Dynamic Programming and Optimal Control

Volume II

Approximate Dynamic Programming

FOURTH EDITION

Dimitri P. Bertsekas

Massachusetts Institute of Technology

WWW site for book information and orders

http://www.athenasc.com

Athena Scientific, Belmont, Massachusetts

Athena Scientific
Post Office Box 805
Nashua, NH 03061-0805
U.S.A.

Email: info@athenasc.com
WWW: http://www.athenasc.com

Cover Design: Ann Gallager, www.gallagerdesign.com
Cover photography: Dimitri and Melina Bertsekas

Publisher's Cataloging-in-Publication Data

Bertsekas, Dimitri P.
Dynamic Programming and Optimal Control
Includes Bibliography and Index
1. Mathematical Optimization. 2. Dynamic Programming. I. Title.
QA402.5 .B465 2012 519.703 01-75941

ISBN-10: 1-886529-44-2, ISBN-13: 978-1-886529-44-1 (Vol. II)
ISBN 1-886529-26-4 (Vol. I)
ISBN 1-886529-08-6 (Two-volume set – latest editions)

Contents

CONTENTS OF VOLUME I

ABOUT THE AUTHOR

Dimitri Bertsekas studied Mechanical and Electrical Engineering at the National Technical University of Athens, Greece, and obtained his Ph.D. in system science from the Massachusetts Institute of Technology. He has held faculty positions with the Engineering-Economic Systems Dept., Stanford University, and the Electrical Engineering Dept. of the University of Illinois, Urbana. Since 1979 he has been teaching at the Electrical Engineering and Computer Science Department of the Massachusetts Institute of Technology (M.I.T.), where he is currently McAfee Professor of Engineering.

His research spans several fields, including optimization, control, large-scale computation, and data communication networks, and is closely tied to his teaching and book authoring activities. He has written numerous research papers, and thirteen books, several of which are used as textbooks in MIT classes. He consults regularly with private industry and has held editorial positions in several journals.

Professor Bertsekas was awarded the INFORMS 1997 Prize for Research Excellence in the Interface Between Operations Research and Computer Science for his book "Neuro-Dynamic Programming" (co-authored with John Tsitsiklis), the 2000 Greek National Award for Operations Research, the 2001 ACC John R. Ragazzini Education Award, and the 2009 INFORMS Expository Writing Award. In 2001, he was elected to the United States National Academy of Engineering.

Preface

This two-volume book is based on a first-year graduate course on dynamic programming and optimal control that I have taught for over twenty years at Stanford University, the University of Illinois, and the Massachusetts Institute of Technology. The course has been typically attended by students from engineering, operations research, economics, and applied mathematics. Accordingly, a principal objective of the book has been to provide a unified treatment of the subject, suitable for a broad audience. In particular, problems with a continuous character, such as stochastic control problems, popular in modern control theory, are simultaneously treated with problems with a discrete character, such as Markovian decision problems, popular in operations research. Furthermore, many applications and examples, drawn from a broad variety of fields, are discussed.

The book may be viewed as a greatly expanded and pedagogically improved version of my 1987 book "Dynamic Programming: Deterministic and Stochastic Models," published by Prentice-Hall. I have included much new material on deterministic and stochastic shortest path problems, as well as a new chapter on continuous-time optimal control problems and the Pontryagin Maximum Principle, developed from a dynamic programming viewpoint. I have also added a fairly extensive exposition of simulation-based approximation techniques for dynamic programming. These techniques, which are often referred to as "neuro-dynamic programming" or "reinforcement learning," represent a breakthrough in the practical application of dynamic programming to complex problems that involve the dual curse of large dimension and lack of an accurate mathematical model. Other material was also augmented, substantially modified, and updated.

With the new material, however, the book grew so much in size that it became necessary to divide it into two volumes: one on finite horizon, and the other on infinite horizon problems. This division was not only natural in terms of size, but also in terms of style and orientation. The first volume is more oriented towards modeling, and the second is more oriented towards mathematical analysis and computation. I have included in the first volume a final chapter that provides an introductory treatment of infinite horizon problems. The purpose is to make the first volume self-

contained for instructors who wish to cover a modest amount of infinite horizon material in a course that is primarily oriented towards modeling, conceptualization, and finite horizon problems,

Many topics in the book are relatively independent of the others. For example Chapter 2 of Vol. I on shortest path problems can be skipped without loss of continuity, and the same is true for Chapter 3 of Vol. I, which deals with continuous-time optimal control. As a result, the book can be used to teach several different types of courses.

(a) A two-semester course that covers both volumes.

(b) A one-semester course primarily focused on finite horizon problems that covers most of the first volume.

(c) A one-semester course focused on stochastic optimal control that covers Chapters 1, 4, 5, and 6 of Vol. I, and Chapters 1, 2, and 4 of Vol. II.

(d) A one-semester course that covers Chapter 1, about 50% of Chapters 2 through 6 of Vol. I, and about 70% of Chapters 1, 2, and 4 of Vol. II. This is the course I usually teach at MIT.

(e) A one-quarter engineering course that covers the first three chapters and parts of Chapters 4 through 6 of Vol. I.

(f) A one-quarter mathematically oriented course focused on infinite horizon problems that covers Vol. II.

The mathematical prerequisite for the text is knowledge of advanced calculus, introductory probability theory, and matrix-vector algebra. A summary of this material is provided in the appendixes. Naturally, prior exposure to dynamic system theory, control, optimization, or operations research will be helpful to the reader, but based on my experience, the material given here is reasonably self-contained.

The book contains a large number of exercises, and the serious reader will benefit greatly by going through them. Solutions to all exercises are compiled in a manual that is available to instructors from the author. Many thanks are due to the several people who spent long hours contributing to this manual, particularly Steven Shreve, Eric Loiederman, Lakis Polymenakos, and Cynara Wu.

Dynamic programming is a conceptually simple technique that can be adequately explained using elementary analysis. Yet a mathematically rigorous treatment of general dynamic programming requires the complicated machinery of measure-theoretic probability. My choice has been to bypass the complicated mathematics by developing the subject in generality, while claiming rigor only when the underlying probability spaces are countable. A mathematically rigorous treatment of the subject is carried out in my monograph "Stochastic Optimal Control: The Discrete Time

Case," Academic Press, 1978,† coauthored by Steven Shreve. This monograph complements the present text and provides a solid foundation for the subjects developed somewhat informally here.

Finally, I am thankful to a number of individuals and institutions for their contributions to the book. My understanding of the subject was sharpened while I worked with Steven Shreve on our 1978 monograph. My interaction and collaboration with John Tsitsiklis on stochastic shortest paths and approximate dynamic programming have been most valuable. Michael Caramanis, Emmanuel Fernandez-Gaucherand, Pierre Humblet, Lennart Ljung, and John Tsitsiklis taught from versions of the book, and contributed several substantive comments and homework problems. A number of colleagues offered valuable insights and information, particularly David Castanon, Eugene Feinberg, and Krishna Pattipati. NSF provided research support. Prentice-Hall graciously allowed the use of material from my 1987 book. Teaching and interacting with the students at MIT have kept up my interest and excitement for the subject.

Dimitri P. Bertsekas
bertsekas@lids.mit.edu
http://web.mit.edu/dimitrib/www/home.html

† Note added in the 2nd edition: This monograph was republished by Athena Scientific in 1996.

Preface to the Second Edition

This second edition of Vol. II should be viewed as a relatively minor revision of the original. The coverage was expanded in a few areas as follows:

(a) In Chapter 1, material was added on variants of the policy iteration method.

(b) In Chapter 2, the material on neuro-dynamic programming methods was updated and expanded to reflect some recent developments.

(c) In Chapter 4, material was added on some new value iteration methods.

(d) In Chapter 5, the material on semi-Markov problems was revised, with a significant portion simplified and shifted to Volume I.

There are also a few miscellaneous additions and improvements scattered throughout the text. Finally, a new internet-based feature was added to the book, which extends its scope and coverage. Many of the theoretical exercises have been solved in detail and their solutions have been posted in the book's www page

http://www.athenasc.com/dpbook.html

These exercises have been marked with the symbol

I would like to express my thanks to the many colleagues who contributed suggestions for improvement of the second edition.

Dimitri P. Bertsekas
bertsekas@lids.mit.edu
http://web.mit.edu/dimitrib/www/home.html

June, 2001

Preface to the Third Edition

This is a major revision of the 2nd edition, and contains a substantial amount of new material, as well as a major reorganization of old material. The length of the text has increased by more than 50%, and more than half of the old material has been restructured and/or revised. Most of the added material is in four areas.

(a) The coverage of the average cost problem of Chapter 4 has greatly increased in scope and depth. In particular, there is now a full analysis of multi-chain problems, as well as a more extensive analysis of infinite-spaces problems (Section 4.6).

(b) The material on approximate dynamic programming has been collected in Chapter 6. It has been greatly expanded to include new research, thereby supplementing the 1996 book "Neuro-Dynamic Programming."

(c) Contraction mappings and their role in various analyses have been highlighted in new material on infinite state space problems (Sections 1.4, 2.5, and 4.6), and in their use in the approximate dynamic programming material of Chapter 6.

(d) The mathematical measure-theoretic issues that must be addressed for a rigorous theory of stochastic dynamic programming have been illustrated and summarized in an appendix for the benefit of the mathematically oriented reader.

Also some exercises were added and a few sections were revised while preserving their essential content.

I would like to express my thanks to many colleagues who contributed valuable comments. I am particularly thankful to Ciamac Moallemi, Steven Shreve, John Tsitsiklis, and Ben Van Roy, who reviewed some of the new material and each contributed several substantial suggestions. I wish to thank especially Janey Yu who read with great care and keen eye large parts of the book, contributed important analysis and many incisive, substantive comments, and also collaborated with me in research that was included in Chapter 6.

Dimitri P. Bertsekas
http://web.mit.edu/dimitrib/www/home.html
Fall 2006

Preface to the Fourth Edition

This is a major revision of Vol. II, and contains a substantial amount of new material, as well as a reorganization of old material. The length has increased by more than 60% from the third edition, and most of the old material has been restructured and/or revised. Volume II now numbers more than 700 pages and is larger in size than Vol. I. It can arguably be viewed as a new book!

Approximate DP has become the central focal point of Vol. II, and occupies more than half of the book (the last two chapters, and large parts of Chapters 1-3). Thus one may also view Vol. II as a followup to my 1996 book "Neuro-Dynamic Programming" (coauthored with John Tsitsiklis). The present book focuses to a great extent on new research that became available after 1996. On the other hand, the textbook style of the book has been preserved, and some mathematically demanding material has been explained at an intuitive or informal level, while referring to the journal literature or the Neuro-Dynamic Programming book for a more mathematical treatment.

In the process of expansion and reorganization, the design of the book became more modular and suitable for classroom use. The core material, which can be covered in about a third to a half of one semester is Chapter 1 (except for the application-specific Sections 1.3 and 1.4), Chapter 2, and Chapter 6, which are self-contained when taken together. This material focuses on discounted problems, and may be supplemented by parts of Chapter 3 and Section 7.1 on stochastic shortest path problems. Indeed, this comprises half of what the author covers in his MIT class (the remaining half comes from Volume I, including Chapter 6 of that volume that deals with finite horizon approximate DP). The material on average cost problems, given in Chapter 5, and Sections 7.2 and 7.4, and the advanced material on positive and negative DP models (Chapter 4), and Monte Carlo linear algebra (Section 7.3) are terminal subjects that may be covered at the instructor's discretion.

As our focus shifted, we have placed increased emphasis on new or recent research in approximate DP and simulation-based methods, as well as on asynchronous iterative methods, in view of the central role of simulation, which is by nature asynchronous. A lot of this material is an outgrowth of the author's research in the six years since the third edition was published. Some of the highlights, in the order appearing in the book, are:

(a) Computational methods for generalized discounted DP (Sections 2.5 and 2.6), including the asynchronous optimistic policy iteration ma-

terial of Section 2.6.3, and its application to game and minimax problems, constrained policy iteration, and Q-learning.

(b) Policy iteration methods (including asynchronous optimistic versions) for stochastic shortest path problems that involve improper policies (Section 3.4).

(c) Extensive new material on various simulation-based, approximate value and policy iteration methods in Sections 6.3-6.6 (projected equation methods, aggregation methods, and Q-learning).

(d) New simulation techniques for multistep methods, such as geometric and free-form sampling (Sections 6.4.1 and 7.3.3).

(e) Extensive new material on Monte Carlo linear algebra in Section 7.3 (primarily the simulation-based and approximate solution of large systems of linear equations), which extends the DP methodology of approximate policy evaluation.

Much of the research in (a)-(d) is based on my work with Janey (Huizhen) Yu, while the research in (e) is based on my work with Janey Yu and Mengdi Wang. My collaboration with Janey and Mengdi has been very productive and is greatly appreciated. The reader is referred to our joint and individual papers, which describe more fully our research, including material that could not be covered in this book.

I want to express my appreciation to colleagues and collaborators in approximate DP research, who contributed to the book in various ways, particularly Vivek Borkar, Angelia Nedić, and Ben Van Roy. A special thanks goes to John Tsitsiklis, with whom I have interacted extensively through collaboration and sharing of ideas on DP and asynchronous algorithms for more than 30 years. I also wish to acknowledge helpful interactions with many colleagues, including Vivek Farias, Eugene Feinberg, Warren Powell, Martin Puterman, Uriel Rothblum, and Bruno Scherrer. Finally, I want to thank the many students in my DP classes of the last decade, who patiently labored with a textbook under development, and contributed their ideas and experiences through their research projects from a broad variety of application fields.

Dimitri P. Bertsekas
http://web.mit.edu/dimitrib/www/home.html
Spring 2012