TECHNICAL RESEARCH REPORT

Institute for Systems Research

Controlled Markov Processes on the Infinite Planning Horizon: Weighted and Overtaking Cost Criteria

by E. Fernández-Gaucherand, M.K. Ghosh, and S.I. Marcus

The Institute for Systems Research is supported by the National Science Foundation Engineering Research Center Program (NSFD CD 8803012), Industry and the University

TR 93-6

CONTROLLED MARKOV PROCESSES ON THE INFINITE PLANNING HORIZON: WEIGHTED AND OVERTAKING COST CRITERIA

Emmanuel Fernández-Gaucherand †
Systems & Industrial Engineering Department
The University of Arizona
Tucson, AZ 85721
Email: emmanuel@sie.arizona.edu.

Mrinal K. Ghosh
Department of Mathematics
Indian Institute of Science
Bangalore, 560012
INDIA.

Steven I. Marcus ‡
Institute for Systems Research
Electrical Engineering Department
The University of Maryland
College Park, MD 20742
Email: marcus@src.umd.edu.

Address to which proofs are to be sent: **First author.**

Running Title: **Weighted and Overtaking Cost Criteria.**

Keywords: **Controlled Markov Processes, Infinite Planning Horizon, Weighted and Overtaking Cost Criteria**

---

† Research partially supported by the National Science Foundation under grant NSF-INT 9201430, and by a grant from the AT&T Foundation.

‡ Research partially supported by the Air Force Office of Scientific Research under Grant F49620-92-J-0045, and in part by the National Science Foundation under Grant CDR-8803012.

CONTROLLED MARKOV PROCESSES ON THE INFINITE PLANNING HORIZON: WEIGHTED AND OVERTAKING COST CRITERIA

Emmanuel Fernández-Gaucherand, Mrinal K. Ghosh

and Steven I. Marcus

SUMMARY

Stochastic control problems for controlled Markov processes models with an infinite planning horizon are considered, under some non-standard cost criteria. The classical discounted and average cost criteria can be viewed as complementary, in the sense that the former captures the short-time and the latter the long-time performance of the system. Thus, we study a cost criterion obtained as weighted combinations of these criteria, extending to a general state and control space framework several recent results by Feinberg and Shwartz, and by Krass et al. In addition, a functional characterization is given for overtaking optimal policies, for problems with countable state spaces and compact control spaces; our approach is based on qualitative properties of the optimality equation for problems with an average cost criterion.

1. Introduction

Controlled Markov Processes (CMP), or Markov decision processes, with an infinite planning horizon are a very important class of stochastic sequential decision processes, with numerous applications in many diverse disciplines; see Bertsekas (1987), Ross (1983), Tijms (1986), and White (1985)-(1987)-(1988). Two important performance criteria usually associated with such problems are the total expected discounted cost (DC), and the long-run expected average cost (AC). When a DC criterion is used, and the one-stage cost function is bounded, the corresponding dynamic programming operator exhibits nice contractive properties which enable the development of a rather complete theory, under very general conditions; see Bertsekas/Shreve (1978), Dynkin/Yushkevich (1979), Hinderer (1970), Hernández-Lerma (1989). In this situation, future costs are discounted at a rate $0 < \beta < 1$, and therefore, if β is not sufficiently close to 1, the asymptotic behavior of the state/cost process may not be important at all. Quite the opposite is the case with the AC criterion, under which all decision epochs are given equal weight and one takes the limit of time-averaged expected costs. Therefore, the finite time evolution of the state/cost process is completely irrelevant in this case, and some sort of asymptotic stable behavior is desired. Thus, this case is mathematically much more involved than the DC case (see Arapostathis et al. (1992)).

Therefore, the AC and DC can be seen as two opposite extremes in the spectrum of possible criteria that can be considered, in the sense that the first one captures the performance of the process at the present and near future, and the second captures the performance at the distant future solely. However, situations often arise where past, present and future are all relevant, and thus it is desirable to introduce a cost criterion which offers a reasonable compromise between the above two criteria. One possibility of accomplishing this is by combining these two criteria in a weighted sum. As an application of more general results, a similar approach was studied by Feinberg (1982), who gave existence results for nearly optimal policies, for various weighted cost criteria in a general setting. In addition, problems with weighted discounted criteria have been studied by Feinberg and Shwartz (1992), for models with countable state space.

Recently, Krass et al. (1992) have studied CMP with a weighted cost (WC), given by a convex combination of the AC and DC criteria, the corresponding weights

being selected depending on whether short-term or long-term behavior of the process is to be emphasized. In Krass et al. (1992), attention is restricted to problems with finite state space and finite action set, and with only one cost function common to both the AC and DC criterion. However, many important applications, for which the WC criterion would constitute a very meaningful measure of performance, cannot be modelled within the framework of Krass et al. (1992). For example, in various problems related to stochastic networks both the AC and DC criteria are of interest, as studied by Borkar (1983) and Stidham (1988), and thus the WC criterion presents itself as a way to capture both aspects of the behavior of the process which are individually measured by the former criteria. These problems are naturally formulated in terms of countable state spaces, e.g. queue lengths, and uncountable action spaces, e.g. service rates taking values in an interval in $\mathbb{R}$. In addition, there are many other problems that are best formulated in terms of general (Borel) state spaces (e.g., see Dynkin/Yushkevich (1979), Hernández-Lerma (1989)), for which the WC criterion should prove useful. For example, conditional probability distributions are used as *information* states for problems with incomplete state information; see Arapostathis et al. (1992), and Bertsekas (1987). Also, in water resource management problems, e.g. reservoir operation, the quantities of interest take values in continua (see Yakowitz (1982)), and both DC and AC criteria are of interest. Furthermore, possibly different one-stage cost functions are discounted at different rates to account for, e.g. water release and power generation, and an average criterion is used to account for, e.g. reservoir level regulation and recreational lake uses.

In this paper, we study CMP with a *generalized* WC criterion, with general state and control spaces and with *several* one-stage cost functions being discounted at *different* rates. Thus, we combine and extend to a much more general context many of the results of Feinberg/Shwartz (1992) and Krass et al. (1992)

In many situations it is of interest to look for control policies that induce an adequate undiscounted total cost. Although an optimal policy for the average cost criterion yields minimal growth rate for finite horizon undiscounted total expected costs, this criterion is totally insensitive to the finite time evolution of the process. The use of the *overtaking criterion* is one way to incorporate such sensitivity, while yielding results about the minimal growth rate of the cost flow. Under this criterion, one looks for policies that yield a smallest finite horizon cost, for all horizons

large enough; this type of policy is called *overtaking optimal*, since for horizons large enough, it overtakes the cost due to any other policy. The overtaking cost (OC) criterion was introduced in the economics literature by Gale (1967) and Von Weizsäcker (1965); Leizarowitz (1987) has studied this problem for finite state CMP. In this paper, we study this problem for countable state space and compact action space. Our approach is entirely different from that of Leizarowitz (1987). In this latter reference, the original stochastic control problem is transformed into a deterministic one, under a restrictive controllability assumption. Then using known results from deterministic control systems, the author has shown existence and characterized a stationary deterministic OC optimal policy for the stochastic problem. Our approach follows that of Ghosh/Marcus (1991) and Leizarowitz (1988), and is based on the qualitative properties of the optimality equations for the AC criterion studied by Borkar/Ghosh (1991); see also Arapostathis et al. (1992).

The paper is organized as follows. In section 2 we present the notation used and some preliminaries. The (generalized) WC criterion is treated in sections 3-5. The OC criterion is studied in section 6. We end with some conclusions

2. Notation and Preliminaries

Given a topological space $\mathbf{W}$, its Borel σ-algebra will be denoted by $\mathcal{B}(\mathbf{W})$; measurability will be always understood as Borel measurability henceforth. A (discrete-time) controlled Markov process is a stochastic dynamical system described by the quadruplet $\langle \mathbf{X}, \mathbf{U}, \mathcal{U}, P \rangle$, where the state space $\mathbf{X}$ is a Borel space, i.e. a Borel subset of a complete separable metric space; $\mathbf{U}$ denotes the control or action set, also taken as a Borel space. To each $x \in \mathbf{X}$, a nonempty compact set $\mathcal{U}(x) \in \mathcal{B}(\mathbf{U})$ of admissible actions is associated. Let $\mathbf{K} := \{(x,u) : x \in \mathbf{X}, u \in \mathcal{U}(x)\}$ denote the space of admissible state-action pairs, which is viewed as a topological subspace of $\mathbf{X} \times \mathbf{U}$. The evolution of the system is governed by the *stochastic kernel* P on $\mathbf{X}$ given $\mathbf{K}$, i.e. $P(B|\cdot)$ is a measurable function on $\mathbf{K}$, for each $B \in \mathcal{B}(\mathbf{X})$, and $P(\cdot\,|x,u)$ is a probability measure on $\mathcal{B}(\mathbf{X})$, for each $(x,u) \in \mathbf{K}$.

In addition, to assess the performance of the system, measurable one-stage cost functions $c : \mathbf{K} \to \mathbb{R}$ are chosen. Thus, at time $t \in \mathbb{N}_0 := \{0, 1, 2, \ldots\}$, the system is observed to be in some state, say $x \in \mathbf{X}$, and a decision $u \in \mathcal{U}(x)$ is taken. Then

a cost $c(x,u)$ is incurred, and by the next decision epoch $t+1$, the state of the system will have evolved to some value in $B \in \mathcal{B}(\mathbf{X})$ with probability $P(B|x,u)$. The available information for decision-making at time $t \in \mathbb{N}_0$ is given by the *history* of the process up to that time $h_t := (x_0, u_0, x_1, u_1, \ldots, u_{t-1}, x_t) \in \mathbf{H}_t$, where

$$\mathbf{H}_0 := \mathbf{X}, \qquad \mathbf{H}_t := \mathbf{H}_{t-1} \times (\mathbf{U} \times \mathbf{X}), \qquad \mathbf{H}_\infty := (\mathbf{X} \times \mathbf{U})^\infty,$$

are the *history spaces*. With respect to their corresponding product topologies, the above are Borel spaces; see Bertsekas/Shreve (1978), Hernández-Lerma (1989). An *admissible control policy*, or *strategy*, is a sequence $\pi = \{\pi_t\}_{t \in \mathbb{N}_0}$ of stochastic kernels π_t on $\mathbf{U}$ given $\mathbf{H}_t$, satisfying the constraint $\pi_t(\mathcal{U}(x) \,|\, h_t) = 1$, for all $h_t = (h_{t-1}, u, x) \in \mathbf{H}_t$. The set of all admissible policies will be denoted by $\mathbf{\Pi}$. A policy $\pi \in \mathbf{\Pi}$ is called a *Markov randomized* policy if $\pi_t(\cdot \,|\, (h_{t-1}, u, x)) = \pi_t(\cdot \,|\, (\overline{h}_{t-1}, \overline{u}, x))$, for all $h_{t-1}, \overline{h}_{t-1} \in \mathbf{H}_{t-1}$, $u, \overline{u} \in \mathbf{U}$, and $t \in \mathbb{N}$. Thus, a Markov randomized policy only depends on the "current" state of the process, and hence we will simply write, e.g. $\pi_t(\cdot \,|\, x)$. The set of all Markov randomized policies is denoted as $\mathbf{\Pi}_{MR}$. The policies $\pi \in \mathbf{\Pi}_{MR}$ for which $\pi_t(\cdot \,|\, x) = \pi_\ell(\cdot \,|\, x)$, for all $t, \ell \in \mathbb{N}_0$, are called *stationary randomized* policies; we will simply write, e.g. $\pi(\cdot \,|\, x)$ for these policies, and set of all such policies is denoted as $\mathbf{\Pi}_{SR}$. Furthermore, if given $\pi \in \mathbf{\Pi}_{SR}$ there exists a measurable (decision) function $f : \mathbf{X} \to \mathbf{U}$ such that $f(x) \in \mathcal{U}(x)$, for all $x \in \mathbf{X}$, and $\pi(\{f(x)\} \,|\, x) = 1$, then π is said to be a *stationary deterministic* policy, and we simply write $\pi(x)$ for the action chosen by such a policy at $x \in \mathbf{X}$. The set of all stationary deterministic policies is denoted as $\mathbf{\Pi}_{SD}$. Similarly, Markov deterministic policies are defined in the obvious way, and its set denoted as $\mathbf{\Pi}_{MD}$. Note that $\mathbf{\Pi}_{SD} \subseteq \mathbf{\Pi}_{SR} \subseteq \mathbf{\Pi}_{MR} \subseteq \mathbf{\Pi}$, and $\mathbf{\Pi}_{SD} \subseteq \mathbf{\Pi}_{MD} \subseteq \mathbf{\Pi}_{MR} \subseteq \mathbf{\Pi}$.

Given the distribution μ of the initial state, and a policy $\pi \in \mathbf{\Pi}$, the corresponding state, control and history processes, $\{X_t\}$, $\{U_t\}$ and $\{H_t\}$ respectively, are random processes defined on the canonical probability space $(\mathbf{H}_\infty, \mathcal{B}(\mathbf{H}_\infty), \mathcal{P}_\mu^\pi)$ via the projections $X_t(h_\infty) := x_t$, $U_t(h_\infty) := u_t$ and $H_t(h_\infty) := h_t$, for each $h_\infty = (x_0, u_0, \ldots, x_t, u_t, \ldots) \in \mathbf{H}_\infty$, where $\mathcal{P}_\mu^\pi$ is uniquely determined; see Arapostathis et al. (1992), Bertsekas/Shreve (1978), Hinderer (1970), Hernández-Lerma (1989). The corresponding expectation operator is denoted by $\mathbb{E}_\mu^\pi$. When μ is concentrated at a point $x \in \mathbf{X}$, we simply write, e.g. $\mathbb{E}_x^\pi$.

For a measurable function $v : \mathbf{W} \to \mathbb{R}$, where $\mathbf{W}$ is a topological space, we define

$$\|v\| := \sup_{w \in \mathbf{W}} \{|v(w)|\}.$$

Correspondingly, the vector space of bounded, measurable functions $v : \mathbf{W} \to \mathbb{R}$ is denoted by

$$\mathcal{M}_b(\mathbf{W}) := \{v : \mathbf{W} \to \mathbb{R} \mid v \text{ is measurable, } \|v\| < \infty\}.$$

Hence for $v \in \mathcal{M}_b(\mathbf{W})$, $\|v\|$ gives the *supremum norm.* Also, $\mathcal{L}(\mathbf{W})$ will denote the collection of lower semicontinuous bounded below functions $f : \mathbf{W} \to \mathbb{R}$, and $\mathcal{L}_b(\mathbf{W}) := \mathcal{L}(\mathbf{W}) \cap \mathcal{M}_b(\mathbf{W})$.

3. The Generalized Weighted Cost Criterion

The following two assumptions will be used subsequently, and are in effect throughout, the second of which is made to guarantee the existence of "measurable selectors;" see Arapostathis et al. (1992), Section 7.5 in Bertsekas /Shreve (1978), and Rieder (1978).

Assumption 3.1: There exists $M \in \mathbb{R}$ such that $|c(x,u)| \leq M$, for all $(x,u) \in \mathbf{K}$.

Assumption 3.2: For each $x \in \mathbf{X}$, $\mathcal{U}(x)$ is a nonempty compact subset of $\mathbf{U}$, and $\mathbf{K}$ is a Borel subset of $\mathbf{X} \times \mathbf{U}$; also $c(x,\cdot) \in \mathcal{L}_b(\mathcal{U}(x))$, and $\int f(y)P(dy \mid x,\cdot) \in \mathcal{L}(\mathcal{U}(x))$ for each $f(\cdot) \in \mathcal{L}(\mathbf{X})$.

Remark 3.1: If for all $x \in \mathbf{X}$, $\mathcal{U}(x)$ is a finite set, then the semicontinuity conditions in Assumption 3.2 are trivially satisfied. In addition, the condition that $\int f(y)P(dy \mid x,\cdot) \in \mathcal{L}(\mathcal{U}(x))$, for each $f(\cdot) \in \mathcal{L}(\mathbf{X})$, is equivalent to P being *weakly continuous*, i.e. $\int g(y)P(dy \mid x,\cdot)$ is a continuous function on $\mathcal{U}(x)$, for all continuous and bounded functions $g : \mathbf{X} \to \mathbb{R}$ (see Dynkin/Yushkevich (1979), p. 52). Also, the assumption that $\mathbf{K} \in \mathcal{B}(\mathbf{X} \times \mathbf{U})$ is equivalent to the *multifunction* $x \mapsto \mathcal{U}(x)$ being measurable; see Arapostathis et al. (1992), Rieder (1978), and references therein.

The following are commonly used criteria to measure the cost incurred by using a policy $\pi \in \mathbf{\Pi}$, when the initial state of the system is $x \in \mathbf{X}$.

Finite Horizon Total Cost (FC): Let $T \in \mathbb{N}_0$; for a policy $\pi \in \mathbf{\Pi}$, the total cost for the finite horizon T is given by

$$J_T(x,\pi) := \mathbb{E}_x^\pi \left[\sum_{t=0}^{T} c(X_t, U_t) \right]. \tag{3.1}$$

Infinite Horizon Total Cost (TC): The total cost incurred by $\pi \in \mathbf{\Pi}$ over the entire planning horizon is given by

$$T(x,\pi) := \mathbb{E}_x^\pi \Big[\sum_{t=0}^{\infty} c(X_t, U_t) \Big]. \tag{3.2}$$

Discounted Cost (DC): For a discount factor $0 < \beta < 1$, the DC incurred by $\pi \in \mathbf{\Pi}$ is given by

$$J_\beta(x,\pi) := \lim_{n\to\infty} \mathbb{E}_x^\pi \left[\sum_{t=0}^{n} \beta^t c(X_t, U_t) \right], \tag{3.3}$$

and the optimal β-discounted *value function* is defined as

$$J_\beta^*(x) := \inf_{\pi\in\mathbf{\Pi}} \big\{ J_\beta(x,\pi) \big\}. \tag{3.4}$$

Note that, due to Assumption 3.1, we have that

$$-\frac{M}{1-\beta} \le J_\beta(x,\pi) \le \frac{M}{1-\beta}, \quad \forall x \in X, \quad \forall \pi \in \mathbf{\Pi}. \tag{3.5}$$

If, given $x \in \mathbf{X}$ and $\varepsilon > 0$, a policy π is such that $J_\beta(x,\pi) \le J_\beta^*(x) + \varepsilon$, then π is said to be DC ε-x-optimal; if π is DC ε-x-optimal for all $x \in \mathbf{X}$, it is simply called DC ε-optimal, and if furthermore it is DC ε-optimal for all $\varepsilon > 0$, it is said to be DC optimal. Similar terminology will be used for other criteria.

Average Cost (AC): The long-run expected AC incurred by $\pi \in \mathbf{\Pi}$ is given by

$$J(x,\pi) := \limsup_{n\to\infty} \frac{1}{N+1} \mathbb{E}_x^\pi \Big[\sum_{t=0}^{N} c(X_t, U_t) \Big], \tag{3.6}$$

and the optimal AC is defined as

$$J^*(x) := \inf_{\pi\in\mathbf{\Pi}} \big\{ J(x,\pi) \big\}. \tag{3.7}$$

Remark 3.2: In view of Assumption 3.1, with no loss in generality, costs may be taken as nonnegative when considering either the DC or AC criterion. Furthermore, when simultaneously considering different one-stage cost functions in the sequel, explicit notation will be used as needed, e.g. $J_\beta(x,\pi,c)$.

As mentioned before, the AC criterion gives a measure of the long-run performance of the state/cost process $\{X_t, c(X_t,U_t)\}$, but completely neglects any finite time behavior of this process. Contrary to this, the initial evolution of the process above is crucial when using a DC criterion, for which the asymptotic behavior is unimportant. On the other hand, the TC criterion gives equal weight to every decision epoch, but in most problems every policy leads to an infinite TC, rendering this criterion useless in order to discriminate among different policies. This, and our previous discussion, motivates the definition of the following criterion; c.f. Feinberg/Shwartz (1992), and Krass et al. (1992).

Generalized Weighted Cost (GWC): Let $K \in \mathbb{N}$ and $0 \le \alpha \le 1$ be given, and let $0 < \beta_K < \beta_{K-1} < \ldots < \beta_2 < \beta_1 < 1$ be given discount factors. In addition, for $k = 1,2,\ldots,K+1$, let $c_k(\cdot,\cdot)$ be given one-stage cost functions, each one satisfying a boundedness condition as in Assumption 3.1. The generalized weighted cost incurred by π is given by

$$W_g(x,\pi) := \alpha(1-\beta_1) J_{\beta_1}(x,\pi,c_d) + (1-\alpha) J(x,\pi,c_{K+1}), \tag{3.8}$$

where, for $t \in \mathbb{N}_0$,

$$c_d(x,u,t) := \sum_{k=1}^{K} \left(\frac{1-\beta_k}{1-\beta_1} \right) \left(\frac{\beta_k}{\beta_1} \right)^t c_k(x,u), \tag{3.9}$$

and the optimal GWC is defined as

$$W_g^*(x) := \inf_{\pi \in \mathbf{\Pi}} \{W_g(x, \pi)\}. \tag{3.10}$$

Remark 3.3: Note that $c_d(x, u, t)$ is (uniformly) bounded in $t \in \mathbb{N}_0$, and

$$J_{\beta_1}(x, \pi, c_d) = \sum_{k=1}^{K} (1 - \beta_k) J_{\beta_k}(x, \pi, c_k), \tag{3.11}$$

and thus the first term on the right-hand side of (3.8) gives a weighted combination of dicounted cost criteria; furthermore the factors $(1 - \beta_k)$ give a DC *per unit time*, and hence the combination of this with $J(x, \pi)$ is more meaningful. Note that $c_d(\cdot, \cdot, \cdot)$ is *nonstationary*, i.e. it depends explicitly on time t. However by considering an augmented state space $\mathbf{X} \times \mathbb{N}_0$, the DC problem using $c_d(\cdot, \cdot, \cdot)$ falls within our original model formulation.

Remark 3.4: When $\alpha = 0$, we recover the criterion in Feinberg/Shwartz (1992); Feinberg (1982) showed, as an application of more general results, that for a given initial state distribution and policy $\pi \in \mathbf{\Pi}$, there is a policy $\pi' \in \mathbf{\Pi}_{MD}$ yielding the same or better GWC performance. When $K = 1$ and $c_1(\cdot, \cdot) = c_2(\cdot, \cdot)$, we recover the criterion in Krass et al. (1992). Moreover, for our results on the GWC criterion in the sequel, $c_{K+1}(\cdot, \cdot)$ *need not* be bounded above; the corresponding AC stochastic control problem with unbounded cost function could then be analized as in Hernández-Lerma/Lasserre (1990), Ritt/Sennott (1992), and Schäl (1992). Solely for ease of exposition, we will continue to assume a uniform boundedness condition on all one-stage cost functions.

The weighted criterion can be interpreted from a different perspective as well. A decision-maker may wish to find a policy which simultaneously minimizes DC and AC criteria. One can give concrete examples (see Feinberg/Shwartz (1992) and Krass et al. (1992)) to show that such a desire is utopian and will not be realized in most situations. Therefore the least the decision maker should look for is a Pareto-optimal solution taking these criteria as multiple objectives. Such a solution will be

unimprovable in the sense that one cannot have strictly better performance relative to both criteria. An optimal policy for the GWC is obviously Pareto-optimal in the above sense. Thus the GWC optimal solutions will give a family of Pareto-optimal solutions, parameterized by α, in the above multi-objective optimization problem; see Ghosh (1990) for relalted work.

Naturally, in order to study the GWC criterion, we must first establish the basic results concerning the AC and DC criteria, which we do next. Let $c(\cdot,\cdot)$ be a (generic) one-stage cost function; the *undiscounted* dynamic programming map $T : \mathcal{L}(\mathbf{X}) \to \mathcal{L}(\mathbf{X})$, is defined as

$$T(f)(x) := \inf_{u \in \mathcal{U}(x)} \left\{ c(x,u) + \int_{\mathbf{X}} f(y) P(dy|x,u) \right\}, \quad \forall x \in \mathbf{X}, \tag{3.12}$$

and for $0 < \beta < 1$,the *discounted* dynamic programming map $T_\beta : \mathcal{L}(\mathbf{X}) \to \mathcal{L}(\mathbf{X})$ is given as

$$T_\beta(f) := T(\beta f). \tag{3.13}$$

These maps, as well as their iterates, are well defined; see Arapostathis et al. (1992), and Bertsekas/Shreve (1978). The following is a well known result; see Arapostathis et al. (1992), Bertsekas/Shreve (1978), Dynkin/Yushkevich (1979), Hernández-Lerma (1989).

Theorem 3.1: Under Assumptions 3.1 and 3.2, we have that:
(i) The *discounted cost optimality equation (DCOE)* holds,

$$\begin{aligned} J^*_\beta(x) &= \inf_{u \in \mathcal{U}(x)} \left\{ c(x,u) + \beta \int_{\mathbf{X}} J^*_\beta(y) P(dy|x,u) \right\} \\ &= T_\beta(J^*_\beta)(x), \quad \forall x \in \mathbf{X}. \end{aligned} \tag{3.14}$$

(ii) Furthermore, a stationary policy $\pi \in \mathbf{\Pi}_{SD}$ is DC optimal *if and only if* $\pi(x)$ attains the infimum in (3.14), for all $x \in \mathbf{X}$, i.e. π is a measurable selector in (3.14), and at least one such policy exists;
(iii) Also, J^*_β is the unique fixed point of T_β in $\mathcal{L}_b(\mathbf{X})$. □

Remark 3.5: Note that when the one-stage cost function is nonstationary, e.g. $c_d(\cdot,\cdot,\cdot)$ in (3.9), then the resulting optimal policy in Theorem 3.1 is Markov (and deterministic).

If there are measurable real-valued functions ρ and h on $\mathbf{X}$, with $h \in \mathcal{L}(\mathbf{X})$, such that

$$\begin{aligned} \rho(x) + h(x) &= \inf_{u \in \mathcal{U}(x)} \left\{ c(x,u) + \int_{\mathbf{X}} h(y) P(dy|x,u) \right\} \\ &= T(h)(x), \qquad \forall x \in \mathbf{X}, \end{aligned} \tag{3.15}$$

then the pair (ρ, h) is said to be a solution to the *average cost optimality equation* (ACOE); see Arapostathis et al. (1992). The interest in (3.15) derives from the following result.

Theorem 3.2: Suppose that (ρ, h) is a solution to the ACOE, and that for each admissible policy $\pi \in \mathbf{\Pi}$ the following holds:

$$\lim_{t \to \infty} \mathbb{E}_x^\pi \left[\frac{h(X_t)}{t} \right] = 0, \qquad \forall x \in \mathbf{X}. \tag{3.16}$$

Then (i)

$$\limsup_{N \to \infty} \frac{1}{N+1} \mathbb{E}_x^\pi \left[\sum_{t=0}^{N} \rho(X_t) \right] \leq J(x, \pi), \tag{3.17}$$

and if $\pi \in \mathbf{\Pi}_{SD}$ is such that $\pi(x)$ attains the infimum in (3.15), equality is attained in (3.17);

(ii) If $\rho(x) = \rho^* \in \mathbb{R}$, for all $x \in \mathbf{X}$, then $\rho^* = J^*(x)$, for all $x \in \mathbf{X}$, any $\pi^* \in \mathbf{\Pi}_{SD}$ such that $\pi^*(x)$ attains the infimum in (3.15) is AC optimal, and one such policy exists. □

The proof of Theorem 3.2 is a simple extension of, e.g. Theorem 2.2 in Hernández-Lerma (1989), and will not be given here. Note that (3.17) above says that if $\rho(\cdot)$ is taken as the one-stage cost function for the CMP $\langle \mathbf{X}, \mathbf{U}, \mathcal{U}, P \rangle$ then, *for any* $\pi \in \mathbf{\Pi}$, the average cost assessed under the cost function $\rho(\cdot)$ does not exceed that under cost function $c(\cdot,\cdot)$. Given the results above, naturally there has

been considerable interest in finding conditions which guarantee the existence of a *bounded* solution (ρ^*, h) to the ACOE, with $\rho^* \in \mathbb{R}$ and $h \in \mathcal{L}_b(\mathbf{X})$, for then (3.16) is satisfied trivially, and (ii) applies, see Arapostathis et al. (1992). Given a bounded solution (ρ^*, h) to the ACOE, properties of policies $\pi^* \in \mathbf{\Pi}_{SD}$ attaining the infimum have been investigated by Yushkevich (1973) (see also Arapostathis et al. (1992)), and it has also been shown by Fernández-Gaucherand et al. (1990) that a boundedness condition, uniform in the discount factor, of differences of discounted value functions is a *necessary* condition for a bounded solution to the ACOE to exist.

4. GWC ε-Optimal Policies

In general, AC optimal policies need not exist (see Ross (1983)). However, for CMP with finite state and finite action spaces, there exist policies in $\mathbf{\Pi}_{SD}$ which are optimal for the AC criterion, and the same is true for the DC criterion; see Derman (1970) and Ross (1983). Furthermore, for this situation, Blackwell optimal policies also exist, i.e. policies in $\mathbf{\Pi}_{SD}$ which are both AC optimal and DC optimal, for all β in a neighborhood of 1. However, even for this simplest of situations, Feinberg and Shwartz, and Krass et al. have given examples which show that: (i) GWC optimal policies need not exist (see Example 2 in Krass et al. (1992)]), and (ii) the strict inequalities

$$\inf_{\pi \in \mathbf{\Pi}_{SD}} \{W_g(x,\pi)\} > \inf_{\pi \in \mathbf{\Pi}_{SR}} \{W_g(x,\pi)\} > W_g^*(x), \tag{4.1}$$

may hold (see Example 2 in Krass et al. (1992)). Thus the simple structure of optimal policies and, as shown in the sequel, of optimality equations is not inherited by the weighted problem from the standard problems. From (3.8) and (3.10), we clearly have

$$W_g^*(x) \geq \alpha(1-\beta_1)J_{\beta_1}^*(x, c_d) + (1-\alpha)J^*(x, c_{K+1}), \quad \forall x \in \mathbf{X}. \tag{4.2}$$

The term on the right in (4.2) is, in general, an ideal yet unachievable level of performance, as mentioned above; for this reason, Krass et al. (1992) call this

term the *utopian* lower bound. As a direct consequence of the definition of $W_g^*(x)$, it follows that there exist ε-x-optimal policies, for each $x \in \mathbf{X}$, $0 \leq \alpha \leq 1$, and $0 < \beta < 1$. To expand on this result, we will use the following assumptions.

Assumption 4.1: There exists an AC optimal policy $\pi_a^* \in \mathbf{\Pi}_{SD}$.

Assumption 4.2: There exists a constant $\rho^* \in \mathbb{R}$ such that $J^*(x, c_{K+1}) = \rho^*$, for all $x \in \mathbf{X}$.

Remark 4.1: By Theorem 3.2(ii), if there is a bounded solution (ρ^*, h) to the ACOE with $\rho^* \in \mathbb{R}$ and $h \in \mathcal{L}_b(\mathbf{X})$, then Assumptions 4.1 and 4.2 are satisfied; see Arapostathis et al. (1992) and references therein.

Theorem 4.1: Let $K \in \mathbb{N}$ and $0 \leq \alpha \leq 1$ be given, and let $0 < \beta_K < \beta_{K-1} < \ldots < \beta_2 < \beta_1 < 1$ be given discount factors. In addition, for $k = 1, 2, \ldots, K+1$, let $c_k(\cdot, \cdot)$ be given one-stage cost functions, each satisfying a boundedness condition as in Assumption 3.1. If Assumption 4.1 holds, then
(i) For each $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbb{N}$, such that for all $x \in \mathbf{X}$ there is a GWC ε-x-optimal policy $\pi^{\varepsilon,x}$, with $\pi_t^{\varepsilon,x} = \pi_a^*$, for all $t \geq N(\varepsilon)$.

If, in addition, Assumption 4.2 holds, then

(ii) For each $\varepsilon > 0$ there exists $N(\varepsilon) \in \mathbb{N}$, such that if $\pi_d^* \in \mathbf{\Pi}_{MD}$ is a DC optimal policy for the one-stage cost function $c_d(\cdot, \cdot, \cdot)$ of (3.9), then for any $N \geq N(\varepsilon)$, the policy π^ε given by

$$\pi_t^\varepsilon := \begin{cases} \pi_{d,t}^*, & \text{if } 0 \leq t < N, \\ \pi_a^*, & \text{if } N \leq t, \end{cases}$$

is GWC ε-optimal.
(iii) The utopian lower bound is attained, i.e.

$$W_g^*(x) = \alpha(1-\beta_1) J_{\beta_1}^*(x, c_d) + (1-\alpha)\rho^*, \qquad \forall x \in \mathbf{X}.$$

Proof: (i) We need to consider only the cases when $0 < \alpha$. Let $\varepsilon > 0$ and $x \in \mathbf{X}$ be given, and without loss of generality, assume that $0 \leq c_d(x, u, t), c_{K+1}(x, u) \leq M$,

for all $(x,u) \in \mathbf{K}$ and all $t \in \mathbb{N}$ (see Remark 3.2). Let $\overline{\pi} \in \mathbf{\Pi}$ be a GWC $(\varepsilon/2)$-x-optimal policy. By the boundedness assumption, there exists an $N(\varepsilon) \in \mathbb{N}$ such that

$$0 \leq \beta_1^t M \leq \frac{\varepsilon}{2\alpha}, \qquad \forall t \geq N(\varepsilon).$$

Let $N \geq N(\varepsilon)$, then for any $\pi \in \mathbf{\Pi}$:

$$\begin{aligned} J_{\beta_1}(x,\pi,c_d) &= \mathbb{E}_x^{\pi}\left\{\sum_{t=0}^{\infty}\beta_1^t c_d(X_t,U_t,t)\right\} \\ &\leq \mathbb{E}_x^{\pi}\left\{\sum_{t=0}^{N-1}\beta_1^t c_d(X_t,U_t,t)\right\} + \beta_1^N\frac{M}{1-\beta_1} \\ &\leq \mathbb{E}_x^{\pi}\left\{\sum_{t=0}^{N-1}\beta_1^t c_d(X_t,U_t,t)\right\} + \frac{\varepsilon}{2\alpha(1-\beta_1)}. \end{aligned}$$

Now, define a policy $\pi^{\varepsilon,x}$ as follows:

$$\pi_t^{\varepsilon,x} := \begin{cases} \overline{\pi}_t, & \text{if } 0 \leq t < N; \\ \pi_a^*, & \text{if } N \leq t. \end{cases}$$

Therefore,

$$\begin{aligned} J_{\beta_1}(x,\pi^{\varepsilon,x},c_d) &\leq \mathbb{E}_x^{\overline{\pi}}\left\{\sum_{t=0}^{N-1}\beta_1^t c_d(X_t,U_t,t)\right\} + \frac{\varepsilon}{2\alpha(1-\beta_1)} \\ &\leq J_{\beta_1}(x,\overline{\pi},c_d) + \frac{\varepsilon}{2\alpha(1-\beta_1)}, \end{aligned} \tag{4.3}$$

and since $J(y,\pi_a^*,c_{K+1}) = J^*(y,c_{K+1})$, for all $y \in \mathbf{X}$, then

$$J(x,\pi^{\varepsilon,x},c_{K+1}) \leq J(x,\overline{\pi},c_{K+1}). \tag{4.4}$$

Therefore,

$$\begin{aligned} W_g(x,\pi^{\varepsilon,x}) &= \alpha(1-\beta_1)J_{\beta_1}(x,\pi^{\varepsilon,x},c_d) + (1-\alpha)J(x,\pi^{\varepsilon,x},c_{K+1}) \\ &\leq \alpha(1-\beta_1)J_{\beta_1}(x,\overline{\pi},c_d) + (1-\alpha)J(x,\overline{\pi},c_{K+1}) + \frac{\varepsilon}{2} \\ &= W_g(x,\overline{\pi}) + \frac{\varepsilon}{2} \\ &\leq W_g^*(x) + \varepsilon, \end{aligned}$$

where the first inequality is obtained from (4.3) and (4.4), and the second inequality follows from $\overline{\pi}$ being GWC $(\varepsilon/2)$-x-optimal.

(ii) Let $N(\varepsilon) \in \mathbb{N}$ be such that

$$\beta_1^t M \leq \frac{\varepsilon}{\alpha}, \qquad \forall t \geq N(\varepsilon),$$

and let π^ε be as in the statement of the Theorem. Similarly as before, we have that

$$J_{\beta_1}(x, \pi^\varepsilon, c_d) \leq J^*_{\beta_1}(x, c_d) + \frac{\varepsilon}{\alpha(1-\beta_1)}, \qquad \forall x \in \mathbf{X}, \tag{4.5}$$

and since $J^*(x, c_{K+1}) = \rho^*$, for all $x \in \mathbf{X}$, then

$$J(x, \pi^\varepsilon, c_{K+1}) = \rho^*, \qquad \forall x \in \mathbf{X}. \tag{4.6}$$

Therefore, for each $x \in \mathbf{X}$,

$$\begin{aligned} W_g(x, \pi^\varepsilon) &\leq \alpha(1-\beta_1) J^*_{\beta_1}(x, c_d) + (1-\alpha)\rho^* + \varepsilon \\ &\leq W_g^*(x) + \varepsilon, \end{aligned} \tag{4.7}$$

where (4.2), (4.5) and (4.6) have been used. Thus π^ε is GWC ε-optimal.

(iii) From (4.2) and (4.7) we have that, for each $x \in \mathbf{X}$,

$$\begin{aligned} \alpha(1-\beta_1) J^*_{\beta_1}(x, c_d) + (1-\alpha)\rho^* &\leq W_g(x, \pi^\varepsilon) \\ &\leq \alpha(1-\beta_1) J^*_{\beta_1}(x, c_d) + (1-\alpha)\rho^* + \varepsilon, \end{aligned}$$

from which the result follows. □

Remark 4.2: The structure of the ε-optimal policy obtained in (ii) above is intuitively appealing: use the best policy for the weighted DC criteria long enough at the beginning, and then switch to the policy which is best in the long run. Likewise, (i) above simply says to eventually use the policy which is best in the long run.

5. WC ε-Optimality Equations

In this section, we consider only a simpler weighted criterion, as in Krass et al. (1992). To this effect, we consider the simpler situation when $K = 1$, and $c_1(\cdot,\cdot) = c_2(\cdot,\cdot)$, and will denote the resulting WC criterion as $W_{\alpha,\beta}(x,\pi)$, making explicit the values of the discount and weight factors. From the definition of the AC, DC, and WC criteria, given $x \in \mathbf{X}$ and $\pi \in \mathbf{\Pi}$, it is easily seen that $\{\mathbb{E}_x^\pi[c(X_t,U_t)]\}$ determines the corresponding costs, e.g.,

$$W_{\alpha,\beta}(x,\pi) = \limsup_{N\to\infty} \sum_{t=0}^{N} \Big\{\alpha(1-\beta)\beta^t + \frac{1-\alpha}{N+1}\Big\}\mathbb{E}_x^\pi\big[c(X_t,U_t)\big].$$

For cost criteria of this type, it can be shown that $\mathbf{\Pi}_{MR}$ is *sufficiently rich*, in the sense that, given *any* $\pi \in \mathbf{\Pi}$, a policy $\pi' \in \mathbf{\Pi}_{MR}$ can be found such that

$$\mathbb{E}_x^\pi\big[c(X_t,U_t)\big] = \mathbb{E}_x^{\pi'}\big[c(X_t,U_t)\big], \tag{5.1}$$

for each $t \in \mathbb{N}_0$. This result follows from an extension of a theorem by Derman and Strauch (1966), and is based on the following fact: given $\pi \in \mathbf{\Pi}$, the probability measure $\nu_t[\pi]$ on $\mathcal{B}(\mathbf{K})$ defined as

$$\nu_t[\pi](\mathcal{K}) := \mathcal{P}_x^\pi\{(X_t,U_t) \in \mathcal{K}\}, \qquad \mathcal{K} \in \mathcal{B}(\mathbf{K}),$$

can be decomposed as

$$\nu_t[\pi](dx,du) = \overline{\nu}_t[\pi](dx)m_t[\pi](du|x),$$

where $\overline{\nu}_t[\pi]$ is the marginal of $\nu_t[\pi]$ on $\mathbf{X}$, and $m_t[\pi]$ is a measurable stochastic kernel on $\mathbf{U}$ given $\mathbf{X}$, satisfying $m_t[\pi](\mathcal{U}(x)|x) = 1$ (see Dynkin/Yushkevich (1979), pp. 88-89). Then, for $\pi' := \{m_0[\pi], m_1[\pi], \ldots\} \in \mathbf{\Pi}_{MR}$, (5.1) is satisfied (see Dynkin/Yushkevich (1979), pp. 96-97 for more details). We thus conclude that when considering any cost criterion, the value of which is determined by $\{\mathbb{E}_x^\pi[c(X_t,U_t)]\}$, like the AC, DC, and WC criterion, then the largest set of policies that needs to be considered is $\mathbf{\Pi}_{MR}$.

We will need the following notation: given $\pi = (\pi_0, \pi_1, \pi_2, \ldots) \in \mathbf{\Pi}_{MR}$, we let $\pi^- := (\pi_1, \pi_2, \pi_3, \ldots) \in \mathbf{\Pi}_{MR}$, and for $\pi^* \in \mathbf{\Pi}_{SR}$, we let $\pi^* \cdot \pi := (\pi^*, \pi_0, \pi_1, \ldots) \in \mathbf{\Pi}_{MR}$. For α_1, α_2 nonnegative numbers, such that $\alpha_1 + \alpha_2 > 0$, we write

$$W(x, \pi; \alpha_1, \alpha_2, \beta) := \alpha_1(1-\beta)J_\beta(x, \pi) + \alpha_2 J(x, \pi), \qquad \forall x \in \mathbf{X}.$$

Note that

$$W(x, \pi; \alpha_1, \alpha_2, \beta) = (\alpha_1 + \alpha_2)W_{\alpha,\beta}(x, \pi),$$

where $\alpha = \alpha_1/(\alpha_1 + \alpha_2)$. Also, $W^*(x, \pi; \alpha_1, \alpha_2, \beta)$ will denote that corresponding infimum over $\mathbf{\Pi}_{MR}$.

Theorem 5.1: Let $\pi \in \mathbf{\Pi}_{MR}$. Then for each $x \in \mathbf{X}$, we have that:

(i) $$W(x, \pi; \alpha_1, \alpha_2, \beta) = \alpha_1(1-\beta)\int_{\mathbf{U}} c(x, u)\pi_0(du \,|\, x) + \int_{\mathbf{X}}\int_{\mathbf{U}} W(y, \pi^-; \beta\alpha_1, \alpha_2, \beta)P(dy \,|\, x, u)\pi_0(du \,|\, x). \quad (5.2)$$

Also, if Assumption 4.2 holds, then:
(ii) There exists $\pi^* \in \mathbf{\Pi}_{SD}$, such that $\pi^*(x)$ attains

$$\inf_{u \in \mathcal{U}(x)} \left\{ \alpha_1(1-\beta)c(x, u) + \int_{\mathbf{X}} W(y, \pi; \beta\alpha_1, \alpha_2, \beta)P(dy \,|\, x, u) \right\};$$

(iii) Furthermore, if for some $\varepsilon > 0$

$$W(x, \pi; \beta\alpha_1, \alpha_2, \beta) \leq W^*(x; \beta\alpha_1, \alpha_2, \beta) + \varepsilon,$$

for all $x \in \mathbf{X}$, then

$$W(x, \pi^* \cdot \pi; \alpha_1, \alpha_2, \beta) \leq W^*(x; \alpha_1, \alpha_2, \beta) + \varepsilon,$$

for all $x \in \mathbf{X}$, i.e. $\pi^* \cdot \pi$ is WC ε-optimal.

Proof: (i) Let $x \in \mathbf{X}$; then by Assumption 3.1, Fatou's lemma and the dominated convergence theorem, we obtain

$$J_\beta(x,\pi) = \int_{\mathbf{U}} c(x,u)\pi_0(du \,|\, x) + \beta \int_{\mathbf{X}} \int_{\mathbf{U}} J_\beta(y,\pi^-)P(dy \,|\, x,u)\pi_0(du \,|\, x),$$

and

$$J(x,\pi) = \int_{\mathbf{X}} \int_{\mathbf{U}} J(y,\pi^-)P(dy \,|\, x,u)\pi_0(du \,|\, x).$$

Hence,

$$\begin{aligned} W(x,\pi;\alpha_1,\alpha_2,\beta) &= \alpha_1(1-\beta)\int_{\mathbf{U}} c(x,u)\pi_0(du \,|\, x) + \int_{\mathbf{X}} \int_{\mathbf{U}} \Big[\beta\alpha_1(1-\beta)J_\beta(y,\pi^-) \\ &\qquad + \alpha_2 J(y,\pi^-)\Big] P(dy \,|\, x,u)\pi_0(du \,|\, x) \\ &= \alpha_1(1-\beta)\int_{\mathbf{U}} c(x,u)\pi_0(du \,|\, x) \\ &\qquad + \int_{\mathbf{X}} \int_{\mathbf{U}} W(y,\pi^-;\beta\alpha_1,\alpha_2,\beta)P(dy \,|\, x,u)\pi_0(du \,|\, x). \end{aligned}$$

(ii) If for each $x \in \mathbf{X}$, $\mathcal{U}(x)$ is a finite set, then the result follows trivially. In more generality, we note that $J_\beta(\cdot\,,\pi) \in \mathcal{L}_b(\mathbf{X})$, for any $\pi \in \mathbf{\Pi}$, due to our Assumption 3.2 (ii) (see Chapter 9 in Bertsekas/Shreve (1978)). Then, by Assumption 4.2, $W(\cdot\,,\pi;\alpha_1,\alpha_2,\beta) \in \mathcal{L}_b(\mathbf{X})$, for any positive α_1,α_2, and $0<\beta<1$. The result then follows by Proposition 7.33 in Bertsekas/Shreve (1978).

(iii) Let $\delta \in \mathbf{\Pi}_{MR}$ be arbitrary, then for each $x \in \mathbf{X}$

$$\begin{aligned} &W(x;\pi^* \cdot \pi;\alpha_1,\alpha_2,\beta) \\ &\quad = \alpha_1(1-\beta)c(x,\pi^*(x)) + \int_{\mathbf{X}} W(y,\pi;\beta\alpha_1,\alpha_2,\beta)P(dy \,|\, x,\pi^*(x)) \\ &\quad = \inf_{u\in\mathcal{U}(x)} \left\{\alpha_1(1-\beta)c(x,u) + \int_{\mathbf{X}} W(y,\pi;\beta\alpha_1,\alpha_2,\beta)P(dy \,|\, x,u)\right\} \\ &\quad \le \inf_{u\in\mathcal{U}(x)} \left\{\alpha_1(1-\beta)c(x,u) + \int_{\mathbf{X}} W(y,\delta^-;\beta\alpha_1,\alpha_2,\beta)P(dy \,|\, x,u)\right\} + \varepsilon \end{aligned}$$

$$\leq \alpha_1(1-\beta)\int_{\mathbf{U}} c(x,u)\delta_0(du\,|\,x)$$

$$+\int_{\mathbf{X}}\int_{\mathbf{U}} W(y,\delta^-;\beta\alpha_1,\alpha_2,\beta)P(dy\,|\,x,u)\delta_0(du\,|\,x)+\varepsilon$$

$$= W(x,\delta;\alpha_1,\alpha_2,\beta)+\varepsilon, \tag{5.3}$$

by (i) above, the definition of π^*, and the ε-optimality of π. Since $\delta \in \mathbf{\Pi}_{MR}$ was arbitrary, and by the sufficiency of $\mathbf{\Pi}_{MR}$, the result follows. □

Remark 5.1: Note that (5.2) resembles the standard (Bellman's) equation satisfied by the value function $J_\beta(\cdot\,,\pi)$; see Bertsekas (1987), Bertsekas/Shreve (1978). However, note that the DC weight is α_1 on the expression on the left, but it is $\beta\alpha_1$ on the right. Also, from (5.3) we see that if $\pi^* \in \mathbf{\Pi}_{SD}$ is such that

$$W(x,\pi^*\cdot\pi;\alpha_1,\alpha_2,\beta)$$

$$= \inf_{u\in\mathcal{U}(x)}\left\{\alpha_1(1-\beta)c(x,u)+\int_{\mathbf{X}} W(y,\pi;\beta\alpha_1,\alpha_2,\beta)P(dy\,|\,x,u)\right\}, \tag{5.4}$$

then $\pi^*\cdot\pi$ inherits the ε-optimality properties of π. Thus (5.4) can be seen as an ε-optimality equation for WC.

6. Overtaking Criterion

Consider now the FC problem; every policy $\pi \in \mathbf{\Pi}$ therefore gives rise to a cost flow $T \mapsto J_T(x,\pi)$. Typically $J_T(x,\pi) \to \infty$, as $T \to \infty$, and therefore one attempts to minimize, e.g. the (limiting) average cost to evaluate the performance of the system under a particular policy. Although an optimal policy for the average cost criterion yields minimal growth rate for $J_T(x,\cdot)$, this criterion is totally insensitive to the finite time evolution of the process. The use of the *overtaking criterion* is another way to incorporate such sensitivity, while yielding results about the minimal growth rate of the cost flow.

We say that a policy $\pi_1 \in \mathbf{\Pi}$ *overtakes* another policy $\pi_2 \in \mathbf{\Pi}$ if

$$J_T(x,\pi_1) \leq J_T(x,\pi_2),$$

for all $x \in \mathbf{X}$, and for all T sufficiently large; note that what is meant by T being "sufficiently large" may depend on the choice of x. A policy is called *overtaking optimal* if it overtakes every other policy. Clearly, overtaking optimality implies average cost optimality, but the converse is not true in general. We take $\mathbf{X} = \mathbb{N}$, $\mathcal{U}(\cdot) = \mathbf{U}$, and work under the following additional assumptions.

Assumption 6.1: The one-stage cost function $c(\cdot, \cdot)$ is bounded and continuous, and $\mathbf{U}$ is compact.

Assumption 6.2: The stochastic kernel (transition law) $P(x' \mid x, \cdot)$ is continuous, for each pair $(x, x') \in \mathbb{N} \times \mathbb{N}$.

Assumption 6.3: Under every $\pi \in \mathbf{\Pi}_{SR}$, the corresponding chain is irreducible and aperiodic.

Remark 6.1: Every OC optimal policy is also AC optimal. Under Assumption 6.3, if the ACOE has a solution (ρ^*, h) satisfying (3.16) and $\rho^* \in \mathbb{R}$, then if a policy $\pi \in \mathbf{\Pi}_{SD}$ is average optimal and the corresponding controlled process $\{X_t\}$ is positive recurrent, equality is attained in the ACOE (3.15); see Arapostathis et al. (1992).

Suppose that (ρ^*, h) is a solution to the ACOE, as in Remark 6.1. Let $\tilde{\mathbf{\Pi}}_{SD} \subseteq \mathbf{\Pi}_{SD}$ be the set of all stationary deterministic policies for which equality is attained in the ACOE. Let $\tilde{\pi} \in \tilde{\mathbf{\Pi}}_{SD}$ and $\pi \in \mathbf{\Pi}$. Then due to Theorem 3.2, $\tilde{\pi}$ overtakes π. Therefore, the search for OC optimal policies can be restricted to $\tilde{\mathbf{\Pi}}_{SD}$. Let $\tilde{\pi} \in \tilde{\mathbf{\Pi}}_{SD}$, then due to Remark 6.1 it follows that

$$J_T(x, \tilde{\pi}) = \rho^* T + h(x) - \mathbb{E}_x^{\tilde{\pi}}\{h(X_T)\}. \tag{6.1}$$

From (6.1), it follows that $\tilde{\pi}$ would be overtaking optimal if $\mathbb{E}_x^{\tilde{\pi}}\{h(X_T)\}$ has the maximal growth rate, as $T \to \infty$. Under a *Lyapunov stability condition* (see Assumption 6.4 below), we will show the existence of a policy $\tilde{\pi} \in \tilde{\mathbf{\Pi}}_{SD}$ as in Remark 6.1, and also having the maximal growth mentioned above.

Assumption 6.4: There exists $w : \mathbf{X} \to \mathbb{R}_+$, a finite set $A \subseteq \mathbf{X}$, and an $\epsilon > 0$ such that:

(i) $0 \in A$, and the set $\{x \in A^c \mid P(y \mid x,u) > 0, \text{ for some } y \in A, u \in \mathbf{U}\}$, is finite.
(ii) $lim_{x\to\infty} w(x) = \infty$.
(iii) Under any $\pi \in \mathbf{\Pi}$, and any $\mu \in \mathbb{P}(\mathbf{X})$

$$\mathbb{E}_\mu^\pi \{[w(X_{t+1}) - w(X_t) + \epsilon]\, \mathbf{1}\{X_t \notin A\} \mid \mathcal{F}_t\} \leq 0, \quad \mathcal{P}_\mu^\pi - a.s.,$$

where $\mathcal{F}_t$ is the σ-algebra generated by the history process $\{H_t\}$ under π, and $\mathbf{1}\{A\}$ denotes the indicator function for the set A.
(iv) There exists a random variable $\mathcal{Z}$ and an scalar $\lambda > 0$ such that $\mathbb{E}\,[exp(\lambda \mathcal{Z})] < \infty$, and for all $b > 0$

$$\mathcal{P}_\mu^\pi \{\mid w(X_{t+1}) - w(X_t) \mid > b \mid \mathcal{F}_t\} \leq P(\mathcal{Z} > b).$$

Under the above Lyapunov condition and Assumption 6.3, all policies $\pi \in \mathbf{\Pi}_{SR}$ are *stable*: the corresponding controlled Markov chain $\{X_t\}$ is positive recurrent; see Arapostathis et al. (1992), Borkar (1991). Let $\pi \in \mathbf{\Pi}_{SD}$, and let $\eta(\pi)$ denote the invariant probability distribution of $\{X_t\}$ under π. The following results are proved by Borkar/Ghosh (1991).

Theorem 6.1: Under Assumptions 6.1-6.3, the following holds:
(i) For any $\pi \in \mathbf{\Pi}_{SD}$

$$\lim_{t\to\infty} \mathbb{E}_x^\pi\{w(X_t)\} = 0, \quad \sum_{x\in\mathbf{X}} w(x)\eta(\pi)(x) < \infty;$$

(ii) The ACOE has a unique solution (ρ^*, h) such that

$$h(0) = 0, \quad h(\cdot) = O(w);$$

(iii) A policy $\pi \in \mathbf{\Pi}_{SD}$ is AC-optimal if and only if it attains equality in the ACOE.

□

The following stability result will play an important role in the sequel.

Lemma 6.1: Let $h(\cdot)$ be as in Theorem 6.1. Then under Assumptions 6.1-6.4, we have that

$$\lim_{t\to\infty} \mathbb{E}_x^\pi\{h(X_t)\} = \sum_{y\in\mathbf{X}} h(y)\eta(\pi)(y) < \infty, \tag{6.2}$$

for all $\pi \in \mathbf{\Pi}_{SR}$, $x \in \mathbf{X}$.

Proof: By (i) in Theorem 6.1,

$$\xi(\pi) := \sum_{y\in\mathbf{X}} w(y)\eta(\pi)(y) < \infty. \tag{6.3}$$

Let $\{t_n\} \subseteq \mathbb{N}$ be a sequence such that $t_n \uparrow \infty$. Let

$$\zeta(\pi, x) := \liminf_{n\to\infty} \mathbb{E}_x^\pi\{w(X_{t_n})\}. \tag{6.4}$$

We claim that $\zeta(\pi, x) = \xi(\pi)$, for all $x \in \mathbf{X}$. By Fatou's Lemma,

$$\xi(\pi) \geq \sum_{y\in\mathbf{X}} \zeta(\pi, y)\eta(\pi)(y). \tag{6.5}$$

Next, let $w_n : \mathbf{X} \to \mathbb{R}_+$ be a sequence of functions with finite support such that $w_n \uparrow w$ pointwise. Then for any $m \geq 1$, we have that

$$\begin{aligned}\liminf_{n\to\infty} \mathbb{E}_x^\pi\{w(X_{t_n})\} &\geq \liminf_{n\to\infty} \mathbb{E}_x^\pi\{w_m(X_{t_n})\} \\ &= \sum_{y\in\mathbf{X}} w_m(y)\eta(\pi)(y).\end{aligned}$$

Letting $m \to \infty$ in the equation above, and using the Monotone Convergence Theorem, it follows that

$$\xi(\pi) \leq \zeta(\pi, x), \forall x \in \mathbf{X}. \tag{6.6}$$

From (6.5), (6.6), and Assumption 6.3 †, it follows that

$$\xi(\pi) = \zeta(\pi, x), \forall x \in \mathbf{X}.$$

Therefore, we have that

$$\lim_{n\to\infty} \mathbb{E}_x^{\pi}\{w(X_t)\} = \sum_{y\in\mathbf{X}} w(y)\eta(\pi)(y). \tag{6.7}$$

Under Assumptions 6.1-6.3, the law of X_t will converge to $\eta(\pi)$ in total variation norm, e.g. see Theorem 3.3, Chapter 1 in Borkar (1991). Now, using the fact that $h(\cdot) = O(w)$, (6.7) and a Generalized Dominated Convergence Theorem as in Royden (1968), p. 232, it follows that

$$\lim_{t\to\infty} \mathbb{E}_x^{\pi}\{h(X_t)\} = \sum_{y\in\mathbf{X}} h(y)\eta(\pi)(y).$$

□

In view of the above result, define a function $\Psi : \tilde{\mathbf{\Pi}}_{SD} \to \mathbb{R}$ as follows:

$$\Psi(\pi) = \sum_{y\in\mathbf{X}} h(y)\eta(\pi)(y). \tag{6.8}$$

Then, the proof of Lemma 3.8 in Borkar/Ghosh (1991) can be easily modified to yield:

$$\sup_{\pi\in\mathbf{\Pi}_{SR}} \sum_{y\in\mathbf{X}} \mid h(y) \mid \eta(\pi)(x) < \infty,$$

and thus,

$$\sup_{\pi\in\tilde{\mathbf{\Pi}}_{SD}} \Psi(\pi) := \Psi^* < \infty, \tag{6.9}$$

and thus we have the following.

Lemma 6.2: There exists a $\pi^* \in \tilde{\mathbf{\Pi}}_{SD}$ such that $\Psi(\pi^*) = \Psi^*$.

† The aperiodicity in Assumption 6.3 is actually not required here.

Proof: Let $\{\pi_n\} \subseteq \mathbf{\Pi_{SD}}$ be a sequence such that $\Psi(\pi_n^*) \uparrow \Psi$. The set $\tilde{\mathbf{\Pi}}_{SD}$ is easily seen to be compact, where $\tilde{\mathbf{\Pi}}_{SD}$ is endowed with the product topology. Thus for some subsequence $\{\pi_{n'}\}$, we have that $\pi_{n'} \to \pi^* \in \tilde{\mathbf{\Pi}}_{SD}$, as $n' \to \infty$; we claim that $\Psi(\pi^*) = \Psi^*$. It can be shown, as in Lemma 3.8 of Borkar/Ghosh (1991), that

$$\sup_{\pi \in \mathbf{\Pi}_{SR}} \sum_{y \in \mathbf{X}} \mid h(y) \mid^2 \eta(\pi)(x) < \infty.$$

Therefore, $h(\cdot)$ is uniformly integrable with respect to the class of measures $\{\eta(\pi) : \pi \in \mathbf{\Pi}_{SR}\}$. Again, since $\pi_{n'} \to \pi^*$, it can be shown (see Chapter 5 in Borkar (1991)) that $\eta(\pi_{n'}) \to \eta(\pi^*)$, in total variation norm. Hence

$$\sum_{y \in \mathbf{X}} h(y)\eta(\pi_n)(y) \underset{n \to \infty}{\longrightarrow} \sum_{y \in \mathbf{X}} h(y)\eta(\pi^*)(y),$$

and therefore we obtain that $\Psi(\pi^*) = \Psi^*$. □

Finally, we show existence of an overtaking optimal policy, and give a functional characterization for this policy.

Theorem 6.2: Let Assumptions 6.1- 6.4 hold. Let $\pi^* \in \tilde{\mathbf{\Pi}}_{SD}$ be such that

$$\Psi(\pi^*) = \Psi^*.$$

Then π^* is overtaking optimal.

Proof: Let $\pi \in \tilde{\mathbf{\Pi}}_{SD}$ be arbitrary. Then using the ACOE, it follows that

$$J_T(\pi, x) = \rho^* T + h(x) - \mathbb{E}_x^{\pi}\{h(X_T)\},$$

where $h(\cdot)$ is as in Theorem 6.1. The desired result then follows by applying Lemmas 6.1 and 6.2. □

Remark 6.2: Here we have worked under a Lyapunov stability condition. We can derive analogous results under a geometric ergodicity condition.

Conclusions

We have studied in this paper CMP with a (generalized) WC criterion, extending (and combining) to our more general setting several results by Feinberg and Shwartz (1992), and by Krass et al. (1992). Even for the case when optimal stationary deterministic policies exist for the AC criterion and and each DC criteria, and the optimal average cost does not depend on the initial state, only ε-optimal policies have been characterized. To give general conditions under which the existence of an optimal GWC stationary deterministic policy can be proved is difficult. This can be intuitively explained due to the complementary properties that such a policy would have to exhibit: it should induce optimal performance both in the short-term and in the long-term performance of the system. Also, the minimization of GWC can also be viewed as a multiobjective problem, as we have observed earlier. The lack of convexity of the feasible domain of policies makes the problem very difficult. If, instead, we have a multiobjective problem where all evaluation criteria are of the same type (e.g., DC or AC), then it can be shown that the feasible domain is convex and the set of all Pareto-optimal solutions can be completely characterized; see Ghosh (1990). Related to this observation is the fact that the AC criterion does not possess certain convexity properties, as studied by Feinberg (1982).

In addition, a functional characterization was given for overtaking optimal policies, for problems with countable state spaces and compact control spaces; our approach is based on qualitative properties of the ACOE, and extends to a much more general setting previous results available in the literature.

References

A. Arapostathis, V.S. Borkar, E. Fernández-Gaucherand, M.K. Ghosh and S.I. Marcus (1992) Discrete-Time controlled Markov processes with an average cost criterion: a survey. Preprint (to appear in SIAM J. Control & Optim.).

D.P. Bertsekas (1987) Dynamic Programming: Deterministic and Stochastic Models. Prentice-Hall, Englewood Cliffs.

D.P. Bertsekas and S.E. Shreve (1978) Stochastic Optimal Control: The Discrete Time Case. Academic Press, New York.

V.S. Borkar (1983) Controlled Markov chains and stochastic networks. SIAM J. Control Optim. 21:652-666.

V.S. Borkar (1991) Topics in Controlled Markov Chains. Pitman Research Notes in Mathematics Series, Longman Scientific & Technical.

V.S. Borkar and M.K. Ghosh (1991) Ergodic and adaptive control of nearest neighbor motions. Math. Control, Signals and Systems 4:81-98.

C. Derman (1970) Finite State Markovian Decision Processes. Academic Press, New York.

C. Derman and R.E. Strauch (1966) A note on memoryless rules for controlling sequential control processes. Ann. Math. Stat. 37:276–278.

E.B. Dynkin and A.A. Yushkevich (1979) Controlled Markov Processes. Springer-Verlag, New York.

E. Fernández-Gaucherand, A. Arapostathis, and S.I. Marcus (1990) Remarks on the existence of solutions to the average cost optimality equation in Markov decision processes. Systems Control Lett. 5:425-432.

E.A. Feinberg (1982) Controlled Markov processes with arbitrary numerical criteria. Theory Prob. Applications 27:486-503.

E.A. Feinberg and A. Shwartz (1992) Markov decision models with weighted discounted criteria. Preprint (to appear in Math. Operat. Res.).

D. Gale (1967) On optimal development in a multisector economy. Rev. Econom. Stud. 34:1-19.

M.K. Ghosh (1990) Markov decision process with multiple costs. Operat. Res. Lett. 9:257-260.

M.K. Ghosh and S.I. Marcus (1991) Infinite horizon controlled diffusion problems with some nonstandard criteria. J. Mathematical Systems and Control 1:45-70.

K. Hinderer (1970) Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameters. Lect. Notes Oper. Res. Math. Syst. #33, Springer-Verlag, Berlin.

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

O. Hernández-Lerma and J.B. Lasserre (1990) Average cost optimal policies for Markov control processes with Borel state space and unbounded costs. Systems Control Lett. 15:349-356.

D. Krass, J.A. Filar and S.S. Sinha (1992) A weighted Markov decision process. Operat. Res. 40:1180-1187.

A. Leizarowitz (1987) Infinite horizon optimization for finite state Markov chain. SIAM J. Control Optim. 25:1601-1618.

A. Leizarowitz (1988) Controlled diffusion processes on infinite horizon with the overtaking criterion. Appl. Math. Optim. 17:61-78.

U. Rieder (1978) Measurable selection theorems for optimization problems. Manuscripta Mathematica 24:115–131.

R.K. Ritt and L.I. Sennott (1992) Optimal stationary policies in general state space Markov decision chains with finite action sets. Math. Operat. Res. 17:901-909.

S.M. Ross (1983) Introduction to Stochastic Dynamic Programming. Academic Press, New York.

H.L. Royden (1968) Real Analysis, 2nd. ed. Macmillan, New York.

M. Schäl (1992) Average optimality in dynamic programming with general state space. Preprint (to appear in Math. Operat. Res.).

S. Stidham (1988) Scheduling, routing, and flow control in stochastic networks. In Stochastic Differential Systems, Stochastic Control Theory and Applications, W. Fleming and P.L. Lions, eds., The IMA Volumes in Mathematics and Its Applications, Springer-Verlag, Berlin, 10:529–561.

H.C. Tijms (1986) Stochastic Modelling and Analysis: A Computational Approach. John Wiley, Chichester.

D.J. White (1985) Real applications of Markov decision processes. Interfaces 15:73–83.

D.J. White (1987) A selective survey of hypothetical applications of Markov decision processes. Dept. of Systems Engineering Report, University of Virginia, Charlottesville, VA.

D.J. White (1988) Further real applications of Markov decision processes. Interfaces 18:55–61.

C.C. Von Weizsäcker (1965) Existence of optimal programs of accumulation for an infinite time horizon. Rev. Econom. Stud. 32:85-164.

S. Yakowitz (1982) Dynamic programming applications in water resources. Water Resour. Res. 18:673-696.

A.A. Yushkevich (1973) On a class of strategies in general Markov decision models. Theory Prob. Applications 18:777-779.