Markov Decision Processes

Markov Decision Processes

Discrete Stochastic Dynamic Programming

MARTIN L. PUTERMAN
University of British Columbia

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication is available.

ISBN 0-471-72782-2

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To my father-in-law
Dr. Fritz Katzenstein
1908–1993
who never lost his love for learning

Contents

Preface

The past decade has seen a notable resurgence in both applied and theoretical research on Markov decision processes. Branching out from operations research roots of the 1950's, Markov decision process models have gained recognition in such diverse fields as ecology, economics, and communications engineering. These new applications have been accompanied by many theoretical advances. In response to the increased activity and the potential for further advances, I felt that there was a need for an up-to-date, unified and rigorous treatment of theoretical, computational, and applied research on Markov decision process models. This book is my attempt to meet this need.

I have written this book with two primary objectives in mind: to provide a comprehensive reference for researchers, and to serve as a text in an advanced undergraduate or graduate level course in operations research, economics, or control engineering. Further, I hope it will serve as an accessible introduction to the subject for investigators in other disciplines. I expect that the material in this book will be of interest to management scientists, computer scientists, economists, applied mathematicians, control and communications engineers, statisticians, and mathematical ecologists. As a prerequisite, a reader should have some background in real analysis, linear algebra, probability, and linear programming; however, I have tried to keep the book self-contained by including relevant appendices. I hope that this book will inspire readers to delve deeper into this subject and to use these methods in research and application.

Markov decision processes, also referred to as stochastic dynamic programs or stochastic control problems, are models for sequential decision making when outcomes are uncertain. The Markov decision process model consists of decision epochs, states, actions, rewards, and transition probabilities. Choosing an action in a state generates a reward and determines the state at the next decision epoch through a transition probability function. Policies or strategies are prescriptions of which action to choose under any eventuality at every future decision epoch. Decision makers seek policies which are *optimal* in some sense. An analysis of this model includes

1. providing conditions under which there exist easily implementable optimal policies;
2. determining how to recognize these policies;
3. developing and enhancing algorithms for computing them; and
4. establishing convergence of these algorithms.

Surprisingly these analyses depend on the criterion used to compare policies. Because of this, I have organized the book chapters on the basis of optimality criterion.

The primary focus of the book is infinite-horizon discrete-time models with discrete state spaces; however several sections (denoted by $*$) discuss models with arbitrary state spaces or other advanced topics. In addition, Chap. 4 discusses finite-horizon models and Chap. 11 considers a special class of continuous-time discrete-state models referred to as semi-Markov decision processes.

This book covers several topics which have received little or no attention in other books on this subject. They include modified policy iteration, multichain models with average reward criterion, and sensitive optimality. Further I have tried to provide an in-depth discussion of algorithms and computational issues. The Bibliographic Remarks section of each chapter comments on relevant historical references in the extensive bibliography. I also have attempted to discuss recent research advances in areas such as countable-state space models with average reward criterion, constrained models, and models with risk sensitive optimality criteria. I include a table of symbols to help follow the extensive notation. As far as possible I have used a common framework for presenting results for each optimality criterion which

- explores the relationship between solutions to the optimality equation and the optimal value function;
- establishes the existence of solutions to the optimality equation;
- shows that it characterizes optimal (stationary) policies;
- investigates solving the optimality equation using value iteration, policy iteration, modified policy iteration, and linear programming;
- establishes convergence of these algorithms;
- discusses their implementation; and
- provides an approach for determining the structure of optimality policies.

With rigor in mind, I present results in a "theorem-proof" format. I then elaborate on them through verbal discussion and examples. The model in Sec. 3.1 is analyzed repeatedly throughout the book, and demonstrates many important concepts. I have tried to use simple models to provide counterexamples and illustrate computation; more significant applications are described in Chap. 1, the Bibliographic Remarks sections, and left as exercises in the Problem sections. I have carried out most of the calculations in this book on a PC using the spreadsheet Quattro Pro (Borland International, Scott's Valley, CA), the matrix language GAUSS (Aptech Systems, Inc., Kent, WA), and Bernard Lamond's package MDPS (Lamond and Drouin, 1992). Most of the numerical exercises can be solved without elaborate coding.

For use as a text, I have included numerous problems which contain applications, numerical examples, computational studies, counterexamples, theoretical exercises, and extensions. For a one-semester course, I suggest covering Chap. 1; Secs. 2.1 and 2.2; Chap. 3; Chap. 4; Chap. 5; Secs. 6.1, 6.2.1–6.2.4, 6.3.1–6.3.2, 6.4.1–6.4.2, 6.5.1–6.5.2, 6.6.1–6.6.7, and 6.7; Secs. 8.1, 8.2.1, 8.3, 8.4.1–8.4.3, 8.5.1–8.5.3, 8.6, and 8.8; and Chap. 11. The remaining material can provide the basis for topics courses, projects and independent study.

This book has its roots in conversations with Nico van Dijk in the early 1980's. During his visit to the University of British Columbia, he used my notes for a course on dynamic programming, and suggested that I expand them into a book. Shortly thereafter, Matt Sobel and Dan Heyman invited me to prepare a chapter on Markov decision processes for *The Handbook on Operations Research: Volume II, Stochastic Models*, which they were editing. This was the catalyst. My first version (180 pages single spaced) was closer to a book than a handbook article. It served as an outline for this book, but has undergone considerable revision and enhancement. I have learned a great deal about this subject since then, and have been encouraged by the breadth and depth of renewed research in this area. I have tried to incorporate much of this recent research.

Many individuals have provided valuable input and/or reviews of portions of this book. Of course, all errors remain my responsibility. I want to thank Hong Chen, Eugene Feinberg, and Bernard Lamond for their input, comments and corrections. I especially want to thank Laurence Baxter, Moshe Haviv, Floske Spieksma and Adam Shwartz for their invaluable comments on several chapters of this book. I am indebted to Floske for detecting several false theorems and unequal equalities. Adam used the first 6 chapters while in proof stage as a course text. My presentation benefited greatly from his insightful critique of this material. Linn Sennott deserves special thanks for her numerous reviews of Sects. 6.10 and 8.10, and I want to thank Pat Kennedy for reviewing my presentation of her research on Cooper's hawk mate desertion, and providing the beautiful slide which appears as Fig. 1.6.1. Bob Foley, Kamal Golabi, Tom McCormick, Evan Porteus, Maurice Queyranne, Matt Sobel, and Pete Veinott have also provided useful input. Several generations of UBC graduate students have read earlier versions of the text. Tim Lauck, Murray Carlson, Peter Roorda, and Kaan Katiriciougulu have all made significant contributions. Tim Lauck wrote preliminary drafts of Sects. 1.4, 1.6, and 8.7.3, provided several problems, and pointed out many inaccuracies and typos. I could not have completed this book without the support of my research assistant, Noel Paul, who prepared all figures and tables, most of the Bibliography, tracked down and copied many of the papers cited in the book, and obtained necessary permissions. I especially wish to thank the Natural Sciences and Engineering Research Council for supporting this project through Operating Grant A5527, The University of British Columbia Faculty of Commerce for ongoing support during the book's development and the Department of Statistics at The University of Newcastle (Australia) where I completed the final version of this book. My sincere thanks also go to Kimi Sugeno of John Wiley and Sons for her editorial assistance and to Kate Roach of John Wiley and Sons who cheerfully provided advice and encouragement.

Finally, I wish to express my appreciation to my wife, Dodie Katzenstein, and my children, Jenny and David, for putting up with my divided attention during this book's six year gestation period.

MARTIN L. PUTERMAN

Markov Decision Processes