

CONVEX STOCHASTIC CONTROL PROBLEMS

Emmanuel Fernández-Gaucherand †, Aristotle Arapostathis †,
and Steven I. Marcus §

I. Introduction

A Controlled Markov Process is a discrete time stochastic dynamical system specified by the five-tuple $\langle \mathbf{X}, \mathbf{U}, \mathcal{U}, P, c \rangle$ where \mathbf{X} is the *state space*; \mathbf{U} is the *action*, or *control space*; $\mathcal{U}(x) \subseteq \mathbf{U}$ is the set of *feasible actions* (or control inputs) when the system is in state $x \in \mathbf{X}$; 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. See [ABFGM], [BS], [HLM1] for more details. 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. $h_t = (X_0, U_0, X_1, \dots, U_{t-1}, X_t)$. A control strategy, or policy, is a rule π for making decisions, based on the available information. Each policy π incurs a stream of costs $\{c(X_0, U_0), c(X_1, U_1), \dots\}$. Depending upon the problem requirements, different criteria can be used to evaluate the performance of the system, under the policy used. The following criteria are frequently used; in the equations below, \mathbb{E}_x^π denotes the expectation operator under the policy π , when $X_0 = x$.

Discounted Cost (DC): For $0 < \beta < 1$, the *discount factor*, and a policy π , the total discounted cost incurred by π is given by

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

the optimal *value function*, i.e. the minimum $J_\beta(x, \pi)$ over all π , is denoted by $J_\beta^*(x)$.

Average Cost (AC): The expected long-run average cost incurred by the policy π is given by

$$J(x, \pi) := \limsup_{N \rightarrow \infty} \mathbb{E}_x^\pi \left[\frac{1}{N} \sum_{t=0}^{N-1} c(X_t, U_t) \right];$$

the optimal average cost is denoted by $J^*(x)$.

II. The Stochastic Control Problem

As it is well known [ABFGM], [BS], [HLM1], the solution of the infinite horizon stochastic control problem under the above criteria, i.e. the functional characterization and computation of optimal values and policies, is related to the following dynamic programming-like functional equations.

The Discounted Cost Optimality Equation (DCOE):

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

where the operator $T_\beta(\cdot)$ is defined in the obvious way.

The Average Cost Optimality Equation (ACOE):

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

where the operator $T(\cdot)$ is defined in the obvious way.

When a (DC) criterion is used, a rather complete theory is available [BE], [BS], [HLM1], [KV]. For the average cost, one looks for conditions under which appropriate solutions to the ACOE exist. A solution is a pair $(\rho(\cdot), h(\cdot))$ of real-valued functions on \mathbf{X} , satisfying the above ACOE. There is a vast literature concerning the problem of existence and functional characterization of average cost optimal policies, when the state space \mathbf{X} is *countable*, and/or the one stage cost function $c(\cdot, \cdot)$ is *bounded* [ABFGM], [BE], [HLM1]. However, this is not the case for the situation when the state space is a *general* (Borel) space, e.g. $\mathbf{X} = \mathbb{R}^n$, and the one-stage cost function is unbounded. Necessary and sufficient conditions for a *bounded* solution to the ACOE have been recently given by the authors [FAM]. However this type of solutions are not natural for problems involving an infinite number of states and an unbounded cost function. Recently, much research activity has been devoted to finding conditions for the functional characterization and existence results for average optimal values and policies, for the case of unbounded cost functions. Sennott [SEN] treated the situation of countable state space and finite action set, and Hernández-Lerma and Lasserre [HLL], [HLM2] extended these results to a general space setting. However, in these references the authors only show existence of solutions to an *average cost optimality inequality* (ACOI):

$$\rho(x) + h(x) \geq \inf_{u \in \mathcal{U}(x)} \left\{ c(x, u) + \int_{\mathbf{X}} h(y) P(dy | x, u) \right\}.$$

Actually, it has been recently shown in [CC] that *strict inequality* is possible under the conditions in [HLL] and [SEN]. The fact that equality is not shown prevents one from, e.g. quantifying the deviation from optimality for a policy π via Mandl's discrepancy function [ABFGM, Theorem 6.3]. Also, policy improvement in a policy iteration algorithm [BE] *cannot* be implemented if only an inequality result is available. Thus, although the ACOI gives a criterion for the *existence* of stationary average cost optimal policies, *it is not useful from an algorithmic standpoint*. Hence, the following question is very relevant:

- What useful properties, shared by large and important problem classes, can be used to further show that an ACOE holds, and how can these properties be exploited to aid in the development of tractable algorithmic solutions?

We address the above question, by concentrating on *structured* solutions to stochastic control models. By a structured solution we mean a model for which value functions and/or optimal policies have some special dependence on the (initial) state. We focus on convexity properties of the value function.

† Systems and Industrial Engineering Department, The University of Arizona, Tucson, Arizona 85721 (emmanuel@sie.arizona.edu).

‡ Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712-1084 (ari@emx.utexas.edu).

§ Systems Research Center & Electrical Engineering Department, The University of Maryland, College Park, Maryland 20742 (marcus@src.umd.edu).

III. Convex Controlled Markov Processes

In [FG] and [FMA], an investigation was initiated on the use of convexity properties of the discounted value function, to give a (partial) answer to the question previously posed. The main idea is to be able to extract an appropriately convergent subsequence, as $\beta \uparrow 1$, of the *differential* discounted value functions

$$h_\beta(x) := J_\beta^*(x) - J_\beta^*(\bar{x}), \quad \forall x \in \mathbf{X},$$

where $\bar{x} \in \mathbf{X}$ (the reference state) is kept fixed, so that the ACOE can be obtained by taking limits in the DCOE (see [ABFGM, Sect.6]). Under a convexity condition of the discounted value functions, *local* uniform boundedness and *local* equicontinuity properties can be shown for $\{h_\beta(\cdot)\}$, and thus the Arzela-Ascoli theorem can be used to obtain the ACOE by taking limits in the DCOE. The framework is the following. Let \mathbf{X} be an open convex subset of \mathbb{R}^n . In particular, \mathbf{X} is a Borel space.

Definition: If the value functions $J_\beta^*(\cdot)$ are convex functions, then we say that $(\mathbf{X}, \mathcal{U}, P, c)$ is a *convex* CMP.

Assumption A: We make the usual *semicontinuity* assumptions on the transition kernel, and nonnegativity of the cost function [ABFGM], [BS]. In addition, we assume that the cost function has the *compact level sets* (CLS) property, i.e. for each $\lambda \in \mathbb{R}$, the set $\{(x, u) | c(x, u) \leq \lambda\}$ is compact. Then, under this conditions it can be shown that: (a) the DCOE holds, (b) a deterministic stationary policy is discount optimal if and only if it attains the infimum in the DCOE, and (c) one such policy exists [ABFGM], [BS], [FG], [HLM1].

The next assumption is quite standard, c.f. [HLL], [SEN].

Assumption B: For each $x \in \mathbf{X}$, $P(B | x, u)$ is continuous in $u \in \mathcal{U}(x)$, for all Borel sets B ; there exists a non-negative, upper semicontinuous function $b : \mathbf{X} \rightarrow \mathbb{R}$, a constant $M \geq 0$, and a sequence $\{\beta_n\} \subseteq (0, 1)$ with $\beta_n \uparrow 1$, such that for all $x \in \mathbf{X}$

- (i) $J_{\beta_n}^*(x) < \infty$;
- (ii) $-M \leq h_{\beta_n}(x) \leq b(x)$;
- (iii) $\int_{\mathbf{X}} b(y)P(dy | x, u) < \infty; \quad \forall u \in \mathcal{U}(x)$.

Finally, we assume the required convexity properties.

Assumption C: $J_{\beta_n}^*(\cdot)$ is a convex function.

Then, the results in [HLL], [HLM2], [SEN] can be strengthened as follows.

Theorem: Under Assumptions A-C, we have that

- (i) there is a constant ρ^* and a convex function $h : \mathbf{X} \rightarrow \mathbb{R}$ with

$$-M \leq h(x) \leq b(x), \quad \forall x \in \mathbf{X},$$

such that the pair (ρ^*, h) is a solution to the ACOE, i.e.,

$$\rho^* + h(x) = \inf_{u \in \mathcal{U}(x)} \left\{ c(x, u) + \int_{\mathbf{X}} h(y)P(dy | x, u) \right\}; \quad (1)$$

- (ii) there exists a (stationary deterministic) policy π^* which is average optimal, and every (stationary deterministic) policy π attaining the infimum in (1) is average optimal;
- (iii) $J^*(x) = \rho^*$, for all $x \in \mathbf{X}$. □

Remark: Examples of CMP exhibiting the necessary convexity condition are: problems with imperfect state information, some gambling models, linear systems with quadratic cost, some inventory control models, problems in Bayesian sequential analy-

sis [BE], [KV]. Also, Dynkin [DYN] introduced the concept of stochastic concave dynamic programming, but his interest was on concavity properties of one-stage cost functions in the control actions, and its implications on the existence of minimizing actions. In addition, convexity of the (discounted) value functions has been used in [HLMR] to obtain monotone approximations.

The assumption that $J_{\beta_n}^*(\cdot)$ is convex is the cornerstone of our developments. This can be established using induction, via the value iteration scheme [ABFGM], [BS], [HLM1], to propagate the desired convexity property; see [HIN].

ACKNOWLEDGEMENTS: This work was supported in part by the Texas Advanced Technology Program under Grants No. 003658-093 and No. 003658-186, in part by the Air Force Office of Scientific Research under Grants AFOSR-91-0033, F49620-92-J-0045, F49620-92-J-0083, and in part by the National Science Foundation under Grants CDR-8803012 and INT-9201430.

REFERENCES

- [ABFGM] A. Arapostathis, V. Borkar, E. Fernández-Gaucherand, M.K. Ghosh and S.I. Marcus, Discrete-Time Controlled Markov Processes with Average Cost Criterion: A Survey, to appear in *SIAM Journal on Control & Optimization*.
- [BE] D.P. Bertsekas, *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, 1987.
- [BS] D.P. Bertsekas and S.E. Shreve, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.
- [CC] R. Cavazos-Cadena, A Counterexample on the Optimality Equation in Markov Decision Chains with the Average Cost Criterion, *Systems & Control Letters* **16** (1991) 387-392.
- [DYN] E.B. Dynkin, Stochastic Concave Dynamic Programming, *Math. USSR Sbornik* **16** (1972) 501-515.
- [FAM] E. Fernández-Gaucherand, A. Arapostathis and S.I. Marcus, Remarks on the Existence of Solutions to the Average Cost Optimality Equation in Markov Decision Processes, *Systems & Control Letters* **15** (1990) 425-432.
- [FG] E. Fernández-Gaucherand, *Controlled Markov Processes on the Infinite Planning Horizon: Optimal & Adaptive Control*, Ph.D. Dissertation, The University of Texas at Austin, August 1991.
- [FMA] E. Fernández-Gaucherand, S.I. Marcus, and A. Arapostathis, Structured Solutions for Stochastic Control Problems, to appear in the *Proceedings of the Stochastic Theory and Adaptive Control Workshop*, Lawrence, Kansas (1991).
- [HIN] K.F. Hinderer, On the Structure of Solutions of Stochastic Dynamic Programs, in: *Proceedings of the 7th Conference on Probability Theory*, Brasov, Romania, (1984) 173-182.
- [HLL] O. Hernández-Lerma and J.B. Lasserre, Average Cost Optimal Policies for Markov Control Processes with Borel State Space and Unbounded Costs, *Systems & Control Letters* **15** (1990) 349-356.
- [HLM1] O. Hernández-Lerma, *Adaptive Markov Control Processes*, Springer Verlag, New York, 1989.
- [HLM2] O. Hernández-Lerma, Average Optimality in Dynamic Programming on Borel Spaces: Unbounded Costs and Controls, *Systems & Control Letters* **17** (1991) 237-242.
- [HLMR] O. Hernández-Lerma and W. J. Runggaldier, Monotone Approximations for Convex Stochastic Control Problems, preprint, 1992.
- [KV] P.R. Kumar and P. Varaiya, *Stochastic Systems: Estimation, Identification and Adaptive Control*, Prentice-Hall, Englewood Cliffs, 1986.
- [SEN] L.I. Sennott, Average Cost Optimal Stationary Policies in Infinite State Markov Decision Processes with Unbounded Costs, *Operations Research* **37** (1989) 626-633.