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

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

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].

References

[A1] R. Ash, "Real Analysis and Probability." Academic Press, New York, 1972.

[A2] K. J. Åström, Optimal control of Markov processes with incomplete state information, *J. Math. Anal. Appl.* **10** (1965), 174–205.

[B1] R. Bellman, "Dynamic Programming." Princeton Univ. Press, Princeton, New Jersey, 1957.

[B2] D. P. Bertsekas, Infinite-time reachability of state-space regions by using feedback control, *IEEE Trans. Automatic Control* **AC-17** (1972), 604–613.

[B3] D. P. Bertsekas, On error bounds for successive approximation methods, *IEEE Trans. Automatic Control* **AC-21** (1976), 394–396.

[B4] D. P. Bertsekas, "Dynamic Programming and Stochastic Control." Academic Press, New York, 1976.

[B5] D. P. Bertsekas, Monotone mappings with application in dynamic programming, *SIAM J. Control Optimization* **15** (1977), 438–464.

[B6] D. P. Bertsekas and S. Shreve, Existence of optimal stationary policies in deterministic optimal control, *J. Math. Anal. Appl.* (to appear).

[B7] P. Billingsley, Invariance principle for dependent random variables, *Trans. Amer. Math. Soc.* **83** (1956), 250–282.

[B8] D. Blackwell, Positive dynamic programming, *Proc. Fifth Berkeley Sympos. Math. Statist. and Probability, 1965*, 415–418.

[B9] D. Blackwell, Discounted dynamic programming, *Ann. Math. Statist.* **36** (1965), 226–235

[B10] D. Blackwell, On stationary policies, *J. Roy. Statist. Soc.* **133A** (1970), 33–37.

[B11] D. Blackwell, Borel-programmable functions, *Ann. Prob.* **6** (1978), 321–324.

[B12] D. Blackwell, D. Freedman, and M. Orkin, The optimal reward operator in dynamic programming, *Ann. Probability* **2** (1974), 926–941.

[B13] N. Bourbaki, "General Topology." Addison–Wesley, Reading, Massachusetts, 1966.

[B14] D. W. Bressler and M. Sion, The current theory of analytic sets, *Canad. J. Math.* **16** (1964), 207–230.

[B15] L. D. Brown and R. Purves, Measurable selections of extrema, *Ann. Statist.* **1** (1973), 902–912.

[C1] D. Cenzer and R. D. Mauldin, Measurable parameterizations and selections, *Trans. Amer. Math. Soc.* (to appear).

[D1] C. Dellacherie, "Ensembles Analytiques, Capacités, Mesures de Hausdorff." Springer-Verlag, Berlin and New York, 1972.

[D2] E. V. Denardo, Contraction mappings in the theory underlying dynamic programming, *SIAM Rev.* **9** (1967), 165–177.

[D3] C. Derman, "Finite State Markovian Decision Processes." Academic Press, New York, 1970.

[D4] J. L. Doob, "Stochastic Processes." Wiley, New York, 1953.

[D5] L. Dubins and D. Freedman, Measurable sets of measures, *Pacific J. Math.* **14** (1964), 1211–1222.

[D6] L. Dubins and L. Savage, "Inequalities for Stochastic Processes (How to Gamble if you Must)." McGraw–Hill, New York, 1965. (Republished by Dover, New York, 1976.)

[D7] J. Dugundji, "Topology." Allyn & Bacon, Rockleigh, New Jersey, 1966.

[D8] E. B. Dynkin and A. A. Juskevic, "Controlled Markov Processes and their Applications." Moscow, 1975. (English translation to be published by Springler-Verlag.)

[F1] D. Freedman, The optimal reward operator in special classes of dynamic programming problems, *Ann. Probability.* **2** (1974), 942–949.

[F2] E. B. Frid, On a problem of D. Blackwell from the theory of dynamic programming, *Theor. Probability Appl.* **15** (1970), 719–722.

[F3] N. Furukawa, Markovian decision processes with compact action spaces, *Ann. Math. Statist.* **43** (1972) 1612–1622.

[F4] N. Furukawa and S. Iwamoto, Markovian decision processes and recursive reward functions, *Bull. Math. Statist.* **15** (1973), 79–91.

[F5] N. Furukawa and S. Iwamoto, Dynamic programming on recursive reward systems, *Bull. Math. Statist.* **17** (1976), 103–126.

[G1] K. Gödel, The consistency of the axiom of choice and of the generalized continuum-hypothesis, *Proc. Nat. Acad. Sci. U.S.A.* **24** (1938), 556–557.

[H1] P. R. Halmos, "Measure Theory." Van Nostrand–Reinhold, Princeton, New Jersey, 1950.

[H2] F. Hausdorff, "Set Theory." Chelsea, Bronx, New York, 1957.

[H3] C. J. Himmelberg, T. Parthasarathy, and F. S. Van Vleck, Optimal plans for dynamic programming problems, *Math. Operations Res.* **1** (1976), 390–394.

[H4] K. Hinderer, "Foundations of Nonstationary Dynamic Programming with Discrete Time Parameter." Springler-Verlag, Berlin and New York, 1970.

[H5] J. Hoffman–Jørgensen, "The Theory of Analytic Spaces." Aarhus Universitet, Aarhus, Denmark, 1970.

[H6] A. Hordijk, "Dynamic Programming and Markov Potential Theory." Mathematical Centre Tracts, Amsterdam, 1974.

[H7] R. Howard, "Dynamic Programming and Markov Processes." MIT Press, Cambridge, Massachusetts, 1960.

[J1] B. Jankov, On the uniformisation of A-sets, *Dokl. Akad. Nauk SSSR* **30** (1941), 591–592 (in Russian).

[J2] W. Jewell, Markov renewal programming I and II, *Operations Res.* **11** (1963), 938–971.

[J3] A. A. Juskevič (Yushkevich), Reduction of a controlled Markov model with incomplete data to a problem with complete information in the case of Borel state and control spaces, *Theor. Probability Appl.* **21** (1976), 153–158.

[K1] L. Kantorovich and B. Livenson, Memoir on analytical operations and projective sets, *Fund. Math.* **18** (1932), 214–279.

[K2] K. Kuratowski, "Topology I." Academic Press, New York, 1966.
[K3] K. Kuratowski, "Topology II." Academic Press, New York, 1968.
[K4] K. Kuratowski and A. Mostowski, "Set Theory." North-Holland, Amsterdam, 1976.
[K5] K. Kuratowski and C. Ryll-Nardzewski, A general theorem on selectors, *Bull. Polish Acad. Sci.* **13** (1965), 397–411.
[K6] H. Kushner, "Introduction to Stochastic Control." Holt, New York, 1971.
[L1] M. Loève, "Probability Theory." Van Nostrand-Reinhold, Princeton, New Jersey, 1963.
[L2] N. Lusin, Sur les ensembles analytiques, *Fund. Math.* **10** (1927), 1–95.
[L3] N. Lusin and W. Sierpinski, Sur quelques propriétés des ensembles (A), *Bull. Acad. Sci. Cracovie* (1918), 35–48.
[M1] G. Mackey, Borel structure in groups and their duals, *Trans. Amer. Math. Soc.* **85** (1957), 134–165.
[M2] A. Maitra, Discounted dynamic programming on compact metric spaces, *Sankhya* **30A** (1968), 211–216.
[M3] J. McQueen, A modified dynamic programming method for Markovian decision problems, *J. Math. Anal. Appl.* **14** (1966), 38–43.
[M4] P. A. Meyer, "Probability and Potentials." Ginn (Blaisdell), Boston, Massachusetts, 1966.
[M5] P. A. Meyer and M. Traki, Reduites et jeux de hasard (Seminaire de Probabilites VII, Universite de Strasbourg, *in* "Lecture Notes in Mathematics," Vol. 321), pp. 155–171. Springer, Berlin, 1973.
[N1] J. von Neumann, On rings of operators. Reduction theory, *Ann. of Math.* **50** (1949), 401–485.
[O1] P. Olsen, Multistage stochastic programming with recourse: The equivalent deterministic problem, *SIAM J. Control Optimization* **14** (1976), 495–517.
[O2] P. Olsen, When is a multistage stochastic programming problem well-defined?, *SIAM J. Control Optimization* **14** (1976), 518–527.
[O3] P. Olsen, Multistage stochastic programming with recourse as mathematical programming in an L_p space, *SIAM J. Control Optimization* **14** (1976), 528–537.
[O4] D. Ornstein, On the existence of stationary optimal strategies, *Proc. Amer. Math. Soc.* **20** (1969), 563–569.
[O5] J. M. Ortega and W. C. Rheinboldt, "Iterative Solutions of Nonlinear Equations in Several Variables." Academic Press, New York, 1970.
[P1] K. Parthasarathy, "Probability Measures on Metric Spaces." Academic Press, New York, 1967.
[P2] Yu. V. Prohorov, Convergence of random processes and limit theorems in probability theory, *Theor. Probability Appl.* **1** (1956), 157–214.
[R1] D. Rhenius, Incomplete information in Markovian decision models, *Ann. Statist.* **2** (1974), 1327–1334.
[R2] R. T. Rockafellar, Integral functionals, normal integrands and measurable selections, *in* "Nonlinear Operators and the Calculus of Variations." Springer-Verlag, Berlin and New York, 1976.
[R3] R. T. Rockafellar and R. Wets, Stochastic convex programming: relatively complete recourse and induced feasibility, *SIAM J. Control Optimization* **14** (1976), 574–589.
[R4] R. T. Rockafellar and R. Wets, Stochastic convex programming: basic duality, *Pacific J. Math.* **62** (1976), 173–195.
[R5] H. L. Royden, "Real Analysis." Macmillan, New York, 1968.
[S1] S. Saks, "Theory of the Integral." Stechert, New York, 1937.
[S2] Y. Sawaragi and T. Yoshikawa, Discrete-time Markovian decision processes with incomplete state information, *Ann. Math. Statist.* **41** (1970), 78–86.

[S3] M. Schäl, On continuous dynamic programming with discrete time parameter, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **21** (1972), 279–288.

[S4] M. Schäl, On dynamic programming: Compactness of the space of policies, *Stochastic Processes Appl.* **3** (1975), 345–364.

[S5] M. Schäl, Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **32** (1975), 179–196.

[S6] E. Selivanovskij, Ob odnom klasse effektivnyh mnozestv (mnozestva C), *Mat. Sb.* **35** (1928), 379–413.

[S7] S. Shreve, *A General Framework for Dynamic Programming with Specializations*, M. S. thesis (1977), Dept. of Elec. Eng., Univ. of Illinois, Urbana.

[S8] S. Shreve, *Dynamic Programming in Complete Separable Spaces*, Ph.D. thesis (1977), Dept. of Math., Univ. of Illinois, Urbana.

[S9] S. Shreve and D. P. Bertsekas, A new theoretical framework for finite horizon stochastic control, *Proc. Fourteenth Annual Allerton Conf. Circuit and System Theory, Allerton Park, Illinois, October, 1976*, 336–343.

[S10] S. Shreve and D. P. Bertsekas, Equivalent stochastic and deterministic optimal control problems, *Proc. 1976 IEEE Conf. Decision and Control, Clearwater Beach, Florida*, 705–709.

[S11] S. Shreve and D. P. Bertsekas, Alternative theoretical frameworks for finite horizon discrete-time stochastic optimal control, *SIAM J. Control Optimization* **16** (1978).

[S12] S. Shreve and D. P. Bertsekas, Universally measurable policies in dynamic programming *Mathematics of Operations Research* (to appear).

[S13] R. Solovay, A model of set-theory in which every set of reals is Lebesgue measurable, *Ann. Math.* **92** (1970), 1–56.

[S14] R. E. Strauch, Negative dynamic programming, *Ann Math. Statist.* **37** (1966), 871–890

[S15] C. Striebel, Sufficient statistics in the optimal control of stochastic systems, *J. Math. Anal. Appl.* **12** (1965), 576–592.

[S16] C. Striebel, "Optimal Control of Discrete Time Stochastic Systems." Springer-Verlag, Berlin and New York, 1975.

[S17] M. Suslin (Souslin), Sur une définition des ensembles measurables B sans nombres transfinis, *C. R. Acad. Sci. Paris* **164** (1917), 88–91.

[V1] V. S. Varadarajan, Weak convergence of measures on separable metric spaces, *Sankhya* **19** (1958), 15–22.

[W1] D. H. Wagner, Survey of measurable selection theorems, *SIAM J. Control Optimization* **15** (1977), 859–903.

[W2] A. Wald, "Statistical Decision Functions." Wiley, New York, 1950.

[W3] H. S. Witsenhausen, A standard form for sequential stochastic control, *Math. Systems Theory* **7** (1973), 5–11.