha \int_X |V^*(y) - V_n^*(y)|Q(dy|x,a) \\ \le& W(x) \int_X |g_n(s) - g(s)|ds + \|V^* - V_n^*\|W(x) \\ &+ \|V^* - V_{n-1}^*\|_W(\beta W(x) + b). \end{aligned} \tag{42}$$

Thus, (31) and Lemma 3.3a yield (37). In addition, there exists a finite constant M such that $\sup_n \Psi_n \le M$, which, combined with (37), yields the convergence in probability

$$\Psi_n \overset{P_x^\pi}{\to} 0 \quad \text{as} \quad n \to \infty.$$

c) But, from (13) we have that

$$\sup_n E_x^\pi \left(W^p(x_n)\Psi_n\right)^p < M^p \sup_n E_x^\pi W^p(x_n) < \infty.$$

We then deduce that $W(x_n)\Psi_n$ is P_x^π-uniformly integrable. Moreover, using Chebychev's inequality,

$$P_x^\pi \left(W(x_n)\psi_n > l_1\right) \leq P_x^\pi \left(\psi_n > \tfrac{l_1}{l_2}\right) + P_x^\pi \left(W(x_n) > l_2\right) \tag{43}$$

$$\leq P_x^\pi \left(\psi_n > \tfrac{l_1}{l_2}\right) + \frac{E_x^\pi W(x_n)}{l_2}. \tag{44}$$

Hence,

$$W(x_n)\psi_n \overset{P_x^\pi}{\to} 0 \quad \text{as } n \to \infty,$$

and as it is P_x^π-uniformly integrable, we get (38).

d) This result is a consequence of Lemma 3.3a and the following inequalities

$$\left|\int_X V_{n-1}^*(y)Q(dy|x,a) - \int_X V^*(y)Q(dy|x,a)\right| \leq \int_X \left|V_{n-1}^*(y) - V^*(y)\right| Q(dy|x,a)$$
$$\leq \|V^* - V_{n-1}^*\|_W \left(\beta W(x) + b\right) \quad \blacksquare$$

3.3 Proof of Theorem 3.1

We prove Theorem 3.1 in the specific case of θ_η, absolutely continuous with respect to the Lebesgue measure, that is, with θ_η, c, $\theta_{\eta,n}$ and c_n given by (29), (30), (32), and (33), respectively. We will show in the Subsection 3.4 that it still holds with an arbitrary distribution θ_η,.

Proof of part a) Observe that by definition of the control policy $\hat{\pi}$ (see (19) and Theorem 3.1) and (34), we have $\Phi_n(.,\hat{\pi}(h_n)) = 0$ for all n. Hence,

$$\begin{aligned}
\Phi(x_n,\hat{\pi}(h_n)) &= \left|\Phi(x_n,\hat{\pi}(h_n)) - \Phi_n(x_n,\hat{\pi}(h_n))\right| \\
&\leq \sup_{a\in A(x)} \left|\Phi(x_n,a) - \Phi_n(x_n,a)\right| \\
&\leq W(x_n) \sup_X \left(W(x)\right)^{-1} \sup_{a\in A(x)} \left|\Phi(x,a) - \Phi_n(x,a)\right| \\
&\leq W(x_n)\Psi_n,
\end{aligned}$$

which, combined with (38), proves the asymptotic discounted optimality of $\hat{\pi}$.

Proof of part b) We fix an arbitrary $x \in X$. Since $f_\infty^{\theta_\eta}$ is an accumulation point of $\left\{f_n^{\theta_{\eta,n}}(x)\right\}$, there exists a subsequence $\{n_i(x)\}$ of $\{n\}$ such that $f_{n_i(x)}^{\theta_{\eta,n_i}} \to f_\infty^{\theta_\eta}(x)$ as $i \to \infty$. In addition,

$$V_{n_i(x)}^*(x) = c_{n_i(x)}(x, f_{n_i(x)}^{\theta_{\eta,n_i}}) + \alpha \int_X V_{n_i(x)-1}^*(y)Q(dy|x, f_{n_i(x)}^{\theta_{\eta,n_i}}). \tag{45}$$

Moreover, from (39), we deduce that

$$\liminf_i E \int_X V_{n_i(x)-1}^*(y)Q(dy|x, f_{n_i(x)}^{\theta_{\eta,n_i}}) = \liminf_i E \int_X V^*(y)Q(dy|x, f_{n_i(x)}^{\theta_{\eta,n_i}})$$
$$\geq \int_X V^*(y)Q(dy|x, f_\infty^{\theta_\eta}) \quad \text{by Fatou's Lemma.}$$

Then, taking the limit infimum in (45) yields

$$c(x, f_\infty^{\theta_\eta}) + \alpha \int_X V^*(y) Q(dy|x, f_\infty^{\theta_\eta}) \leq V^*(x).$$

As x was arbitrary, the equality holds for every $x \in X$ and so $f_\infty^{\theta_\eta}$ is optimal for the control model $\mathcal{M}_\eta$. ■

3.4 Cost estimation with the empirical distribution of θ_η

In this part, we suppose the disturbance space S and the distribution θ_η arbitrary. To estimate θ_η we use the empirical distribution $\{\theta_{\eta,t}\} \subset \mathbb{P}(S)$ of the disturbance process $\{\eta_t\}$, defined as follows. Let $\nu \in \mathbb{P}(S)$ be a given arbitrary probability measure. Then

$$\begin{aligned} \theta_{\eta,0} &:= \nu, \\ \theta_{\eta,n}(B) &:= \frac{1}{n} \sum_{i=0}^{n-1} 1_B(\eta_i), \quad \text{for all } n \geq 1 \text{ and } B \in \mathcal{B}(S). \end{aligned} \tag{46}$$

Under this context, we have (see (4) and (18))

$$c(x, a) = \int_S \bar{c}(x, a, s) \theta_\eta(ds) \tag{47}$$

and

$$c_n(x, a) = \int_S \bar{c}(x, a, s) \theta_{\eta,n}(ds) := \frac{1}{n} \sum_{i=0}^{n-1} \bar{c}(x, a, \eta_i). \tag{48}$$

Clearly, from the law of large numbers, for each $(x, a) \in \mathbb{K}$, $c_n(x, a) \to c(x, a)$ *a.s.*, as $n \to \infty$. However, to our objectives, we need uniform convergence on (*x, a*) of the costs, for which we impose the following conditions.

Assumption 3.4 *a) The family of functions*

$$\mathcal{C}_W := \left\{ \frac{\bar{c}(x, a, \cdot)}{W(x)} : (x, a) \in \mathbb{K} \right\}$$

is equicontinuous on S.
b) The function

$$\varphi(s) := \sup_{(x,a) \in \mathbb{K}} [W(x)]^{-1} \bar{c}(x, a, s)$$

is continuous on S.

Remark 3.5 *a) Observe that from Assumption 2.3(b), $\varphi(s) \leq 1$, $s \in S$. Hence $E[\varphi(\eta_0)]^r < \infty$, for all r > 0. Then, from Assumption 3.4 and applying Theorem 6.4 in (Ranga Rao, 1962), we get, as $n \to \infty$,*

$$\sup_{(x,a) \in \mathbb{K}} \left| \int_S \frac{\bar{c}(x, a, s)}{W(x)} \theta_\eta(ds) - \int_S \frac{\bar{c}(x, a, s)}{W(x)} \theta_{\eta,n}(ds) \right| \to 0 \quad a.s. \tag{49}$$

b) The function φ in Assumption 3.4 might be non continuous. In such case we replace Assumption 3.4(b) by supposing the existence of a continuous majorant $\bar{\varphi}$ of φ such that $E\,[\bar{\varphi}(\eta_0)]^r < \infty$ for some $r > 1$.

Let $\{V_n^*\}$, $\{\Phi_n\}$, and $\{\Psi_n\}$ be the sequences of functions defined in (17), (34), and (35), respectively, corresponding to the cost functions (47) and (48). According to Subsection 3.3 and the proof of Lemma 3.3, to prove the Theorem 3.1 under the empirical estimator, it is sufficient to state the following results.

Lemma 3.6 *Under Assumptions 2.3 and 3.4,*

a) $\lim_n \|V_n^* - V^*\|_W = 0 \;\; a.s.$;

b) $\lim_n \Psi_n = 0 \;\; a.s.$;

c) for each $x \in X$ *and* $\pi \in \Pi$, $\lim_n E_x^\pi\,(W(x_n)\Psi_n) = 0$;

d) $\lim_n \left|\int_X V_{n-1}^*(y)Q(dy|x,a) - \int_X V^*(y)Q(dy|x,a)\right| = 0 \;\; a.s.$

Proof. The part a) is a consequence of the following inequality. From (10), (20), (47), (48), and (49), for each $x \in X$,

$$\|TV^* - T_nV^*\|_W \leq \sup_{(x,a)\in\mathbb{K}} \left| \int_S \frac{\bar{c}(x,a,s)}{W(x)}\theta_\eta(ds) - \int_S \frac{\bar{c}(x,a,s)}{W(x)}\theta_{\eta,n}(ds) \right| \to 0 \;\; a.s.$$

Thus, (28) and (23) yield the part a).

On the other hand, observe that (see (42))

$$\Psi_n \leq \sup_{(x,a)\in\mathbb{K}} \left| \int_S \frac{\bar{c}(x,a,s)}{W(x)}\theta_\eta(ds) - \int_S \frac{\bar{c}(x,a,s)}{W(x)}\theta_{\eta,n}(ds) \right| + \|V^* - V_n^*\|_W + \|V^* - V_{n-1}^*\|_W(\beta + b).$$

Therefore, the part b) follows from (49) and the part a).

Finally, the parts c) and d) are obtained by applying similar arguments as for proving (38) and (39). ■

4. Concluding remarks

A general scheme to construct adaptive policies in control models as (2) is to combine statistical estimation methods of the unknown distribution with control procedures. Such policies have optimality properties provided that the estimators are consistent in an appropriate sense. In this paper we studied two cases of adaptive control models, model $\mathcal{M}_\xi$ who has an unknown system disturbance distribution, and model $\mathcal{M}_\eta$ where the cost function depends on an exogenous variable of unknown distribution. We stated two ways of estimating θ_ξ or θ_η which yielded two different asymptotically discounted optimal adaptive policies. In the first one, it is assumed that the distribution possesses a density function g on $\Re^k$. The estimation process in this case is based on the estimation of g, which can be done in a number of nice ways (see, e.g., Devroye (1987), Devroye & Lugosi (2001)), but has the disadvantage of excluding the case when the distribution is discrete.

The construction of adaptive policies using the empirical distribution as estimator is very general in the sense that the disturbance space S as well as the distribution θ_ξ or θ_η can be arbitrary. This approach has the disadvantage that it requires restrictive equicontinuity conditions (see Assumption 3.4) which is the price we have to pay for nothing assuming on the unknown distribution. However, this assumption is satisfied in important cases. For

instance, an obvious sufficient condition for Assumption 3.4(a) is that S is countable. Also, this assumption holds in the case of additive-noise cost function of the form $\bar{c}(x,a,s) = G(x,a) + s$, where G is a continuous function.

5. References

Devroye, L. (1987). *A Course in Density Estimation,* Birkhauser, Boston.

Devroye, L. *&* Lugosi, G. (2001). *Combinatorial Methods in Density Estimation,* Springer, New York.

Gordienko, E.I. & Minjárez-Sosa, J.A. (1998). Adaptive control for discrete-time Markov processes with unbounded costs: discounted criterion. *Kybernetika,* Vol. 34, 217-234.

Hernández-Lerma, O. (1989). *Adaptive Markov Control Processes,* Springer-Verlag, New York.

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

Hernández-Lerma, O. & Lasserre, J.B. (1999). *further Topics on Discrete-Time Markov Control Processes,* Springer-Verlag, New York.

Hilgert, N. & Minjárez-Sosa, J.A. (2006). Adaptive Control of Stochastic Systems with Unknown Disturbance Distribution: Discounted Criteria. *Math. Methods Oper. Res.,* Vol. 63, No. 3, 443-460.

Kurano M. (1972). Discrete-time markovian decision processes with an unknown parameter - average return criterion. *J. Oper. Res. Soc. Japan,* Vol. 15, 67-76.

Mandl P. (1974). Estimation and control in Markov chains. *Adv. Appl. Probab.,* Vol. 6, 40-60.

Minjárez-Sosa, J.A. (1999). Nonparametric adaptive control for discrete-time Markov processes with unbounded costs under average criterion. *Appl. Math.,* Vol. 26, No. 3, 267-280.

Minjárez-Sosa, J.A. (2008). Empirical estimation in average Markov control processes. *Applied Mathematics Letters,* Vol. 21, No. 5, 459-464.

Ranga Rao, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statistics,* Vol. 33, 659-680.

Schäl, M. (1975). Conditions for optimality and for the limit of n-stage optimal policies to be optimal. *Z. Wahrs. Verw. Gerb.,* Vol. 32,179-196.

Schäl, M. (1987). Estimation and control in discounted stochastic dynamic programming. *Stochastics,* Vol. 20, 51-71.

Stockey N.L.& Lucas R.E. Jr (1989). *Recursive Methods in Economic Dynamics.* Harvard University Press, Cambridge, MA.

Van Nunen, J.A.E.E. *&* Wessels, J. (1978). A note on dynamic programming with unbounded rewards. *Manag. Sci.,* Vol. 24, 576-580.

Frontiers in Adaptive Control

Edited by Shuang Cong

ISBN 978-953-7619-43-5
Hard cover, 334 pages
Publisher InTech
Published online 01, January, 2009
Published in print edition January, 2009

The objective of this book is to provide an up-to-date and state-of-the-art coverage of diverse aspects related to adaptive control theory, methodologies and applications. These include various robust techniques, performance enhancement techniques, techniques with less a-priori knowledge, nonlinear adaptive control techniques and intelligent adaptive techniques. There are several themes in this book which instance both the maturity and the novelty of the general adaptive control. Each chapter is introduced by a brief preamble providing the background and objectives of subject matter. The experiment results are presented in considerable detail in order to facilitate the comprehension of the theoretical development, as well as to increase sensitivity of applications in practical problems

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen