

- 13. E. Fernández-Gaucherand**, “Non-Standard Optimality Criteria for Controlled Markov Processes.” *ZAMM: Zeitschrift für Angewandte Mathematik und Mechanik*, Special issue on Applied Stochastics and Optimization, (1996) 423-424.

EMMANUEL FERNÁNDEZ-GAUCHERAND

Non-standard Optimality Criteria for Controlled Markov Processes

Stochastic optimal control methods find significant applications in areas as management and finance, manufacturing and production, communication/computer networks, military and service industry logistics, etc. Several of these areas fall in the general category of Discrete Event Stochastic Dynamic Systems (DESDS), an area of research at the intersection of systems and control, operations research, and knowledge-based systems. DESDS dynamic evolution is driven by the (random) occurrence of controlled and uncontrolled discrete events, e.g., system failures, customer arrivals, etc [7]. The common objective is to obtain rules for operating the system, i.e., control policies or decision rules, which optimize an appropriate performance measure. Appropriate mathematical descriptions for the state evolution and control of these processes fall within the domain of controlled Markov processes (CMP). However, standard optimality criteria, e.g., expected discounted or averaged costs, frequently fail to capture important aspects associated with the stochastic control problems that arise in many applications. Thus the need arises for new paradigms, which lead to the formulation of some non-standard performance criteria that address some of the shortcomings of, and give alternatives to, standard criteria. In this paper, recent and new results in this area, as well as some open questions, are surveyed [1], [3], [4], [5], [6].

1. Controlled Markov Processes

A CMP is a discrete-time, discrete-event stochastic dynamic system specified by the four-tuple $(\mathbf{X}, \mathbf{U}, P, c)$, where \mathbf{X} is the *state space*; \mathbf{U} is the *action, or control space*; each pair (x, u) in $\mathbf{X} \times \mathbf{U}$ determines a *transition law* $P(\cdot | x, u)$; and $c : \mathbf{X} \times \mathbf{U} \rightarrow \mathbb{R}$ is the one-stage cost function. A control strategy, or policy, is a rule π for making decisions, based on the available information. At a given time t , the available information is the set h_t of observed states and actions taken up to that time, i.e., $h_t = (X_0, U_0, X_1, \dots, U_{t-1}, X_t)$. Each policy π incurs a stream of costs $\{c(X_0, U_0), c(X_1, U_1), \dots\}$. The two standard criteria used are the following. *Discounted Cost (DC)*: For $0 < \beta < 1$, the *discount factor*, and a policy π , the total discounted cost is given by $J_\beta(x, \pi) := E_x^\pi [\sum_{t=0}^{\infty} \beta^t c(X_t, U_t)]$; *Average Cost (AC)*: Given a policy π , we have that $J(x, \pi) := \limsup_{N \rightarrow \infty} E_x^\pi \left[\frac{1}{N} \sum_{t=0}^{N-1} c(X_t, U_t) \right]$.

2. Non-Standard Criteria

The AC and DC criteria suffer from several shortcomings. Among these are: (a) They measure the *expected* sum of (discounted or averaged) cost, and thus are *insensitive* to, e.g., the variance of the sum of costs; (b) They can be seen as two opposite extremes in the spectrum of possible criteria; (c) In many situations, the optimal policy has to meet stringent robustness requirements, e.g., optimality for *almost all* sample paths, not obtained with AC or DC criteria; (d) Stronger criteria than AC that exhibit selectivity with respect to long but finite horizons is desirable. In this paper, we survey several promising alternatives that address the above shortcomings.

2.a: Risk-Sensitive Criteria

Consider the risk-sensitive optimal control problem for hidden Markov models (HMM), i.e., controlled Markov chains where state information is only available to the controller via an output (message) process. Here $\mathbf{X} = \{1, 2, \dots, N_X\}$; $\mathbf{U} = \{1, 2, \dots, N_U\}$; $P(u) := [p_{i,j}(u)]$ is the $N_X \times N_X$ state transition matrix. In addition, $\mathbf{Y} = \{1, 2, \dots, N_Y\}$ is the set of observations (or messages), and $Q(u) := [q_{x,y}(u)]$ is the $N_X \times N_Y$ state/message matrix, i.e., $q_{x,y}(u)$ is the probability of receiving message y when the state is x and action u has been selected. Hence, based on $\mathcal{I}_t := (U_0, Y_1, U_1, Y_2, \dots, U_t, Y_{t+1})$, a new decision U_{t+1} is selected at time $t + 1$.

Let $C_M := \sum_{t=0}^{M-1} c(X_t, U_t)$ be the sum of costs for the finite horizon M . The *risk-sensitive optimal control* problem is that of finding a control policy $\pi = \{\pi_0, \pi_1, \dots, \pi_{M-1}\}$, with $\mathcal{I}_t \mapsto \pi_t(\mathcal{I}_t) \in \mathbf{U}$, such that the following criterion is minimized: $J^\gamma(\pi, X_0) := \text{sgn}(\gamma) E^\pi [\exp(\gamma \cdot C_M)]$, where $\gamma \neq 0$ is the *risk-factor*, and $\text{sgn}(\gamma)$ is the sign of γ . If $\gamma > 0$, then the controller is *risk-averse* or *pessimistic*, whereas if $\gamma < 0$ then the controller is *risk-preferring* or *optimistic*; see [5], and [9] for details.

Controlled Markov chains with full state information and a risk-sensitive performance criterion have received some attention; see [5], [9]. Whittle and others [9] have extensively studied the risk-sensitive optimal control of partially-observable linear exponential quadratic Gaussian (LEQG) systems. Recently, James, Baras, and Elliott [2], [8] have derived information states for nonlinear, discrete-time partially observable stochastic systems, and in particular for HMM, under an “exponential of additive costs” criterion. They have also given dynamic programming equations from which optimal values and controls can be computed, for problems with a finite horizon. Nevertheless, the following remains mostly an open question: **How does risk-sensitivity manifest itself in a policy?**

In [5], several general results on the risk-sensitive control of HMM were obtained, e.g., piecewise linearity of value functions with respect to the information state. Moreover, via the examination of a popular benchmark problem, the above question was addressed. A threshold structure was obtained for the optimal controller. The cases when $\gamma \rightarrow 0$ and $\gamma \rightarrow \infty$ were also studied.

2.b: Weighted, Overtaking, Strong Average, and Sample Path Criteria

One possibility of obtaining a reasonable compromise of AC and DC criteria is by combining these in a weighted sum. CMP with a *generalized* WC criterion and with general state and control spaces and with *several* one-stage cost functions being discounted at *different* rates, were studied in [6]. The existence of ϵ -optimal policies, as well as dynamic programming-like equations, were obtained in [6] for the above GWC criterion.

The use of the *overtaking criterion* (OC) is one way to incorporate sensitivity to finite time behavior, while preserving results obtained under an AC criterion. A policy π_1 is said to *overtake* another policy π_2 if $J_T(x, \pi_1) \leq J_T(x, \pi_2)$, for all $x \in \mathbf{X}$, and for all T sufficiently large; $J_T(\cdot, \cdot)$ denotes the total expected cost up to time T . A policy is called *overtaking optimal* if it overtakes every other policy. In [6], we studied models with an OC, countable state space and compact action space. Our approach was based on qualitative properties of the optimality equations for the AC criterion [1]. In [6] it was shown that under a *Lyapunov Function Condition* [1], OC optimal policies exists and can be characterized as the maximizers in a certain functional equation.

The *strong average cost* (SAC) criterion is also introduced to assess the performance of a policy over long but finite horizons, as well as in the long-run average sense. We say that policy π^* is *strong average cost* (SAC) optimal if $\frac{1}{n+1} [J_n(x, \pi^*) - J_n^*(x)] \xrightarrow{n \rightarrow \infty} 0, \forall x \in \mathbf{X}$. Note that such a π^* induces good performance for long but finite horizons, and that every policy that is SAC optimal is also AC optimal. However the opposite is not necessarily true. It was shown in [4] that for bounded one-stage cost functions, conditions introduced by Sennott (see [1]) are sufficient to guarantee that *every* AC optimal policy is also SAC optimal. On the other hand, a detailed counterexample is given that shows that this result does not extend to the case of unbounded cost functions.

Departing from the use of *expected* values of costs, in [3] we focused on a sample path analysis of the stream of costs. Under a Lyapunov Function Condition [1], we showed that stationary policies obtained from the average cost optimality equation are not only expected average cost optimal, but indeed sample path average cost optimal.

3. References

- 1 ARAPOSTATHIS, V. S. BORKAR, E. FERNÁNDEZ-GAUCHERAND, M. K. GHOSH, AND S. I. MARCUS: Discrete-Time Controlled Markov Processes with Average Cost Criterion: A Survey, *SIAM Journal of Control and Optimization* **31** (1993) 282-344.
- 2 J.S. BARAS AND M.R. JAMES: Robust and Risk-sensitive Output Feedback Control for Finite State Machines and Hidden Markov Models, preprint, February 1993 and September 1994.
- 3 CAVAZOS-CADENA AND E. FERNÁNDEZ-GAUCHERAND: “Denumerable Controlled Markov Chains with Average Reward Criterion: Sample Path Optimality,” to appear in *ZOR: Methods and Models in Operations Research*. See also *Proc. 33rd IEEE Conference on Decision and Control*, Orlando, FL, (1994) 162-167.
- 4 R. CAVAZOS-CADENA AND E. FERNÁNDEZ-GAUCHERAND: “Denumerable Controlled Markov Chains with Strong Average Optimality Criterion: Bounded & Unbounded Costs,” to appear in *ZOR: Methods and Models in Operations Research*. See also *Proc. 33rd IEEE Conference on Decision and Control*, Orlando, FL, (1994) 1456-1461.
- 5 E. FERNÁNDEZ-GAUCHERAND AND S.I.MARCUS: Risk-Sensitive Optimal Control of Hidden Markov Models: A Case Study, in *Proc. 33rd IEEE Conference on Decision and Control*, Orlando, FL, (1994) 1657-1662.
- 6 E. FERNÁNDEZ-GAUCHERAND, M.K. GHOSH AND S.I. MARCUS: Controlled Markov Processes on the Infinite Planning Horizon: Weighted and Overtaking Cost Criteria, *ZOR: Methods and Models in Operations Research* **39** (1994) 131-155.
- 7 Y.C. HO: Dynamics of Discrete Event Systems, *Proc. IEEE*, **77** (1989) 3-6.
- 8 M.R. JAMES, J.S. BARAS AND R.J. ELLIOTT: Risk-sensitive control and dynamic games for Partially Observed Discrete-time Nonlinear Systems, *IEEE Transactions on Automatic Control* **39** (1994) 780-792.
- 9 P. WHITTLE: *Risk-sensitive Optimal Control*, Wiley, England, 1990.

Address: DR. EMMANUEL FERNÁNDEZ-GAUCHERAND, Systems & Industrial Engineering Department, The University of Arizona, Tucson, AZ 85721. Email: emmanuel@sie.arizona.edu.