is then given to reveal the relationship between state-space models and transfer functions in Chapter 3. Following that, some special systems with constraints are discussed with the same mathematical style in Chapter 4.

Part II deals with finite-dimensional nonlinear systems. With a structure almost parallel to that of Part I, the three chapters in Part II are presented corresponding to the first three chapters of Part I to introduce results on controllability, stabilizability, and realization, respectively, for nonlinear systems. Material on optimal control of finite-dimensional systems follows in Part III. Dynamic programming is first presented in Chapter 1, with the optimal control of impulsive systems introduced in Chapter 2. The maximum principle is presented in Chapter 3, followed by the Fillipov theorem on the existence of optimal strategies in Chapter 4.

Part IV handles the control of infinite-dimensional systems using operator-based approaches. Chapter 1 in this part lays the related mathematical groundwork and is followed by basic concepts such as controllability, stabilizability, and optimality in the three chapters that follow for infinite-dimensional systems.

DISCUSSION

The contents of this well-organized book mainly include the analysis of control properties and optimization. I enjoyed reading the concise mathematical description with the clean logical structure. I also learned several new things or reviewed some materials from new angles, for example, the topological stabilizability criteria for control systems in Chapter 2 of Part II and the Filipov-based existence results in Chapter 4 of Part III. However, due to space limitations, some general discussions and detailed analysis are missing in this monograph. For example, various versions of controllability in nonlinear or infinite-dimensional systems and their relationships are not fully displayed. The readers who are interested in the extended or generalized results may have to consult other books such as [2]. Moreover, the feedback control synthesis problems, such as robust and adaptive control design, which have attracted more and more attention in recent years, are absent. This omission may make the book less attractive to readers who are eager to learn new advances in feedback design.

In some sense, *Mathematical Control Theory* is not suitable for a textbook since it is too difficult for general postgraduate students unless they have a strong mathematical background. Moreover, the presentation style of the book does not facilitate the understanding of practical control meaning behind the mathematics. Many students will find the chapters full of mathematical formulas and equations without enough introduction of physical intuition or engineering background. As such, the book may not be a good choice for lower level classes, for newcomers to the control science and technology field, or for those to be trained as practicing control engineers. However, I recommend the book to readers who are interested in the rigorous mathematical buildup of control systems and problems. Indeed, for mathematicians who look for the basic ideas or a general picture about the main branches of control theory, I believe this book can provide an excellent bridge to this area. Finally, for students who are ready for a more rigorous approach after grasping suitable mathematical preliminaries and control engineering background, this book can be helpful owing to its theoretical beauty and clarity.

REFERENCES

[1] M. Athans and P.L. Falb, *Optimal Control: An Introduction to Theory and Its Applications*. New York: McGraw-Hill, 1966.

[2] R. Curtain and H. Zwart, *An Introduction to Infinite-Dimensional Linear Systems Theory*, New York: Springer, 1995.

[3] H. Khalil, *Nonlinear Systems*. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 2002.

[4] W. Wonham, *Linear Multivariable Control: A Geometric Approch*. 3rd ed. New York: Springer-Verlag, 1985.

REVIEWER INFORMATION

Yiguang Hong received the B.S. and M.S. degrees from Peking University, China, and the Ph.D. degree from the Chinese Academy of Sciences (CAS). He is currently a professor at the Institute of Systems Science, Academy of Mathematics and Systems Science, CAS. He is the 1997 recipient of a Guangzhaozhi Award of the Chinese Control Conference, the Young Author Prize at the 1999 IFAC World Congress, and a Youth Award for Science and Technology of China in 2006. His research interests include nonlinear dynamics and control, multiagent systems, robotics, and system reliability.

Stochastic Learning and Optimization: A Sensitivity-Based Approach

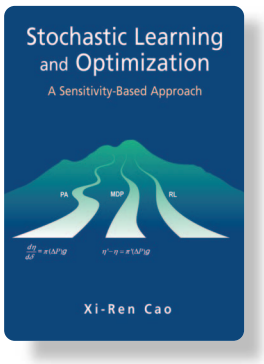
by XI-REN CAO

Reviewed by Pravin Varaiya

The typical graduate control curriculum introduces stochastic control through the linear-quadratic-Gaussian (LQG) problem, using only the basic properties of discrete-time linear systems and quadratic forms as well as elementary probability. State estimation can be formulated as weighted least squares, leading to the Kalman filter. The LQG control is obtained by means of the separation theorem and the feedback control for minimum quadratic cost. A more advanced course might deal with spectral factorization, linear system identification, and adaptive control, requiring a firmer background in probability and statistics. Twenty years ago a student with this background could read the current literature and begin research in stochastic control.

Digital Object Identifier 10.1109/MCS.2008.929812

1066-033X/08/$25.00©2008IEEE



Springer, 2007,
119 illustrations and
212 problems
ISBN-13: 978-0-387-36787-3,
e-ISBN-13: 987-0-387-69082-7,
US$129, 566 pages.

Today, a grounding in linear stochastic systems is inadequate for formulating problems of sequential decision-making in nonlinear and discrete-event stochastic systems or to become familiar with developments over the past 20 years in analytical and computational techniques, results, and applications of stochastic learning and optimization. The instructor faces three interdependent questions in preparing a course that covers these developments. Can prerequisites be limited to basic stochastic processes, such as Markov chains and applications such as queuing networks? What should be covered in terms of models, techniques, and applications? What reading material should be selected from the array of specialized books and articles?

One attractive answer to these questions is offered by *Stochastic Learning and Optimization: A Sensitivity-Based Approach*. This book presents a unique synthesis of research over the past two decades in stochastic optimization and learning, including Markov decision problems (MDPs), perturbation analysis (PA), reinforcement learning (RL), and identification and adaptive control (I&AC). The book is divided into three parts. Part I, chapters 2–7, covers the above-mentioned topics. Part II, chapters 8–9, presents the author's original work on event-based optimization, extending the discussion in Part I to systems in which, instead of observing the state, the system is observed through "events." Part III covers the necessary background in Markov chains and queuing theory. Parts I and III can be comfortably covered in one semester. The text flows effortlessly, the figures enhance intuition, and the problems at the end of each chapter test understanding and further explore the topics covered.

SENSITIVITY-BASED APPROACH

The author's sensitivity-based approach provides both a unifying perspective for optimization and a sample path-based physical interpretation for learning. Here are the key ideas. Consider a discrete-time MDP with state $X_l \in \mathcal{S}$, action or control $A_l \in \mathcal{A}$, and decision rule or feedback policy $d : \mathcal{S} \to \mathcal{A}$. System performance is measured by the long-term average reward

$$\eta^d = \lim_{L\to\infty} \frac{1}{L} \sum_{l=0}^{L-1} Ef(X_l, d(X_l)) = \lim_{L\to\infty} \frac{1}{L} \sum_{l=0}^{L-1} f(X_l, d(X_l)), \quad (1)$$

where $f(i, \alpha)$ is the one-period reward and E denotes expectation. Consider two policies, d and h. Let P be the transition probability matrix under d, π the steady-state probability row vector, f the one-step reward column vector, and η the average reward. The corresponding quantities for h are distinguished by primes, for example, P', π'. Let $\Delta P = P' - P$, $\Delta f = f' - f$, $P_\delta = P + \delta(\Delta P)$, and $0 \le \delta \le 1$. The sensitivity-based approach refers to the search for good or optimal policies through the analysis and computation of the performance *difference*

Final Touchdown

However, in reality the situation may be more complicated. Namely, if the system is subject to unknown but bounded disturbances, or if the system parameters are not exactly known it may become necessary to compute the set of states reachable despite the disturbances or data uncertainty or, if exact reachability is impossible, to find guaranteed errors for reachability. These questions have implicitly been present in traditional studies on feedback control under uncertainty for continuous-time systems. They have had very serious motivations from problems like calculation of landing ranges during the liquidation of space station "Mir".

—"National achievements in control theory; The aerospace perspective," By V.F. Krotov, and A.B. Kurzhanski, *Annual Reviews in Control*, Vol. 29, 2005, p. 26.

$$\Delta\eta = \eta' - \eta, \tag{2}$$

or the performance *gradient* (η_δ is the performance for P_δ)

$$\frac{d\eta_\delta}{d\delta} = \lim_{\delta \to 0} \frac{\eta_\delta - \eta}{\delta}. \tag{3}$$

Relations (1)–(3) structure the organization and limit the scope of the book.

Evidently, the average reward $\eta = \pi f$. The *potential* g satisfies the Poisson equation

$$(I - P)g + \eta e = f, \tag{4}$$

in which e is the column vector of all 1s. The solution g of (4) is unique up to an additive constant. The performance difference (2) is given by

$$\eta' - \eta = \pi'[(\Delta P)g + \Delta f], \tag{5}$$

and the performance gradient (3) is given by

$$\frac{d\eta_\delta}{d\delta} = \pi[(\Delta P)g + \Delta f]. \tag{6}$$

The book shows how many results in optimization and stochastic learning are derived or explained from these two fundamental sensitivity formulas.

OPTIMIZATION

Formula (6) allows one to conduct a gradient search to improve the current policy d. Observe that (6) depends only on π and g, which are determined by d. Further, (6) implies that η_δ is differentiable arbitrarily often, and the book develops a MacLaurin series expansion of η_δ, which can be used to compute η_δ. The performance difference (5) on the other hand depends on both d and h, which makes it less immediately useful for policy improvement. However, the *policy iteration* algorithm follows directly from this equation. From (5), we obtain

$$[\eta' - \eta = \pi'[(f' + P'g) - (f + Pg)],$$

which leads to the following policy iteration algorithm: At step k, choose

$$d_{k+1} \in \arg\max_d \left[f^d + P^d g^{d_k} \right]; \tag{7}$$

and the optimality condition in which d is optimal if and only if for all policies h

$$f + Pg \geq f' + P'g. \tag{8}$$

The book shows how the performance difference leads to the solution of various optimization problems, including the bias and n-bias optimality and Blackwell optimality, for the multi-chain case.

LEARNING

Using (3) or (7) requires calculating π and g, which may be computationally difficult or even impossible if P is not known. Learning refers to estimation of the potential (or the transition probabilities P) from a sample path, which may be obtained from a simulation or online measurements. The key observation is

$$\begin{aligned} g(i) &= E\left\{ \sum_{k=0}^{\infty} [f(X_{l+k}) - \eta] \,|\, X_l = i \right\} \\ &= E\{[f(X_l) - \eta] + g(X_{l+1}) \,|\, X_l = i\}, \end{aligned} \tag{9}$$

so that, in the spirit of stochastic approximation, one obtains an update rule for the estimate of the potential

$$\begin{aligned} g(X_l) &\leftarrow g(X_l) + -\kappa_l\{g(X_l) - [f(X_l) - \eta + g(X_{l+1})]\} \\ &= g(X_l) + \kappa_l \delta_l, \end{aligned} \tag{10}$$

wherein the *temporal difference (TD)* is defined as

$$\delta_l = [f(X_l) - \eta + g(X_{l+1}) - g(X_l)], \;\; l = 0, 1, \ldots .$$

In addition, η can be estimated by

$$\eta_{l+1} = \eta_l - \kappa_{l+1}[\eta_l - f(X_{l+1})].$$

The step sizes κ_l must be appropriately selected for convergence. The book presents several other algorithms, including variants of the TD method above, for estimating the potential. These estimates can be combined with policy iteration (7) to search for better policies, without explicitly calculating the potential by means of the Poisson equation.

If P is unknown, the potential estimates (10) cannot be used in the policy iteration (7). Instead one can estimate the right-hand side of (7) as the *Q-factors*,

$$Q^{d_k}(i, \alpha) = f(i, \alpha) + \sum_{j \in \mathcal{S}} p(j \,|\, i, \alpha) g^{d_k}(j), \tag{11}$$

and replace the iteration (7) by

$$\alpha_{k+1} \in \arg\max_{\alpha' \in \mathcal{A}} Q^{d_k}(i, \alpha').$$

As with (10), (11) must be replaced by a stochastic approximation

algorithm to estimate the Q-factors. Again, the author presents many variants of such methods, which collectively have become known as reinforcement learning. One feature of this book is its emphasis on learning algorithms for performance gradients, which extend the scope of RL and link PA and RL together.

QUEUING SYSTEMS

Although the discussion above is formulated in terms of discrete-time MDPs, it carries over with small changes to continuous-time MDPs. Some problems in the control of queuing systems can be posed as continuous-time MDPs. However, problems in which the decision variable d changes the service time of a queuing system cannot be treated as MDPs. These problems can be analyzed by means of *perturbation analysis*, which can be illustrated by a single-server queue with arrival process a_t and virtual service times y_k, resulting in x_t queued customers. Performance is measured by

$$\eta = \lim_{T\to\infty} \frac{1}{T}\int_0^T f(x_t)dt.$$

Suppose the service-time distribution is perturbed slightly and suppose that x'_t is the resulting queue. One key idea of PA is to slightly perturb the service times y_k to obtain y'_k corresponding to the perturbed distribution, so that x'_t is a small perturbation of x_t and the change in performance can be estimated by

$$\Delta\eta = \eta' - \eta \approx \frac{1}{T}\int_0^T [f(x'_t) - f(x_t)]dt.$$

Since the samples of y'_k are obtained from those of y_k, this estimate can be computed from the latter; moreover, the estimate above has small variance. The subtleties in PA arise when a small perturbation in y_k causes a large change in the busy period or when there is a network of queues. The author provides systematic algorithms for tracing the effects of the initial perturbation. The performance difference estimates can then be used to obtain the performance gradient in optimization.

EVENT-BASED OPTIMIZATION

Part II of the book is devoted to the author's most recent research. The idea of event-based control can be introduced as follows. Consider a Markov chain. An *event* $E \subset \mathcal{S}\times\mathcal{S}$ is a subset of state transitions. Let $\mathcal{E} = \{E_1, E_2, \dots\}$ be a collection of disjoint events with $\cup_i E_i = \mathcal{S}\times\mathcal{S}$. At any time l the controller observes one of the events, say E_l, with the understanding that $(X_l, X_{l+1}) \in E_l$. The controller does not directly observe X_l. Based on the observation E_l, the controller must select an action $\alpha \in \mathcal{A}$, which determines the probability of a transition belonging to a subset of E_l, $P^\alpha\{(X_l, X_{l+1}) \in E \mid (X_l, X_{l+1}) \in E_l\}$, $E \subset E_l$. An event-based decision policy d is thus a map from $\mathcal{E}$ into $\mathcal{A}$. The performance is given by (1) as above. The book gives conditions on the structure of these conditional probabilities under which the performance difference (5) can be expressed in terms of an aggregated potential $(g(E), E \in \mathcal{E})$. The optimization and learning algorithms described in Part I carry over with suitable modifications.

Event-based control is natural in settings in which a sensor or alarm signals the occurrence of an impending event, such as the presence of a vehicle awaiting service at a traffic light or the arrival of a call at a telephone switch. The author explores in detail the example of a network with K queues, whose state is $\bar{n} = (n_1, \dots, n_K)$, where n_k is the number of customers in queue k. Suppose the controller can observe the number of customers $n = \sum n_k$ and the arrival of a customer but not which queue the customer joins. This event comprises all transitions $(\bar{n}, \bar{n}')$ in which, for exactly one k, $n'_k - n_k = 1$ and $n'_j = n_j$, $j \neq k$. Thus there are far fewer events than states. More importantly, it is not generally possible to express event-based control as an MDP even though the process is Markov.

CONCLUSIONS

The book contains many other interesting results, including MDPs with continuous state space, needed to formulate the LQG problem as an MDP; and bias optimality for selecting the best among many non-unique policies that maximize the reward (1). Most importantly, this review does not convey the deep physical insight based on sample paths that the author employs to derive all of the results discussed above (briefly, the performance gradient and difference for Markov systems can be constructed by using potentials $g(i), i \in \mathcal{S}$ as building blocks). The concluding chapter of Part II indicates the way in which such insight can be used to extend the sensitivity-based approach to new problems.

In conclusion, this book introduces stochastic learning and optimization and original research in these areas within a unified framework. It fits well the need for a textbook for students in control and optimization to gain an overview beyond the linear stochastic control.

REVIEWER INFORMATION

Pravin Varaiya is a professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. His research concerns communication networks, transportation, and hybrid systems. He has held a Guggenheim Fellowship and a Miller Research Professorship. He has received two honorary doctorates, as well as the Field Medal and Bode Prize of the IEEE Control Systems Society. He is a Fellow of IEEE, a member of the National Academy of Engineering, and a Fellow of the American Academy of Arts and Science.