Abstract Dynamic Programming

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

Publisher's Cataloging-in-Publication Data

Bertsekas, Dimitri P.
Abstract Dynamic Programming
Includes bibliographical references and index
1. Mathematical Optimization. 2. Dynamic Programming. I. Title.

QA402.5 .B465 2013 519.703 01-75941

ISBN-10: 1-886529-42-6, ISBN-13: 978-1-886529-42-7

Contents

Preface

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://www.mit.edu:8001/people/dimitrib/books.htm

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 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

dimitrib@mit.edu

Spring 2013