Stochastic Optimal Control: The Discrete-Time Case

Dimitri P. Bertsekas and Steven E. Shreve

WWW site for book information and orders
http://world.std.com/~athenasc/

Athena Scientific, Belmont, Massachusetts

Athena Scientific
Post Office Box 391
Belmont, Mass. 02178-9998
U.S.A.

Email: athenasc@world.std.com
WWW information and orders: http://world.std.com/~athenasc/

Cover Design: *Ann Gallager*

Originally published by Academic Press, Inc., in 1978

OPTIMIZATION AND NEURAL COMPUTATION SERIES

1. Dynamic Programming and Optimal Control, Vols. I and II, by Dimitri P. Bertsekas, 1995
2. Nonlinear Programming, by Dimitri P. Bertsekas, 1995
3. Neuro-Dynamic Programming, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1996
4. Constrained Optimization and Lagrange Multiplier Methods, by Dimitri P. Bertsekas, 1996
5. Stochastic Optimal Control: The Discrete-Time Case by Dimitri P. Bertsekas and Steven E. Shreve, 1996

Publisher's Cataloging-in-Publication Data

Bertsekas, Dimitri P.
Stochastic Optimal Control: The Discrete-Time Case
Includes bibliographical references and index
1. Dynamic Programming. 2. Stochastic Processes. 3. Measure Theory. I. Shreve, Steven E., joint author. II. Title.
T57.83.B49 1996 519.7'03 96-80191

ISBN 1-886529-03-5

To
Joanna
and
Steve's Mom and Dad

Contents

Preface

This monograph is the outgrowth of research carried out at the University of Illinois over a three-year period beginning in the latter half of 1974. The objective of the monograph is to provide a unifying and mathematically rigorous theory for a broad class of dynamic programming and discrete-time stochastic optimal control problems. It is divided into two parts, which can be read independently.

Part I provides an analysis of dynamic programming models in a unified framework applicable to deterministic optimal control, stochastic optimal control, minimax control, sequential games, and other areas. It resolves the *structural questions* associated with such problems, i.e., it provides results that draw their validity exclusively from the sequential nature of the problem. Such results hold for models where measurability of various objects is of no essential concern, for example, in deterministic problems and stochastic problems defined over a countable probability space. The starting point for the analysis is the mapping defining the dynamic programming algorithm. A single abstract problem is formulated in terms of this mapping and counterparts of nearly all results known for deterministic optimal control problems are derived. A new stochastic optimal control model based on outer integration is also introduced in this

part. It is a broadly applicable model and requires no topological assumptions. We show that all the results of Part I hold for this model.

Part II resolves the *measurability questions* associated with stochastic optimal control problems with perfect and imperfect state information. These questions have been studied over the past fifteen years by several researchers in statistics and control theory. As we explain in Chapter 1, the approaches that have been used are either limited by restrictive assumptions such as compactness and continuity or else they are not sufficiently powerful to yield results that are as strong as their structural counterparts. These deficiencies can be traced to the fact that the class of policies considered is not sufficiently rich to ensure the existence of everywhere optimal or ϵ-optimal policies except under restrictive assumptions. In our work we have appropriately enlarged the space of admissible policies to include *universally measurable policies*. This guarantees the existence of ϵ-optimal policies and allows, for the first time, the development of a general and comprehensive theory which is as powerful as its deterministic counterpart.

We mention, however, that the class of universally measurable policies is not the smallest class of policies for which these results are valid. The smallest such class is the class of *limit measurable policies* discussed in Section 11.1. The σ-algebra of limit measurable sets (or C-sets) is defined in a constructive manner involving transfinite induction that, from a set theoretic point of view, is more satisfying than the definition of the universal σ-algebra. We believe, however, that the majority of readers will find the universal σ-algebra and the methods of proof associated with it more understandable, and so we devote the main body of Part II to models with universally measurable policies.

Parts I and II are related and complement each other. Part II makes extensive use of the results of Part I. However, the special forms in which these results are needed are also available in other sources (e.g., the textbook by Bertsekas [B4]). Each time we make use of such a result, we refer to both Part I and the Bertsekas textbook, so that Part II can be read independently of Part I. The developments in Part II show also that stochastic optimal control problems with measurability restrictions on the admissible policies can be embedded within the framework of Part I, thus demonstrating the broad scope of the formulation given there.

The monograph is intended for applied mathematicians, statisticians, and mathematically oriented analysts in engineering, operations research, and related fields. We have assumed throughout that the reader is familiar with the basic notions of measure theory and topology. In other respects, the monograph is self-contained. In particular, we have provided all necessary background related to Borel spaces and analytic sets.

Acknowledgments

This research was begun while we were with the Coordinated Science Laboratory of the University of Illinois and concluded while Shreve was with the Departments of Mathematics and Statistics of the University of California at Berkeley. We are grateful to these institutions for providing support and an atmosphere conducive to our work, and we are also grateful to the National Science Foundation for funding the research. We wish to acknowledge the aid of Joseph Doob, who guided us into the literature on analytic sets, and of John Addison, who pointed out the existing work on the limit σ-algebra. We are particularly indebted to David Blackwell, who inspired us by his pioneering work on dynamic programming in Borel spaces, who encouraged us as our own investigation was proceeding, and who showed us Example 9.2. Chapter 9 is an expanded version of our paper "Universally Measurable Policies in Dynamic Programming" published in *Mathematics of Operations Research*. The permission of The Institute of Management Sciences to include this material is gratefully acknowledged. Finally we wish to thank Rose Harris and Dee Wrather for their excellent typing of the manuscript.

Chapter 1

Introduction

1.1 Structure of Sequential Decision Models

Sequential decision models are mathematical abstractions of situations in which decisions must be made in several stages while incurring a certain cost at each stage. Each decision may influence the circumstances under which future decisions will be made, so that if total cost is to be minimized, one must balance his desire to minimize the cost of the present decision against his desire to avoid future situations where high cost is inevitable.

A classical example of this situation, in which we treat profit as negative cost, is portfolio management. An investor must balance his desire to achieve immediate return, possibly in the form of dividends, against a desire to avoid investments in areas where low long-run yield is probable. Other examples can be drawn from inventory management, reservoir control, sequential analysis, hypothesis testing, and, by discretizing a continuous problem, from control of a large variety of physical systems subject to random disturbances. For an extensive set of sequential decision models, see Bellman [B1], Bertsekas [B4], Dynkin and Juskevič [D8], Howard [H7], Wald [W2], and the references contained therein.

Dynamic programming (DP for short) has served for many years as the principal method for analysis of a large and diverse group of sequential

decision problems. Examples are deterministic and stochastic optimal control problems, Markov and semi-Markov decision problems, minimax control problems, and sequential games. While the nature of these problems may vary widely, their underlying structures turn out to be very similar. In all cases, the cost corresponding to a policy and the basic iteration of the DP algorithm may be described by means of a certain mapping which differs from one problem to another in details which to a large extent are inessential. Typically, this mapping summarizes all the data of the problem and determines all quantities of interest to the analyst. Thus, in problems with a finite number of stages, this mapping may be used to obtain the optimal cost function for the problem as well as to compute an optimal or ε-optimal policy through a finite number of steps of the DP algorithm. In problems with an infinite number of stages, one hopes that the sequence of functions generated by successive application of the DP iteration converges in some sense to the optimal cost function for the problem. Furthermore, all basic results of an analytical and computational nature can be expressed in terms of the underlying mapping defining the DP algorithm. Thus by taking this mapping as a starting point one can provide powerful analytical results which are applicable to a large collection of sequential decision problems.

To illustrate our viewpoint, let us consider formally a deterministic optimal control problem. We have a discrete-time system described by the system equation

$$x_{k+1} = f(x_k, u_k), \tag{1}$$

where x_k and x_{k+1} represent a state and its succeeding state and will be assumed to belong to some state space S; u_k represents a control variable chosen by the decisionmaker in some constraint set $U(x_k)$, which is in turn a subset of some control space C. The cost incurred at the kth stage is given by a function $g(x_k, u_k)$. We seek a finite sequence of control functions $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ (also referred to as a *policy*) which minimizes the total cost over N stages. The functions μ_k map S into C and must satisfy $\mu_k(x) \in U(x)$ for all $x \in S$. Each function μ_k specifies the control $u_k = \mu_k(x_k)$ that will be chosen when at the kth stage the state is x_k. Thus the total cost corresponding to a policy $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ and initial state x_0 is given by

$$J_{N,\pi}(x_0) = \sum_{k=0}^{N-1} g[x_k, \mu_k(x_k)], \tag{2}$$

where the states $x_1, x_2, \ldots, x_{N-1}$ are generated from x_0 and π via the system equation

$$x_{k+1} = f[x_k, \mu_k(x_k)], \qquad k = 0, \ldots, N-2. \tag{3}$$

Corresponding to each initial state x_0 and policy π, there is a sequence of control variables $u_0, u_1, \ldots, u_{N-1}$, where $u_k = \mu_k(x_k)$ and x_k is generated by

(3). Thus an alternative formulation of the problem would be to select a sequence of control variables minimizing $\sum_{k=0}^{N-1} g(x_k, u_k)$ rather than a policy π minimizing $J_{N,\pi}(x_0)$. The formulation we have given here, however, is more consistent with the DP framework we wish to adopt.

As is well known, the DP algorithm for the preceding problem is given by

$$J_0(x) = 0, \tag{4}$$

$$J_{k+1}(x) = \inf_{u \in U(x)} \{g(x,u) + J_k[f(x,u)]\}, \qquad k = 0, \ldots, N-1, \tag{5}$$

and the optimal cost $J^*(x_0)$ for the problem is obtained at the Nth step, i.e.,

$$J^*(x_0) = \inf_{\pi} J_{N,\pi}(x_0) = J_N(x_0).$$

One may also obtain the value $J_{N,\pi}(x_0)$ corresponding to any $\pi = (\mu_0, \mu_1, \ldots, \mu_{N-1})$ at the Nth step of the algorithm

$$J_{0,\pi}(x) = 0, \tag{6}$$

$$J_{k+1,\pi}(x) = g[x, \mu_{N-k-1}(x)] + J_{k,\pi}[f(x, \mu_{N-k-1}(x))], \qquad k = 0, \ldots, N-1. \tag{7}$$

Now it is possible to formulate the previous problem as well as to describe the DP algorithm (4)–(5) by means of the mapping H given by

$$H(x,u,J) = g(x,u) + J[f(x,u)]. \tag{8}$$

Let us define the mapping T by

$$T(J)(x) = \inf_{u \in U(x)} H(x,u,J) \tag{9}$$

and, for any function $\mu: S \to C$, define the mapping T_μ by

$$T_\mu(J)(x) = H[x, \mu(x), J]. \tag{10}$$

Both T and T_μ map the set of real-valued (or perhaps extended real-valued) functions on S into itself. Then in view of (6)–(7), we may write the cost functional $J_{N,\pi}(x_0)$ of (2) as

$$J_{N,\pi}(x_0) = (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})(J_0)(x_0), \tag{11}$$

where J_0 is the zero function on S $[J_0(x) = 0 \ \forall x \in S]$ and $(T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})$ denotes the composition of the mappings $T_{\mu_0}, T_{\mu_1}, \ldots, T_{\mu_{N-1}}$. Similarly the DP algorithm (4)–(5) may be described by

$$J_{k+1}(x) = T(J_k)(x), \qquad k = 0, \ldots, N-1, \tag{12}$$

and we have

$$\inf_{\pi} J_{N,\pi}(x_0) = T^N(J_0)(x_0),$$

where T^N is the composition of T with itself N times. Thus *both the problem and the major algorithmic procedure relating to it can be expressed in terms of the mappings T and T_μ.*

One may also consider an infinite horizon version of the problem whereby we seek a sequence $\pi = (\mu_0, \mu_1, \ldots)$ that minimizes

$$J_\pi(x_0) = \lim_{N\to\infty} \sum_{k=0}^{N-1} g[x_k, \mu_k(x_k)] = \lim_{N\to\infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})(J_0)(x_0) \quad (13)$$

subject to the system equation constraint (3). In this case one needs, of course, to make assumptions which ensure that the limit in (13) is well defined for each π and x_0. Under appropriate assumptions, the optimal cost function defined by

$$J^*(x) = \inf_{\pi} J_\pi(x)$$

can be shown to satisfy Bellman's functional equation given by

$$J^*(x) = \inf_{u\in U(x)} \{g(x,u) + J^*[f(x,u)]\}.$$

Equivalently

$$J^*(x) = T(J^*)(x) \qquad \forall x \in S,$$

i.e., J^* is a fixed point of the mapping T. Most of the infinite horizon results of analytical interest center around this equation. Other questions relate to the existence and characterization of optimal policies or nearly optimal policies and to the validity of the equation

$$J^*(x) = \lim_{N\to\infty} T^N(J_0)(x) \qquad \forall x \in S, \quad (14)$$

which says that the DP algorithm yields in the limit the optimal cost function for the problem. Again the problem and the basic analytical and computational results relating to it can be expressed in terms of the mappings T and T_μ.

The deterministic optimal control problem just described is representative of a plethora of sequential optimization problems of practical interest which may be formulated in terms of mappings similar to the mapping H of (8). As shall be described in Chapter 2, one can formulate in the same manner stochastic optimal control problems, minimax control problems, and others. *The objective of Part I is to provide a common analytical frame-*

work for all these problems and derive in a broadly applicable form all the results which draw their validity exclusively from the basic sequential structure of the decision-making process. This is accomplished by taking as a starting point a mapping H such as the one of (8) and deriving all major analytical and computational results within a generalized setting. The results are subsequently specialized to five particular models described in Section 2.3: *deterministic optimal control problems, three types of stochastic optimal control problems (countable disturbance space, outer integral formulation, and multiplicative cost functional), and minimax control problems.*

1.2 Discrete-Time Stochastic Optimal Control Problems—Measurability Questions

The theory of Part I is not adequate by itself to provide a complete analysis of stochastic optimal control problems, the treatment of which is the major objective of this book. The reason is that when such problems are formulated over uncountable probability spaces nontrivial measurability restrictions must be placed on the admissible policies unless we resort to an outer integration framework.

A discrete-time stochastic optimal control problem is obtained from the deterministic problem of the previous section when the system includes a stochastic disturbance w_k in its description. Thus (1) is replaced by

$$x_{k+1} = f(x_k, u_k, w_k) \tag{15}$$

and the cost per stage becomes $g(x_k, u_k, w_k)$. The disturbance w_k is a member of some probability space $(W, \mathscr{F})$ and has distribution $p(dw_k|x_k, u_k)$. Thus the control variable u_k exercises influence over the transition from x_k to x_{k+1} in two places, once in the system equation (15) and again as a parameter in the distribution of the disturbance w_k. Likewise, the control u_k influences the cost at two points. This is a redundancy in the system equation model given above which will be eliminated in Chapter 8 when we introduce the transition kernel and reduced one-stage cost function and thereby convert to a model frequently adopted in the statistics literature (see, e.g., Blackwell [B9]; Strauch [S14]). The system equation model is more common in engineering literature and generally more convenient in applications, so we are taking it as our starting point. The transition kernel and reduced one-stage cost function are technical devices which eliminate the disturbance space $(W, \mathscr{F})$ from consideration and make the model more suitable for analysis. We take pains initially to point out how properties of the original system carry over into properties of the transition kernel and reduced one-stage cost function (see the remarks following Definitions 8.1 and 8.7).

Stochastic optimal control is distinguished from its deterministic counterpart by the concern with when information becomes available. In deterministic control, to each initial state and policy there corresponds a sequence of control variables $(u_0, \ldots, u_{N-1})$ which can be specified beforehand, and the resulting states of the system are determined by (1). In contrast, if the control variables are specified beforehand for a stochastic system, the decisionmaker may realize in the course of the system evolution that unexpected states have appeared and the specified control variables are no longer appropriate. Thus it is essential to consider *policies* $\pi = (\mu_0, \ldots, \mu_{N-1})$, where μ_k is a function from history to control. If x_0 is the initial state, $u_0 = \mu_0(x_0)$ is taken to be the first control. If the states and controls $(x_0, u_0, \ldots, u_{k-1}, x_k)$ have occurred, the control

$$u_k = \mu_k(x_0, u_0, \ldots, u_{k-1}, x_k) \tag{16}$$

is chosen. We require that the control constraint

$$\mu_k(x_0, u_0, \ldots, u_{k-1}, x_k) \in U(x_k)$$

be satisfied for every $(x_0, u_0, \ldots, u_{k-1}, x_k)$ and k. In this way the decisionmaker utilizes the full information available to him at each stage. Rather than choosing a sequence of control variables, the decisionmaker attempts to choose a policy which minimizes the total expected cost of the system operation. Actually, we will show that for most cases it is sufficient to consider only *Markov policies*, those for which the corresponding controls u_k depend only on the current state x_k rather than the entire history $(x_0, u_0, \ldots, u_{k-1}, x_k)$. This is the type of policy encountered in Section 1.1.

The analysis of the stochastic decision model outlined here can be fairly well divided into two categories—*structural considerations* and *measurability considerations*. Structural analysis consists of all those results which can be obtained if measurability of all functions and sets arising in the problem is of no real concern; for example, if the model is deterministic or, more generally, if the disturbance space W is countable. In Part I structural results are derived using mappings H, T_μ, and T of the kind considered in the previous section. Measurability analysis consists of showing that the structural results remain valid even when one places nontrivial measurability restrictions on the set of admissible policies. The work in Part II consists primarily of measurability analysis relying heavily on structural results developed in Part I as well as in other sources (e.g., Bertsekas [B4]).

One can best illustrate this dichotomy of analysis by the finite horizon DP algorithm considered by Bellman [B1]:

$$J_0(x) = 0, \tag{17}$$

$$J_{k+1}(x) = \inf_{u \in U(x)} E\{g(x, u, w) + J_k[f(x, u, w)]\}, \qquad k = 0, \ldots, N-1, \tag{18}$$

where the expectation is with respect to $p(dw|x,u)$. This is the stochastic counterpart of the deterministic DP algorithm (4)–(5).

It is reasonable to expect that $J_k(x)$ is the optimal cost of operating the system over k stages when the initial state is x, and that if $\mu_k(x)$ achieves the infimum in (18) for every x and $k = 0,\ldots,N-1$, then $\pi = (\mu_0,\ldots,\mu_{N-1})$ is an optimal policy for every initial state x. If there are no measurability considerations, this is indeed the case under very mild assumptions, as shall be shown in Chapter 3. Yet it is a major task to properly formulate the stochastic control problem and demonstrate that the DP algorithm (17)–(18) makes sense in a measure-theoretic framework. One of the difficulties lies in showing that the expression in curly braces in (18) is measurable in some sense. Thus we must establish measurability properties for the functions J_k. Related to this is the need to balance the measurability of policies (necessary so the expected cost corresponding to a policy can be defined) against a desire to be able to select at or near the infimum in (18). We illustrate these difficulties by means of a simple two-stage example.

TWO-STAGE PROBLEM Consider the following sequence of events:

(a) An initial state $x_0 \in R$ is generated (R is the real line).

(b) Knowing x_0, the decisionmaker selects a control $u_0 \in R$.

(c) A state $x_1 \in R$ is generated according to a known probability measure $p(dx_1|x_0,u_0)$ on $\mathcal{B}_R$, the Borel subsets of R, depending on x_0, u_0. [In terms of our earlier model, this corresponds to a system equation of the form $x_1 = w_0$ and $p(dw_0|x_0,u_0) = p(dx_1|x_0,u_0)$.]

(d) Knowing x_1, the decisionmaker selects a control $u_1 \in R$.

Given $p(dx_1|x_0,u_0)$ for every $(x_0,u_0) \in R^2$ and a function $g: R^2 \to R$, the problem is to find a policy $\pi = (\mu_0,\mu_1)$ consisting of two functions $\mu_0: R \to R$ and $\mu_1: R \to R$ that minimizes

$$J_\pi(x_0) = \int g[x_1, \mu_1(x_1)]\, p(dx_1|x_0, \mu_0(x_0)). \tag{19}$$

We temporarily postpone a discussion of restrictions (if any) that must be placed on g, μ_0, and μ_1 in order for the integral in (19) to be well defined. In terms of our earlier model, the function g gives the cost for the second stage while we assume no cost for the first stage.

The DP algorithm associated with the problem is

$$J_1(x_1) = \inf_{u_1} g(x_1,u_1), \tag{20}$$

$$J_2(x_0) = \inf_{u_0} \int J_1(x_1)\, p(dx_1|x_0,u_0), \tag{21}$$

and, assuming that $J_2(x_0) > -\infty$, $J_1(x_1) > -\infty$ for all $x_0 \in R$, $x_1 \in R$, the

results one expects to be true are:

R.1 There holds

$$J_2(x_0) = \inf_{\pi} J_\pi(x_0) \qquad \forall x_0 \in R.$$

R.2 Given $\varepsilon > 0$, there is an (everywhere) ε-optimal policy, i.e., a policy π_ε such that

$$J_{\pi_\varepsilon}(x_0) \leq \inf_{\pi} J_\pi(x_0) + \varepsilon \qquad \forall x_0 \in R.$$

R.3 If the infimum in (20) and (21) is attained for all $x_1 \in R$ and $x_0 \in R$, then there exists a policy that is optimal for every $x_0 \in R$.

R.4 If $\mu_1^*(x_1)$ and $\mu_0^*(x_0)$, respectively, attain the infimum in (20) and (21) for all $x_1 \in R$ and $x_0 \in R$, then $\pi^* = (\mu_0^*, \mu_1^*)$ is optimal for every $x_0 \in R$, i.e.,

$$J_{\pi^*}(x_0) = \inf_{\pi} J_\pi(x_0) \qquad \forall x_0 \in R.$$

A formal derivation of R.1 consists of the following steps:

$$\inf_{\pi} J_\pi(x_0) = \inf_{\mu_0} \inf_{\mu_1} \int g[x_1, \mu_1(x_1)] p(dx_1 | x_0, \mu_0(x_0)) \tag{22a}$$

$$= \inf_{\mu_0} \int \left\{ \inf_{u_1} g(x_1, u_1) \right\} p(dx_1 | x_0, \mu_0(x_0)) \tag{22b}$$

$$= \inf_{\mu_0} \int J_1(x_1) p(dx_1 | x_0, \mu_0(x_0))$$

$$= \inf_{u_0} \int J_1(x_1) p(dx_1 | x_0, u_0) = J_2(x_0).$$

Similar formal derivations can be given for R.2, R.3, and R.4.

The following points need to be justified in order to make the preceding derivation meaningful and mathematically rigorous.

(a) In (22a), g and μ_1 must be such that $g[x_1, \mu_1(x_1)]$ can be integrated in a well-defined manner.

(b) In (22b), the interchange of infimization and integration must be legitimate. Furthermore g must be such that $J_1(x_1)\,[= \inf_{u_1} g(x_1, u_1)]$ can be integrated in a well-defined manner.

We first observe that if, for each (x_0, u_0), $p(dx_1 | x_0, u_0)$ has *countable support*, i.e., is concentrated on a countable number of points, then integration in (22a) and (22b) reduces to infinite summation. Thus there is no need to impose measurability restrictions on g, μ_0, and μ_1, and the interchange of infimization and integration in (22b) is justified in view of the assumption

$\inf_{u_1} g(x_1, u_1) > -\infty$ for all $x_1 \in R$. (For $\varepsilon > 0$, take $\mu_\varepsilon : R \to R$ such that

$$g[x_1, \mu_\varepsilon(x_1)] \leq \inf_{u_1} g(x_1, u_1) + \varepsilon \qquad \forall x_1 \in R. \tag{23}$$

Then

$$\begin{aligned} \inf_{\mu_1} \int g[x_1, \mu_1(x_1)] p(dx_1 | x_0, \mu_0(x_0)) &\leq \int g[x_1, \mu_\varepsilon(x_1)] p(dx_1 | x_0, \mu_0(x_0)) \\ &\leq \int \inf_{u_1} g(x_1, u_1) p(dx_1 | x_0, \mu_0(x_0)) + \varepsilon. \end{aligned} \tag{24}$$

Since $\varepsilon > 0$ is arbitrary, it follows that

$$\inf_{\mu_1} \int g[x_1, \mu_1(x_1)] p(dx_1 | x_0, \mu_0(x_0)) \leq \int \left\{ \inf_{u_1} g(x_1, u_1) \right\} p(dx_1 | x_0, \mu_0(x_0)).$$

The reverse inequality is clear, and the result follows.) A similar argument proves R.2, while R.3 and R.4 are trivial in view of the fact that there are no measurability restrictions on μ_0 and μ_1.

If $p(dx_1 | x_0, u_0)$ does not have countable support, there are two main approaches. The first is to *expand the notion of integration*, and the second is to *restrict g, μ_0, and μ_1 to be appropriately measurable.*

Expanding the notion of integration can be achieved by interpreting the integrals in (22a) and (22b) as *outer integrals* (see Appendix A). Since the outer integral can be defined for any function, measurable or not, there is no need to require that g, μ_0, and μ_1 are measurable in any sense. As a result, (22a) and (22b) make sense and an argument such as the one beginning with (23) goes through. This approach is discussed in detail in Part I, where we show that all the basic results for finite and infinite horizon problems of perfect state information carry through within an outer integration framework. However, there are inherent limitations in this approach centering around the pathologies of outer integration. Difficulties also occur in the treatment of imperfect information problems using sufficient statistics.

The major alternative approach was initiated in more general form by Blackwell [B9] in 1965. Here we assume at the outset that g is Borel- measurable, and furthermore, for each $B \in \mathcal{B}_R$ ($\mathcal{B}_R$ is the Borel σ-algebra on R), the function $p(B | x_0, u_0)$ is Borel-measurable in (x_0, u_0). In the initial treatment of the problem, the functions μ_0 and μ_1 were restricted to be Borel-measurable. With these assumptions, $g[x_1, \mu_1(x_1)]$ is Borel-measurable in x_1 when μ_1 is Borel-measurable, and the integral in (22a) is well defined.

A major difficulty occurs in (22b) since it is not necessarily true that $J_1(x_1) = \inf_{u_1} g(x_1, u_1)$ is Borel-measurable, even if g is. The reason can be traced to the fact that the orthogonal projection of a Borel set in R^2 on one

of the axes need not be Borel-measurable (see Section 7.6). Since we have for $c \in R$

$$\{x_1 | J_1(x_1) < c\} = \operatorname{proj}_{x_1} \{(x_1, u_1) | g(x_1, u_1) < c\},$$

where $\operatorname{proj}_{x_1}$ denotes projection on the x_1-axis, it can be seen that $\{x_1 | J_1(x_1) < c\}$ need not be Borel, even though $\{(x_1, u_1) | g(x_1, u_1) < c\}$ is. The difficulty can be overcome in part by showing that J_1 is a lower semi-analytic and hence also universally measurable function (see Section 7.7). Thus J_1 can be integrated with respect to any probability measure on $\mathscr{B}_R$.

Another difficulty stems from the fact that one cannot in general find a Borel-measurable ε-optimal selector μ_ε satisfying (23), although a weaker result is available whereby, given a probability measure p on $\mathscr{B}_R$, the existence of a Borel-measurable selector μ_ε satisfying

$$g[x_1, \mu_\varepsilon(x_1)] \leq \inf_{u_1} g(x_1, u_1) + \varepsilon$$

for p almost every $x_1 \in R$ can be ascertained. This result is sufficient to justify (24) and thus prove result R.1 ($J_2 = \inf_\pi J_\pi$). However, results R.2 and R.3 cannot be proved when μ_0 and μ_1 are restricted to be Borel-measurable except in a weaker form involving the notion of p-optimality (see [S14]; [H4]).

The objective of Part II is to resolve the measurability questions in stochastic optimal control in such a way that almost every result can be proved in a form as strong as its structural counterpart. This is accomplished by enlarging the set of admissible policies to include all *universally measurable policies.* In particular, we show the existence of policies within this class that are optimal or nearly optimal for *every* initial state.

A great many authors have dealt with measurability in stochastic optimal control theory. We describe three approaches taken and how their aims and results relate to our own. A fourth approach, due to Blackwell *et al.* [B12] and based on analytically measurable policies, is discussed in the next section and in Section 11.2.

I The General Model

If the state, control, and disturbance spaces are arbitrary measure spaces, very little can be done. One attempt in this direction is the work of Striebel [S16] involving p-essential infima. Geared toward giving meaning to the dynamic programming algorithm, this work replaces (18) by

$$J_{k+1}(x) = p_k\text{-essential} \inf_\mu E\{g[x, \mu(x), w] + J_k[f(x, \mu(x), w)]\}, \qquad (25)$$

$k = 0, \ldots, N-1$, where the p-essential infimum is over all measurable μ from state space S to control space C satisfying any constraints which may have been imposed. The functions J_k are measurable, and if the probability measures $p_0, \ldots, p_{N-1}$ are properly chosen and the so-called countable ε-lattice property holds, this modified dynamic programming algorithm generates the optimal cost function and can be used to obtain policies which are optimal or nearly optimal for p_{N-1} almost all initial states. The selection of the proper probability measures $p_0, \ldots, p_{N-1}$, however, is at least as difficult as executing the dynamic programming algorithm, and the verification of the countable ε-lattice property is equivalent to proving the existence of an ε-optimal policy.

II The Semicontinuous Models

Considerable attention has been directed toward models in which the state and control spaces are Borel spaces or even R^n and the reduced cost function

$$h(x, u) = \int g(x, u, w) p(dw|x, u)$$

has semicontinuity and/or convexity properties. A companion assumption is that the mapping

$$x \to U(x)$$

is a measurable closed-valued multifunction [R2]. In the latter case there exists a Borel-measurable selector $\mu: S \to C$ such that $\mu(x) \in U(x)$ for every state x (Kuratowski and Ryll–Nardzewski [K5]). This is of course necessary if any Borel-measurable policy is to exist at all.

The main fact regarding models of this type is that under various combinations of semicontinuity and compactness assumptions, the functions J_k defined by (17) and (18) are semicontinuous. In addition, it is often possible to show that the infimum in (18) is achieved for every x and k, and there are Borel-measurable selectors $\mu_0, \ldots, \mu_{N-1}$ such that $\mu_k(x)$ achieves this infimum (see Freedman [F1], Furukawa [F3], Himmelberg, *et al.* [H3], Maitra [M2], Schäl [S3], and the references contained therein). Such a policy $(\mu_0, \ldots, \mu_{N-1})$ is optimal, and the existence of this optimal policy is an additional benefit of imposing topological conditions to ensure that the problem is well defined. In Section 9.5 we show that lower semicontinuity and compactness conditions guarantee convergence of the dynamic programming algorithm over an infinite horizon to the optimal cost function, and that this algorithm can be used to generate an optimal stationary policy.

Continuity and compactness assumptions are integral to much of the work that has been done in stochastic programming. This work differs from

our own in both its aims and its framework. First, in the usual stochastic programming model, the controls cannot influence the distribution of future states (see Olsen [O1–O3], Rockafellar and Wets [R3–R4], and the references contained therein). As a result, the model does not include as special cases many important problems such as, for example, the classical linear quadratic stochastic control problem [B4, Section 3.1]. Second, assumptions of convexity, lower semicontinuity, or both are made on the cost function, the model is designed for the Kuratowski–Ryll–Nardzewski selection theorem, and the analysis is carried out in a finite-dimensional Euclidean state space. All of this is for the purpose of overcoming measurability problems. Results are not readily generalizable beyond Euclidean spaces (Rockafellar [R2]). The thrust of the work is toward convex programming type results, i.e., duality and Kuhn–Tucker conditions for optimality, and so a narrow class of problems is considered and powerful results are obtained.

III The Borel Models

The Borel space framework was introduced by Blackwell [B9] and further refined by Strauch, Dynkin, Juskevič, Hinderer, and others. The state and control spaces S and C were assumed to be Borel spaces, and the functions defining the model were assumed to be Borel-measurable. Initial efforts were directed toward proving the existence of "nice" optimal or nearly optimal policies in this framework. Policies were required to be Borel-measurable. For this model it is possible to prove the universal measurability of the optimal cost function and the existence for every $\varepsilon > 0$ and probability measure p on S of a p–ε-optimal policy (Strauch [S14, Theorems 7.1 and 8.1]). A p–ε-optimal policy is one which leads to a cost differing from the optimal cost by less than ε for p almost every initial state. As discussed earlier, even over a finite horizon the optimal cost function need not be Borel-measurable and there need not exist an everywhere ε-optimal policy (Blackwell [B9, Example 2]). The difficulty arises from the inability to choose a Borel-measurable function $\mu_k : S \to C$ which nearly achieves the infimum in (18) uniformly in x. The nonexistence of such a function interferes with the construction of optimal policies via the dynamic programming algorithm (17) and (18), since one must first determine at each stage the measure p with respect to which it is satisfactory to nearly achieve the infimum in (18) for p almost every x. This is essentially the same problem encountered with (25). The difficulties in constructing nearly optimal policies over an infinite horizon are more acute. Furthermore, from an applications point of view, a p–ε-optimal policy, even if it can be constructed, is a much less appealing object than an everywhere ε-optimal policy, since in many situations the distribution p is unknown or may change when the system is

operated repetitively, in which case a new p–ε-optimal policy must be computed.

In our formulation, the class of admissible policies in the Borel model is enlarged to include all universally measurable policies. We show in Part II that this class is sufficiently rich to ensure that *there exist everywhere ε-optimal policies and, if the infimum in the DP algorithm (18) is attained for every x and k, then an everywhere optimal policy exists.* Thus the notion of p-optimality can be dispensed with. The basic reason why optimal and nearly optimal policies can be found within the class of universally measurable policies may be traced to the selection theorem of Section 7.7. Another advantage of working with the class of universally measurable functions is that this class is closed under certain basic operations such as integration with respect to a universally measurable stochastic kernel and composition.

Our method of proof of infinite horizon results is based on an equivalence of stochastic and deterministic decision models which is worked out in Sections 9.1–9.3. The conversion is carried through only for the infinite horizon model, as it is not necessary for the development in Chapter 8. It is also done only under assumptions (P), (N), or (D) of Definition 9.1, although the models make sense under conditions similar to the (F^+) and (F^-) assumptions of Section 8.1. The relationship between the stochastic and the deterministic models is utilized extensively in Sections 9.4–9.6, where structural results proved in Part I are applied to the deterministic model and then transferred to the stochastic model. The analysis shows how results for stochastic models with measurability restrictions on the set of admissible policies can be obtained from the general results on abstract dynamic programming models given in Part I and provides the connecting link between the two parts of this work.

1.3 The Present Work Related to the Literature

This section summarizes briefly the contents of each chapter and points out relations with existing literature. During the course of our research, many of our results were reported in various forms (Bertsekas [B3–B5]; Shreve [S7–S8]; Shreve and Bertsekas [S9–S12]). Since the present monograph is the culmination of our joint work, we report particular results as being new even though they may be contained in one or more of the preceding references.

Part I

The objective of Part I is to provide a unifying framework for finite and infinite horizon dynamic programming models. We restrict our attention to

three types of infinite horizon models, which are patterned after the discounted and positive models of Blackwell [B8–B9] and the negative model of Strauch [S14]. It is an open question whether the framework of Part I can be effectively extended to cover other types of infinite horizon models such as the average cost model of Howard [H7] or convergent dynamic programming models of the type considered by Dynkin and Juskevič [D8] and Hordijk [H6].

The problem formulation of Part I is new. The work that is most closely related to our framework is the one by Denardo [D2], who considered an abstract dynamic programming model under contraction assumptions. Most of Denardo's results have been incorporated in slightly modified form in Chapter 4. Denardo's problem formulation is predicated on his contraction assumptions and is thus unsuitable for finite horizon models such as the one in Chapter 3 and infinite horizon models such as the ones in Chapter 5. This fact provided the impetus for our different formulation.

Most of the results of Part I constitute generalizations of results known for specific classes of problems such as, for example, deterministic and stochastic optimal control problems. We make an effort to identify the original sources, even though in some cases this is quite difficult. Some of the results of Part I have not been reported earlier even for a specific class of problems, and they will be indicated as new.

Chapter 2 Here we formulate the basic abstract sequential optimization problem which is the subject of Part I. Several classes of problems of practical interest are described in Section 2.3 and are shown to be special cases of the abstract problem. All these problems have received a great deal of attention in the literature with the exception of the stochastic optimal control model based on outer integration (Section 2.3.3). This model, as well as the results in subsequent chapters relating to it, is new. A stochastic model based on outer integration has also been considered by Denardo [D2], who used a different definition of outer integration. His definition works well under contraction assumptions such as the one in Chapter 4. However, many of the results of Chapters 3 and 5 do not hold if Denardo's definition of outer integral is adopted. By contrast, all the basic results of Part I are valid when specialized to the model of Section 2.3.3.

Chapter 3 This chapter deals with the finite horizon version of our abstract problem. The central results here relate to the validity of the dynamic programming algorithm, i.e., the equation $J_N^* = T^N(J_0)$. The validity of this equation is often accepted without scrutiny in the engineering literature, while in mathematical works it is usually proved under assumptions that are stronger than necessary. While we have been unable to locate an appropriate source, we feel certain that the results of Proposition 3.1 are known

for stochastic optimal control problems. The notion of a sequence of policies exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality and the corresponding existence result (Proposition 3.2) are new.

Chapter 4 Here we treat the infinite horizon version of our abstract problem under a contraction assumption. The developments in this chapter overlap considerably with Denardo's work [D2]. Our contraction assumption C is only slightly different from the one of Denardo. Propositions 4.1, 4.2, 4.3 (a), and 4.3 (c) are due to Denardo [D2], while Proposition 4.3 (b) has been shown by Blackwell [B9] for stochastic optimal control problems. Proposition 4.4 is new. Related compactness conditions for existence of a stationary optimal policy in stochastic optimal control problems were given by Maitra [M2], Kushner [K6], and Schäl [S5]. Propositions 4.6 and 4.7 improve on corresponding results by Denardo [D2] and McQueen [M3]. The modified policy iteration algorithm and the corresponding convergence result (Proposition 4.9) are new in the form given here. Denardo [D2] gives a somewhat less general form of policy iteration. The idea of policy iteration for deterministic and stochastic optimal control problems dates, of course, to the early days of dynamic programming (Bellman [B1]; Howard [H7]). The mathematical programming formulation of Section 4.3.3 is due to Denardo [D2].

Chapter 5 Here we consider infinite horizon versions of our abstract model patterned after the positive and negative models of Blackwell [B8, B9] and Strauch [S14]. When specialized to stochastic optimal control problems, most of the results of this chapter have either been shown by these authors or can be trivially deduced from their work. The part of Proposition 5.1 dealing with existence of an ε-optimal stationary policy is new, as is the last part of Proposition 5.2. Forms of Propositions 5.3 and 5.5 specialized to certain gambling problems have been shown by Dubins and Savage [D6], whose monograph provided the impetus for much of the subsequent work on dynamic programming. Propositions 5.9–5.11 are new. Results similar to those of Proposition 5.10 have been given by Schäl [S5] for stochastic optimal control problems under semicontinuity and compactness assumptions.

Chapter 6 The analysis in this chapter is new. It is motivated by the fact that the framework and the results of Chapters 2–5 are primarily applicable to problems where measurability issues are of no essential concern. While it is possible to apply the results to problems where policies are subject to measurability restrictions, this can be done only after a fairly elaborate reformulation (see Chapter 9). Here we generalize our framework so that problems in which measurability issues introduce genuine complications can be dealt with directly. However, only a portion of our earlier results carry

through within the generalized framework—primarily those associated with finite horizon models and infinite horizon models under contraction assumptions.

Part II

The objective of Part II is to develop in some detail the discrete-time stochastic optimal control problem (additive cost) in Borel spaces. The measurability questions are addressed explicitly. This model was selected from among the specialized models of Part I because it is often encountered and also because it can serve as a guide in the resolution of measurability difficulties in a great many other decision models.

In Chapter 7 we present the relevant topological properties of Borel spaces and their probability measures. In particular, the properties of analytic sets are developed. Chapter 8 treats the finite horizon stochastic optimal control problem, and Chapter 9 is devoted to the infinite horizon version. Chapter 10 deals with the stochastic optimal control problem when only a "noisy" measurement of the state of the system is possible. Various extensions of the theory of Chapters 8 and 9 are given in Chapter 11.

Chapter 7 The properties presented for metrizable spaces are well known. The material on Borel spaces can be found in Chapter 1 of Parthasarathy [P1] and is also available in Kuratowski [K2–K3]. A discussion of the weak topology can be found in Parthasarathy [P1]. Propositions 7.20, 7.21, and 7.23 are due to Prohorov [P2], but their presentation here follows Varadarajan [V1]. Part of Proposition 7.21 also appears in Billingsley [B7]. Proposition 7.25 is an extension of a result for compact X found in Dubins and Freedman [D5]. Versions of Proposition 7.25 have been used in the literature for noncompact X (Strauch [S14]; Blackwell *et al.* [B12]), the authors evidently intending an extension of the compact result by using Urysohn's theorem to embed X in a compact metric space. Proposition 7.27 is reported by Rhenius [R1], Juskevič [J3] and Striebel [S16]. We give Striebel's proof. Propositions 7.28 and 7.29 appear in some form in several texts on probability theory. A frequently cited reference is Loève [L1]. Propositions 7.30 and 7.31 are easily deduced from Maitra [M2] or Schäl [S4], and much of the rest of the discussion of semicontinuous functions is found in Hausdorff [H2]. Proposition 7.33 is due to Dubins and Savage [D6]. Proposition 7.34 is taken from Freedman [F1].

The investigation of analytic sets in Borel spaces began several years ago, but has been given additional impetus recently by the discovery of their applications to stochastic processes. Suslin schemes and analytic sets first appear in a paper by M. Suslin (or Souslin) in 1917 [S17], although the idea is generally attributed to Alexandroff. Suslin pointed out that every Borel

subset of the real line could be obtained as the nucleus of a Suslin scheme for the closed intervals, and non-Borel sets could be obtained this way as well. He also noted that the analytic subsets of R were just the projections on an axis of the Borel subsets of R^2. The universal measurability of analytic sets (Corollary 7.42.1) was proved by Lusin and Sierpinski [L3] in 1918. (See also Lusin [L2].) Our proof of this fact is taken from Saks [S1]. We have also taken material on analytic sets from Kuratowski [K2], Dellacherie [D1], Meyer [M4], Bourbaki [B13], Parthasarathy [P1], and Bressler and Sion [B14]. Proposition 7.43 is due to Meyer and Traki [M5], but our proof is original. The proofs given here of Propositions 7.47 and 7.49 are very similar to those found in Blackwell *et al.* [B12]. The basic result of Proposition 7.49 is due to Jankov [J1], but was also worked out about the same time and published later by von Neumann [N1, Lemma 5, p. 448]. The Jankov–von Neumann result was strengthened by Mackey [M1, Theorem 6.3]. The history of this theorem is related by Wagner [W1, pp. 900–901]. Proposition 7.50(a) is due to Blackwell *et al.* [B12]. Proposition 7.50(b) together with its strengthened version Proposition 11.4 generalize a result by Brown and Purves [B15], who proved existence of a universally measurable φ for the case where f is Borel measurable.

Chapter 8 The finite horizon stochastic optimal control model of Chapter 8 is essentially a finite horizon version of the models considered by Blackwell [B8, B9], Strauch [S14], Hinderer [H4], Dynkin and Juskevič [D8], Blackwell *et al.* [B12], and others. With the exception of [B12], all these works consider Borel-measurable policies and obtain existence results of a p–ε-optimal nature (see the discussion of the previous section). We allow universally measurable policies and thereby obtain everywhere ε-optimal existence results. While in Chapters 8 and 9 we concentrate on proving results that hold everywhere, the previously available results which allow only Borel-measurable policies and hold p almost everywhere can be readily obtained as corollaries. This follows from the following fact, whose proof we sketch shortly:

(F) *If X and Y are Borel spaces, $p_0, p_1, \ldots$ is a sequence of probability measures on X, and μ is a universally measurable map from X to Y, then there is a Borel measurable map μ' from X to Y such that*

$$\mu(x) = \mu'(x)$$

for p_k almost every x, $k = 0, 1, \ldots$.

As an example of how this observation can be used to obtain p almost everywhere existence results from ours, consider Proposition 9.19. It states in part that if $\varepsilon > 0$ and the discount factor α is less than one, then an ε-optimal nonrandomized stationary policy exists, i.e., a policy $\pi = (\mu, \mu, \ldots)$,

where μ is a universally measurable mapping from S to C. Given p_0 on S, this policy generates a sequence of measures $p_0, p_1, \ldots$ on S, where p_k is the distribution of the kth state when the initial state has distribution p_0 and the policy π is used. Let $\mu' : S \to C$ be Borel-measurable and equal to μ for p_k almost every x, $k = 0, 1, \ldots$. Let $\pi' = (\mu', \mu', \ldots)$. Then it can be shown that for p_0 almost every initial state, the cost corresponding to π' equals the cost corresponding to π, so π' is a p_0–ε-optimal nonrandomized stationary *Borel*-measurable policy. The existence of such a π' is a new result. This type of argument can be applied to all the existence results of Chapters 8 and 9.

We now sketch a proof of (F). Assume first that Y is a Borel subset of $[0, 1]$. Then for $r \in [0, 1]$, r rational, the set

$$U(r) = \{x | \mu(x) \leq r\}$$

is universally measurable. For every k, let $p_k^*[U(r)]$ be the outer measure of $U(r)$ with respect to p_k and let $B_{k1}, B_{k2}, \ldots$ be a decreasing sequence of Borel sets containing $U(r)$ such that

$$p_k^*[U(r)] = p_k\left[\bigcap_{j=1}^{\infty} B_{kj}\right].$$

Let $B(r) = \bigcap_{k=1}^{\infty} \bigcap_{j=1}^{\infty} B_{kj}$. Then

$$p_k^*[U(r)] = p_k[B(r)], \qquad k = 0, 1, \ldots,$$

and the argument of Lemma 7.27 applies. If Y is an arbitrary Borel space, it is Borel isomorphic to a Borel subset of $[0, 1]$ (Corollary 7.16.1), and (F) follows.

Proposition 8.1 is due to Strauch [S14], and Proposition 8.2 is contained in Theorem 14.4 of Hinderer [H4]. Example 8.1 is taken from Blackwell [B9]. Proposition 8.3 is new, the strongest previous result along these lines being the existence of an analytically measurable ε-optimal policy when the one-stage cost function is nonpositive [B12]. Propositions 8.4 and 8.5 are new, as are the corollaries to Proposition 8.5. Lower semicontinuous models have received much attention in the literature (Maitra [M2]; Furukawa [F3]; Schäl [S3–S5]; Freedman [F1]; Himmelberg *et al.* [H3]). Our lower semicontinuous model differs somewhat from those in the literature, primarily in the form of the control constraint. Proposition 8.6 is closely related to the analysis in several of the previously mentioned references. Proposition 8.7 is due to Freedman [F1].

Chapter 9 Example 9.1 is a modification of Example 6.1 of Strauch [S14], and Proposition 9.1 is taken from Strauch [S14]. The conversion of the stochastic optimal control problem to the deterministic one was suggested

by Witsenhausen [W3] in a different context and carried out systematically for the first time here. This results in a simple proof of the lower semianalyticity of the infinite horizon optimal cost function (cf. Corollary 9.4.1 and Strauch [S14, Theorem 7.1]). Propositions 9.8 and 9.9 are due to Strauch [S14], as are the (D) and (N) parts of Proposition 9.10. The (P) part of Proposition 9.10 is new. Proposition 9.12 appears as Theorem 5.2.2 of Schäl [S5], but Corollary 9.12.1 is new. Proposition 9.14 is a special case of Theorem 14.5 of Hinderer [H4]. Propositions 9.15–9.17 and the corollaries to Proposition 9.17 are new, although Corollary 9.17.2 is very close to Theorem 13.3 of Schäl [S5]. Propositions 9.18–9.20 are new. Proposition 9.21 is an infinite horizon version of a finite horizon result due to Freedman [F1], except that the nonrandomized ε-optimal policy Freedman constructs may not be semi-Markov.

Chapter 10 The use of the conditional distribution of the state given the available information as a basis for controlling systems with imperfect state information has been explored by several authors under various assumptions (see, for example, Åström [A2], Striebel [S15], and Sawaragi and Yoshikawa [S2]). The treatment of imperfect state information models with uncountable Borel state and action spaces, however, requires the existence of a regular conditional distribution with a measurable dependence on a parameter (Proposition 7.27), and this result is quite recent (Rhenius [R1]; Juskevič [J3]; Striebel [S16]). Chapter 10 is related to Chapter 3 of Striebel [S16] in that the general concept of a statistic sufficient for control is defined. We use such a statistic to construct a perfect state information model which is equivalent in the sense of Propositions 10.2 and 10.3 to the original imperfect state information model. From this equivalence the validity of the dynamic programming algorithm and the existence of ε-optimal policies under the mild conditions of Chapters 8 and 9 follow. Striebel justifies use of a statistic sufficient for control by showing that under a very strong hypothesis [S16, Theorem 5.5.1] the dynamic programming algorithm is valid and an ε-optimal policy can be based on the sufficient statistic. The strong hypothesis arises from the need to specify the null sets in the range spaces of the statistic in such a way that this specification is independent of the policy employed. This need results from the inability to deal with the pointwise partial infima of multivariate functions without the machinery of universally measurable policies and lower semianalytic functions. Like Striebel, we show that the conditional distributions of the states based on the available information constitute a statistic sufficient for control (Proposition 10.5), as do the vectors of available information themselves (Proposition 10.6).

The treatments of Rhenius [R1] and Juskevič [J3] are like our own in that perfect state information models which are equivalent to the original

one are defined. In his perfect state information model, Rhenius bases control on the observations and conditional distributions of the states, i.e., these objects are the states of his perfect state information model. It is necessary in Rhenius' framework for the controller to know the most recent observation, since this tells him which controls are admissible. We show in Proposition 10.5 that if there are no control constraints, then there is nothing to be gained by remembering the observations. In the model of Juskevič [J3], there are no control constraints and control is based on the past controls and conditional distributions. In this case, ε-optimal control is possible without reference to the past controls (Propositions 10.5, 8.3, 9.19, and 9.20), so our formulation is somewhat simpler and just as effective.

Chapter 10 differs from all the previously mentioned works in that simple conditions which guarantee the existence of a statistic sufficient for control are given, and once this existence is established, all the results of Chapters 8 and 9 can be brought to bear on the imperfect state information model.

Chapter 11 The use in Section 11.1 of limit measurability in dynamic programming is new. In particular, Proposition 11.3 is new, and as discussed earlier in regard to Proposition 7.50(b), a result by Brown and Purves [B15] is generalized in Proposition 11.4. Analytically measurable policies were introduced by Blackwell *et al.* [B12], whose work is referenced in Section 11.2. Borel space models with multiplicative cost fall within the framework of Furukawa and Iwamoto [F4–F5], and in [F5] the dynamic programming algorithm and a characterization of uniformly N-stage optimal policies are given. The remainder of Proposition 11.7 is new.

Appendix A Outer integration has been used by several authors, but we have been unable to find a systematic development.

Appendix B Proposition B.6 was first reported by Suslin [S17], but the proof given here is taken from Kuratowski [K2, Section 38VI]. According to Kuratowski and Mostowski [K4, p. 455], the limit σ-algebra $\mathscr{L}_X$ was introduced by Lusin, who called its members the "C-sets." A detailed discussion of the σ-algebra was given by Selivanovskij [S6] in 1928. Propositions B.9 and B.10 are fairly well known among set theorists, but we have been unable to find an accessible treatment. Proposition B.11 is new. Cenzer and Mauldin [C1] have also shown independently that $\mathscr{L}_X$ is closed under composition of functions, which is part of the result of Proposition B.11. Proposition B.12 is new.

It seems plausible that there are an infinity of distinct σ-algebras between the limit σ-algebra and the universal σ-algebra that are suitable for dynamic programming. One promising method of constructing such σ-algebras involves the R-operator of descriptive set theory (see Kantorovitch and

Livenson [K1]). In a recent paper [B11], Blackwell has employed a different method to define the "Borel-programmable" σ-algebra and has shown it to have many of the same properties we establish in Appendix B for the limit σ-algebra. It is not known, however, whether the Borel-programmable σ-algebra satisfies a condition like Proposition B.12 and is thereby suitable for dynamic programming. It is easily seen that the limit σ-algebra is contained in Blackwell's Borel-programmable σ-algebra, but whether the two coincide is also unknown.

Appendix C A detailed discussion of the exponential topology on the set of closed subsets of a topological space can be found in Kuratowski [K2–K3]. Properties of semicontinuous (K) functions are also proved there, primarily in Section 43 of [K3]. The Hausdorff metric is discussed in Section 38 of [H2].

Part I

Analysis of Dynamic Programming Models

Chapter 2

Monotone Mappings Underlying Dynamic Programming Models[†]

This chapter formulates the basic abstract sequential optimization problem which is the subject of Part I. It also provides examples of special cases which include wide classes of problems of practical interest.

2.1 Notation and Assumptions

Our usage of mathematical notation is fairly standard. For the reader's convenience we mention here that we use R to denote the real line and R^* to denote the extended real line, i.e., $R^* = R \cup \{-\infty, \infty\}$. The sets $(-\infty, \infty] = R \cup \{\infty\}$ and $[-\infty, \infty) = R \cup \{-\infty\}$ will be written out explicitly. We will assume throughout that R is equipped with the usual topology generated by the open intervals (α, β), $\alpha, \beta \in R$, and with the (Borel) σ-algebra generated by this topology. Similarly R^* is equipped with the topology generated by the open intervals (α, β), $\alpha, \beta \in R$, together with the sets $(\gamma, \infty]$, $[-\infty, \gamma)$, $\gamma \in R$, and with the σ-algebra generated by this topology. The Cartesian product of sets $X_1, X_2, \ldots, X_n$ is denoted $X_1 X_2 \cdots X_n$.

[†] Parts I and II can be read independently. The reader may proceed directly to Part II if he so wishes.

The following definitions and conventions will apply throughout Part I.

(1) S and C are two given sets referred to as the *state space* and *control space*, respectively.

(2) For each $x \in S$, there is given a nonempty subset $U(x)$ of C referred to as the *control constraint set at* x.

(3) We denote by M the set of all functions $\mu: S \to C$ such that $\mu(x) \in U(x)$ for all $x \in S$. We denote by Π the set of all sequences $\pi = (\mu_0, \mu_1, \ldots)$ such that $\mu_k \in M$ for all k. Elements of Π are referred to as *policies*. Elements of Π of the form $\pi = (\mu, \mu, \ldots)$, where $\mu \in M$, are referred to as *stationary policies*.

(4) We denote:

F the set of all extended real-valued functions $J: S \to R^*$;

B the Banach space of all bounded real-valued functions $J: S \to R$ with the supremum norm $\|\cdot\|$ defined by

$$\|J\| = \sup_{x \in S} |J(x)| \qquad \forall J \in B.$$

(5) For all $J, J' \in F$ we write

$$J = J' \quad \text{if} \quad J(x) = J'(x) \qquad \forall x \in S,$$
$$J \leq J' \quad \text{if} \quad J(x) \leq J'(x) \qquad \forall x \in S.$$

For all $J \in F$ and $\varepsilon \in R$, we denote by $J + \varepsilon$ the function taking the value $J(x) + \varepsilon$ at each $x \in S$, i.e.,

$$(J + \varepsilon)(x) = J(x) + \varepsilon \qquad \forall x \in S.$$

(6) Throughout Part I the analysis is carried out within the set of extended real numbers R^*. We adopt the usual conventions regarding ordering, addition, and multiplication in R^* except that we take

$$\infty - \infty = -\infty + \infty = \infty,$$

and we take the product of zero and infinity to be zero. In this way the sum and the product of any two extended real numbers is well defined. Division by zero or ∞ does not appear in our analysis. In particular, we adopt the following rules in calculations involving ∞ and $-\infty$:

$$\alpha + \infty = \infty + \alpha = \infty \qquad \text{for} \quad -\infty \leq \alpha \leq \infty,$$
$$\alpha - \infty = -\infty + \alpha = -\infty \qquad \text{for} \quad -\infty \leq \alpha < \infty;$$
$$\alpha\infty = \infty\alpha = \infty, \qquad \alpha(-\infty) = (-\infty)\alpha = -\infty \qquad \text{for} \quad 0 < \alpha \leq \infty,$$
$$\alpha\infty = \infty\alpha = -\infty, \qquad \alpha(-\infty) = (-\infty)\alpha = \infty \qquad \text{for} \quad -\infty \leq \alpha < 0;$$
$$0\infty = \infty 0 = 0 = 0(-\infty) = (-\infty)0, \qquad -(-\infty) = \infty;$$
$$\inf \varnothing = +\infty, \qquad \sup \varnothing = -\infty,$$

where $\varnothing$ is the empty set.

Under these rules the following laws of arithmetic are still valid:

$$\alpha_1 + \alpha_2 = \alpha_2 + \alpha_1, \qquad (\alpha_1 + \alpha_2) + \alpha_3 = \alpha_1 + (\alpha_2 + \alpha_3),$$
$$\alpha_1\alpha_2 = \alpha_2\alpha_1, \qquad (\alpha_1\alpha_2)\alpha_3 = \alpha_1(\alpha_2\alpha_3).$$

We also have

$$\alpha(\alpha_1 + \alpha_2) = \alpha\alpha_1 + \alpha\alpha_2$$

if either $\alpha \geq 0$ or else $(\alpha_1 + \alpha_2)$ is not of the form $+\infty - \infty$.

(7) For any sequence $\{J_k\}$ with $J_k \in F$ for all k, we denote by $\lim_{k\to\infty} J_k$ the pointwise limit of $\{J_k\}$ (assuming it is well defined as an extended real-valued function) and by $\limsup_{k\to\infty} J_k$ ($\liminf_{k\to\infty} J_k$) the pointwise limit superior (inferior) of $\{J_k\}$. For any collection $\{J_\alpha | \alpha \in A\} \subset F$ parameterized by the elements of a set A, we denote by $\inf_{\alpha\in A} J_\alpha$ the function taking the value $\inf_{\alpha\in A} J_\alpha(x)$ at each $x \in S$.

The Basic Mapping

We are given a function H which maps SCF (Cartesian product of S, C, and F) into R^*, and we define for each $\mu \in M$ the mapping $T_\mu : F \to F$ by

$$T_\mu(J)(x) = H[x, \mu(x), J] \qquad \forall x \in S. \tag{1}$$

We define also the mapping $T : F \to F$ by

$$T(J)(x) = \inf_{u\in U(x)} H(x, u, J) \qquad \forall x \in S. \tag{2}$$

We denote by T^k, $k = 1, 2, \ldots$, the composition of T with itself k times. For convenience we also define $T^0(J) = J$ for all $J \in F$. For any $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$ we denote by $(T_{\mu_0} T_{\mu_1} \cdots T_{\mu_k})$ the composition of the mappings $T_{\mu_0}, \ldots, T_{\mu_k}$, $k = 0, 1, \ldots$.

The following assumption will be in effect throughout Part I.

Monotonicity Assumption For every $x \in S$, $u \in U(x)$, $J, J' \in F$, we have

$$H(x, u, J) \leq H(x, u, J') \qquad \text{if} \quad J \leq J'. \tag{3}$$

The monotonicity assumption implies the following relations:

$$J \leq J' \Rightarrow T(J) \leq T(J') \qquad \forall J, J' \in F,$$
$$J \leq J' \Rightarrow T_\mu(J) \leq T_\mu(J') \qquad \forall J, J' \in F, \quad \mu \in M.$$

These relations in turn imply the following facts for all $J \in F$:

$$J \leq T(J) \Rightarrow T^k(J) \leq T^{k+1}(J), \qquad k = 0, 1, \ldots,$$
$$J \geq T(J) \Rightarrow T^k(J) \geq T^{k+1}(J), \qquad k = 0, 1, \ldots,$$
$$J \leq T_\mu(J) \quad \forall \mu \in M \Rightarrow (T_{\mu_0} \cdots T_{\mu_k})(J) \leq (T_{\mu_0} \cdots T_{\mu_{k+1}})(J),$$
$$k = 0, 1, \ldots, \quad \pi = (\mu_0, \mu_1, \ldots) \in \Pi,$$
$$J \geq T_\mu(J) \quad \forall \mu \in M \Rightarrow (T_{\mu_0} \cdots T_{\mu_k})(J) \geq (T_{\mu_0} \cdots T_{\mu_{k+1}})(J),$$
$$k = 0, 1, \ldots, \quad \pi = (\mu_0, \mu_1, \ldots) \in \Pi.$$

Another fact that we shall be using frequently is that for each $J \in F$ and $\varepsilon > 0$, there exists a $\mu_\varepsilon \in M$ such that

$$T_{\mu_\varepsilon}(J)(x) \leq \begin{cases} T(J)(x) + \varepsilon & \text{if} \quad T(J)(x) > -\infty, \\ -1/\varepsilon & \text{if} \quad T(J)(x) = -\infty. \end{cases}$$

In particular, if J is such that $T(J)(x) > -\infty$ for $\forall x \in S$, then for each $\varepsilon > 0$, there exists a $\mu_\varepsilon \in M$ such that

$$T_{\mu_\varepsilon}(J) \leq T(J) + \varepsilon.$$

2.2 Problem Formulation

We are given a function $J_0 \in F$ satisfying

$$J_0(x) > -\infty \qquad \forall x \in S, \tag{4}$$

and we consider for every policy $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$ and positive integer N the functions $J_{N,\pi} \in F$ and $J_\pi \in F$ defined by

$$J_{N,\pi}(x) = (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})(J_0)(x) \qquad \forall x \in S, \tag{5}$$

$$J_\pi(x) = \lim_{N \to \infty} (T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{N-1}})(J_0)(x) \qquad \forall x \in S. \tag{6}$$

For every result to be shown, appropriate assumptions will be in effect which guarantee that the function J_π is well defined (i.e., the limit in (6) exists for all $x \in S$). We refer to $J_{N,\pi}$ as the *N-stage cost function for* π and to J_π as the *cost function for* π. Note that $J_{N,\pi}$ depends only on the first N functions in π while the remaining functions are superfluous. Thus we could have considered policies consisting of finite sequences of functions in connection with the N-stage problem, and this is in fact done in Chapter 8. However, there are notational advantages in using a common type of policy in finite and infinite horizon problems, and for this reason we have adopted such a notation for Part I.

Throughout Part I we will be concerned with the N-stage optimization problem

$$\begin{aligned} &\text{minimize} \quad J_{N,\pi}(x) \\ &\text{subject to} \quad \pi \in \Pi, \end{aligned} \tag{F}$$

and its infinite horizon version

$$\begin{aligned} &\text{minimize} \quad J_\pi(x) \\ &\text{subject to} \quad \pi \in \Pi. \end{aligned} \tag{I}$$

We refer to problem (F) as the *N-stage finite horizon problem* and to problem (I) as the *infinite horizon problem.*

For a fixed $x \in S$, we denote by $J_N^*(x)$ and $J^*(x)$ the optimal costs for these problems, i.e.,

$$J_N^*(x) = \inf_{\pi \in \Pi} J_{N,\pi}(x) \qquad \forall x \in S, \tag{7}$$

$$J^*(x) = \inf_{\pi \in \Pi} J_\pi(x) \qquad \forall x \in S. \tag{8}$$

We refer to the function J_N^* as the *N-stage optimal cost function* and to the function J^* as the *optimal cost function.*

We say that a policy $\pi^* \in \Pi$ is *N-stage optimal at* $x \in S$ if $J_{N,\pi^*}(x) = J_N^*(x)$ and *optimal at* $x \in S$ if $J_{\pi^*}(x) = J^*(x)$. We say that $\pi^* \in \Pi$ is *N-stage optimal* (respectively *optimal*) if $J_{N,\pi^*} = J_N^*$ (respectively $J_{\pi^*} = J^*$). A policy $\pi^* = (\mu_0^*, \mu_1^*, \ldots)$ will be called *uniformly N-stage optimal* if the policy $(\mu_i^*, \mu_{i+1}^*, \ldots)$ is $(N - i)$-stage optimal for all $i = 0, 1, \ldots, N - 1$. Thus if a policy is uniformly N-stage optimal, it is also N-stage optimal, but not conversely. *For a stationary policy* $\pi = (\mu, \mu, \ldots) \in \Pi$, *we write* $J_\pi = J_\mu$. Thus a stationary policy $\pi^* = (\mu^*, \mu^*, \ldots)$ is optimal if $J^* = J_{\mu^*}$.

Given $\varepsilon > 0$, we say that a policy $\pi_\varepsilon \in \Pi$ is *N-stage ε-optimal* if

$$J_{N,\pi_\varepsilon}(x) \leq \begin{cases} J_N^*(x) + \varepsilon & \text{if } J_N^*(x) > -\infty, \\ -1/\varepsilon & \text{if } J_N^*(x) = -\infty. \end{cases}$$

We say that $\pi_\varepsilon \in \Pi$ is *ε-optimal* if

$$J_{\pi_\varepsilon}(x) \leq \begin{cases} J^*(x) + \varepsilon & \text{if } J^*(x) > -\infty, \\ -1/\varepsilon & \text{if } J^*(x) = -\infty. \end{cases}$$

If $\{\varepsilon_n\}$ is a sequence of positive numbers with $\varepsilon_n \downarrow 0$, we say that a sequence of policies $\{\pi_n\}$ exhibits *$\{\varepsilon_n\}$-dominated convergence to optimality* if

$$\lim_{n \to \infty} J_{N,\pi_n} = J_N^*,$$

and, for $n = 2, 3, \ldots,$

$$J_{N,\pi_n}(x) \leq \begin{cases} J_N^*(x) + \varepsilon_n & \text{if } J_N^*(x) > -\infty, \\ J_{N,\pi_{n-1}}(x) + \varepsilon_n & \text{if } J_N^*(x) = -\infty. \end{cases}$$

2.3 Application to Specific Models

A large number of sequential optimization problems of practical interest may be viewed as special cases of the abstract problems (F) and (I). In this section we shall describe several such problems that will be of continuing interest to us throughout Part I. Detailed treatments of some of these problems can be found in DPSC.†

† We denote by DPSC the textbook by Bertsekas, "Dynamic Programming and Stochastic Control." Academic Press, New York, 1976.

2.3.1 Deterministic Optimal Control

Consider the mapping $H: SCF \to R^*$ defined by

$$H(x,u,J) = g(x,u) + \alpha J[f(x,u)] \qquad \forall x \in S, \quad u \in C, \quad J \in F. \tag{9}$$

Our standing assumptions throughout Part I relating to this mapping are:

(1) The functions g and f map SC into $[-\infty, \infty]$ and S, respectively.
(2) The scalar α is positive.

The mapping H clearly satisfies the monotonicity assumption. Let J_0 be identically zero, i.e.,

$$J_0(x) = 0 \qquad \forall x \in S.$$

Then the corresponding N-stage optimization problem (F) can be written as

$$\begin{aligned} &\text{minimize} \quad J_{N,\pi}(x_0) = \sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k)] \\ &\text{subject to} \quad x_{k+1} = f[x_k, \mu_k(x_k)], \qquad \mu_k \in M, \quad k = 0, \ldots, N-1. \end{aligned} \tag{10}$$

This is a finite horizon deterministic optimal control problem. The scalar α is known as the *discount factor*. The infinite horizon problem (I) can be written as

$$\begin{aligned} &\text{minimize} \quad J_\pi(x_0) = \lim_{N \to \infty} \sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k)] \\ &\text{subject to} \quad x_{k+1} = f[x_k, \mu_k(x_k)], \qquad \mu_k \in M, \quad k = 0, 1, \ldots. \end{aligned} \tag{11}$$

This limit exists if any one of the following three conditions is satisfied:

$$g(x,u) \geq 0 \qquad \forall x \in S, \quad u \in U(x), \tag{12}$$

$$g(x,u) \leq 0 \qquad \forall x \in S, \quad u \in U(x), \tag{13}$$

$$\alpha < 1, \qquad 0 \leq g(x,u) \leq b \qquad \text{for some } b \in (0, \infty) \text{ and all } x \in S, u \in U(x). \tag{14}$$

Every result to be shown for problem (11) will explicitly assume one of these three conditions. Note that the requirement $0 \leq g(x,u) \leq b$ in (14) is no more strict than the usual requirement $|g(x,u)| \leq b/2$. This is true because adding the constant $b/2$ to g increases the cost corresponding to every policy by $b/2(1-\alpha)$ and the problem remains essentially unaffected.

Deterministic optimal control problems such as (10) and (11) and their stochastic counterparts under the countability assumption of the next subsection have been studied extensively in DPSC (Chapters 2, 6, and 7). They are given here in their stationary form in the sense that the state and control spaces S and C, the control constraint $U(\cdot)$, the system function f, and the

cost per stage g do not change from one stage to the next. When this is not the case, we are faced with a *nonstationary problem.* Such a problem, however, may be converted to a stationary problem by using a procedure described in Section 10.1 and in DPSC (Section 6.7). For this reason, we will not consider further nonstationary problems in Part I. Notice that within our formulation it is possible to handle state constraints of the form $x_k \in X$, $k = 0, 1, \ldots$, by defining $g(x, u) = \infty$ whenever $x \notin X$. This is our reason for allowing g to take the value ∞. Generalized versions of problems (10) and (11) are obtained if the scalar α is replaced by a function $\alpha: SC \to R^*$ with $0 \leq \alpha(x, u)$ for all $x \in S$, $u \in U(x)$, so that the discount factor depends on the current state and control. It will become evident to the reader that our general results for problems (F) and (I) are applicable to these more general deterministic problems.

2.3.2 Stochastic Optimal Control—Countable Disturbance Space

Consider the mapping $H: SCF \to R^*$ defined by

$$H(x, u, J) = E\{g(x, u, w) + \alpha J[f(x, u, w)] | x, u\}, \tag{15}$$

where the following are assumed:

(1) The parameter w takes values in a *countable* set W with given probability distribution $p(dw|x, u)$ depending on x and u, and $E\{\cdot|x, u\}$ denotes expected value with respect to this distribution. (See a detailed definition below.)

(2) The functions g and f map SCW into $[-\infty, \infty]$ and S, respectively.

(3) The scalar α is positive.

Our usage of expected value in (15) is consistent with the definition of the usual integral (Section 7.4.4) and the outer integral (Appendix A), where the σ-algebra on W is taken to be the set of all subsets of W. Thus if w^i, $i = 1, 2, \ldots$, are the elements of W, $(p^1, p^2, \ldots)$ any probability distribution on W, and $z: W \to R^*$ a function, we define

$$E\{z(w)\} = \sum_{i=1}^{\infty} p^i z^+(w_i) - \sum_{i=1}^{\infty} p^i z^-(w_i),$$

where

$$z^+(w_i) = \max\{0, z(w_i)\}, \qquad i = 1, 2, \ldots,$$
$$z^-(w_i) = \max\{0, -z(w_i)\}, \qquad i = 1, 2, \ldots.$$

In view of our convention $\infty - \infty = \infty$, the expected value $E\{z(w)\}$ is well defined for every function $z: W \to R^*$ and every probability distribution $(p^1, p^2, \ldots)$ on W. In particular, if we denote by $(p^1(x, u), p^2(x, u), \ldots)$ the

probability distribution $p(dw|x,u)$ on $W=\{w^1,w^2,\ldots\}$, then (15) can be written as

$$H(x,u,J)=\sum_{i=1}^{\infty}p^i(x,u)\max\{0,g(x,u,w^i)+\alpha J[f(x,u,w^i)]\}$$
$$-\sum_{i=1}^{\infty}p^i(x,u)\max\{0,-[g(x,u,w^i)+\alpha J[f(x,u,w^i)]]\}.$$

A point where caution is necessary in the use of expected value defined this way is that for two functions $z_1: W\to R^*$ and $z_2: W\to R^*$, the equality

$$E\{z_1(w)+z_2(w)\}=E\{z_1(w)\}+E\{z_2(w)\} \tag{16}$$

need not always hold. It is guaranteed to hold if (a) $E\{z_1^+(w)\}<\infty$ and $E\{z_2^+(w)\}<\infty$, or (b) $E\{z_1^-(w)\}<\infty$ and $E\{z_2^-(w)\}<\infty$, or (c) $E\{z_1^+(w)\}<\infty$ and $E\{z_1^-(w)\}<\infty$ (see Lemma 7.11). We always have, however,

$$E\{z_1(w)+z_2(w)\}\le E\{z_1(w)\}+E\{z_2(w)\}.$$

It is clear that the mapping H of (15) satisfies the monotonicity assumption. Let J_0 be identically zero, i.e.,

$$J_0(x)=0 \qquad \forall x\in S.$$

Then if $g(x,u,w)>-\infty$ for all x,u,w, the N-stage cost function can be written as

$$\begin{aligned}J_{N,\pi}(x_0)&=\underset{w_0}{E}\{g[x_0,\mu_0(x_0),w_0]+\underset{w_1}{E}\{\alpha g[x_1,\mu_1(x_1),w_1]+\underset{w_2}{E}\{\cdots\\&\quad+\underset{w_{N-1}}{E}\{\alpha^{N-1}g[x_{N-1},\mu_{N-1}(x_{N-1}),w_{N-1}]|x_{N-1},\\&\quad\mu_{N-1}(x_{N-1})\}|\cdots\}|x_0,\mu_0(x_0)\}\\&=\underset{w_0}{E}\left\{E\left\{\cdots\underset{w_{N-1}}{E}\left\{\sum_{k=0}^{N-1}\alpha^k g[x_k,\mu_k(x_k),w_k]|x_{N-1},\right.\right.\right.\\&\quad\left.\left.\left.\mu_{N-1}(x_{N-1})\right\}|\cdots\right\}|x_0,\mu_0(x_0)\right\},\end{aligned} \tag{17}$$

where the states $x_1,x_2,\ldots,x_{N-1}$ satisfy

$$x_{k+1}=f[x_k,\mu_k(x_k),w_k], \qquad k=0,\ldots,N-2. \tag{18}$$

The interchange of expectation and summation in (17) is valid, since $g(x,u,w)>-\infty$ for all x,u,w, and we have for any measure space $(\Omega,\mathcal{F},\nu)$,

measurable $h:\Omega \to R^*$, and $\lambda \in (-\infty, +\infty]$,

$$\lambda + \int h\, dv = \int (\lambda + h)\, dv.$$

When Eq. (18) is used successively to express the states $x_1, x_2, \ldots, x_{N-1}$ exclusively in terms of w_0, $w_1, \ldots, w_{N-1}$ and x_0, one can see from (17) that $J_{N,\pi}(x_0)$ is given in terms of successive iterated integration over $w_{N-1}, \ldots, w_0$. For each $x_0 \in S$ and $\pi \in \Pi$ the probability distributions $p^i(x_0, \mu_0(x_0)), \ldots, p^i(x_{N-1}, \mu_{N-1}(x_{N-1}))$, $i = 1, 2, \ldots$, over W specify, by the product measure theorem [A1, Theorem 2.6.2], a unique product measure on the cross product W^N of N copies of W. If Fubini's theorem [A1, Theorem 2.6.4] is applicable, then from (17) the N-stage cost function $J_{N,\pi}(x_0)$ can be alternatively expressed as

$$J_{N,\pi}(x_0) = E\left\{\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\}, \tag{19}$$

where this expectation is taken with respect to the product measure on W^N and the states $x_1, x_2, \ldots, x_{N-1}$ are expressed in terms of $w_0, w_1, \ldots, w_{N-1}$ and x_0 via (18). Fubini's theorem can be applied if the expected value in (19) is not of the form $\infty - \infty$, i.e., if either

$$E\left\{\max\left\{0, \sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\}\right\} < \infty$$

or

$$E\left\{\max\left\{0, -\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\}\right\} < \infty.$$

In particular, this is true if either

$$E\{\max\{0, g[x_k, \mu_k(x_k), w_k]\}\} < \infty, \qquad k = 0, \ldots, N-1,$$

or

$$E\{\max\{0, -g[x_k, \mu_k(x_k), w_k]\}\} < \infty, \qquad k = 0, \ldots, N-1$$

or if g is uniformly bounded above or below by a real number. If $J_{N,\pi}(x_0)$ can be expressed as in (19) for each $x_0 \in S$ and $\pi \in \Pi$, then the N-stage problem can be written as

$$\text{minimize} \quad J_{N,\pi}(x_0) = E\left\{\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\}$$

$$\text{subject to} \quad x_{k+1} = f[x_k, \mu_k(x_k), w_k], \qquad \mu_k \in M, \quad k = 0, \ldots, N-1,$$

which is the traditional form of an N-stage stochastic optimal control problem and is also the starting point for the N-stage model of Part II (Definition 8.3).

The corresponding infinite horizon problem is (cf. Definition 9.3)

$$\begin{aligned} &\text{minimize} \quad J_\pi(x_0) = \lim_{N\to\infty} E\left\{\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\} \\ &\text{subject to} \quad x_{k+1} = f[x_k, \mu_k(x_k), w_k], \qquad \mu_k \in M, \quad k = 0, 1, \ldots . \end{aligned} \tag{20}$$

This limit exists under any one of the conditions:

$$g(x, u, w) \geq 0 \qquad \forall x \in S, \quad u \in U(x), \quad w \in W, \tag{21}$$

$$g(x, u, w) \leq 0 \qquad \forall x \in S, \quad u \in U(x), \quad w \in W, \tag{22}$$

$$\alpha < 1, \qquad 0 \leq g(x, u, w) \leq b \qquad \text{for some } b \in (0, \infty) \text{ and all } x \in S,\ u \in U(x),\ w \in W. \tag{23}$$

Every result to be shown for problem (20) will explicitly assume one of these three conditions.

Similarly as for the deterministic problem, a generalized version of the stochastic problem is obtained if the scalar α is replaced by a function $\alpha: SCW \to R^*$ satisfying $0 \leq \alpha(x, u, w)$ for all (x, u, w). The mapping H takes the form

$$H(x, u, J) = E\{g(x, u, w) + \alpha(x, u, w) J[f(x, u, w)] | x, u\}.$$

This case covers certain semi-Markov decision problems (see [J2]). We will not be further concerned with this mapping and will leave it to the interested reader to obtain specific results relating to the corresponding problems (F) and (I) by specializing abstract results obtained subsequently in Part I. Also, nonstationary versions of the problem may be treated by reduction to the stationary case (see Section 10.1 or DPSC, Section 6.7).

The countability assumption on W is satisfied for many problems of interest. For example, it is satisfied in stochastic control problems involving Markov chains with a finite or countable number of states (see, e.g., [D3], [K6]). When the set W is not countable, then matters are complicated by the need to define the expected value

$$E\{g[x, \mu(x), w] + \alpha J[f(x, \mu(x), w)] | x, u\}$$

for every $\mu \in M$. There are two approaches that one can employ to overcome this difficulty. One possibility is to define the expected value as an outer integral, as we do in the next subsection. The other approach is the subject of Part II where we impose an appropriate measurable space structure on S, C, and W and require that the functions $\mu \in M$ be measurable. Under these circumstances a reformulation of the stochastic optimal control problem into the form of the abstract problems (F) or (I) is not straightforward. Nonetheless, such a reformulation is possible as well as useful as we will demonstrate in Chapter 9.

2.3.3 *Stochastic Optimal Control—Outer Integral Formulation*

Consider the mapping $H: SCF \to R^*$ defined by

$$H(x, u, J) = E^*\{g(x, u, w) + \alpha J[f(x, u, w)] | x, u\}, \tag{24}$$

where the following are assumed:

(1) The parameter w takes values in a measurable space $(W, \mathscr{F})$. For each fixed $(x, u) \in SC$, a probability measure $p(dw|x, u)$ on $(W, \mathscr{F})$ is given and $E^*\{\cdot|x, u\}$ in (24) denotes the outer integral (see Appendix A) with respect to that measure. Thus we may write, in the notation of Appendix A,

$$H(x, u, J) = \int^* \{g(x, u, w) + \alpha J[f(x, u, w)]\} p(dw|x, u).$$

(2) The functions g and f map SCW into $[-\infty, \infty]$ and S, respectively.
(3) The scalar α is positive.

We note that mappings (9) and (15) of the previous two subsections are special cases of the mapping H of (24). The mapping (9) (deterministic problem) is obtained from (24) when the set W consists of a single element. The mapping (15) (stochastic problem with countable disturbance space) is the special case of (24) where W is a countable set and $\mathscr{F}$ is the σ-algebra consisting of all subsets of W. For this reason, in our subsequent analysis we will not further consider the mappings (9) and (15), but will focus attention on the mapping (24).

Clearly H as defined by (24) satisfies the monotonicity assumption. Just as for the models of the previous two sections, we take

$$J_0(x) = 0 \qquad \forall x \in S$$

and consider the corresponding N-stage and infinite horizon problems (F) and (I).

If appropriate measurability assumptions are placed on S, C, f, g, and p, then the N-stage cost

$$J_{N,\pi}(x) = (T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(x)$$

can be rewritten in terms of ordinary integration for every policy $\pi = (\mu_0, \mu_1, \ldots)$ for which μ_k, $k = 0, 1, \ldots$, is appropriately measurable. To see this, suppose that S has a σ-algebra $\mathscr{S}$, C has a σ-algebra $\mathscr{C}$, and $\mathscr{B}$ is the Borel σ-algebra on R^*. Suppose f is $(\mathscr{S}\mathscr{C}\mathscr{F}, \mathscr{S})$-measurable and g is $(\mathscr{S}\mathscr{C}\mathscr{F}, \mathscr{B})$-measurable, where $\mathscr{S}\mathscr{C}\mathscr{F}$ denotes the product σ-algebra on SCW. Assume that for each fixed $B \in \mathscr{F}$, $p(B|x, u)$ is $\mathscr{S}\mathscr{C}$-measurable in (x, u) and consider a policy $\pi = (\mu_0, \mu_1, \ldots)$, where μ_k is $(\mathscr{S}, \mathscr{C})$-measurable for all k. These conditions guarantee that $T_{\mu_k}(J)$ given by

$$T_{\mu_k}(J)(x) = \int \{g[x, \mu_k(x), w] + \alpha J[f(x, \mu_k(x), w)]\} p(dw|x, u)$$

is $\mathscr{S}$-measurable for all k and $J \in F$ that are $\mathscr{S}$-measurable. Just as in the previous section, for a fixed $x_0 \in S$ and $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$, the probability measures $p(\cdot|x_0, \mu_0(x_0)), \ldots, p(\cdot|x_{N-1}, \mu_{N-1}(x_{N-1}))$ together with the system equation

$$x_{k+1} = f[x_k, \mu_k(x_k), w_k], \qquad k = 0, \ldots, N-2, \tag{25}$$

define a unique product measure $p(d(w_0, \ldots, w_{N-1})|x_0, \pi)$ on the cross product W^N of N copies of W. [Note that x_k, $k = 0, 1, \ldots, N-1$, can be expressed as a measurable function of $(w_0, \ldots, w_{N-1})$ via (25)]. Using the calculation of the previous section, we have that if $g(x, u, w) > -\infty$ for all x, u, w, and Fubini's theorem is applicable, then

$$\begin{aligned} J_{N,\pi}(x_0) &= E\left\{\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\} \\ &= \int_{W^N} \left\{\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\} p(d(w_0, \ldots, w_{N-1})|x_0, \pi), \end{aligned}$$

where $x_1, x_2, \ldots, x_{N-1}$ are expressed in terms of w_0, $w_1, \ldots, w_{N-1}$ and x_0 via (25). Also, as in the previous section, Fubini's theorem applies if either

$$E\left\{\max\left\{0, \sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\}\right\} < \infty$$

or

$$E\left\{\max\left\{0, -\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\}\right\} < \infty.$$

Thus if appropriate measurability conditions are placed on S, C, W, f, g, and $p(dw|x, u)$ and Fubini's theorem applies, then the N-stage cost $J_{N,\pi}$ corresponding to measurable π reduces to the traditional form

$$J_{N,\pi}(x_0) = E\left\{\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\}.$$

This observation is significant in view of the fact that

$$\inf_{\pi \in \Pi} J_{N,\pi}(x) \leq \inf_{\pi \in \tilde{\Pi}} J_{N,\pi}(x) \qquad \forall x \in S,$$

where

$$\tilde{\Pi} = \{\pi \in \Pi | \pi = (\mu_0, \mu_1, \ldots), \mu_k \in M \text{ is } (\mathscr{S}, \mathscr{C})\text{-measurable}, k = 0, 1, \ldots\}.$$

Thus, if an optimal (ε-optimal) policy π^* can be found for problem (F) and

$\pi^* \in \tilde{\Pi}$ (i.e., is measurable), then π^* is optimal (ε-optimal) for the problem

$$\begin{aligned} &\text{minimize} && J_{N,\pi}(x) \\ &\text{subject to} && \pi \in \tilde{\Pi}, \end{aligned}$$

which is a traditional stochastic optimal control problem.

These remarks illustrate how one can utilize the outer integration framework in an initial formulation of a particular problem and subsequently show via further (and hopefully simple) analysis that attention can be restricted to the class of measurable policies $\tilde{\Pi}$ for which the cost function admits a traditional interpretation. The main advantage that the outer integral formulation offers is simplicity. One does not need to introduce an elaborate topological and measure-theoretic structure such as the one of Part II in an initial formulation of the problem. In addition the policy iteration algorithm of Chapter 4 is applicable to the problem of this section but cannot be justified for the corresponding model of Part II. The outer integral formulation has, however, important limitations which become apparent in the treatment of problems with imperfect state information by means of sufficient statistics (Chapter 10).

2.3.4 *Stochastic Optimal Control—Multiplicative Cost Functional*

Consider the mapping $H: SCF \to R^*$ defined by

$$H(x, u, J) = E\{g(x, u, w)J[f(x, u, w)]|x, u\}. \tag{26}$$

We make the same assumptions on w, g, and f as in Section 2.3.2, i.e., w takes values in a *countable* set W with a given probability distribution depending on x and u. We assume further that

$$g(x, u, w) \geq 0 \qquad \forall x \in S, \quad u \in U(x), \quad w \in W. \tag{27}$$

In view of (27), the mapping H of (26) satisfies the monotonicity assumption.

We take

$$J_0(x) = 1 \quad \forall x \in S$$

and consider the problems (F) and (I). Problem (F) corresponds to the stochastic optimal control problem

$$\begin{aligned} &\text{minimize} && J_{N,\pi}(x_0) = E\{g[x_0, \mu_0(x_0), w_0] \cdots g[x_{N-1}, \mu_{N-1}(x_{N-1}), w_{N-1}]\} \\ &\text{subject to} && x_{k+1} = f[x_k, \mu_k(x_k), w_k], \qquad \mu_k \in M, \quad k = 0, 1, \ldots, \end{aligned} \tag{28}$$

and problem (I) corresponds to the infinite horizon version of (28). The limit as $N \to \infty$ in (28) exists if $g(x, u, w) \geq 1$ for every x, u, w or $0 \leq g(x, u, w) \leq 1$

for every x, u, w. A special case of (28) is the exponential cost functional problem

$$\begin{aligned} &\text{minimize} \quad E\left\{\exp\left[\sum_{k=0}^{N-1} g'[x_k, \mu_k(x_k), w_k]\right]\right\} \\ &\text{subject to} \quad x_{k+1} = f[x_k, \mu_k(x_k), w_k], \qquad \mu_k \in M, \quad k = 0, 1, \ldots, \end{aligned}$$

where g' is some function mapping SCW into $(-\infty, \infty]$.

2.3.5 *Minimax Control*

Consider the mapping $H: SCF \to R^*$ defined by

$$H(x, u, J) = \sup_{w \in W(x,u)} \{g(x, u, w) + \alpha J[f(x, u, w)]\} \tag{29}$$

where the following are assumed:

(1) The parameter w takes values in a set W and $W(x, u)$ is a nonempty subset of W for each $x \in S$, $u \in U(x)$.

(2) The functions g and f map SCW into $[-\infty, \infty]$ and S respectively.

(3) The scalar α is positive.

Clearly the monotonicity assumption is satisfied.

We take

$$J_0(x) = 0 \qquad \forall x \in S.$$

If $g(x, u, w) > -\infty$ for all x, u, w, the corresponding N-stage problem (F) can also be written as

$$\begin{aligned} &\text{minimize} \quad J_{N,\pi}(x_0) = \sup_{w_k \in W[x_k, \mu_k(x_k)]} \left\{\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\} \\ &\text{subject to} \quad x_{k+1} = f[x_k, \mu_k(x_k), w_k], \qquad \mu_k \in M, \quad k = 0, 1, \ldots, \end{aligned} \tag{30}$$

and this is an N-stage minimax control problem. The infinite horizon version is

$$\begin{aligned} &\text{minimize} \quad J_\pi(x_0) = \lim_{N \to \infty} \sup_{w_k \in W[x_k, \mu_k(x_k)]} \left\{\sum_{k=0}^{N-1} \alpha^k g[x_k, \mu_k(x_k), w_k]\right\} \\ &\text{subject to} \quad x_{k+1} = f[x_k, \mu_k(x_k), w_k], \qquad \mu_k \in M, \quad k = 0, 1, \ldots. \end{aligned} \tag{31}$$

The limit in (31) exists under any one of the conditions (21), (22), or (23). This problem contains as a special case the problem of infinite time reachability examined in Bertsekas [B2]. Problems (30) and (31) arise also in the analysis of sequential zero-sum games.

Chapter 3

Finite Horizon Models

3.1 General Remarks and Assumptions

Consider the N-stage optimization problem

$$\begin{aligned} &\text{minimize} \quad J_{N,\pi}(x) = (T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(x) \\ &\text{subject to} \quad \pi = (\mu_0, \mu_1, \ldots) \in \Pi, \end{aligned}$$

where for every $\mu \in M$, $J \in F$, and $x \in S$ we have

$$T_\mu(J)(x) = H[x, \mu(x), J], \qquad T(J)(x) = \inf_{u \in U(x)} H(x, u, J).$$

Experience with a large variety of sequential optimization problems suggests that the N-stage optimal cost function J_N^* satisfies

$$J_N^* = \inf_{\pi \in \Pi} J_{N,\pi} = T^N(J_0),$$

and hence is obtained after N steps of the DP algorithm. In our more general setting, however, we shall need to place additional conditions on H in order to guarantee this equality. Consider the following two assumptions.

Assumption F.1 If $\{J_k\} \subset F$ is a sequence satisfying $J_{k+1} \leq J_k$ for all k and $H(x, u, J_1) < \infty$ for all $x \in S$, $u \in U(x)$, then

$$\lim_{k \to \infty} H(x, u, J_k) = H\left(x, u, \lim_{k \to \infty} J_k\right) \qquad \forall x \in S, \quad u \in U(x).$$

Assumption F.2 There exists a scalar $\alpha \in (0, \infty)$ such that for all scalars $r \in (0, \infty)$ and functions $J \in F$, we have

$$H(x, u, J) \leq H(x, u, J + r) \leq H(x, u, J) + \alpha r \qquad \forall x \in S, \quad u \in U(x).$$

We will also consider the following assumption, which is admittedly somewhat complicated. It will enable us to obtain a stronger result on the existence of nearly optimal policies (Proposition 3.2) than can be obtained under F.2. The assumption is satisfied for the stochastic optimal control problem of Section 2.3.3, as we show in the last section of this chapter.

Assumption F.3 There is a scalar $\beta \in (0, \infty)$ such that if $J \in F$, $\{J_n\} \subset F$, and $\{\varepsilon_n\} \subset R$ satisfy

$$\sum_{n=1}^{\infty} \varepsilon_n < \infty, \qquad \varepsilon_n > 0, \quad n = 1, 2, \ldots,$$

$$J = \lim_{n \to \infty} J_n, \qquad J \leq J_n, \quad n = 1, 2, \ldots,$$

$$J_n(x) \leq \begin{cases} J(x) + \varepsilon_n, & n = 1, 2, \ldots \text{ and } x \in S \text{ with } J(x) > -\infty, \\ J_{n-1}(x) + \varepsilon_n, & n = 2, 3, \ldots \text{ and } x \in S \text{ with } J(x) = -\infty, \end{cases}$$

$$H(x, u, J_1) < \infty, \qquad \forall x \in S, \quad u \in U(x),$$

then there exists a sequence $\{\mu_n\} \subset M$ such that

$$\lim_{n \to \infty} T_{\mu_n}(J_n) = T(J),$$

$$T_{\mu_n}(J_n)(x) \leq \begin{cases} T(J)(x) + \beta\varepsilon_n, & n = 1, 2, \ldots, \ x \in S \text{ with } T(J)(x) > -\infty, \\ T_{\mu_{n-1}}(J_{n-1})(x) + \beta\varepsilon_n, & n = 2, 3, \ldots, \ x \in S \text{ with } T(J)(x) = -\infty. \end{cases}$$

Each of our results will require *at most* one of the preceding assumptions. As we show in Section 3.3, at least one of these assumptions is satisfied by every specific model considered in Section 2.3.

3.2 Main Results

The central question regarding the finite horizon problem is whether $J_N^* = T^N(J_0)$, in which case the N-stage optimal cost function J_N^* can be obtained via the DP algorithm that successively computes $T(J_0), T^2(J_0), \ldots$. A related question is whether optimal or nearly optimal policies exist. The results of this section provide conditions under which the answer to these questions is affirmative.

Proposition 3.1 (a) Let F.1 hold and assume that $J_{k,\pi}(x) < \infty$ for all $x \in S$, $\pi \in \Pi$, and $k = 1, 2, \ldots, N$. Then

$$J_N^* = T^N(J_0).$$

(b) Let F.2 hold and assume that $J_k^*(x) > -\infty$ for all $x \in S$ and $k = 1, 2, \ldots, N$. Then

$$J_N^* = T^N(J_0),$$

and for every $\varepsilon > 0$, there exists an N-stage ε-optimal policy, i.e., a $\pi_\varepsilon \in \Pi$ such that

$$J_N^* \leq J_{N,\pi_\varepsilon} \leq J_N^* + \varepsilon.$$

Proof (a) For each $k = 0, 1, \ldots, N-1$, consider a sequence $\{\mu_k^i\} \subset M$ such that

$$\lim_{i \to \infty} T_{\mu_k^i}[T^{N-k-1}(J_0)] = T^{N-k}(J_0), \qquad k = 0, \ldots, N-1,$$

$$T_{\mu_k^i}[T^{N-k-1}(J_0)] \geq T_{\mu_k^{i+1}}[T^{N-k-1}(J_0)], \qquad k = 0, \ldots, N-1, \quad i = 0, 1, \ldots.$$

By using F.1 and the assumption that $J_{k,\pi}(x) < \infty$, we have

$$\begin{aligned} J_N^* &\leq \inf_{i_0} \cdots \inf_{i_{N-1}} (T_{\mu_0^{i_0}} \cdots T_{\mu_{N-1}^{i_{N-1}}})(J_0) \\ &= \inf_{i_0} \cdots \inf_{i_{N-2}} (T_{\mu_0^{i_0}} \cdots T_{\mu_{N-2}^{i_{N-2}}}) \left[\inf_{i_{N-1}} T_{\mu_{N-1}^{i_{N-1}}}(J_0) \right] \\ &= \inf_{i_0} \cdots \inf_{i_{N-2}} (T_{\mu_0^{i_0}} \cdots T_{\mu_{N-2}^{i_{N-2}}})[T(J_0)] \\ &= T^N(J_0), \end{aligned}$$

where the last equality is obtained by repeating the process used to obtain the previous equalities. On the other hand, it is clear from the definitions of Chapter 2 that $T^N(J_0) \leq J_N^*$, and hence $J_N^* = T^N(J_0)$.

(b) We use induction. The result clearly holds for $N = 1$. Assume that it holds for $N = k$, i.e., $J_k^* = T^k(J_0)$ and for a given $\varepsilon > 0$, there is a $\pi_\varepsilon \in \Pi$ with $J_{k,\pi_\varepsilon} \leq J_k^* + \varepsilon$. Using F.2 we have for all $\mu \in M$,

$$J_{k+1}^* \leq T_\mu(J_{k,\pi_\varepsilon}) \leq T_\mu(J_k^*) + \alpha\varepsilon.$$

Hence $J_{k+1}^* \leq T(J_k^*)$, and by using the induction hypothesis we obtain $J_{k+1}^* \leq T^{k+1}(J_0)$. On the other hand, we have clearly $T^{k+1}(J_0) \leq J_{k+1}^*$, and hence $T^{k+1}(J_0) = J_{k+1}^*$. For any $\bar{\varepsilon} > 0$, let $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1, \ldots)$ be such that $J_{k,\bar{\pi}} \leq J_k^* + (\bar{\varepsilon}/2\alpha)$, and let $\bar{\mu} \in M$ be such that $T_{\bar{\mu}}(J_k^*) \leq T(J_k^*) + (\bar{\varepsilon}/2)$. Consider the policy $\bar{\pi}_{\bar{\varepsilon}} = (\bar{\mu}, \bar{\mu}_0, \bar{\mu}_1, \ldots)$. Then

$$J_{k+1,\bar{\pi}_{\bar{\varepsilon}}} = T_{\bar{\mu}}(J_{k,\bar{\pi}}) \leq T_{\bar{\mu}}(J_k^*) + (\bar{\varepsilon}/2) \leq T(J_k^*) + \bar{\varepsilon} = J_{k+1}^* + \bar{\varepsilon}.$$

The induction is complete. Q.E.D.

Proposition 3.1(a) may be strengthened by using the following assumption in place of F.1.

Assumption F.1′ The function J_0 satisfies

$$J_0(x) \geq H(x, u, J_0) \qquad \forall x \in S, \quad u \in U(x),$$

and if $\{J_k\} \subset F$ is a sequence satisfying $J_{k+1} \leq J_k \leq J_0$ for all k, then

$$\lim_{k \to \infty} H(x, u, J_k) = H\left(x, u, \lim_{k \to \infty} J_k\right) \qquad \forall x \in S, \quad u \in U(x).$$

The following corollary is obtained by verbatim repetition of the proof of Proposition 3.1(a).

Corollary 3.1.1 Let F.1′ hold. Then

$$J_N^* = T^N(J_0).$$

Proposition 3.1 and Corollary 3.1.1 may fail to hold if their assumptions are slightly relaxed.

COUNTEREXAMPLE 1 Take $S = \{0\}$, $C = U(0) = (-1, 0]$, $J_0(0) = 0$, $H(0, u, J) = u$ if $-1 < J(0)$, $H(0, u, J) = J(0) + u$ if $J(0) \leq -1$. Then $(T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(0) = \mu_0(0)$ and $J_N^*(0) = -1$, while $T^N(J_0)(0) = -N$ for every N. Here the assumptions $J_{k,\pi}(0) < \infty$ and $J_k^*(0) > -\infty$ are satisfied, but F.1, F.1′, and F.2 are violated.

COUNTEREXAMPLE 2 Take $S = \{0, 1\}$, $C = U(0) = U(1) = (-\infty, 0]$, $J_0(0) = J_0(1) = 0$, $H(0, u, J) = u$ if $J(1) = -\infty$, $H(0, u, J) = 0$ if $J(1) > -\infty$, and $H(1, u, J) = u$. Then $(T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(0) = 0$, $(T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(1) = \mu_0(1)$ for all $N \geq 1$. Hence, $J_N^*(0) = 0$, $J_N^*(1) = -\infty$. On the other hand, we have $T^N(J_0)(0) = T^N(J_0)(1) = -\infty$ for all $N \geq 2$. Here F.2 is satisfied, but F.1, F.1′, and the assumptions $J_{k,\pi}(x) < \infty$ and $J_k^*(x) > -\infty$ for $\forall x \in S$ are all violated.

The following counterexample is a stochastic optimal control problem with countable disturbance space as discussed in Section 2.3.2. We use the notation introduced there.

COUNTEREXAMPLE 3 Let $N = 2$, $S = \{0, 1\}$, $C = U(0) = U(1) = R$, $W = \{2, 3, \ldots\}$, $p(w = k | x, u) = k^{-2}(\sum_{n=2}^{\infty} n^{-2})^{-1}$ for $k = 2, 3, \ldots$, $x \in S$, $u \in C$, $f(0, u, w) = f(1, u, w) = 1$ for $\forall u \in C$, $w \in W$, $g(0, u, w) = w$, $g(1, u, w) = u$ for $\forall u \in C$, $w \in W$. Then a straightforward calculation shows that $J_2^*(0) = \infty$, $J_2^*(1) = -\infty$, while $T^2(J_0)(0) = -\infty$, $T^2(J_0)(1) = -\infty$. Here F.1 and F.2 are satisfied, but F.1′ and the assumptions $J_{k,\pi}(x) < \infty$ for all x, π, k, and $J_k^*(x) > -\infty$ for all x and k are all violated.

The next counterexample is a deterministic optimal control problem as discussed in Section 2.3.1. We use the notation introduced there.

COUNTEREXAMPLE 4 Let $N = 2$, $S = \{0, 1, \ldots\}$, $C = U(x) = (0, \infty)$ for $\forall x \in S$, $f(x, u) = 0$ for $\forall x \in S$, $u \in C$, $g(0, u) = -u$ for $\forall u \in U(0)$, $g(x, u) = x$ for $\forall u \in U(x)$ if $x \neq 0$. Then for $\pi \in \Pi$ and $x \neq 0$, we have $J_{2,\pi}(x) = x - \mu_1(0)$, so that $J_2^*(x) = -\infty$ for all $x \in S$. On the other hand clearly there is no two-stage ε-optimal policy for any $\varepsilon > 0$. Here F.1, F.2, and the assumption $J_{k,\pi}(x) < \infty$ for all x, π, k are satisfied, and indeed we have $J_2^*(x) = T^2(J_0)(x) = -\infty$ for $\forall x \in S$. However, the assumption $J_k^*(x) > -\infty$ for all x and k is violated.

As Counterexample 4 shows there may not exist an N-stage ε-optimal policy if we have $J_k^*(x) = -\infty$ for some k and $x \in S$. The following proposition establishes, under appropriate assumptions, the existence of a sequence of nearly optimal policies whose cost functions converge to the optimal cost function.

Proposition 3.2 Let F.3 hold and assume $J_{k,\pi}(x) < \infty$ for all $x \in S$, $\pi \in \Pi$, and $k = 1, 2, \ldots, N$. Then

$$J_N^* = T^N(J_0).$$

Furthermore, if $\{\varepsilon_n\}$ is a sequence of positive numbers with $\varepsilon_n \downarrow 0$, then there exists a sequence of policies $\{\pi_n\}$ exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality. In particular, if in addition $J_N^*(x) > -\infty$ for all $x \in S$, then for every $\varepsilon > 0$ there exists an ε-optimal policy.

Proof We will prove by induction that for $K \leq N$ we have $J_K^* = T^K(J_0)$, and furthermore, given K and $\{\varepsilon_n\}$ with $\varepsilon_n \downarrow 0$, $\varepsilon_n > 0$ for $\forall n$, there exists a sequence $\{\pi_n\} \subset \Pi$ such that for all n,

$$\lim_{n \to \infty} J_{K,\pi_n} = J_K^*, \tag{1}$$

$$J_{K,\pi_n}(x) \leq \begin{cases} J_K^*(x) + \varepsilon_n & \forall x \in S \quad \text{with} \quad J_K^*(x) > -\infty, \quad (2) \\ J_{K,\pi_{n-1}}(x) + \varepsilon_n & \forall x \in S \quad \text{with} \quad J_K^*(x) = -\infty. \quad (3) \end{cases}$$

We show that this holds for $K = 1$. We have

$$J_1^*(x) = \inf_{\pi \in \Pi} J_{1,\pi}(x) = \inf_{\mu \in M} H[x, \mu(x), J_0] = T(J_0)(x) \qquad \forall x \in S.$$

It is also clear that, given $\{\varepsilon_n\}$, there exists a sequence $\{\pi_n\} \subset \Pi$ satisfying (1)–(3) for $K = 1$.

Assume that the result is true for $K = N - 1$. Let β be the scalar specified in F.3. Consider a sequence $\{\varepsilon_n\} \subset R$ with $\varepsilon_n > 0$ for $\forall n$ and $\lim_{n \to \infty} \varepsilon_n = 0$, and let $\{\hat{\pi}_n\} \subset \Pi$, $\hat{\pi}_n = (\mu_1^n, \mu_2^n, \ldots)$, be such that

$$\lim_{n \to \infty} J_{N-1,\hat{\pi}_n} = J_{N-1}^*, \tag{4}$$

$$J_{N-1,\hat{\pi}_n}(x) \leq \begin{cases} J_{N-1}^*(x) + \beta^{-1}\varepsilon_n & \forall x \in S \quad \text{with} \quad J_{N-1}^*(x) > -\infty, \quad (5) \\ J_{N-1,\hat{\pi}_{n-1}}(x) + \beta^{-1}\varepsilon_n & \forall x \in S \quad \text{with} \quad J_{N-1}^*(x) = -\infty. \quad (6) \end{cases}$$

The assumption $J_{k,\pi}(x) < \infty$ for all $x \in S$, $\pi \in \Pi$, $k = 1, 2, \ldots, N$, guarantees that we have

$$H(x, u, J_{N-1,\hat{\pi}_1}) < \infty \qquad \forall x \in S, \quad u \in U(x). \tag{7}$$

Without loss of generality we assume that $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. Then Assumption F.3 together with (4) implies that there exists a sequence $\{\mu_0^n\} \subset M$ such that, for all n,

$$\lim_{n\to\infty} T_{\mu_0^n}(J_{N-1,\hat{\pi}_n}) = T(J_{N-1}^*), \tag{8}$$

$$T_{\mu_0^n}(J_{N-1,\hat{\pi}_n})(x) \leq \begin{cases} T(J_{N-1}^*)(x) + \varepsilon_n & \text{if} \quad T(J_{N-1}^*)(x) > -\infty, \quad (9) \\ T_{\mu_0^{n-1}}(J_{N-1,\pi_{\hat{n}-1}})(x) + \varepsilon_n & \text{if} \quad T(J_{N-1}^*)(x) = -\infty. \quad (10) \end{cases}$$

We have by the induction hypothesis $J_{N-1}^* = T^{N-1}(J_0)$, and it is clear that $T^N(J_0) \leq J_N^*$. Hence,

$$T(J_{N-1}^*) = T^N(J_0) \leq J_N^*. \tag{11}$$

We also have

$$J_N^* \leq \lim_{n\to\infty} T_{\mu_0^n}(J_{N-1,\hat{\pi}_n}) \tag{12}$$

Combining (8), (11), and (12), we obtain

$$J_N^* = T(J_{N-1}^*) = T^N(J_0). \tag{13}$$

Let $\pi_n = (\mu_0^n, \mu_1^n, \mu_2^n, \ldots)$. Then from (8)–(10) and (13), we obtain, for all n,

$$\lim_{n\to\infty} J_{N,\pi_n} = J_N^*,$$

$$J_{N,\pi_n}(x) \leq \begin{cases} J_N^*(x) + \varepsilon_n & \forall x \in S \quad \text{with} \quad J_N^*(x) > -\infty, \\ J_{N,\pi_{n-1}}(x) + \varepsilon_n & \forall x \in S \quad \text{with} \quad J_N^*(x) = -\infty, \end{cases}$$

and the induction argument is complete. Q.E.D.

Despite the need for various assumptions in order to guarantee $J_N^* = T^N(J_0)$, the following result, which establishes the validity of the DP algorithm as a means for constructing optimal policies, requires no assumption other than monotonicity of H.

Proposition 3.3 A policy $\pi^* = (\mu_0^*, \mu_1^*, \ldots)$ is uniformly N-stage optimal if and only if

$$(T_{\mu_k^*} T^{N-k-1})(J_0) = T^{N-k}(J_0), \qquad k = 0, \ldots, N-1. \tag{14}$$

Proof Let (14) hold. Then we have, for $k = 0, 1, \ldots, N-1$,

$$(T_{\mu_k^*} \cdots T_{\mu_{N-1}^*})(J_0) = T^{N-k}(J_0).$$

On the other hand, we have $J^*_{N-k} \leq (T_{\mu^*_k} \cdots T_{\mu^*_{N-1}})(J_0)$, while $T^{N-k}(J_0) \leq J^*_{N-k}$. Hence, $J^*_{N-k} = (T_{\mu^*_k} \cdots T_{\mu^*_{N-1}})(J_0)$ and π^* is uniformly N-stage optimal.

Conversely, let π^* be uniformly N-stage optimal. Then

$$T(J_0) = J^*_1 = T_{\mu^*_{N-1}}(J_0)$$

by definition. We also have for every $\mu \in M$, $(T_\mu T)(J_0) = (T_\mu T_{\mu^*_{N-1}})(J_0)$, which implies that

$$T^2(J_0) = \inf_{\mu \in M} (T_\mu T)(J_0) = \inf_{\mu \in M} (T_\mu T_{\mu^*_{N-1}})(J_0)$$

$$\geq J^*_2 = (T_{\mu^*_{N-2}} T_{\mu^*_{N-1}})(J_0) \geq T^2(J_0).$$

Therefore

$$T^2(J_0) = J^*_2 = (T_{\mu^*_{N-2}} T_{\mu^*_{N-1}})(J_0) = (T_{\mu^*_{N-2}} T)(J_0).$$

Proceeding similarly, we show all the equations in (14). Q.E.D.

As a corollary of Proposition 3.3, we have the following.

Corollary 3.3.1 (a) There exists a uniformly N-stage optimal policy if and only if the infimum in the relation

$$T^{k+1}(J_0)(x) = \inf_{u \in U(x)} H[x, u, T^k(J_0)] \tag{15}$$

is attained for each $x \in S$ and $k = 0, 1, \ldots, N-1$.

(b) If there exists a uniformly N-stage optimal policy, then

$$J^*_N = T^N(J_0).$$

We now turn to establishing conditions for existence of a uniformly N-stage optimal policy. For this we need compactness assumptions. If C is a Hausdorff topological space, we say that a subset U of C is compact if every collection of open sets that covers U has a finite subcollection that covers U. The empty set in particular is considered to be compact. Any sequence $\{u_n\}$ belonging to a compact set $U \subset C$ has at least one accumulation point $\bar{u} \in U$, i.e., a point $\bar{u} \in U$ every (open) neighborhood of which contains an infinite number of elements of $\{u_n\}$. Furthermore, all accumulation points of $\{u_n\}$ belong to U. If $\{U_n\}$ is a sequence of nonempty compact subsets of C and $U_n \supset U_{n+1}$ for all n, then the intersection $\bigcap_{n=1}^{\infty} U_n$ is nonempty and compact. This yields the following lemma, which will be useful in what follows.

Lemma 3.1 Let C be a Hausdorff space, $f: C \to R^*$ a function, and U a subset of C. Assume that the set $U(\lambda)$ defined by

$$U(\lambda) = \{u \in U \mid f(u) \leq \lambda\}$$

is compact for each $\lambda \in R$. Then f attains a minimum over U.

Proof If $f(u) = \infty$ for all $u \in U$, then every $u \in U$ attains the minimum. If $f^* = \inf\{f(u) | u \in U\} < \infty$, let $\{\lambda_n\}$ be a scalar sequence such that $\lambda_n > \lambda_{n+1}$ for all n and $\lambda_n \to f^*$. Then the sets $U(\lambda_n)$ are nonempty, compact, and satisfy $U(\lambda_n) \supset U(\lambda_{n+1})$ for all n. Hence, the intersection $\bigcap_{n=1}^{\infty} U(\lambda_n)$ is nonempty and compact. Let u^* be any point in the intersection. Then $u^* \in U$ and $f(u^*) \leq \lambda_n$ for all n, and it follows that $f(u^*) \leq f^*$. Hence, f attains its minimum over U at u^*. Q.E.D.

Direct application of Corollary 3.3.1 and Lemma 3.1 yields the following proposition.

Proposition 3.4 Let the control space C be a Hausdorff space and assume that for each $x \in S$, $\lambda \in R$, and $k = 0, 1, \ldots, N-1$, the set

$$U_k(x, \lambda) = \{u \in U(x) | H[x, u, T^k(J_0)] \leq \lambda\} \tag{16}$$

is compact. Then

$$J_N^* = T^N(J_0),$$

and there exists a uniformly N-stage optimal policy.

The compactness of the sets $U_k(x, \lambda)$ of (16) may be verified in a number of important special cases. As an illustration, we state two sets of assumptions which guarantee compactness of $U_k(x, \lambda)$ in the case of the mapping

$$H(x, u, J) = g(x, u) + \alpha(x, u) J[f(x, u)]$$

corresponding to a deterministic optimal control problem (Section 2.3.1).

Assume that $0 \leq \alpha(x, u)$, $b \leq g(x, u) < \infty$ for some $b \in R$ and all $x \in S$, $u \in U(x)$, and take $J_0 \equiv 0$. Then compactness of $U_k(x, \lambda)$ is guaranteed if:

(a) $S = R^n$ (n-dimensional Euclidean space), $C = R^m$, $U(x) \equiv C$, f, g, and α are continuous in (x, u), and g satisfies $\lim_{k \to \infty} g(x_k, u_k) = \infty$ for every bounded sequence $\{x_k\}$ and every sequence $\{u_k\}$ for which $|u_k| \to \infty$ ($|\cdot|$ is a norm on R^m);

(b) $S = R^n$, $C = R^m$, f, g, and α are continuous, $U(x)$ is compact and nonempty for each $x \in R^n$, and $U(\cdot)$ is a continuous point-to-set mapping from R^n to the space of all nonempty compact subsets of R^m. The metric on this space is given by (3) of Appendix C.

The proof consists of verifying that the functions $T^k(J_0)$, $k = 0, 1, \ldots, N-1$, are continuous, which in turn implies compactness of the sets $U_k(x, \lambda)$ of (16). Additional results along the lines of Proposition 3.4 will be given in Part II (cf. Corollary 8.5.2 and Proposition 8.6).

3.3 Application to Specific Models

We will now apply the results of the previous section to the models described in Section 2.3.

Stochastic Optimal Control—Outer Integral Formulation

Proposition 3.5 The mapping

$$H(x,u,J) = E^*\{g(x,u,w) + \alpha J[f(x,u,w)]|x,u\} \tag{17}$$

of Section 2.3.3 satisfies Assumptions F.2 and F.3.

Proof We have

$$H(x,u,J) = \int^*\{g(x,u,w) + \alpha J[f(x,u,w)]\}p(dw|x,u),$$

where $\int^*$ denotes the outer integral as in Appendix A. From Lemma A.3(b) we obtain for all $x \in S$, $u \in C$, $J \in F$, $r > 0$,

$$H(x,u,J) \le H(x,u,J+r) \le H(x,u,J) + 2\alpha r.$$

Hence, F.2 is satisfied.

We now show F.3. Let $J \in F$, $\{J_n\} \subset F$, $\{\varepsilon_n\} \subset R$ satisfy $\sum_{n=1}^{\infty} \varepsilon_n < \infty$, $\varepsilon_n > 0$, and for all n,

$$J = \lim_{n\to\infty} J_n, \qquad J \le J_n, \tag{18}$$

$$J_n(x) \le \begin{cases} J(x) + \varepsilon_n & \text{if} \quad J(x) > -\infty, \\ J_{n-1}(x) + \varepsilon_n & \text{if} \quad J(x) = -\infty, \end{cases} \tag{19, 20}$$

$$H(x,u,J_1) < \infty, \qquad \forall x \in S, \quad u \in U(x). \tag{21}$$

Let $\{\bar{\mu}_n\} \subset M$ be such that for all n,

$$T_{\bar{\mu}_n}(J)(x) \le \begin{cases} T(J)(x) + \varepsilon_n & \text{if} \quad T(J)(x) > -\infty, \\ -1/\varepsilon_n & \text{if} \quad T(J)(x) = -\infty, \end{cases} \tag{22, 23}$$

$$T_{\bar{\mu}_n}(J) \le T_{\bar{\mu}_{n-1}}(J). \tag{24}$$

Consider the set

$$A(J) = \{x \in S | \text{there exists } u \in U(x) \text{ with } p^*(\{w|J[f(x,u,w)] = -\infty\}|x,u) > 0\},$$

where p^* denotes p-outer measure (see Appendix A). Let $\bar{\mu} \in M$ be such that

$$p^*(\{w|J[f(x,\bar{\mu}(x),w)] = -\infty\}|x,\bar{\mu}(x)) > 0 \qquad \forall x \in A(J). \tag{25}$$

Define for all n

$$\mu_n(x) = \begin{cases} \bar{\mu}(x) & \text{if } x \in A(J), \\ \bar{\mu}_n(x) & \text{if } x \notin A(J). \end{cases} \tag{26}$$

We will show that $\{\mu_n\}$ thus defined satisfies the requirement of F.3 with $\beta = 1 + 2\alpha$.

For $x \in A(J)$, we have, from Corollary A.1.1 and (18)–(21),

$$\begin{aligned} \limsup_{n\to\infty} T_{\mu_n}(J_n)(x) &= \limsup_{n\to\infty} T_{\bar{\mu}}(J_n)(x) \\ &= \limsup_{n\to\infty} \int^* \{g[x, \bar{\mu}(x), w] + \alpha J_n[f(x, \bar{\mu}(x), w)]\} \\ &\quad \times p(dw|x, \bar{\mu}(x)) \\ &= \int^* \{g[x, \bar{\mu}(x), w] + \alpha J[f(x, \bar{\mu}(x), w)]\} p(dw|x, \bar{\mu}(x)). \end{aligned}$$

It follows from Lemma A.3(g) and the fact that $T_{\bar{\mu}}(J)(x) < \infty$ [cf. (18) and (21)] that

$$\limsup_{n\to\infty} T_{\mu_n}(J_n)(x) = -\infty \le T(J)(x). \tag{27}$$

For $x \notin A(J)$, we have, for all n,

$$p^*(\{w|J[f(x, \mu_n(x), w)] = -\infty\}|x, \mu_n(x)) = 0.$$

Take $B_n \in \mathcal{F}$ to contain $\{w|J[f(x, \mu_n(x), w)] = -\infty\}$ and satisfy

$$p(B_n|x, \mu_n(x)) = 0 \qquad \forall n.$$

Using Lemma A.3(e) and (b) and (19), we have

$$\begin{aligned} T_{\mu_n}(J_n)(x) &= \int^* \chi_{W-B_n}(w)\{g[x, \mu_n(x), w] + \alpha J_n[f(x, \mu_n(x), w)]\} p(dw|x, \mu_n(x)) \\ &\le \int^* \chi_{W-B_n}(w)\{g[x, \mu_n(x), w] + \alpha J[f(x, \mu_n(x), w)]\} \\ &\quad \times p(dw|x, \mu_n(x)) + 2\alpha\varepsilon_n \\ &= T_{\mu_n}(J)(x) + 2\alpha\varepsilon_n. \end{aligned} \tag{28}$$

Hence, for $x \notin A(J)$ we have from (28), (22), and (23) that

$$\limsup_{n\to\infty} T_{\mu_n}(J_n)(x) \le \limsup_{n\to\infty} T_{\mu_n}(J)(x) = T(J)(x).$$

Combining (27) and this relation we obtain

$$\limsup_{n\to\infty} T_{\mu_n}(J_n)(x) \le T(J)(x) \qquad \forall x \in S,$$

and since $T_{\mu_n}(J_n) \geq T(J)$ for all n, it follows that

$$\lim_{n\to\infty} T_{\mu_n}(J_n) = T(J). \tag{29}$$

If x is such that $T(J)(x) > -\infty$, it follows from (27) and (29) that we must have $x \notin A(J)$. Hence, from (28), (22), and Lemma A.3(b),

$$T_{\mu_n}(J_n)(x) \leq T_{\mu_n}(J)(x) + 2\alpha\varepsilon_n \leq T(J)(x) + (1+2\alpha)\varepsilon_n \qquad \text{if} \quad T(J)(x) > -\infty. \tag{30}$$

If x is such that $T(J)(x) = -\infty$, there are two possibilities:

(a) $x \notin A(J)$ and
(b) $x \in A(J)$.

If $x \notin A(J)$, it follows from (28), (24), and (18) that

$$\begin{aligned} T_{\mu_n}(J_n)(x) &\leq T_{\mu_n}(J)(x) + 2\alpha\varepsilon_n \leq T_{\mu_{n-1}}(J)(x) + 2\alpha\varepsilon_n \\ &\leq T_{\mu_{n-1}}(J_{n-1})(x) + 2\alpha\varepsilon_n. \end{aligned} \tag{31}$$

If $x \in A(J)$, then by (18)–(20) and Lemma A.3(b),

$$\begin{aligned} T_{\mu_n}(J_n)(x) &= \int^* \{g[x, \bar{\mu}(x), w] + \alpha J_n[f(x, \bar{\mu}(x), w)]\} p(dw|x, \bar{\mu}(x)) \\ &\leq \int^* \{g[x, \bar{\mu}(x), w] + \alpha J_{n-1}[f(x, \bar{\mu}(x), w)]\} p(dw|x, \bar{\mu}(x)) + 2\alpha\varepsilon_n \\ &= T_{\mu_{n-1}}(J_{n-1})(x) + 2\alpha\varepsilon_n. \end{aligned} \tag{32}$$

It follows now from (29)–(32) that $\{\mu_n\}$ satisfies the requirement of F.3 with $\beta = 1 + 2\alpha$. Q.E.D.

As mentioned earlier, mapping (17) contains as special cases the mappings of Sections 2.3.1 and 2.3.2. In fact, for those mappings F.1 is satisfied as well, as the reader may easily verify by using the monotone covergence theorem for ordinary integration.

Direct application of the results of the previous section and Proposition 3.5 yields the following.

Corollary 3.5.1 Let H be mapping (17) and let $J_0(x) = 0$ for $\forall x \in S$.

(a) If $J_{k,\pi}(x) < \infty$ for all $x \in S$, $\pi \in \Pi$, and $k = 1, 2, \ldots, N$, then $J_N^* = T^N(J_0)$ and for each sequence $\{\varepsilon_n\}$ with $\varepsilon_n \downarrow 0$, $\varepsilon_n > 0$ for $\forall n$, there exists a sequence of policies $\{\pi_n\}$ exhibiting $\{\varepsilon_n\}$-dominated convergence to optimality. In particular, if in addition $J_N^*(x) > -\infty$ for all $x \in S$, then for every $\varepsilon > 0$ there exists an ε-optimal policy.

(b) If $J_k^*(x) > -\infty$ for all $x \in S$, $k = 1, 2, \ldots, N$, then $J_N^* = T^N(J_0)$ and for each $\varepsilon > 0$ there exists an N-stage ε-optimal policy.

(c) Propositions 3.3 and 3.4 and Corollary 3.3.1 apply.

As Counterexample 3 in the previous section shows, it is possible to have $J_N^* \neq T^N(J_0)$ in the stochastic optimal control problem if the assumptions of parts (a) and (b) of Corollary 3.5.1 are not satisfied. Naturally for special classes of problems it may be possible to guarantee the equality $J_N^* = T^N(J_0)$ in other ways. For example, if the problem is such that existence of a uniformly N-stage optimal policy is assured, then we obtain $J_N^* = T^N(J_0)$ via Corollary 3.3.1(b). An important special case where we have $J_N^* = T^N(J_0)$ without any further assumptions is the deterministic optimal control problem of Section 2.3.1. This fact can be easily verified by the reader by using essentially the same argument as the one used to prove Proposition 3.1(a). However, if $J_N^*(x) = -\infty$ for some $x \in S$, even in the deterministic problem there may not exist an N-stage ε-optimal policy for a given ε (see Counterexample 4).

Stochastic Optimal Control—Multiplicative Cost Functional

Proposition 3.6 The mapping

$$H(x, u, J) = E\{g(x, u, w)J[f(x, u, w)]|x, u\} \tag{33}$$

of Section 2.3.4 satisfies F.1. If there exists a $b \in R$ such that $0 \leq g(x, u, w) \leq b$ for all $x \in S$, $u \in U(x)$, $w \in W$, then H satisfies F.2.

Proof Assumption F.1 is satisfied by virtue of the monotone convergence theorem for ordinary integration (recall that W is countable). Also, if $0 \leq g(x, u, w) \leq b$, we have for every $J \in F$ and $r > 0$,

$$\begin{aligned} H(x, u, J + r) &= E\{g(x, u, w)(J[f(x, u, w)] + r)|x, u\} \\ &= E\{g(x, u, w)J[f(x, u, w)]|x, u\} + rE\{g(x, u, w)|x, u\}. \end{aligned}$$

Thus F.2 is satisfied with $\alpha = b$. Q.E.D.

By combining Propositions 3.6 and 3.1, we obtain the following.

Corollary 3.6.1 Let H be the mapping (33) and $J_0(x) = 1$ for $\forall x \in S$.

(a) If $J_{k,\pi}(x) < \infty$ for all $x \in S$, $\pi \in \Pi$, $k = 1, 2, \ldots, N$, then $J_N^* = T^N(J_0)$.

(b) If there exists a $b \in R$ such that $0 \leq g(x, u, w) \leq b$ for all $x \in S$, $u \in U(x)$, $w \in W$, then $J_N^* = T^N(J_0)$ and there exists an N-stage ε-optimal policy.

(c) Propositions 3.3 and 3.4 and Corollary 3.3.1 apply.

We now provide two counterexamples showing that the conclusions of parts (a) and (b) of Corollary 3.6.1 may fail to hold if the corresponding assumptions are relaxed.

COUNTEREXAMPLE 5 Let everything be as in Counterexample 3 except that $C = (0, \infty)$ instead of $C = R$ (and, of course, $J_0(0) = J_0(1) = 1$ instead

of $J_0(0) = J_0(1) = 0$). Then a straightforward calculation shows that $J_2^*(0) = \infty$, $J_2^*(1) = 0$, while $T^2(J_0)(0) = T^2(J_0)(1) = 0$. Here the assumption that $J_{k,\pi}(x) < \infty$ for all x, π, k is violated, and g is unbounded above.

COUNTEREXAMPLE 6 Let everything be as in Counterexample 4 except for the definition of g. Take $g(0,u) = u$ for $\forall u \in U(0)$ and $g(x,u) = x$ for $\forall u \in U(x)$ if $x \neq 0$. Then for every $\pi \in \Pi$ we have $J_{2,\pi}(x) = x\mu_1(0)$ for every $x \neq 0$, and $J_2^*(x) = 0$ for $\forall x \in S$. On the other hand, there is no two-stage ε-optimal policy for any $\varepsilon > 0$. Here the assumption $J_{k,\pi}(x) < \infty$ for all x, π, k is satisfied, and indeed we have $J_2^*(x) = T^2(J_0)(x) = 0$ for $\forall x \in S$. However, g is unbounded above.

Minimax Control

Proposition 3.7 The mapping

$$H(x,u,J) = \sup_{w \in W(x,u)} \{g(x,u,w) + \alpha J[f(x,u,w)]\} \tag{34}$$

of Section 2.3.5 satisfies F.2.

Proof We have for $r > 0$ and $J \in F$,

$$H(x,u,J+r) = \sup_{w \in W(x,u)} \{g(x,u,w) + \alpha J[f(x,u,w)] + \alpha r\}$$

$$= H(x,u,J) + \alpha r. \qquad \text{Q.E.D.}$$

Corollary 3.7.1 Let H be mapping (34) and $J_0(x) = 0$ for $\forall x \in S$.

(a) If $J_k^*(x) > -\infty$ for all $x \in S$, $k = 1, 2, \ldots, N$, then $J_N^* = T^N(J_0)$, and for each $\varepsilon > 0$ there exists an N-stage ε-optimal policy.

(b) Propositions 3.3 and 3.4 and Corollary 3.3.1 apply.

If we have $J_k^*(x) = -\infty$ for some $x \in S$, then it is clearly possible that there exists no N-stage ε-optimal policy for a given $\varepsilon > 0$, since this is true even for deterministic optimal control problems (Counterexample 4). It is also possible to construct examples very similar to Counterexample 3 which show that it is possible to have $J_N^* \neq T^N(J_0)$ if $J_k^*(x) = -\infty$ for some x and k.

Chapter 4

Infinite Horizon Models under a Contraction Assumption

4.1 General Remarks and Assumptions

Consider the infinite horizon problem

$$\text{minimize } J_\pi(x) = \lim_{N\to\infty}(T_{\mu_0}T_{\mu_1}\cdots T_{\mu_{N-1}})(J_0)(x)$$

$$\text{subject to } \pi = (\mu_0, \mu_1, \ldots) \in \Pi.$$

The following assumption is motivated by the contraction property of the mapping associated with discounted stochastic optimal control problems with bounded cost per stage (cf. DPSC, Chapter 6).

Assumption C (Contraction Assumption) There is a closed subset $\bar{B}$ of the space B (Banach space of all bounded real-valued functions on S with the supremum norm) such that $J_0 \in \bar{B}$, and for all $J \in \bar{B}$, $\mu \in M$, the functions $T(J)$ and $T_\mu(J)$ belong to $\bar{B}$. Furthermore, for every $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$, the limit

$$\lim_{N\to\infty}(T_{\mu_0}T_{\mu_1}\cdots T_{\mu_{N-1}})(J_0)(x) \tag{1}$$

exists and is a real number for each $x \in S$. In addition, there exists a positive integer m and scalars ρ, α, with $0 < \rho < 1$, $0 < \alpha$, such that

$$\|T_\mu(J) - T_\mu(J')\| \le \alpha\|J - J'\| \qquad \forall \mu \in M, \quad J, J' \in B, \tag{2}$$

$$\|(T_{\mu_0}T_{\mu_1}\cdots T_{\mu_{m-1}})(J) - (T_{\mu_0}T_{\mu_1}\cdots T_{\mu_{m-1}})(J')\| \le \rho\|J - J'\|$$
$$\forall \mu_0, \ldots, \mu_{m-1} \in M, \quad J, J' \in \bar{B}. \tag{3}$$

Condition (3) implies that the mapping $(T_{\mu_0}T_{\mu_1}\cdots T_{\mu_{m-1}})$ is a contraction mapping in $\bar{B}$ for all $\mu_k \in \mathrm{M}$, $k = 0, 1, \ldots, m-1$. When $m = 1$, the mapping T_μ is a contraction mapping for each $\mu \in M$. Note that (2) is required to hold on a possibly larger set of functions than (3). It is often convenient to take $\bar{B} = B$. This is the case for the problems of Sections 2.3.1, 2.3.2, and 2.3.5 assuming that $\alpha < 1$ and g is uniformly bounded above and below. We will demonstrate this fact in Section 4.4. In other problems such as, for example, the one of Section 2.3.3, the contraction property (3) can be verified only on a strict subset $\bar{B}$ of B.

4.2 Convergence and Existence Results

We first provide some preliminary results in the following proposition.

Proposition 4.1 Let Assumption C hold. Then:

(a) For every $J \in \bar{B}$ and $\pi \in \Pi$, we have

$$J_\pi = \lim_{N\to\infty}(T_{\mu_0}\cdots T_{\mu_{N-1}})(J_0) = \lim_{N\to\infty}(T_{\mu_0}\cdots T_{\mu_{N-1}})(J).$$

(b) For each positive integer N and each $J \in \bar{B}$, we have

$$\inf_{\pi\in\Pi}(T_{\mu_0}\cdots T_{\mu_{N-1}})(J) = T^N(J)$$

and, in particular,

$$J_N^* = \inf_{\pi\in\Pi}(T_{\mu_0}\cdots T_{\mu_{N-1}})(J_0) = T^N(J_0).$$

(c) The mappings T^m and T_μ^m, $\mu \in M$, are contraction mappings in $\bar{B}$ with modulus ρ, i.e.,

$$\|T^m(J) - T^m(J')\| \le \rho\|J - J'\| \qquad \forall J, J' \in \bar{B},$$
$$\|T_\mu^m(J) - T_\mu^m(J')\| \le \rho\|J - J'\| \qquad \forall J, J' \in \bar{B}, \quad \mu \in M.$$

Proof (a) For any integer $k \ge 0$, write $k = nm + q$, where q, n are nonnegative integers and $0 \le q < m$. Then for any $J, J' \in \bar{B}$, using (2) and (3), we obtain

$$\|(T_{\mu_0}\cdots T_{\mu_{k-1}})(J) - (T_{\mu_0}\cdots T_{\mu_{k-1}})(J')\| \le \rho^n\alpha^q\|J - J'\|,$$

from which, by taking the limit as k (and hence also n) tends to infinity, we have

$$\lim_{k\to\infty}(T_{\mu_0}\cdots T_{\mu_{k-1}})(J_0)=\lim_{k\to\infty}(T_{\mu_0}\cdots T_{\mu_{k-1}})(J)\qquad \forall J\in\bar{B}.$$

(b) Since $T^k(J)\in\bar{B}$ for all k by assumption, we have $T^k(J)(x)>-\infty$ for all $x\in S$ and k. For any $\varepsilon>0$, let $\bar{\mu}_k\in M$, $k=0,1,\ldots,N-1$, be such that

$$\begin{aligned}T_{\bar{\mu}_{N-1}}(J)&\le T(J)+\varepsilon,\\ (T_{\bar{\mu}_{N-2}}T)(J)&\le T^2(J)+\varepsilon,\\ &\vdots\\ (T_{\bar{\mu}_0}T^{N-1})(J)&\le T^N(J)+\varepsilon.\end{aligned}$$

Using (2) we obtain

$$\begin{aligned}T^N(J)&\ge (T_{\bar{\mu}_0}T^{N-1})(J)-\varepsilon\\ &\ge T_{\bar{\mu}_0}[(T_{\bar{\mu}_1}T^{N-2})(J)-\varepsilon]-\varepsilon\\ &\ge (T_{\bar{\mu}_0}T_{\bar{\mu}_1}T^{N-2})(J)-\alpha\varepsilon-\varepsilon\\ &\vdots\\ &\ge (T_{\bar{\mu}_0}T_{\bar{\mu}_1}\cdots T_{\bar{\mu}_{N-1}})(J)-\left(\sum_{k=0}^{N-1}\alpha^k\varepsilon\right)\\ &\ge \inf_{\pi\in\Pi}(T_{\mu_0}\cdots T_{\mu_{N-1}})(J)-\left(\sum_{k=0}^{N-1}\alpha^k\varepsilon\right).\end{aligned}$$

Since $\varepsilon>0$ is arbitrary, it follows that

$$T^N(J)\ge\inf_{\pi\in\Pi}(T_{\mu_0}\cdots T_{\mu_{N-1}})(J).$$

The reverse inequality clearly holds and the result follows.

(c) The fact that T_μ^m is a contraction mapping is immediate from (3). We also have from (3) for all $\mu_k\in M$, $k=0,\ldots,m-1$, and $J,J'\in\bar{B}$,

$$(T_{\mu_0}\cdots T_{\mu_{m-1}})(J)\le(T_{\mu_0}\cdots T_{\mu_{m-1}})(J')+\rho\|J-J'\|.$$

Taking the infimum of both sides over $\mu_k\in M$, $k=0,1,\ldots,m-1$, and using part (b) we obtain

$$T^m(J)\le T^m(J')+\rho\|J-J'\|.$$

A symmetric argument yields

$$T^m(J')\le T^m(J)+\rho\|J-J'\|.$$

Combining the two inequalities, we obtain $\|T^m(J)-T^m(J')\|\le\rho\|J-J'\|$.

Q.E.D.

In what follows we shall make use of the following fixed point theorem. (See [O5, p. 383]—the proof found there can be generalized to Banach spaces.)

Fixed Point Theorem If $\bar{B}$ is a closed subset of a Banach space with norm denoted by $\|\cdot\|$ and $L: \bar{B} \to \bar{B}$ is a mapping such that for some positive integer m and scalar $\rho \in (0, 1)$, $\|L^m(z) - L^m(z')\| \leq \rho \|z - z'\|$ for all $z, z' \in \bar{B}$, then L has a unique fixed point in $\bar{B}$, i.e., there exists a unique vector $z^* \in \bar{B}$ such that $L(z^*) = z^*$. Furthermore, for every $z \in \bar{B}$, we have

$$\lim_{N \to \infty} \|L^N(z) - z^*\| = 0.$$

The following proposition characterizes the optimal cost function J^* and the cost function J_μ corresponding to any stationary policy $(\mu, \mu, \ldots) \in \Pi$. It also shows that these functions can be obtained in the limit via successive application of T and T_μ on any $J \in \bar{B}$.

Proposition 4.2 Let Assumption C hold. Then:

(a) The optimal cost function J^* belongs to $\bar{B}$ and is the unique fixed point of T within $\bar{B}$, i.e., $J^* = T(J^*)$, and if $J' \in \bar{B}$ and $J' = T(J')$, then $J' = J^*$. Furthermore, if $J' \in \bar{B}$ is such that $T(J') \leq J'$, then $J^* \leq J'$, while if $J' \leq T(J')$, then $J' \leq J^*$.

(b) For every $\mu \in M$, the function J_μ belongs to $\bar{B}$ and is the unique fixed point of T_μ within $\bar{B}$.

(c) There holds

$$\lim_{N \to \infty} \|T^N(J) - J^*\| = 0 \qquad \forall J \in \bar{B},$$

$$\lim_{N \to \infty} \|T_\mu^N(J) - J_\mu\| = 0 \qquad \forall J \in \bar{B}, \quad \mu \in M.$$

Proof From part (c) of Proposition 4.1 and the fixed point theorem, we have that T and T_μ have unique fixed points in $\bar{B}$. The fixed point of T_μ is clearly J_μ, and hence part (b) is proved. Let $\tilde{J}^*$ be the fixed point of T. We have $\tilde{J}^* = T(\tilde{J}^*)$. For any $\bar{\varepsilon} > 0$, take $\bar{\mu} \in M$ such that

$$T_{\bar{\mu}}(\tilde{J}^*) \leq \tilde{J}^* + \bar{\varepsilon}.$$

From (2) it follows that $T_{\bar{\mu}}^2(\tilde{J}^*) \leq T_{\bar{\mu}}(\tilde{J}^*) + \alpha\bar{\varepsilon} \leq \tilde{J}^* + (1 + \alpha)\bar{\varepsilon}$. Continuing in the same manner, we obtain

$$T_{\bar{\mu}}^m(\tilde{J}^*) \leq \tilde{J}^* + (1 + \alpha + \cdots + \alpha^{m-1})\bar{\varepsilon}.$$

Using (3) we have

$$\begin{aligned} T_{\bar{\mu}}^{2m}(\tilde{J}^*) &\leq T_{\bar{\mu}}^m(\tilde{J}^*) + \rho(1 + \alpha + \cdots + \alpha^{m-1})\bar{\varepsilon} \\ &\leq \tilde{J}^* + (1 + \rho)(1 + \alpha + \cdots + \alpha^{m-1})\bar{\varepsilon}. \end{aligned}$$

Proceeding similarly, we obtain, for all $k \geq 1$,

$$T_{\bar{\mu}}^{km}(\tilde{J}^*) \leq \tilde{J}^* + (1 + \rho + \cdots + \rho^{k-1})(1 + \alpha + \cdots + \alpha^{m-1})\bar{\varepsilon}.$$

Taking the limit as $k \to \infty$ and using the fact that $J_{\bar{\mu}} = \lim_{k\to\infty} T_{\bar{\mu}}^{km}(\tilde{J}^*)$, we have

$$J_{\bar{\mu}} \leq \tilde{J}^* + \frac{1}{1-\rho}(1 + \alpha + \cdots + \alpha^{m-1})\bar{\varepsilon}. \tag{4}$$

Taking $\bar{\varepsilon} = (1-\rho)(1 + \alpha + \cdots + \alpha^{m-1})^{-1}\varepsilon$, we obtain

$$J_{\bar{\mu}} \leq \tilde{J}^* + \varepsilon.$$

Since $J^* \leq J_{\bar{\mu}}$ and $\varepsilon > 0$ is arbitrary, we see that $J^* \leq \tilde{J}^*$. We also have

$$J^* = \inf_{\pi\in\Pi} \lim_{N\to\infty} (T_{\mu_0} \cdots T_{\mu_{N-1}})(\tilde{J}^*) \geq \lim_{N\to\infty} T^N(\tilde{J}^*) = \tilde{J}^*.$$

Hence $J^* = \tilde{J}^*$ and J^* is the unique fixed point of T. Part (c) follows immediately from the fixed point theorem. The remaining part of (a) follows easily from part (c) and the monotonicity of the mapping T. Q.E.D.

The next proposition relates to the existence and characterization of stationary optimal policies.

Proposition 4.3 Let Assumption C hold. Then:

(a) A stationary policy $\pi^* = (\mu^*, \mu^*, \ldots) \in \Pi$ is optimal if and only if

$$T_{\mu^*}(J^*) = T(J^*).$$

Equivalently, π^* is optimal if and only if

$$T_{\mu^*}(J_{\mu^*}) = T(J_{\mu^*}).$$

(b) If for each $x \in S$ there exists a policy which is optimal at x, then there exists a stationary optimal policy.

(c) For any $\varepsilon > 0$, there exists a stationary ε-optimal policy, i.e., a $\pi_\varepsilon = (\mu_\varepsilon, \mu_\varepsilon, \ldots) \in \Pi$ such that

$$\|J^* - J_{\mu_\varepsilon}\| \leq \varepsilon.$$

Proof (a) If π^* is optimal, then $J_{\mu^*} = J^*$ and the result follows from parts (a) and (b) of Proposition 4.2. If $T_{\mu^*}(J^*) = T(J^*)$, then $T_{\mu^*}(J^*) = J^*$, and hence $J_{\mu^*} = J^*$ by part (b) of Proposition 4.2. If $T_{\mu^*}(J_{\mu^*}) = T(J_{\mu^*})$, then $J_{\mu^*} = T(J_{\mu^*})$ and $J_{\mu^*} = J^*$ by part (a) of Proposition 4.2.

(b) Let $\pi_x^* = (\mu_{0,x}^*, \mu_{1,x}^*, \ldots)$ be a policy which is optimal at $x \in S$. Then using part (a) of Proposition 4.1 and part (a) of Proposition 4.2, we have

$$\begin{aligned} J^*(x) = J_{\pi_x^*}(x) &= \lim_{k\to\infty} (T_{\mu_{0,x}^*} \cdots T_{\mu_{k,x}^*})(J_0)(x) \\ &= \lim_{k\to\infty} (T_{\mu_{0,x}^*} \cdots T_{\mu_{k,x}^*})(J^*)(x) \\ &\geq \lim_{k\to\infty} (T_{\mu_{0,x}^*} T^k)(J^*)(x) = T_{\mu_{0,x}^*}(J^*)(x) \geq T(J^*)(x) = J^*(x). \end{aligned}$$

Hence $T_{\mu_{0,x}^*}(J^*)(x) = T(J^*)(x)$ for each x. Define $\mu^* \in M$ by means of $\mu^*(x) = \mu_{0,x}^*(x)$. Then $T_{\mu^*}(J^*) = T(J^*)$ and the stationary policy $(\mu^*, \mu^*, \ldots)$ is optimal by part (a).

(c) This part was proved earlier in the proof of part (a) of Proposition 4.2 [cf. (4)]. Q.E.D.

Part (a) of Proposition 4.3 shows that there exists a stationary optimal policy if and only if the infimum is attained for every $x \in S$ in the optimality equation

$$J^*(x) = T(J^*)(x) = \inf_{u \in U(x)} H(x, u, J^*).$$

Thus if the set $U(x)$ is a finite set for each $x \in S$, then there exists a stationary optimal policy. The following proposition strengthens this result and also shows that stationary optimal policies may be obtained in the limit from finite horizon optimal policies via the DP algorithm, which for any given $J \in \bar{B}$ successively computes $T(J), T^2(J), \ldots$.

Proposition 4.4 Let Assumption C hold and assume that the control space C is a Hausdorff space. Assume further that for some $J \in \bar{B}$ and some positive integer $\bar{k}$, the sets

$$U_k(x, \lambda) = \{u \in U(x) \mid H[x, u, T^k(J)] \leq \lambda\} \tag{5}$$

are compact for all $x \in S$, $\lambda \in R$, and $k \geq \bar{k}$. Then:

(a) There exists a policy $\pi^* = (\mu_0^*, \mu_1^*, \ldots) \in \Pi$ attaining the infimum for all $x \in S$ and $k \geq \bar{k}$ in the DP algorithm with initial function J, i.e.,

$$(T_{\mu_k^*} T^k)(J) = T^{k+1}(J) \qquad \forall k \geq \bar{k}. \tag{6}$$

(b) There exists a stationary optimal policy.

(c) For every policy π^* satisfying (6), the sequence $\{\mu_k^*(x)\}$ has at least one accumulation point for each $x \in S$.

(d) If $\mu^*: S \to C$ is such that $\mu^*(x)$ is an accumulation point of $\{\mu_k^*(x)\}$ for each $x \in S$, then the stationary policy $(\mu^*, \mu^*, \ldots)$ is optimal.

Proof (a) We have

$$T^{k+1}(J)(x) = \inf_{u \in U(x)} H[x, u, T^k(J)],$$

and the result follows from compactness of sets (5) and Lemma 3.1.

(b) This part will follow immediately once we prove (c) and (d).

(c) Let $\pi^* = (\mu_0^*, \mu_1^*, \ldots)$ satisfy (6) and define

$$\varepsilon_k = \sup\{\|T^i(J) - J^*\| \mid i \geq k\}, \qquad k = 0, 1, \ldots.$$

We have from (2), (6), and the fact that $T(J^*) = J^*$,

$$\begin{aligned}\|(T_{\mu_n^*}T^n)(J) - J^*\| &= \|T^{n+1}(J) - T(J^*)\| \\ &\le \alpha\|T^n(J) - J^*\| \qquad \forall n \ge \bar{k},\\ \|(T_{\mu_n^*}T^n)(J) - (T_{\mu_n^*}T^k)(J)\| &\le \alpha\|T^n(J) - T^k(J)\| \\ &\le \alpha\|T^n(J) - J^*\| + \alpha\|T^k(J) - J^*\| \\ &\qquad\qquad \forall n \ge \bar{k}, \quad k = 0, 1, \ldots.\end{aligned}$$

From these two relations we obtain

$$\begin{aligned}H[x, \mu_n^*(x), T^k(J)] &\le H[x, \mu_n^*(x), T^n(J)] + 2\alpha\varepsilon_k \\ &\le J^*(x) + 3\alpha\varepsilon_k \qquad \forall n \ge k, \quad k \ge \bar{k}.\end{aligned}$$

It follows that $\mu_n^*(x) \in U_k[x, J^*(x) + 3\alpha\varepsilon_k]$ for all $n \ge k$ and $k \ge \bar{k}$, and $\{\mu_n^*(x)\}$ has an accumulation point by the compactness of $U_k[x, J^*(x) + 3\alpha\varepsilon_k]$.

(d) If $\mu^*(x)$ is an accumulation point of $\{\mu_n^*(x)\}$, then $\mu^*(x) \in U_k[x, J^*(x) + 3\alpha\varepsilon_k]$ for all $k \ge \bar{k}$, or equivalently,

$$(T_{\mu^*}T^k)(J)(x) \le J^*(x) + 3\alpha\varepsilon_k \qquad \forall x \in S, \quad k \ge \bar{k}.$$

By using (2), we have, for all k,

$$\|(T_{\mu^*}T^k)(J) - T_{\mu^*}(J^*)\| \le \alpha\|T^k(J) - J^*\| \le \alpha\varepsilon_k.$$

Combining the preceding two inequalities, we obtain

$$T_{\mu^*}(J^*)(x) \le J^*(x) + 4\alpha\varepsilon_k \qquad \forall x \in S, \quad k \ge \bar{k}.$$

Since $\varepsilon_k \to 0$ [cf. Proposition 4.2(c)], we obtain $T_{\mu^*}(J^*) \le J^*$. Using the fact that $J^* = T(J^*) \le T_{\mu^*}(J^*)$, we obtain $T_{\mu^*}(J^*) = J^*$, which implies by Proposition 4.3 that the stationary policy $(\mu^*, \mu^*, \ldots)$ is optimal. Q.E.D.

Examples where compactness of sets (5) can be verified were given at the end of Section 3.2. Another example is the lower semicontinuous stochastic optimal control model of Section 8.3.

4.3 Computational Methods

There are a number of computational methods which can be used to obtain the optimal cost function J^* and optimal or nearly optimal stationary policies. Naturally, these methods will be useful in practice only if they require a finite number of arithmetic operations. Thus, while "theoretical" algorithms which require an infinite number of arithmetic operations are of

some interest, in practice we must modify these algorithms so that they become computationally implementable. In the algorithms we provide, we assume that for any $J \in \bar{B}$ and $\varepsilon > 0$ there is available a computational method which determines in a finite number of arithmetic operations functions $J_\varepsilon \in \bar{B}$ and $\mu_\varepsilon \in M$ such that

$$J_\varepsilon \leq T(J) + \varepsilon, \qquad T_{\mu_\varepsilon}(J) \leq T(J) + \varepsilon.$$

For many problems of interest, S is a compact subset of a Euclidean space, and such procedures may be based on discretization of the state space or the control space (or both) and piecewise constant approximations of various functions (see e.g., DPSC, Section 5.2). Based on this assumption (the limitations of which we fully realize), we shall provide computationally implementable versions of all "theoretical" algorithms we consider.

4.3.1 Successive Approximation

The successive approximation method consists of choosing a starting function $J \in \bar{B}$ and computing successively $T(J), T^2(J), \ldots, T^k(J), \ldots$. By part (c) of Proposition 4.2, we have $\lim_{k\to\infty} \|T^k(J) - J^*\| = 0$, and hence we obtain in the limit the optimal cost function J^*. Subsequently, stationary optimal policies (if any exist) may be obtained by minimization for each $x \in S$ in the optimality equation

$$J^*(x) = \inf_{x \in U(x)} H(x, u, J^*).$$

If this minimization cannot be carried out exactly or if only an approximation to J^* is available, then nearly optimal stationary policies can still be obtained, as the following proposition shows.

Proposition 4.5 Let Assumption C hold and assume that $\tilde{J}^* \in \bar{B}$ and $\mu \in M$ are such that

$$\|\tilde{J}^* - J^*\| \leq \varepsilon_1, \qquad T_\mu(\tilde{J}^*) \leq T(\tilde{J}^*) + \varepsilon_2,$$

where $\varepsilon_1 \geq 0$, $\varepsilon_2 \geq 0$ are scalars. Then

$$J^* \leq J_\mu \leq J^* + [(2\alpha\varepsilon_1 + \varepsilon_2)(1 + \alpha + \cdots + \alpha^{m-1})/(1 - \rho)].$$

Proof Using (2) we obtain

$$T_\mu(J^*) - \alpha\varepsilon_1 \leq T_\mu(\tilde{J}^*) \leq T(\tilde{J}^*) + \varepsilon_2 \leq T(J^*) + (\alpha\varepsilon_1 + \varepsilon_2),$$

and it follows that

$$T_\mu(J^*) \leq J^* + (2\alpha\varepsilon_1 + \varepsilon_2).$$

Using this inequality and an argument identical to the one used to prove (4) in Proposition 4.2, we obtain our result. Q.E.D.

An interesting corollary of this proposition is the following.

Corollary 4.5.1 Let Assumption C hold and assume that S is a finite set and $U(x)$ is a finite set for each $x \in S$. Then the successive approximation method yields an optimal stationary policy after a finite number of iterations in the sense that, for a given $J \in \bar{B}$, if $\pi^* = (\mu_0^*, \mu_1^*, \ldots) \in \Pi$ is such that

$$(T_{\mu_k^*} T^k)(J) = T^{k+1}(J), \qquad k = 0, 1, \ldots,$$

then there exists an integer $\bar{k}$ such that the stationary policy $(\mu_k^*, \mu_k^*, \ldots)$ is optimal for every $k \geq \bar{k}$.

Proof Under our finiteness assumptions, the set M is a finite set. Hence there exists a scalar $\varepsilon^* > 0$ such that $J_\mu \leq J^* + \varepsilon^*$ implies that $(\mu, \mu, \ldots)$ is optimal. Take $\bar{k}$ sufficiently large so that $\|T^k(J) - J^*\| \leq \bar{\varepsilon}$ for all $k \geq \bar{k}$, where $\bar{\varepsilon}$ satisfies $2\alpha\bar{\varepsilon}(1 + \alpha + \cdots + \alpha^{m-1})(1 - \rho)^{-1} \leq \varepsilon^*$, and use Proposition 4.5. Q.E.D.

The successive approximation scheme can be sharpened considerably by making use of the monotonic error bounds of the following proposition.

Proposition 4.6 Let Assumption C hold and assume that for all scalars $r \neq 0$, $J \in B$, and $x \in S$, we have

$$\alpha_1 \leq [T^m(J + r)(x) - T^m(J)(x)]/r \leq \alpha_2, \tag{7}$$

where α_1, α_2 are two scalars satisfying $0 \leq \alpha_1 \leq \alpha_2 < 1$. Then for all $J \in \bar{B}$, $x \in S$, and $k = 1, 2, \ldots$, we have

$$\begin{aligned} T^{km}(J)(x) + b_k &\leq T^{(k+1)m}(J)(x) + b_{k+1} \\ &\leq J^*(x) \leq T^{(k+1)m}(J)(x) + \bar{b}_{k+1} \leq T^{km}(J)(x) + \bar{b}_k, \end{aligned} \tag{8}$$

where

$$b_k = \min\left[\frac{\alpha_1}{1 - \alpha_1} d_k, \frac{\alpha_2}{1 - \alpha_2} d_k\right], \qquad \bar{b}_k = \max\left[\frac{\alpha_1}{1 - \alpha_1} \bar{d}_k, \frac{\alpha_2}{1 - \alpha_2} \bar{d}_k\right],$$

$$d_k = \inf_{x \in S}[T^{km}(J)(x) - T^{(k-1)m}(J)(x)], \qquad \bar{d}_k = \sup_{x \in S}[T^{km}(J)(x) - T^{(k-1)m}(J)(x)].$$

Note If $B = \bar{B}$ we can always take $\alpha_2 = \rho$, $\alpha_1 = 0$, but sharper bounds are obtained if scalars α_1 and α_2 with $0 < \alpha_1$ and/or $\alpha_2 < \rho$ are available.

Proof It is sufficient to prove (8) for $k = 1$, since the result for $k > 1$ then follows by replacing J by $T^{(k-1)m}(J)$. In order to simplify the notation, we assume $m = 1$. In order to prove the result for the general case simply

replace T by T^m in the following arguments. We also use the notation

$$d_1 = d, \qquad \bar{d}_1 = \bar{d}, \qquad d_2 = d', \qquad \bar{d}_2 = \bar{d}'.$$

Relation (7) may also be written (for $m = 1$) as

$$T(J) + \min[\alpha_1 r, \alpha_2 r] \leq T(J + r) \leq T(J) + \max[\alpha_1 r, \alpha_2 r]. \tag{9}$$

We have for all $x \in S$,

$$J(x) + d \leq T(J)(x). \tag{10}$$

Applying T on both sides of (10) and using (9) and (10), we obtain

$$\begin{aligned} J(x) + \min[d + \alpha_1 d, d + \alpha_2 d] &\leq T(J)(x) + \min[\alpha_1 d, \alpha_2 d] \\ &\leq T(J + d)(x) \leq T^2(J)(x). \end{aligned} \tag{11}$$

By adding $\min[\alpha_1^2 d, \alpha_2^2 d]$ to each side of these inequalities, using (9) (with J replaced by $T(J)$ and $r = \min[\alpha_1 d, \alpha_2 d]$), and then again (11), we obtain

$$\begin{aligned} J(x) + \min[d + \alpha_1 d + \alpha_1^2 d, d + \alpha_2 d + \alpha_2^2 d] &\leq T(J)(x) + \min[\alpha_1 d + \alpha_1^2 d, \alpha_2 d + \alpha_2^2 d] \\ &\leq T^2(J)(x) + \min[\alpha_1^2 d, \alpha_2^2 d] \\ &\leq T[T(J) + \min[\alpha_1 d, \alpha_2 d]](x) \\ &\leq T^3(J)(x). \end{aligned}$$

Proceeding similarly, we have for every $k = 1, 2, \ldots,$

$$\begin{aligned} J(x) + \min\left[\sum_{i=0}^{k} \alpha_1^i d, \sum_{i=0}^{k} \alpha_2^i d\right] &\leq T(J)(x) + \min\left[\sum_{i=1}^{k} \alpha_1^i d, \sum_{i=1}^{k} \alpha_2^i d\right] \\ &\leq \cdots \leq T^k(J)(x) + \min[\alpha_1^k d, \alpha_2^k d] \\ &\leq T^{k+1}(J)(x). \end{aligned}$$

Taking the limit as $k \to \infty$, we have

$$\begin{aligned} J(x) + \min\left[\frac{1}{1 - \alpha_1} d, \frac{1}{1 - \alpha_2} d\right] &\leq T(J)(x) + \min\left[\frac{\alpha_1}{1 - \alpha_1} d, \frac{\alpha_2}{1 - \alpha_2} d\right] \\ &\leq T^2(J)(x) + \min\left[\frac{\alpha_1^2}{1 - \alpha_1} d, \frac{\alpha_2^2}{1 - \alpha_2} d\right] \\ &\leq J^*(x). \end{aligned} \tag{12}$$

Also, we have from (11) that

$$\min[\alpha_1 d, \alpha_2 d] \leq T^2(J)(x) - T(J)(x),$$

and by taking the infinum over $x \in S$, we see that

$$\min[\alpha_1 d, \alpha_2 d] \leq d'.$$

It is easy to see that this relation implies

$$\min\left[\frac{\alpha_1^2}{1-\alpha_1}d, \frac{\alpha_2^2}{1-\alpha_2}d\right] \leq \min\left[\frac{\alpha_1}{1-\alpha_1}d', \frac{\alpha_2}{1-\alpha_2}d'\right]. \tag{13}$$

Combining (12) and (13) and using the definition of b_1 and b_2, we obtain

$$T(J)(x) + b_1 \leq T^2(J)(x) + b_2.$$

Also from (12) we have $T(J)(x) + b_1 \leq J^*(x)$, and an identical argument shows that $T^2(J)(x) + b_2 \leq J^*(x)$. Hence the left part of (8) is proved for $k=1, m=1$. The right part follows by an entirely similar argument. Q.E.D.

Notice that the scalars b_k and $\bar{b}_k$ in (8) are readily available as a byproduct of the computation. Computational examples and further discussion of the error bounds of Proposition 4.6 may be found in DPSC, Section 6.2.

By using the error bounds of Proposition 4.6, we can obtain J^* to an arbitrary prespecified degree of accuracy in a finite number of iterations of the successive approximation method. However, we still do not have an implementable algorithm, since Proposition 4.6 requires the exact values of the functions $T^k(J)$. Approximations to $T^k(J)$ may, however, be obtained in a computationally implementable manner as shown in the following proposition, which also yields error bounds similar to those of Proposition 4.6.

Proposition 4.7 Let Assumption C hold. For a given $J \in \bar{B}$ and $\varepsilon > 0$, consider a sequence $\{J_k\} \subset \bar{B}$ satisfying

$$T(J) \leq J_1 \leq T(J) + \varepsilon,$$
$$T(J_k) \leq J_{k+1} \leq T(J_k) + \varepsilon, \qquad k = 1, 2, \ldots.$$

Then

$$\|T^{km}(J) - J_{km}\| \leq \bar{\varepsilon}, \qquad k = 0, 1, \ldots, \tag{14}$$

where

$$\bar{\varepsilon} = \varepsilon(1 + \alpha + \cdots + \alpha^{m-1})/(1-\rho)$$

Furthermore, if the assumptions of Proposition 4.6 hold, then for all $x \in S$ and $k = 1, 2, \ldots$

$$J_{km}(x) + \beta_k \leq J^*(x) \leq J_{km}(x) + \bar{\beta}_k,$$

where

$$\beta_k = \min\left[\frac{\alpha_1}{1-\alpha_1}\delta_k, \frac{\alpha_2}{1-\alpha_2}\delta_k\right] - \bar{\varepsilon}, \qquad \bar{\beta}_k = \max\left[\frac{\alpha_1}{1-\alpha_1}\bar{\delta}_k, \frac{\alpha_2}{1-\alpha_2}\bar{\delta}_k\right] + \bar{\varepsilon},$$

$$\delta_k = \inf_{x \in S}[J_{km}(x) - J_{(k-1)m}(x)] - 2\bar{\varepsilon}, \qquad \bar{\delta}_k = \sup_{x \in S}[J_{km}(x) - J_{(k-1)m}(x)] + 2\bar{\varepsilon}.$$

Proof We have

$$\begin{aligned} J_m \le T(J_{m-1}) + \varepsilon &\le T\big[T(J_{m-2}) + \varepsilon\big] + \varepsilon \\ &\le T^2(J_{m-2}) + (1+\alpha)\varepsilon \\ &\le T^2\big[T(J_{m-3}) + \varepsilon\big] + (1+\alpha)\varepsilon \\ &\le T^3(J_{m-3}) + (1+\alpha+\alpha^2)\varepsilon \\ &\vdots \\ &\le T^{m-1}(J_1) + (1+\alpha+\cdots+\alpha^{m-2})\varepsilon \\ &\le T^m(J) + (1+\alpha+\cdots+\alpha^{m-1})\varepsilon. \end{aligned}$$

An identical argument yields

$$J_{2m} \le T^m(J_m) + (1+\alpha+\cdots+\alpha^{m-1})\varepsilon,$$

and we also have

$$\|T^m(J_m) - T^{2m}(J)\| \le \rho\|J_m - T^m(J)\|.$$

Using the preceding three inequalities we obtain

$$\begin{aligned} \|J_{2m} - T^{2m}(J)\| &\le \|J_{2m} - T^m(J_m)\| + \|T^m(J_m) - T^{2m}(J)\| \\ &\le (1+\rho)\varepsilon(1+\alpha+\cdots+\alpha^{m-1}). \end{aligned}$$

Proceeding similarly we obtain, for $k = 1, 2, \ldots,$

$$\|J_{km} - T^{km}(J)\| \le (1+\rho+\cdots+\rho^{k-1})\varepsilon(1+\alpha+\cdots+\alpha^{m-1}),$$

and (14) follows. The remaining part of the proposition follows by using (14) and the error bounds of Proposition 4.6. Q.E.D.

Proposition 4.7 provides the basis for a computationally feasible algorithm to determine J^* to an arbitrary degree of accuracy, and nearly optimal stationary policies can be obtained using the result of Proposition 4.5.

4.3.2 Policy Iteration

The policy iteration algorithm in its theoretical form proceeds as follows. An initial function $\mu_0 \in M$ is chosen, the corresponding cost function J_{μ_0} is computed, and a new function $\mu_1 \in M$ satisfying $T_{\mu_1}(J_{\mu_0}) = T(J_{\mu_0})$ is obtained. More generally, given $\mu_k \in M$, one computes J_{μ_k} and a function $\mu_{k+1} \in M$ satisfying $T_{\mu_{k+1}}(J_{\mu_k}) = T(J_{\mu_k})$, and the process is repeated. When S is a finite set and $U(x)$ is a finite set for each $x \in S$, one can often compute J_{μ_k} in a finite number of arithmetic operations, and the algorithm can be carried out in a computationally implementable manner. Under these circumstances, one obtains an optimal stationary policy in a finite number of iterations, as the following proposition shows.

Proposition 4.8 Let Assumption C hold and assume that S is a finite set and $U(x)$ is a finite set for each $x \in S$. Then for any starting function $\mu_0 \in M$, the policy iteration algorithm yields a stationary optimal policy after a finite number of iterations, i.e., if $\{\mu_k\}$ is the generated sequence, there exists an integer $\bar{k}$ such that $(\mu_k, \mu_k, \ldots)$ is optimal for all $k \geq \bar{k}$.

Proof We have, for all k,

$$T_{\mu_{k+1}}(J_{\mu_k}) = T(J_{\mu_k}) \leq T_{\mu_k}(J_{\mu_k}) = J_{\mu_k}.$$

Applying $T_{\mu_{k+1}}$ repeatedly on both sides, we obtain

$$\begin{aligned} T^N_{\mu_{k+1}}(J_{\mu_k}) \leq T^{N-1}_{\mu_{k+1}}(J_{\mu_k}) \leq \cdots \leq T_{\mu_{k+1}}(J_{\mu_k}) = T(J_{\mu_k}) \\ \leq J_{\mu_k}, \qquad N = 1, 2, \ldots \end{aligned} \tag{15}$$

By Proposition 4.2,

$$\lim_{N \to \infty} T^N_{\mu_{k+1}}(J_{\mu_k}) = J_{\mu_{k+1}}, \tag{16}$$

so $J_{\mu_{k+1}} \leq J_{\mu_k}$

If $(\mu_k, \mu_k, \ldots)$ is an optimal policy, then $J_{\mu_{k+1}} = J_{\mu_k} = J^*$ and $(\mu_{k+1}, \mu_{k+1}, \ldots)$ is also optimal. Otherwise, we must have $J_{\mu_{k+1}}(x) < J_{\mu_k}(x)$ for some $x \in S$, for if $J_{\mu_{k+1}} = J_{\mu_k}$, then from (15) and (16) we have $T(J_{\mu_k}) = J_{\mu_k}$, which implies the optimality of $(\mu_k, \mu_k, \ldots)$. Hence, either $(\mu_k, \mu_k, \ldots)$ is optimal or else $(\mu_{k+1}, \mu_{k+1}, \ldots)$ is a strictly better policy. Since the set M is finite under our assumptions, the result follows. Q.E.D.

When S and $U(x)$ are not finite sets, the policy iteration algorithm must be modified for a number of reasons. First, given μ_k, there may not exist a μ_{k+1} such that $T_{\mu_{k+1}}(J_{\mu_k}) = T(J_{\mu_k})$. Second, even if such a μ_{k+1} exists, it may not be possible to obtain $T_{\mu_{k+1}}(J_{\mu_k})$ and $J_{\mu_{k+1}}$ in a computationally implementable manner. For these reasons we consider the following *modified policy iteration algorithm.*

Step 1 Choose a function $\mu_0 \in M$ and positive scalars γ, δ, and ε.

Step 2 Given $\mu_k \in M$, find $\tilde{J}_{\mu_k} \in \bar{B}$ such that $\|\tilde{J}_{\mu_k} - J_{\mu_k}\| \leq \gamma \rho^k$.

Step 3 Find $\mu_{k+1} \in M$ such that $\|T_{\mu_{k+1}}(\tilde{J}_{\mu_k}) - T(\tilde{J}_{\mu_k})\| \leq \delta \rho^k$. If

$$\|T_{\mu_{k+1}}(\tilde{J}_{\mu_k}) - \tilde{J}_{\mu_k}\| \leq \varepsilon,$$

stop. Otherwise, replace μ_k by μ_{k+1} and return to Step 2.

Notice that Steps 2 and 3 of the algorithm are computationally implementable. The next proposition establishes the validity of the algorithm.

Proposition 4.9 Let Assumption C hold. Then the modified policy iteration algorithm terminates in a finite, say $\bar{k}$, number of iterations, and

the final function $\mu_{\bar{k}}$ satisfies

$$\|J_{\mu_{\bar{k}}} - J^*\| \leq \gamma\rho^{\bar{k}} + \frac{(\varepsilon + \delta\rho^{\bar{k}})(1 + \alpha + \cdots + \alpha^{m-1})}{1 - \rho}. \tag{17}$$

Proof We first show that if the algorithm terminates at the $\bar{k}$th iteration, then (17) holds. Indeed we have

$$\|\tilde{J}_{\mu_{\bar{k}}} - J_{\mu_{\bar{k}}}\| \leq \gamma\rho^{\bar{k}}, \tag{18}$$

$$\|T_{\mu_{\bar{k}+1}}(\tilde{J}_{\mu_{\bar{k}}}) - T(\tilde{J}_{\mu_{\bar{k}}})\| \leq \delta\rho^{\bar{k}}, \tag{19}$$

$$\|T_{\mu_{\bar{k}+1}}(\tilde{J}_{\mu_{\bar{k}}}) - \tilde{J}_{\mu_{\bar{k}}}\| \leq \varepsilon. \tag{20}$$

For any positive integer n, we have

$$\begin{aligned}\|\tilde{J}_{\mu_{\bar{k}}} - J^*\| \leq{}& \|\tilde{J}_{\mu_{\bar{k}}} - T^m(\tilde{J}_{\mu_{\bar{k}}})\| + \|T^m(\tilde{J}_{\mu_{\bar{k}}}) - T^{2m}(\tilde{J}_{\mu_{\bar{k}}})\| + \cdots \\ &+ \|T^{(n-1)m}(\tilde{J}_{\mu_{\bar{k}}}) - T^{nm}(\tilde{J}_{\mu_{\bar{k}}})\| + \|T^{nm}(\tilde{J}_{\mu_{\bar{k}}}) - J^*\|.\end{aligned}$$

From this relation we obtain, for all $n \geq 1$,

$$\|\tilde{J}_{\mu_{\bar{k}}} - J^*\| \leq (1 + \rho + \cdots + \rho^{n-1})\|\tilde{J}_{\mu_{\bar{k}}} - T^m(\tilde{J}_{\mu_{\bar{k}}})\| + \|T^{nm}(\tilde{J}_{\mu_{\bar{k}}}) - J^*\|. \tag{21}$$

We also have

$$\begin{aligned}\|\tilde{J}_{\mu_{\bar{k}}} - T^m(\tilde{J}_{\mu_{\bar{k}}})\| \leq{}& \|\tilde{J}_{\mu_{\bar{k}}} - T(\tilde{J}_{\mu_{\bar{k}}})\| + \|T(\tilde{J}_{\mu_{\bar{k}}}) - T^2(\tilde{J}_{\mu_{\bar{k}}})\| + \cdots \\ &+ \|T^{m-1}(\tilde{J}_{\mu_{\bar{k}}}) - T^m(\tilde{J}_{\mu_{\bar{k}}})\|,\end{aligned}$$

from which we obtain, by using (2),

$$\|\tilde{J}_{\mu_{\bar{k}}} - T^m(\tilde{J}_{\mu_{\bar{k}}})\| \leq (1 + \alpha + \cdots + \alpha^{m-1})\|\tilde{J}_{\mu_{\bar{k}}} - T(\tilde{J}_{\mu_{\bar{k}}})\|. \tag{22}$$

Combining (21) and (22), we obtain, for all $n \geq 1$,

$$\begin{aligned}\|\tilde{J}_{\mu_{\bar{k}}} - J^*\| \leq{}& (1 + \rho + \cdots + \rho^{n-1})(1 + \alpha + \cdots + \alpha^{m-1})\|\tilde{J}_{\mu_{\bar{k}}} - T(\tilde{J}_{\mu_{\bar{k}}})\| \\ &+ \|T^{nm}(\tilde{J}_{\mu_{\bar{k}}}) - J^*\|.\end{aligned}$$

Taking the limit as $n \to \infty$, we obtain

$$\|\tilde{J}_{\mu_{\bar{k}}} - J^*\| \leq (1 + \alpha + \cdots + \alpha^{m-1})\|\tilde{J}_{\mu_{\bar{k}}} - T(\tilde{J}_{\mu_{\bar{k}}})\|/(1 - \rho). \tag{23}$$

Using (18), we also have

$$\|J_{\mu_{\bar{k}}} - J^*\| \leq \|J_{\mu_{\bar{k}}} - \tilde{J}_{\mu_{\bar{k}}}\| + \|\tilde{J}_{\mu_{\bar{k}}} - J^*\| \leq \gamma\rho^{\bar{k}} + \|\tilde{J}_{\mu_{\bar{k}}} - J^*\|. \tag{24}$$

From (19) and (20) we obtain

$$\|\tilde{J}_{\mu_{\bar{k}}} - T(\tilde{J}_{\mu_{\bar{k}}})\| \leq \varepsilon + \delta\rho^{\bar{k}}. \tag{25}$$

By combining (23)–(25), we obtain (17).

To show that the algorithm will terminate in a finite number of iterations, assume the contrary, i.e., assume we have $\|T_{\mu_{k+1}}(\tilde{J}_{\mu_k}) - \tilde{J}_{\mu_k}\| > \varepsilon$ for all k,

and the algorithm generates an infinite sequence $\{\mu_k\} \subset M$. We have, for all k,

$$\begin{aligned}\|T_{\mu_{k+1}}(J_{\mu_k}) - T(J_{\mu_k})\| &\le \|T_{\mu_{k+1}}(J_{\mu_k}) - T_{\mu_{k+1}}(\tilde{J}_{\mu_k})\| \\ &\quad + \|T_{\mu_{k+1}}(\tilde{J}_{\mu_k}) - T(\tilde{J}_{\mu_k})\| + \|T(\tilde{J}_{\mu_k}) - T(J_{\mu_k})\| \\ &\le (\delta + 2\alpha\gamma)\rho^k.\end{aligned}$$

This relation yields, for all k,

$$\begin{aligned}T_{\mu_{k+1}}(J_{\mu_k}) \le T(J_{\mu_k}) + (\delta + 2\alpha\gamma)\rho^k &\le T_{\mu_k}(J_{\mu_k}) + (\delta + 2\alpha\gamma)\rho^k \\ &= J_{\mu_k} + (\delta + 2\alpha\gamma)\rho^k. \qquad (26)\end{aligned}$$

Applying $T_{\mu_{k+1}}$ to both sides of (26) and using (26) again, we obtain

$$\begin{aligned}T^2_{\mu_{k+1}}(J_{\mu_k}) &\le T_{\mu_{k+1}}(J_{\mu_k}) + \alpha(\delta + 2\alpha\gamma)\rho^k \le T(J_{\mu_k}) + (1 + \alpha)(\delta + 2\alpha\gamma)\rho^k \\ &\le J_{\mu_k} + (1 + \alpha)(\delta + 2\alpha\gamma)\rho^k.\end{aligned}$$

Proceeding similarly, we obtain, for all k,

$$\begin{aligned}T^m_{\mu_{k+1}}(J_{\mu_k}) &\le T(J_{\mu_k}) + (1 + \alpha + \cdots + \alpha^{m-1})(\delta + 2\alpha\gamma)\rho^k \\ &\le J_{\mu_k} + (1 + \alpha + \cdots + \alpha^{m-1})(\delta + 2\alpha\gamma)\rho^k.\end{aligned}$$

Applying $T^m_{\mu_{k+1}}$ repeatedly to both sides, we obtain, for all n and k,

$$T^{nm}_{\mu_{k+1}}(J_{\mu_k}) \le T(J_{\mu_k}) + (1 + \rho + \cdots + \rho^{n-1})(1 + \alpha + \cdots + \alpha^{m-1})(\delta + 2\alpha\gamma)\rho^k. \qquad (27)$$

Denote

$$\lambda = (1 + \alpha + \cdots + \alpha^{m-1})(\delta + 2\alpha\gamma)/(1 - \rho).$$

Then by taking the limit in (27) as $n \to \infty$, we obtain

$$J_{\mu_{k+1}} \le T(J_{\mu_k}) + \lambda\rho^k, \qquad k = 0, 1, \ldots.$$

By repeatedly applying T to both sides, we obtain

$$J_{\mu_{nm}} \le T^m(J_{\mu_{(n-1)m}}) + \lambda(\alpha^{m-1} + \alpha^{m-2}\rho + \cdots + \rho^{m-1})\rho^{(n-1)m}. \qquad (28)$$

Let $\bar{\lambda} = \lambda(\alpha^{m-1} + \alpha^{m-2}\rho + \cdots + \rho^{m-1})$. Then (28) can be written as

$$J_{\mu_{nm}} \le T^m(J_{\mu_{(n-1)m}}) + \bar{\lambda}\rho^{(n-1)m}, \qquad n = 1, 2, \ldots. \qquad (29)$$

Using (29) repeatedly, we have, for all n,

$$\begin{aligned}J_{\mu_{nm}} &\le T^m(J_{\mu_{(n-1)m}}) + \bar{\lambda}\rho^{(n-1)m} \\ &\le T^m\big[T^m(J_{\mu_{(n-2)m}}) + \bar{\lambda}\rho^{(n-2)m}\big] + \bar{\lambda}\rho^{(n-1)m} \\ &\le T^{2m}(J_{\mu_{(n-2)m}}) + \bar{\lambda}\big[\rho^{(n-1)m} + \rho\rho^{(n-2)m}\big] \\ &\;\;\vdots \\ &\le T^{nm}(J_{\mu_0}) + \bar{\lambda}\big[\rho^{(n-1)m} + \rho\rho^{(n-2)m} + \rho^2\rho^{(n-3)m} + \cdots + \rho^{n-1}\big].\end{aligned}$$

Since $\rho^k \rho^{(n-k-1)m} \leq \rho^{n-1}$ for all $k = 0, 1, \ldots, n-1$, this inequality yields

$$J^* \leq J_{\mu_{nm}} \leq T^{nm}(J_{\mu_0}) + n\rho^{n-1}\bar{\lambda}, \qquad n = 1, 2, \ldots.$$

Since $\lim_{n\to\infty}(n\rho^{n-1}) = 0$ and $\lim_{n\to\infty} \|T^{nm}(J_{\mu_0}) - J^*\| = 0$, the right side tends to J^* as $n \to \infty$, and it follows that

$$\lim_{n\to\infty} \|J_{\mu_{nm}} - J^*\| = 0.$$

Since by construction

$$\begin{aligned}\|T_{\mu_{nm+1}}(\tilde{J}_{\mu_{nm}}) - \tilde{J}_{\mu_{nm}}\| &\leq \|T_{\mu_{nm+1}}(\tilde{J}_{\mu_{nm}}) - T(\tilde{J}_{\mu_{nm}})\| \\ &\quad + \|T(\tilde{J}_{\mu_{nm}}) - T(J_{\mu_{nm}})\| + \|T(J_{\mu_{nm}}) - T(J^*)\| \\ &\quad + \|J^{*\prime} - J_{\mu_{nm}}\| + \|J_{\mu_{nm}} - \tilde{J}_{\mu_{nm}}\| \\ &\leq (\delta + \alpha\gamma + \gamma)\rho^{nm} + (1+\alpha)\|J_{\mu_{nm}} - J^*\|,\end{aligned}$$

we conclude that

$$\lim_{n\to\infty} \|T_{\mu_{nm+1}}(\tilde{J}_{\mu_{nm}}) - \tilde{J}_{\mu_{nm}}\| = 0.$$

This contradicts our assumption that

$$\|T_{\mu_{k+1}}(\tilde{J}_{\mu_k}) - \tilde{J}_{\mu_k}\| > \varepsilon$$

for every k. Q.E.D.

4.3.3 Mathematical Programming

Let the state space S be a finite set denoted by

$$S = \{x_1, x_2, \ldots, x_n\},$$

and assume $\bar{B} = B$. From part (a) of Proposition 4.2, we have that whenever $J \in B$ and $J \leq T(J)$, then $J \leq J^*$. Hence the values $J^*(x_1), \ldots, J^*(x_n)$ solve the mathematical programming problem

$$\begin{aligned}&\text{maximize} \quad \sum_{i=1}^{n} \lambda_i \\ &\text{subject to} \quad \lambda_i \leq H(x_i, u, J_\lambda), \qquad i = 1, \ldots, n, \quad u \in U(x_i),\end{aligned}$$

where J_λ is the function taking values $J_\lambda(x_i) = \lambda_i, i = 1, \ldots, n$. If $U(x_i)$ is a finite set for each i, then this problem is a finite-dimensional (possibly nonlinear) programming problem having a finite number of inequality constraints. In fact, for the stochastic optimal control problem of Section 2.3.2, this problem is a linear programming problem, as the reader can easily verify (see also DPSC, Section 6.2). This linear program can be solved in a finite number of arithmetic operations.

4.4 Application to Specific Models

The results of the preceding sections apply in their entirety to the problems of Sections 2.3.3 and 2.3.5 if $\alpha < 1$ and g is a nonnegative bounded function. Under these circumstances Assumption C is satisfied, as we now show.

Stochastic Optimal Control—Outer Integral Formulation

Proposition 4.10 Consider the mapping

$$H(x, u, J) = E^*\{g(x, u, w) + \alpha J[f(x, u, w)]|x, u\} \tag{30}$$

of Section 2.3.3 and let $J_0(x) = 0$ for $\forall x \in S$. Assume that $\alpha < 1$ and for some $b \in R$ there holds

$$0 \leq g(x, u, w) \leq b \qquad \forall x \in S, \quad u \in U(x), \quad w \in W.$$

Then Assumption C is satisfied with $\bar{B}$ equal to the set of all nonnegative functions $J \in B$, the scalars in (2) and (3) equal to 2α and α, respectively, and $m = 1$.

Note If the special cases of the mappings of Sections 2.3.1 and 2.3.2 are considered, then $\bar{B}$ can be taken equal to B, and the scalars in (2) and (3) can both be taken equal to α.

Proof Clearly $J_0 \in \bar{B}$ and $T(J)$, $T_\mu(J) \in \bar{B}$ for all $J \in \bar{B}$ and $\mu \in M$. We also have, for any $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$,

$$J_0 \leq T_{\mu_0}(J_0) \leq \cdots \leq (T_{\mu_0} \cdots T_{\mu_k})(J_0) \leq (T_{\mu_0} \cdots T_{\mu_{k+1}})(J_0) \leq \cdots,$$

and hence $\lim_{N\to\infty}(T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(x)$ exists for all $x \in S$. It is also easy to verify inductively using Lemma A.2 that

$$(T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(x) \leq \sum_{k=0}^{N-1} \alpha^k b \leq b/(1-\alpha) \qquad \forall x \in S, \quad N = 1, 2, \ldots.$$

Hence $\lim_{N\to\infty}(T_{\mu_0} \cdots T_{\mu_{N-1}})(J_0)(x)$ is a real number for every x.

We have for all $x \in S$, $J, J' \in B$, $\mu \in M$, and $w \in W$,

$$g[x, \mu(x), w] + \alpha J[f(x, \mu(x), w)] \leq g[x, \mu(x), w] + \alpha J'[f(x, \mu(x), w)] + \alpha\|J - J'\|. \tag{31}$$

Hence, using Lemma A.3(b),

$$E^*\{g[x, \mu(x), w] + \alpha J[f(x, \mu(x), w)]|x, u\} \leq E^*\{g[x, \mu(x), w] + \alpha J'[f(x, \mu(x), w)]|x, u\} + 2\alpha\|J - J'\|. \tag{32}$$

Hence

$$T_\mu(J)(x) - T_\mu(J')(x) \leq 2\alpha\|J - J'\|.$$

A symmetric argument yields $T_\mu(J')(x) - T_\mu(J)(x) \leq 2\alpha\|J - J'\|$. Therefore,

$$|T_\mu(J)(x) - T_\mu(J')(x)| \leq 2\alpha\|J - J'\| \qquad \forall x \in S, \qquad \mu \in M.$$

Taking the supremum of the left side over $x \in S$, we have

$$\|T_\mu(J) - T_\mu(J')\| \leq 2\alpha\|J - J'\| \qquad \forall \mu \in M, \quad J, J' \in B, \tag{33}$$

which shows that (2) holds.

If $J, J' \in \bar{B}$, then from (31), Lemma A.2, and Lemma A.3(a), we obtain in place of (32)

$$E^*\{g[x, \mu(x), w] + \alpha J[f(x, \mu(x), w)]|x, u\}$$
$$\leq E^*\{g[x, \mu(x), w] + \alpha J'[f(x, \mu(x), w)]|x, u\} + \alpha\|J - J'\|,$$

and proceeding as before, we obtain in place of (33)

$$\|T_\mu(J) - T_\mu(J')\| \leq \alpha\|J - J'\| \qquad \forall \mu \in M, \quad J, J' \in \bar{B}.$$

This shows that (3) holds with $\rho = \alpha$. Q.E.D.

Minimax Control

Proposition 4.11 Consider the mapping

$$H(x, u, J) = \sup_{w \in W(x,u)} \{g(x, u, w) + \alpha J[f(x, u, w)]\} \tag{34}$$

of Section 2.3.5 and let $J_0(x) = 0$ for $\forall x \in S$. Assume that $\alpha < 1$ and for some $b \in R$, there holds

$$0 \leq g(x, u, w) \leq b \qquad \forall x \in S, \quad u \in U(x), \quad w \in W.$$

Then Assumption C is satisfied with $\bar{B}$ equal to B, $m = 1$, and the scalars in (2) and (3) both equal to α.

Proof The proof is entirely similar to the one of Proposition 4.10. Q.E.D.

For additional problems where the theory of this chapter is applicable, we refer the reader to DPSC. An example of an interesting problem where Assumption C is satisfied with $m > 1$ is the first passage problem described in Section 7.4 of DPSC.