Abstract Dynamic Programming

THIRD EDITION

Dimitri P. Bertsekas

Arizona State University

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: Dimitri Bertsekas

Publisher's Cataloging-in-Publication Data

Bertsekas, Dimitri P.
Abstract Dynamic Programming: Third Edition
Includes bibliographical references and index
1. Mathematical Optimization. 2. Dynamic Programming. I. Title.
QA402.5 .B465 2022 519.703 01-75941

ISBN-10: 1-886529-47-7, ISBN-13: 978-1-886529-47-2

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 Department, Stanford University, and the Electrical Engineering Department of the University of Illinois, Urbana. From 1979 to 2019 he was in the faculty of the Electrical Engineering and Computer Science Department of the Massachusetts Institute of Technology (M.I.T.). In 2019, he joined the School of Computing and Augmented Intelligence at the Arizona State University, Tempe, AZ, as Fulton Professor of Computational Decision Making.

Professor Bertsekas' teaching and research have spanned several fields, including deterministic optimization, dynamic programming and stochastic control, large-scale and distributed computation, artificial intelligence, and data communication networks. He has authored or coauthored numerous research papers and nineteen books, several of which are currently used as textbooks in ASU and MIT classes, including "Dynamic Programming and Optimal Control," "Data Networks," "Introduction to Probability," "Nonlinear Programming," and "Reinforcement Learning and Optimal Control."

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 2001 AACC John R. Ragazzini Education Award, the 2009 INFORMS Expository Writing Award, the 2014 AACC Richard Bellman Heritage Award, the 2014 INFORMS Khachiyan Prize for Life-Time Accomplishments in Optimization, the 2015 MOS/SIAM George B. Dantzig Prize, and the 2022 IEEE Control Systems Award. In 2018 he shared with his coauthor, John Tsitsiklis, the 2018 INFORMS John von Neumann Theory Prize for the contributions of the research monographs "Parallel and Distributed Computation" and "Neuro-Dynamic Programming." Professor Bertsekas was elected in 2001 to the United States National Academy of Engineering for "pioneering contributions to fundamental research, practice and education of optimization/control theory, and especially its application to data communication networks."

ATHENA SCIENTIFIC

OPTIMIZATION AND COMPUTATION SERIES

1. Abstract Dynamic Programming, 3rd Edition, by Dimitri P. Bertsekas, 2022, ISBN 978-1-886529-47-2
2. Rollout, Policy Iteration, and Distributed Reinforcement Learning, by Dimitri P. Bertsekas, 2020, ISBN 978-1-886529-07-6
3. Reinforcement Learning and Optimal Control, by Dimitri P. Bertsekas, 2019, ISBN 978-1-886529-39-7
4. Dynamic Programming and Optimal Control, Two-Volume Set, by Dimitri P. Bertsekas, 2017, ISBN 1-886529-08-6
5. Nonlinear Programming, 3rd Edition, by Dimitri P. Bertsekas, 2016, ISBN 1-886529-05-1
6. Convex Optimization Algorithms, by Dimitri P. Bertsekas, 2015, ISBN 978-1-886529-28-1
7. Convex Optimization Theory, by Dimitri P. Bertsekas, 2009, ISBN 978-1-886529-31-1, 256 pages
8. Introduction to Probability, 2nd Edition, by Dimitri P. Bertsekas and John N. Tsitsiklis, 2008, ISBN 978-1-886529-23-6
9. Convex Analysis and Optimization, by Dimitri P. Bertsekas, Angelia Nedić, and Asuman E. Ozdaglar, 2003, ISBN 1-886529-45-0
10. Network Optimization: Continuous and Discrete Models, by Dimitri P. Bertsekas, 1998, ISBN 1-886529-02-7
11. Network Flows and Monotropic Optimization, by R. Tyrrell Rockafellar, 1998, ISBN 1-886529-06-X
12. Introduction to Linear Optimization, by Dimitris Bertsimas and John N. Tsitsiklis, 1997, ISBN 1-886529-19-1
13. Parallel and Distributed Computation: Numerical Methods, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1997, ISBN 1-886529-01-9
14. Neuro-Dynamic Programming, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1996, ISBN 1-886529-10-8
15. Constrained Optimization and Lagrange Multiplier Methods, by Dimitri P. Bertsekas, 1996, ISBN 1-886529-04-3
16. Stochastic Optimal Control: The Discrete-Time Case, by Dimitri P. Bertsekas and Steven E. Shreve, 1996, ISBN 1-886529-03-5

Contents

Preface of the First Edition

This book aims at a unified and economical development of the core theory and algorithms of total cost sequential decision problems, based on the strong connections of the subject with fixed point theory. The analysis focuses on the abstract mapping that underlies dynamic programming (DP for short) and defines the mathematical character of the associated problem. Our discussion centers on two fundamental properties that this mapping may have: *monotonicity* and (weighted sup-norm) *contraction*. It turns out that the nature of the analytical and algorithmic DP theory is determined primarily by the presence or absence of these two properties, and the rest of the problem's structure is largely inconsequential.

In this book, with some minor exceptions, we will assume that monotonicity holds. Consequently, we organize our treatment around the contraction property, and we focus on four main classes of models:

(a) **Contractive models**, discussed in Chapter 2, which have the richest and strongest theory, and are the benchmark against which the theory of other models is compared. Prominent among these models are discounted stochastic optimal control problems. The development of these models is quite thorough and includes the analysis of recent approximation algorithms for large-scale problems (neuro-dynamic programming, reinforcement learning).

(b) **Semicontractive models**, discussed in Chapter 3 and parts of Chapter 4. The term "semicontractive" is used qualitatively here, to refer to a variety of models where some policies have a regularity/contraction-like property but others do not. A prominent example is stochastic shortest path problems, where one aims to drive the state of a Markov chain to a termination state at minimum expected cost. These models also have a strong theory under certain conditions, often nearly as strong as those of the contractive models.

(c) **Noncontractive models**, discussed in Chapter 4, which rely on just monotonicity. These models are more complex than the preceding ones and much of the theory of the contractive models generalizes in weaker form, if at all. For example, in general the associated Bellman equation need not have a unique solution, the value iteration method may work starting with some functions but not with others, and the policy iteration method may not work at all. Infinite horizon examples of these models are the classical positive and negative DP problems, first analyzed by Dubins and Savage, Blackwell, and

Strauch, which are discussed in various sources. Some new semicontractive models are also discussed in this chapter, further bridging the gap between contractive and noncontractive models.

(d) **Restricted policies and Borel space models**, which are discussed in Chapter 5. These models are motivated in part by the complex measurability questions that arise in mathematically rigorous theories of stochastic optimal control involving continuous probability spaces. Within this context, the admissible policies and DP mapping are restricted to have certain measurability properties, and the analysis of the preceding chapters requires modifications. Restricted policy models are also useful when there is a special class of policies with favorable structure, which is "closed" with respect to the standard DP operations, in the sense that analysis and algorithms can be confined within this class.

We do not consider average cost DP problems, whose character bears a much closer connection to stochastic processes than to total cost problems. We also do not address specific stochastic characteristics underlying the problem, such as for example a Markovian structure. Thus our results apply equally well to Markovian decision problems and to sequential minimax problems. While this makes our development general and a convenient starting point for the further analysis of a variety of different types of problems, it also ignores some of the interesting characteristics of special types of DP problems that require an intricate probabilistic analysis.

Let us describe the research content of the book in summary, deferring a more detailed discussion to the end-of-chapter notes. A large portion of our analysis has been known for a long time, but in a somewhat fragmentary form. In particular, the contractive theory, first developed by Denardo [Den67], has been known for the case of the unweighted sup-norm, but does not cover the important special case of stochastic shortest path problems where all policies are proper. Chapter 2 transcribes this theory to the weighted sup-norm contraction case. Moreover, Chapter 2 develops extensions of the theory to approximate DP, and includes material on asynchronous value iteration (based on the author's work [Ber82], [Ber83]), and asynchronous policy iteration algorithms (based on the author's joint work with Huizhen (Janey) Yu [BeY10a], [BeY10b], [YuB11a]). Most of this material is relatively new, having been presented in the author's recent book [Ber12a] and survey paper [Ber12b], with detailed references given there. The analysis of infinite horizon noncontractive models in Chapter 4 was first given in the author's paper [Ber77], and was also presented in the book by Bertsekas and Shreve [BeS78], which in addition contains much of the material on finite horizon problems, restricted policies models, and Borel space models. These were the starting point and main sources for our development.

The new research presented in this book is primarily on the semi-

contractive models of Chapter 3 and parts of Chapter 4. Traditionally, the theory of total cost infinite horizon DP has been bordered by two extremes: discounted models, which have a contractive nature, and positive and negative models, which do not have a contractive nature, but rely on an enhanced monotonicity structure (monotone increase and monotone decrease models, or in classical DP terms, positive and negative models). Between these two extremes lies a gray area of problems that are not contractive, and either do not fit into the categories of positive and negative models, or possess additional structure that is not exploited by the theory of these models. Included are stochastic shortest path problems, search problems, linear-quadratic problems, a host of queueing problems, multiplicative and exponential cost models, and others. Together these problems represent an important part of the infinite horizon total cost DP landscape. They possess important theoretical characteristics, not generally available for positive and negative models, such as the uniqueness of solution of Bellman's equation within a subset of interest, and the validity of useful forms of value and policy iteration algorithms.

Our semicontractive models aim to provide a unifying abstract DP structure for problems in this gray area between contractive and noncontractive models. The analysis is motivated in part by stochastic shortest path problems, where there are two types of policies: *proper*, which are the ones that lead to the termination state with probability one from all starting states, and *improper*, which are the ones that are not proper. Proper and improper policies can also be characterized through their Bellman equation mapping: for the former this mapping is a contraction, while for the latter it is not. In our more general semicontractive models, policies are also characterized in terms of their Bellman equation mapping, through a notion of *regularity*, which generalizes the notion of a proper policy and is related to classical notions of asymptotic stability from control theory.

In our development a policy is regular within a certain set if its cost function is the unique asymptotically stable equilibrium (fixed point) of the associated DP mapping within that set. *We assume that some policies are regular while others are not*, and impose various assumptions to ensure that attention can be focused on the regular policies. From an analytical point of view, this brings to bear the theory of fixed points of monotone mappings. From the practical point of view, this allows application to a diverse collection of interesting problems, ranging from stochastic shortest path problems of various kinds, where the regular policies include the proper policies, to linear-quadratic problems, where the regular policies include the stabilizing linear feedback controllers.

The definition of regularity is introduced in Chapter 3, and its theoretical ramifications are explored through extensions of the classical stochastic shortest path and search problems. In Chapter 4, semicontractive models are discussed in the presence of additional monotonicity structure, which brings to bear the properties of positive and negative DP models. With the

aid of this structure, the theory of semicontractive models can be strengthened and can be applied to several additional problems, including risk-sensitive/exponential cost problems.

The book has a theoretical research monograph character, but requires a modest mathematical background for all chapters except the last one, essentially a first course in analysis. Of course, prior exposure to DP will definitely be very helpful to provide orientation and context. A few exercises have been included, either to illustrate the theory with examples and counterexamples, or to provide applications and extensions of the theory. Solutions of all the exercises can be found in Appendix D, at the book's internet site

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

and at the author's web site

http://web.mit.edu/dimitrib/www/home.html

Additional exercises and other related material may be added to these sites over time.

I would like to express my appreciation to a few colleagues for interactions, recent and old, which have helped shape the form of the book. My collaboration with Steven Shreve on our 1978 book provided the motivation and the background for the material on models with restricted policies and associated measurability questions. My collaboration with John Tsitsiklis on stochastic shortest path problems provided inspiration for the work on semicontractive models. My collaboration with Janey (Huizhen) Yu played an important role in the book's development, and is reflected in our joint work on asynchronous policy iteration, on perturbation models, and on risk-sensitive models. Moreover Janey contributed significantly to the material on semicontractive models with many insightful suggestions. Finally, I am thankful to Mengdi Wang, who went through portions of the book with care, and gave several helpful comments.

Dimitri P. Bertsekas

Spring 2013

Preface to the Second Edition

The second edition aims primarily to amplify the presentation of the semicontractive models of Chapter 3 and Chapter 4, and to supplement it with a broad spectrum of research results that I obtained and published in journals and reports since the first edition was written. As a result, the size of this material more than doubled, and the size of the book increased by about 40%.

In particular, I have thoroughly rewritten Chapter 3, which deals with semicontractive models where stationary regular policies are sufficient. I expanded and streamlined the theoretical framework, and I provided new analyses of a number of shortest path-type applications (deterministic, stochastic, affine monotonic, exponential cost, and robust/minimax), as well as several types of optimal control problems with continuous state space (including linear-quadratic, regulation, and planning problems).

In Chapter 4, I have extended the notion of regularity to nonstationary policies (Section 4.4), aiming to explore the structure of the solution set of Bellman's equation, and the connection of optimality with other structural properties of optimal control problems. As an application, I have discussed in Section 4.5 the relation of optimality with classical notions of stability and controllability in continuous-spaces deterministic optimal control. In Section 4.6, I have similarly extended the notion of a proper policy to continuous-spaces stochastic shortest path problems.

I have also revised Chapter 1 a little (mainly with the addition of Section 1.2.5 on the relation between proximal algorithms and temporal difference methods), added to Chapter 2 some analysis relating to λ-policy iteration and randomized policy iteration algorithms (Section 2.5.3), and I have also added several new exercises (with complete solutions) to Chapters 1-4. Additional material relating to various applications can be found in some of my journal papers, reports, and video lectures on semicontractive models, which are posted at my web site.

In addition to the changes in Chapters 1-4, I have also eliminated from the second edition the analysis that deals with restricted policies (Chapter 5 and Appendix C of the first edition). This analysis is motivated in part by the complex measurability questions that arise in mathematically rigorous theories of stochastic optimal control with Borel state and control spaces. This material is covered in Chapter 6 of the monograph by Bertsekas and Shreve [BeS78], and followup research on the subject has been limited. Thus, I decided to just post Chapter 5 and Appendix C of the first

edition at the book's web site (40 pages), and omit them from the second edition. As a result of this choice, the entire book now requires only a modest mathematical background, essentially a first course in analysis and in elementary probability.

The range of applications of dynamic programming has grown enormously in the last 25 years, thanks to the use of approximate simulation-based methods for large and challenging problems. Because approximations are often tied to special characteristics of specific models, their coverage in this book is limited to general discussions in Chapter 1 and to error bounds given in Chapter 2. However, much of the work on approximation methods so far has focused on finite-state discounted, and relatively simple deterministic and stochastic shortest path problems, for which there is solid and robust analytical and algorithmic theory (part of Chapters 2 and 3 in this monograph). As the range of applications becomes broader, I expect that the level of mathematical understanding projected in this book will become essential for the development of effective and reliable solution methods. In particular, much of the new material in this edition deals with infinite-state and/or complex shortest path type-problems, whose approximate solution will require new methodologies that transcend the current state of the art.

Dimitri P. Bertsekas

January 2018

Preface to the Third Edition

The third edition is based on the same theoretical framework as the second edition, but contains two major additions. The first is to highlight the central role of abstract DP methods in the conceptualization of reinforcement learning and approximate DP methods, as described in the author's recent book "Lessons from AlphaZero for Optimal, Model Predictive, and Adaptive Control," Athena Scientific, 2022. The main idea here is that approximation in value space with one-step lookahead amounts to a step of Newton's method for solving the abstract Bellman's equation. This material is included in summary form in view of its strong reliance on abstract DP visualization. Our presentation relies primarily on geometric illustrations rather than mathematical analysis, and is given in Section 1.3.

The second addition is a new Chapter 5 on abstract DP methods for minimax and zero sum game problems, which is based on the author's recent paper [Ber21c]. A primary motivation here is the resolution of some long-standing convergence difficulties of the "natural" policy iteration algorithm, which have been known since the Pollatschek and Avi-Itzhak method [PoA69] for finite-state Markov games. Mathematically, this "natural" algorithm is a form of Newton's method for solving the corresponding Bellman's equation, but Newton's method, contrary to the case of single-player DP problems, is not globally convergent in the case of a minimax problem, because the Bellman operator may have components that are neither convex nor concave. Our approach in Chapter 5 has been to introduce a special type of abstract Bellman operator for minimax problems, and modify the standard PI algorithm along the lines of the asynchronous optimistic PI algorithm of Section 2.6.3, which involves a parametric contraction mapping with a uniform fixed point.

The third edition also contains a number of small corrections and editorial changes. The author wishes to thank the contributions of several colleagues in this regard, and particularly Yuchao Li, who proofread with care large portions of the book.

Dimitri P. Bertsekas

February 2022