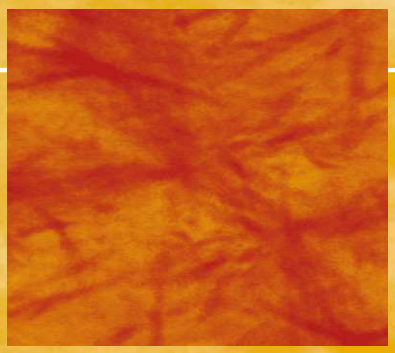
Atle Seierstad

Stochastic Control in Discrete and Continuous Time



Springer

Stochastic Control in Discrete and Continuous Time

Atle Seierstad

Stochastic Control in Discrete and Continuous Time

Springer

Atle Seierstad

Department of Economics
University of Oslo
1095 Blindern
0317 Oslo
Norway
Atle.Seierstad@econ.uio.no

ISBN 978-0-387-76616-4 e-ISBN 978-0-387-76617-1
DOI: 10.1007/978-0-387-76617-1

Library of Congress Control Number: 2008938655

Mathematics Subject Classification (2000): 90C40, 93E20

Printed on acid-free paper

springer.com

This book is dedicated to a lynx I saw in Maridalen in the vicinity of Oslo in April 2006.

Without having read this book (so far), yet maybe it hunts according to an optimized piecewise deterministic process.

Preface

This book contains an introduction to three topics in stochastic control: discrete time stochastic control, i.e., stochastic dynamic programming (Chapter 1), piecewise deterministic control problems (Chapter 3), and control of Ito diffusions (Chapter 4). The chapters include treatments of optimal stopping problems. An Appendix recalls material from elementary probability theory and gives heuristic explanations of certain more advanced tools in probability theory.

The book will hopefully be of interest to students in several fields: economics, engineering, operations research, finance, business, mathematics. In economics and business administration, graduate students should readily be able to read it, and the mathematical level can be suitable for advanced undergraduates in mathematics and science. The prerequisites for reading the book are only a calculus course and a course in elementary probability. (Certain technical comments may demand a slightly better background.)

As this book perhaps (and hopefully) will be read by readers with widely differing backgrounds, some general advice may be useful: Don't be put off if paragraphs, comments, or remarks contain material of a seemingly more technical nature that you don't understand. Just skip such material and continue reading, it will surely not be needed in order to understand the main ideas and results.

The presentation avoids the use of measure theory. At certain points, mainly in Chapter 4, certain measure theoretical concepts are used, they are then explained in a heuristic manner, with more detailed but still heuristic explanations appearing in Appendix. The chosen manner of exposition is quite standard, except in Chapter 3, where a slightly unusual treatment is given that can be useful for the elementary types of problems studied there. In all chapters, problems with terminal restrictions are included.

One might doubt if Ito-diffusions can at all be presented in a useful way without using measure theory. One then has to strike a balance between being completely intuitive and at least giving some ideas about where problems lie hidden, how proofs can be constructed, and directions in which a more advanced treatment must move. I hope that my choices in this respect are not too bad.

A small chapter (Chapter 2) treats deterministic control problems and has been included because it makes possible a very simple exposition of ideas that later on reappear in Chapters 3 and 4. In addition, and more formally, certain proofs in Chapter 3 make use of proven results in Chapter 2.

The level of rigor varies greatly. Formal results should preferably be stated will full rigor, but certain compromises have been unavoidable, especially in Chapter 4, but in fact in all chapters. The degree of rigor in the proofs varies even more. Some proofs are completely heuristic (or even omitted), other ones are nearly, or essentially, rigorous. Quite frequently, first nonrigorous proofs are presented, and then, perhaps annoyingly often, some comment on what is lacking in the proofs, or how they might be improved upon, are added.

Hopefully, what might be called introductory proofs of the most central results are easy to read. Other proofs of more technical material may be quite compact, and then more difficult to read. So the reader may feel that readability varies a lot. In a book of this type and length, it was difficult to avoid this variability in the manner of exposition.

Altogether, there are a great number of remarks in the text giving refinements or extension of results. On (very) first reading, it is advisable to skip most of the remarks and concentrate on the main theory and the examples. Asterisks, usually one (*), are used to indicate material that can be jumped over at first reading; when two asterisks (**) are used, it indicates in addition that somewhat more advanced mathematical tools are used.

Solved examples, examples with analytical (or closed form) solutions, play a big role in the text. The aim is to give the reader a firmer understanding of the theoretical results. It will also equip him or her with a better knowledge of how to solve similar, simple problems, and an idea of how solutions may look like in slightly more complicated problems where analytical solutions cannot be found. The reader should get to know, however, that most problems cannot be solved analytically, they need numerical methods, not treated in the current book.

On the whole, sufficient conditions for optimality are proved with greater rigor (sometimes even with full rigor) compared with proofs of necessary conditions. Even if the latter proofs are heuristic, the necessary conditions that they seem to provide can be compared with the sufficient conditions established, and the former (slightly imprecise) conditions can tell how useful the latter conditions are, in other words how frequently we can hope that the sufficient conditions can help us solve the optimization problems we consider.

A number of exercises, with answers provided (except for a few theoretical problems), have been included in the book.

For the reader wanting to continue studying some of the themes of this book, or who wants to consult alternative expositions of the theory, a small selection of books and articles are provided in the References. These works are also referred to at the ends of the chapters. The few titles provided can only give some hints as to where one can seek more information; for more extensive lists of references, one should look into more specialized works.

Sections 1.1 and 1.2 have, with some changes, appeared in the book by K. Sydsæter et al. (2005), *Further Mathematics for Economic Analysis* by Prentice Hall.

Early versions of the chapters have been tried out in courses for PhD students in economics, from whom useful feedback is gratefully acknowledged.

Peter Hammond has read early versions of the chapters in the book and given a tremendous amount of useful advice, much more than he surely remembers.

Knut Sydsæter and Arne Strøm have read and given comments on parts of the book and also helped in technical matters related to presentation and layout. Also, Tore Schweder has helped me at certain points. For all this help, I am extremely grateful.

All errors and other shortcomings are entirely my own.

The Department of Economics at University of Oslo has over the years made available an excellent work environment and Knut Sydsæter and Arne Strøm, my co-mathematicians at this department, have carried out more than their share of work related to teaching and advising students of all types, which has given me more time to work on this book, among other things. Thanks again.

Finally, I am very grateful for the excellent technical support provided by Springer concerning the preparation of the final version of the manuscript.

Oslo, Norway *Atle Seierstad*
October 2007

Contents

Chapter 1
Stochastic Control over Discrete Time

This chapter describes the optimal governing of certain discrete time stochastic processes over time. First, solution tools for finite horizon problems are presented, the most important being the dynamic programming equation, but also a stochastic maximum principle is rendered. In three sections, infinite horizon problems are treated. Optimal stopping problems are discussed, where when to stop is *a* — or *the* — central question, both for a finite and infinite horizon. Problems of incomplete observations, where we learn more the longer the process runs, are also discussed, and we end this part by presenting stochastic control with Kalman filtering. Finally, some approximation methods are briefly discussed, and an extension to stochastic time periods is presented.

1.1 Stochastic Dynamic Programming

What is the best way of controlling a system governed by a difference equation that is subject to random disturbances? Stochastic dynamic programming is a central tool for tackling this question.

In deterministic dynamic programming, the state develops according to a difference equation $x_{t+1} = f(t_1, x_t, u_t)$, controlled by appropriate choices of the control variables u_t. In the current chapter, the function f is also influenced by random disturbances, so that x_{t+1} is a stochastic quantity. Following common practice, we often (but not always) use capital letters instead of lower-case letters for stochastic quantities, e.g., X_t instead of x_t.

Suppose then that the *state equation* is of the form

$$X_{t+1} = f(t, X_t, u_t, V_{t+1}), \quad X_0 = x_0, V_0 = v_0, x_0, v_0 \text{ given }, u_t \in U, \tag{1.1}$$

where, for each t, V_{t+1} is a random variable that takes values in a finite set $\mathcal{V}$. The probability that $V_{t+1} = v \in \mathcal{V}$ is written $P_t(v|v_t)$; it is assumed that it may depend on the outcome v_t at time t, as well as explicitly on time t. We may allow V_{t+1} to

A. Seierstad, *Stochastic Control in Discrete and Continuous Time*,
DOI: 10.1007/978-0-387-76617-1_1,

be a continuous variable that takes values anywhere in $\mathbb{R}$. Then the distributions of V_{t+1} are often given by densities $p_t(v|v_t)$, separately piecewise continuous in v and v_t. Mostly, we speak as if the V_t's are discrete random variables. However, the solution tools presented can also be used for continuous stochastic variables. We assume that $t = 0,1,\ldots,T$, T a given positive integer, that x_t belongs to $\mathbb{R}^n$, and that u_t is required to belong to a given subset U of $\mathbb{R}^r$. The vectors u_t are subject to choice, and these choices, as well as the stochastic disturbances V_{t+1} determine the development of the state X_t.

Example 1.1. Suppose that $Z_1, Z_2, \ldots$ are independently distributed stochastic variables that take a finite number of positive values (or a continuum of positive values) with specified probabilities independent of both the state and the control. The state X_t develops according to:

$$X_{t+1} = Z_{t+1}(X_t - u_t), \qquad u_t \in [0,\infty). \tag{i}$$

Here u_t is consumption, $X_t - u_t$ is investment, and Z_{t+1} is the return per invested dollar. Moreover, the utility of the terminal state x_T is $\beta^T B x_T^{1-\gamma}$ and the utility of the current consumption is $\beta^t u_t^{1-\gamma}$ for $t < T$, where β is a discount factor, and $0 < \gamma < 1$. The development of the state x_t is now uncertain (stochastic). The objective function to be maximized is the sum of expected discounted utility, given by

$$\sum_{t=0}^{T-1} \beta^t E u_t^{1-\gamma} + \beta^T B E X_T^{1-\gamma}. \tag{ii}$$

□

Let us, for a moment, consider a two-stage decision problem. Assume that one wants to maximize the criterion:

$$E\{f_0(0,X_0,u_0) + f_0(1,X_1,u_1)\} = f_0(0,X_0,u_0) + E f_0(1,X_1,u_1),$$

where E denotes expectation and f_0 is some given function. Here the initial state $X_0 = x_0$ and an initial outcome v_0 are given and X_1 is determined by the difference equation (1.1), i.e., $X_1 = f(0,x_0,u_0,V_1)$. To find the maximum, the following method works: We can first maximize with respect to u_1, and then with respect to u_0. When choosing u_1, we simply maximize $f_0(1,X_1,u_1)$, assuming that X_1 is known before the maximization is carried out. The maximum point u_1^* becomes a function $u_1^*(X_1)$ of X_1. Imagine that this function is inserted for u_1 in the criterion, and that the two occurrences of X_1 are replaced by $f(0,x_0,u_0,V_1)$. Then the criterion becomes equal to

$$f_0(0,X_0,u_0) + E\{f_0(1,f(0,x_0,u_0,V_1),u_1^*(f(0,x_0,u_0,V_1)))\},$$

i.e., u_0 occurs in both terms in the criterion. A maximizing value of u_0 is then chosen, taking both these occurrences into account.

When there are more than two stages, this process is continued backwards, as we shall see.

To see why it matters that we can observe X_1 before choosing u_1, consider the following problem: Let $T = 1$, $f_0(0,x_1,u_1) = 0, f_0(1,x_1,u_1) = X_1u_1$, $X_1 = V_1$, where V_1 takes the values 1 and -1 with probabilities $1/2$, and where u can take the values 1 and -1. Then, $EX_1u_1 = 0$ if we have to choose u_1 before observing X_1 (hence a constant u_1), but if we can first observe X_1, then we can let u_1 depend on X_1. If we choose $u_1 = u_1(X_1) = X_1$, then $EX_1u_1 = 1$, which yields a better value of the objective. In Sections 1.1–1.6, we shall assume that X_t, in fact both X_t and V_t, can be observed before choosing u_t.

Let us turn to the general problem. The process X_t, determined by (1.1) and the random variables V_t, is to be controlled in the best possible manner by appropriate choices of the variables u_t. The *objective function* is now the expectation

$$\sum_{t=0}^{T} E[f_0(t,X_t,u_t(X_t,V_t))]. \tag{1.2}$$

Here several things have to be explained. Each control u_t, $t = 0,1,\ldots,T$ is a function, $u_t(x_t,v_t)$, of the current state x_t and the outcome v_t. Such a function is called a *policy* (or more specifically a *Markov policy* or a *Markov control*). For a large class of stochastic optimization problems, including the one we are now studying, this is the natural class of controls to consider in order to achieve an optimum. Both V_t and X_t are random variables, the X_t's arising from the state equation when the functions $u_s(X_s,V_s)$ are inserted in the equation. The letter E, as before, denotes expectation. For completeness, a detailed description of its calculation follows in the next paragraph, but because it will not be much used later on, readers may want to skip reading it.

To compute the expectation requires specifying the probabilities that are needed in the calculation of the expectation. Given v_0, recall that the probability for the events $V_1 = v_1$ and $V_2 = v_2$ jointly to occur equals the conditional probability for $V_2 = v_2$ to occur, given $V_1 = v_1$, times the probability for $V_1 = v_1$ to occur, given $V_0 = v_0$. Hence it equals $P_1(v_2|v_1)$ times $P_0(v_1|v_0)$. Similarly, given v_0, the probability of the joint event $V_1 = v_1, V_2 = v_2, \ldots, V_t = v_t$, is given by

$$p^*(v_1,\ldots,v_t) := P_0(v_1|v_0)\cdot P_1(v_2|v_1)\cdot\ldots\cdot P_{t-1}(v_t|v_{t-1}). \tag{1.3}$$

(This is actually a conditional probability, v_0 given.) Now, given the policies $u_t(x_t,v_t)$, the sequence X_t, $t = 1,\ldots,T$, in (1.2) is the solution of (1.1), found by calculating, successively, $X_1,X_2,\ldots$, when, successively, $V_1,V_2,\ldots.$ and $u_1 = u_1(X_1,V_1)$, $u_2 = u_2(X_2,V_2),\ldots$ are inserted. Hence X_t depends on V_1, …, V_t and, for each t, the expectation $E f_0(t,X_t,u_t(X_t,V_t))$ is calculated by means of the probabilities specified in (1.3).

Though not always necessary, we shall assume that f_0 and f are continuous in (x,u), (in (x,u,v) if V_t takes values in a nondiscrete set).

The optimization problem is to find a sequence of policies $u_0^*(x_0,v_0),\ldots,$ $u_T^*(x_T,v_T)$ that gives the expression in (1.2) the largest possible value. Such a policy sequence is called an *optimal policy sequence*.

We now define the *optimal value function*

$$J(t,x_t,v_t) = \max E\left[\sum_{s=t}^{T} f_0(s,X_s,u_s(X_s,V_s)) \mid x_t,v_t\right], \tag{1.4}$$

where the maximum is taken over all policy sequences $u_s = u_s(x_s,v_s), s = t,\dots,T$, given v_t and given that we "start the equation" (1.1) at the state x_t at time t, as indicated by "$\mid x_t,v_t$" in (1.4) and apply the controls u_s from the sequence when using (1.1) to calculate all the X_s's. The computation of the expectation in (1.4), (or expectations, when E is taken inside the sum), is now based on conditional probabilities of the form $p^*(v_{t+1},\dots,v_s|v_t) = P_t(v_{t+1}|v_t)\cdot\ldots\cdot P_{s-1}(v_s|v_{s-1})$.

The central tool in solving optimization problems of the type (1.1), (1.2) is the following *optimality* (or *dynamic programming*) equation:

$$J(t-1,x_{t-1},v_{t-1}) = \max_{u_{t-1}}\left\{f_0(t-1,x_{t-1},u_{t-1}) + E\left[J(t,X_t,V_t) \mid x_{t-1},v_{t-1}\right]\right\} \tag{1.5}$$

where $X_t = f(t-1,x_{t-1},u_{t-1},V_t)$ is to be inserted. The "x_{t-1}" in the symbol "$|x_{t-1},v_{t-1}]$" is just a reminder that x_{t-1} occurs in the expression to be inserted. After the insertion, the equation becomes $J(t-1,x_{t-1},v_{t-1}) =$

$$\max_{u_{t-1}}\left\{f_0(t-1,x_{t-1},u_{t-1}) + E\left[J(t,f(t-1,x_{t-1},u_{t-1},V_t),V_t) \mid v_{t-1}\right]\right\},$$

$t = 1,\dots,T$. Moreover, at time T, we have

$$J(T,x_T,v_T) = J(T,x_T) = \max_{u_T} f_0(T,x_T,u_T). \tag{1.6}$$

The equations (1.5), (1.6) are, essentially, both necessary and sufficient. They are sufficient in the sense that *if* $u^*_{t-1}(x_{t-1},v_{t-1})$ maximizes the right-hand side of (1.5) for $t = 1,\dots,T$ and the right-hand side of (1.6) for $t = T+1$, then $u^*_{t-1}(x_{t-1},v_{t-1})$, $t = 1,\dots,T+1$, are optimal policies. On the other hand, they are necessary in the sense that, for every x_{t-1},v_{t-1}, an optimal control $u^*_{t-1}(x_{t-1},v_{t-1}), t = 1,\dots,T$, yields a maximum on the right-hand side of (1.5), and, for $t = T+1$, on the right-hand side of (1.6). To be a little more precise, it is necessary that the optimal control $u^*_{t-1}(x_{t-1},v_{t-1})$ yields a maximum on the right-hand side of (1.5), (1.6) for all values of x_{t-1},v_{t-1} that can occur with positive probability, given $\{u^*_s\}_s$.

The solution method is thus as follows: The relation (1.6) is used to find the functions $u^*_T(x_T,v_T)$ and $J(T,x_T,v_T)$, and then (1.5) is used to find first $u^*_{T-1}(x_{T-1},v_{T-1})$ and $J(T-1,x_{T-1},v_{T-1})$ (then $J(T,x_T,v_T)$ is needed), and then $u^*_{T-2}(x_{T-2},v_{T-2})$ and $J(T-2,x_{T-2},v_{T-2})$ (then $J(T-1,x_{T-1},v_{T-1})$ is needed), and so on, going backwards in time until $u_0(x_0,v_0)$ and $J(0,x_0,v_0)$ have been constructed. At any time t, the optimal control to use, given that (x_t,v_t) has been observed, is then $u_t(x_t,v_t)$.

The intuitive argument for (1.5) is as follows: Suppose the system is in a given state x_{t-1}, and v_{t-1} is given. For a given u_{t-1}, the "instantaneous" reward is $f_0(t,x_{t-1},u_{t-1})$. In addition, the maximal expected sum of rewards at all later

times is $E[J(t,X_t,V_t)|x_{t-1},v_{t-1}]$ when $X_t = f(t-1,x_{t-1},u_{t-1},V_t)$. When using u_{t-1}, the total expected maximum value gained over all future time points (now including even $t-1$), is the sum in (1.5). The largest expected gain comes from choosing u_{t-1} to maximize this sum.

Note that when $P_t(v|v_t)$ does not depend on v_t, then v_t can be dropped in the functions $J_t(x_t,v_t), u_t(x_t,v_t)$, and in (1.5),(1.6). Then in (1.5) the conditioning on v_{t-1} drops out, and $J(t-1,x_{t-1},v_{t-1})$, and the maximizing vector $u_{t-1} = u_{t-1}(x_{t-1},v_{t-1})$ will not depend on v_{t-1}. (In some later sections, $f_0(t,.,.)$ will depend also on v_t and then this simplification does not hold.) In examples below, this simplification is employed.

Example 1.2. Consider the following example

$$\max E\left[\sum_{t=0}^{T-1}(1/2)^t((1-u_t)x_t)^{1/2} + (1/2)^T 2^{1/2}(X_T)^{1/2}\right],$$

subject to

$$X_{t+1} = u_t X_t V_{t+1},\ X_0 = 1, V_t \in \{0,8\},\ \Pr[V_t = 8] = 1/2,\ u_t \in [0,1].$$

This problem is closely related to Example 1.1.

Solution. Evidently, $J(T,x_T) = (1/2)^T 2^{1/2}(x_T)^{1/2}$.

Next, let us find the optimal $u = u_{T-1}$ and $J(T-1,x_{T-1})$, where $J(T-1,x_{T-1}) =$

$$\max_u\{(1/2)^{T-1}((1-u)x_{T-1})^{1/2} + E[(1/2)^T 2^{1/2}(ux_{T-1}V_T)^{1/2}]\} =$$
$$\max_u\{(1/2)^{T-1}((1-u)x_{T-1})^{1/2} + (1/2)^T 2^{1/2}(1/2)8^{1/2}(ux_{T-1})^{1/2}\} =$$
$$\max_u\{(1/2)^{T-1}(x_{T-1})^{1/2}[(1-u)^{1/2} + u^{1/2}]\}.$$

When differentiating to obtain the maximum point (we have a concave function in u), we get

$$(1/2)^{T-1}(x_{T-1})^{1/2}[-(1/2)(1-u)^{-1/2} + (1/2)u^{-1/2}] = 0,$$

which gives $(1-u)^{-1/2} = u^{-1/2}$, or $1-u = u$, i.e., $u = u_{T-1} = 1/2$. Inserting in the maximand, we get

$$J(T-1,x_{T-1}) = (1/2)^{T-1}(x_{T-1})^{1/2}2(1/2)^{1/2} = (1/2)^{T-1}2^{1/2}(x_{T-1})^{1/2}.$$

We now guess that, generally, $J(t,x_t) = (1/2)^t 2^{1/2}(x_t)^{1/2}$. Let us try this guess, hence let us find $J(t-1,x_{t-1})$ and the optimal $u = u_{t-1}$ from the optimality equation (we now see that we can repeat the above calculations for T replaced by t):

$$\begin{aligned} J(t-1,x_{t-1}) &= \max_u\{(1/2)^{t-1}((1-u)x_{t-1})^{1/2} + E[(1/2)^t 2^{1/2}(ux_{t-1}V_t)^{1/2}]\} \\ &= \max_u\{(1/2)^{t-1}(x_{t-1})^{1/2}[(1-u)^{1/2} + u^{1/2}]\} \\ &= (1/2)^{t-1}2^{1/2}(x_{t-1})^{1/2}, \end{aligned}$$

the last equality because when differentiating to obtain the maximum point, we get $(1/2)^{t-1}(x_{t-1})^{1/2}[(-1/2)(1-u)^{-1/2} + (1/2)u^{-1/2}] = 0$, which gives $(1-u)^{-1/2} = u^{-1/2}$, i.e., $u = u_{t-1} = 1/2$ again.

In this example, incidentally, u_{t-1} came out as independent of x_{t-1}. □

In the next example, the outcome of the stochastic variable depends on its value one period earlier.

Example 1.3. We want to solve the problem

$$\begin{aligned} \max \mathrm{E}[X_T + V_T],\ X_{t+1} &= u_t X_t V_{t+1} + (1-u_t)X_t(1-V_{t+1}), \\ X_0 &= 1, u_t \in [0,1], \\ V_t \in \{0,1\}, \Pr[V_{t+1} = 1 | V_t = 1] &= 3/4, \Pr[V_{t+1} = 1 | V_t = 0] = 1/4. \end{aligned}$$

Solution. Formally, we need to work with a second state variable, say Y_t governed by $Y_{t+1} = V_{t+1}$. Then, $f_0(T, x_T, y_T) = x_T + y_T$, while $f_0(t,.,.)$ vanishes for $t < T$. However, below we write $x_T + v_T$ and $J(t, x_t, v_t)$ instead of $x_T + y_T$ and $J(t, x_t, y_t, v_t)$. Note that, by necessity, $X_t \geq 0$ for all t.

Now, $J(T, x_T, v_T) = x_T + v_T$. Let us next find $J(T-1, x_{T-1}, v_{T-1})$.
For $v_{T-1} = 1$, $J(T-1, x_{T-1}, v_{T-1}) =$

$$\begin{aligned} &\max_u E\{ux_{T-1}V_T + (1-u)x_{T-1}(1-V_T) + V_T | v_{T-1} = 1\} \\ &\quad = \max_u\{(3/4)ux_{T-1} + 3/4 + (1/4)(1-u)x_{T-1}\} = (3/4)x_{T-1} + 3/4. \end{aligned}$$

Here, $u = u_{T-1} = 1$ is optimal.

For $v_{T-1} = 0$, $J(T-1, x_{T-1}, v_{T-1}) =$

$$\begin{aligned} &\max_u E\{ux_{T-1}V_T + (1-u)x_{T-1}(1-V_T) + V_T | v_{T-1} = 0\} \\ &\quad = \max_u\{(1/4)ux_{T-1} + 1/4 + (3/4)(1-u)x_{T-1}\} = (3/4)x_{T-1} + 1/4. \end{aligned}$$

Here, $u = u_{T-1} = 0$ is optimal.

Let us now find $J(T-2, x_{T-2}, v_{T-2})$. We can write $J(T-1, x_{T-1}, v_{T-1}) = (3/4)x_{T-1} + 3v_{T-1}/4 + (1-v_{T-1})/4$.

For $v_{T-2} = 1$, $J(T-2, x_{T-2}, v_{T-2}) =$

$$\begin{aligned} &\max_u\{E[(3/4)(ux_{T-2}V_{T-1} + (1-u)x_{T-2}(1-V_{T-1})) + 3V_{T-1}/4 \\ &\qquad + (1-V_{T-1})/4 | v_{T-2}]\} \\ &\quad = \max_u\{(3/4)[(3/4)ux_{T-2} + 3/4] + (1/4)[(3/4)(1-u)x_{T-2} + 1/4]\} \end{aligned}$$

$$= (3/4)^2 x_{T-2} + (3/4)^2 + (1/4)^2$$
$$= (3/4)^2 x_{T-2} + 10/16.$$

Here, $u = u_{T-2} = 1$ is optimal.

For $v_{T-2} = 0$, $J(T-2, x_{T-2}, v_{T-2}) =$

$$\max_u \{E[(3/4)(ux_{T-2}V_{T-1} + (1-u)x_{T-2}(1-V_{T-1})) + 3V_{T-1}/4$$
$$+(1-V_{T-1})/4 | v_{T-2}]\}$$
$$= \max_u \{(1/4)[(3/4)ux_{T-2} + 3/4] + (3/4)[(3/4)(1-u)x_{T-2} + 1/4]\}$$
$$= (3/4)^2 x_{T-2} + 6/16.$$

Here, $u = u_{T-2} = 0$ is optimal.

We now guess that $J(t, x_t, v_t)$ is of the form $J(t, x_t, v_t) = (3/4)^{T-t} x_t + a_t$ when $v_t = 1, J(t, x_t, v_t) = (3/4)^{T-t} x_t + b_t$ when $v_t = 0$. We can write $J(t, x_t, v_t) = (3/4)^{T-t} x_t + a_t v_t + b_t(1-v_t)$. Then, $J(t-1, x_{t-1}, v_{t-1}) =$

$$\max_u E\{(3/4)^{T-t}(ux_{t-1}V_t + (1-u)x_{t-1}(1-V_t)) + a_t V_t + b_t(1-V_t) | v_{t-1}\}.$$

For $v_{t-1} = 1$ this expression equals

$$\max_u \{(3/4)[(3/4)^{T-t} ux_{t-1} + a_t] + (1/4)[3/4)^{T-t}(1-u)x_{t-1} + b_t]\}$$
$$= (3/4)^{T-(t-1)} x_{t-1} + (3/4)a_t + (1/4)b_t,$$

with $u = u_{t-1} = 1$ optimal, and for $v_{t-1} = 0$, we get $J(t-1, x_{t-1}, v_{t-1}) =$

$$\max_u \{(1/4)[(3/4)^{T-t} ux_{t-1} + a_t] + (3/4)[3/4)^{T-t}(1-u)x_{t-1} + b_t]\}$$
$$= (3/4)^{T-(t-1)} x_{t-1} + (1/4)a_t + (3/4)b_t,$$

with $u = u_{t-1} = 0$ optimal.

Note that for all t, the optimal u_t equals v_t.

The entities a_t and b_t are governed by the backwards difference equations $a_{t-1} = (3/4)a_t + (1/4)b_t$, $b_{t-1} = (1/4)a_t + (3/4)b_t$, $a_T = 1$, $b_T = 0$, and so are known. In fact, it is easy to find a formula for them. Adding the right-hand side of the equations, we see that $a_{t-1} + b_{t-1} = a_t + b_t$, so using $a_T + b_T = 1$ yields $a_t + b_t \equiv 1$. So $a_{t-1} = (1/2)a_t + 1/4$, which has the solution $a_t = (1/2)^{T-(t-1)} + 1/2$, while $b_t = 1 - a_t = 1/2 - (1/2)^{T-(t-1)}$. □

Remark 1.4 (State- and time-dependent control region).* The theory above holds also if the control region depends on t, x in the manner that $U = U(t,x) = \{u : h_i(t,x,u) \geq 0, i = 1, \ldots, i^*\}$, for some given functions h_i's that are continuous in (x,u). If $U(t,x)$ is empty, then, by convention, the maximum over $U(t,x)$ is set equal

to $-\infty$. Hence, now $u_t(x_t, v_t)$ has to take values in $U(t,x_t)$, and the maximization in (1.5), respectively (1.6), is carried out over $U(t-1,x_{t-1})$, respectively, $U(T,x_T)$ $\square$

An additional comment is perhaps needed to make quite clear what the problem now is: A maximum of the criterion is sought in the set of all pairs of sequences $\{X_s\}_s$, $\{u_s(x,v)\}_s$ that satisfy the state equation and the condition $u_s(X_s,V_s) \in U(s,X_s)$ a.s. for $s = 0,\dots,T$. If the set of such pairs is empty, the problem has no solution.

Example 1.5. Let us solve the problem in Example 1.1:

$$\max E\left[\sum_{t=0}^{T-1} \beta^t u_t^{1-\gamma} + \beta^T B X_T^{1-\gamma}\right], \tag{i}$$

$$X_{t+1} = Z_{t+1}(X_t - u_t), \qquad u_t \in (0,x_t), \tag{ii}$$

where $0<\gamma<1$, $0<\beta<1$, $B>0$, and Z_t, $t=0,1,\dots$ are independently distributed non-negative random variables, $EZ_t^{1-\gamma} < \infty$.

Solution. Here $J(T,x_T) = \beta^T B x_T^{1-\gamma}$. To find $J(T-1,x_{T-1})$, we use the optimality equation

$$J(T-1,x_{T-1}) = \max_u \left(\beta^{T-1}u^{1-\gamma} + E\left[\beta^T B(Z_T(x_{T-1}-u))^{1-\gamma}\right]\right). \tag{iii}$$

The expectation must be calculated by using the probability distribution for Z_T. Now the expectation in (iii) is equal to

$$\beta^T B D_T (x_{T-1}-u)^{1-\gamma}, \qquad D_t = E\left[Z_t^{1-\gamma}\right]. \tag{iv}$$

Hence, the expression to be maximized in (iii) is $\beta^{T-1}u^{1-\gamma} + \beta^T B D_T (x_{T-1}-u)^{1-\gamma}$. If we put $u = wx$, $w \in (0,1)$, and let $\varphi(w) := w^{1-\gamma} + h(1-w)^{1-\gamma}$, where $h = \beta B D_T$, then $J(T-1,x_{T-1}) = \beta^{T-1}x^{1-\gamma}\max_w \varphi(w)$, and we see that we need to solve the maximization problem

$$\max_{w\in(0,1)} \varphi(w) = \max_{w\in(0,1)} \left[w^{1-\gamma} + h(1-w)^{1-\gamma}\right]. \tag{v}$$

We find the maximum of the concave function φ, by solving

$$\varphi'(w) = (1-\gamma)w^{-\gamma} - (1-\gamma)h(1-w)^{-\gamma} = 0,$$

which yields $w^{-\gamma} = h(1-w)^{-\gamma}$. Solving for w yields

$$w = \frac{1}{1+h^{1/\gamma}}. \tag{vi}$$

Inserting this in φ gives its maximal value

$$\max_w \varphi(w) = 1/(1+h^{1/\gamma})^{1-\gamma} + h[h^{1/\gamma}/(1+h^{1/\gamma})]^{1-\gamma} = (1+h^{1/\gamma})^{\gamma}. \tag{vii}$$

Define $C_T := B, C_{T-1}^{1/\gamma} := 1 + (\beta B D_T)^{1/\gamma} = 1 + h^{1/\gamma}$, and generally,

$$C_t^{1/\gamma} := 1 + (\beta C_{t+1} D_{t+1})^{1/\gamma}. \tag{viii}$$

Then, the optimal $u_{T-1} = w x_{T-1} = x_{T-1}/C_{T-1}^{1/\gamma}$ and $J(T-1, x_{T-1}) = \beta^{T-1} x_{T-1}^{1-\gamma} \max \varphi(w) = \beta^{T-1} C_{T-1} x_{T-1}^{1-\gamma}$. As $J(T-1, x_{T-1})$ has the same form as $J(T, x_T)$, then, to find the optimal u_{T-2} and $J(T-2, x_{T-2})$, (vi) and (vii) are used for $h = \beta C_{T-1} D_{T-1}$. This yields $u_{T-2} = x_{T-2}/C_{T-2}^{1/\gamma}$ and $J(T-2, x_{T-2}) = \beta^{T-2} C_{T-2} x_{T-2}^{1-\gamma}$. This continues backwards, so evidently we obtain generally $u_t = x_t/C_t^{1/\gamma} \in (0, x_t)$, $(C_t^{1/\gamma} > 1)$, and $J(t, x_t) = \beta^t C_t x_t^{1-\gamma}$.

Note that C_t is a known sequence; it is determined by $C_T = B$ and backwards recursion, using (viii). □

1.2 Infinite Horizon

Suppose that $P_t(v_{t+1}|v_t)$ and f are independent of t, and that f_0 can be written $f_0(t, x, u) = g(x, u)\alpha^t$, $\alpha \in (0, 1]$ (f and g continuous). The problem is often called stationary, or autonomous, if these properties hold. Put $\pi = ((u_0(x_0, v_0), u_1(x_1, v_1) \ldots)$. The problem is now

$$\max_{\pi} E\left[\sum_{t=0}^{\infty} \alpha^t g(X_t, u_t(X_t, V_t))\right],\ u_t(X_t, V_t) \in U, \quad \Pr[V_{t+1} = v | v_t] = P(v|v_t), \tag{1.7}$$

where X_t is governed by the stochastic difference equation

$$X_{t+1} = f(X_t, u_t(X_t, V_t), V_{t+1}), \tag{1.8}$$

with X_0, V_0 given, $(g, f, P(v|v_t)$ given entities, the control u_t subject to choice in U, U given). Again, V_{t+1} can also be allowed to be a continuous random variable, governed by a density $p(v|v_t)$, separately piecewise continuous in v and v_t, but independent of t. The maximum in (1.7) is sought when considering all sequences $\pi := (u_0(x_0, v_0), u_1(x_1, v_1), \ldots)$ and selecting the best. We base our discussion upon condition (1.11) below, or P., or N. in Remark 1.6 below, implying that the infinite sum in (1.7) always exists in $[-\infty, \infty]$, for condition (1.11), the sum belongs to $(-\infty, \infty)$. For a given sequence $\pi := (u_0(x_0, v_0), u_1(x_1, v_1), \ldots)$, let us write

$$J_\pi(s, x_s, v_s) = E\left[\sum_{t=s}^{\infty} \alpha^t g(X_t, u_t(X_t, V_t)) \,\middle|\, x_s, v_s\right], \tag{1.9}$$

where we now start the difference (state) equation at $X_s = x_s$. Let $J(s, x_s, v_s) = \sup_\pi J_\pi(s, x_s, v_s)$. We now prove that $J(1, x_0, v_0) = \alpha J(0, x_0, v_0)$. The intuitive argument is as follows. Let $J_\pi^k(x, v) = \sum_{t=k}^{\infty} E[\alpha^{t-k} g(X_t, u_t(X_t, V_t))|x, v]$ and let $J^k(x, v) =$

$\sup_\pi J_\pi^k(x,v)$. Then $J^k(x,v)$ is the maximal expected present value of future rewards discounted back to $t=k$, given that the process starts at (x,v) at time $t=k$. When starting at (x,v) at time $t=0$, and discounting back to $t=0$, the corresponding maximal expected value is $J^0(x,v)=J(0,x,v)$. Because time does not enter explicitly in $P(v|v_t), g$ and f, the future looks exactly the same at times $t=0$, and $t=k$, hence $J^k(x,v)=J(0,x,v)$. As $J_\pi(k,x_0,v_0)=\alpha^k J_\pi^k(x_0,v_0)$ (in the definition of $J_\pi(k,x_0,v_0)$ we discount back to $t=0$) and hence $J(k,x_0,v_0)=\alpha^k J^k(x_0,v_0)=\alpha^k J(0,x_0,v_0)$ and in particular $J(1,x_0,v_0)=\alpha J(0,x_0,v_0)$.

The heuristic argument for the optimality equation can be repeated in the infinite horizon case. So (1.5) still holds. Using (1.5) for $t=1$, and then inserting $\alpha J(0,x,v))=J(1,x,v)$ and writing $J(x,v)=J(0,x,v), x=x_0, v=v_0$, gives the following *optimality equation*, or *equilibrium optimality equation* or *Bellman equation*

$$J(x,v)=\max_u\big\{g(x,u)+\alpha E[J(X_1,V_1)\mid x,v]\big\}, \tag{1.10}$$

where $X_1=f(x,u,V_1)$.

Observe that (1.10) is a "functional equation," an equation (hopefully) determining the unknown function J (J occurs on both sides of the equality sign). Once J is known, the optimal Markov control is obtained from the maximization in the optimality equation. Evidently, the maximization yields a control function $u(x,v)$, not dependent on t, and this is what we should expect of an optimal control function: If we have observed x,v at time 0 and time t, the optimal choice of control should be the same in the two situations, because then the future looks exactly the same at these two points in time.

It can be shown that the optimal value function $J(x,v)=\sup_\pi J_\pi(0,x,v)$ is defined and satisfies the equilibrium optimality equation in three cases to be discussed below, (1.11), as well as P. and N. in Remark 1.6. (At least this is so when "max" is replaced by "sup" in the equation.) Let us first consider the following case.

$$M_1\le g(x,u)\le M_2 \text{ for all } (x,u)\in\mathbb{R}^n\times U, \tag{1.11}$$

where M_1 and M_2 are given numbers. In case of the boundedness condition (1.11), it is known that the equilibrium optimality equation has a unique bounded solution $J(x,v)$ (when "max" is replaced by "sup," if necessary). Furthermore, $J(x,v)$ is automatically the optimal value function in the problem, and a control $u(x,v)$ giving maximum in the optimality equation, given $J(x,v)$, is the optimal control.

Remark 1.6 (Alternative boundedness conditions).* Complications arise when the boundedness condition (1.11) fails to hold. Then we cannot know for sure that the optimal value function is bounded, so we may have to look for unbounded solutions of the Bellman equation. But then false solutions can occur (bounded or not bounded), not equal to the optimal value function. Even in the case where (1.11) holds, allowing unbounded solutions may lead to nonunique solutions, even a plethora of solutions (see Exercise 1.48 and the following even simpler problem, where $g=0$, $f=x/\alpha$, and where $J(x)=ax$ satisfies the Bellman equation for all a).

We shall consider two cases, P. and N., where some results can be obtained. In both cases, we must allow for infinite values for the optimal value function $J(x,v)$, $+\infty$ in case P, and $-\infty$ in case N.

P. *Either* $g(x,u) \geq 0$ *for all* $(x,u) \in \mathbb{R}^n \times U$, *and* $\alpha = 1$, *or for some negative number* γ, $g(x,u) \geq \gamma$ *for all* $(x,u) \in \mathbb{R}^n \times U$ *and* $\alpha \in (0,1)$.

Let $J^u(x,v)$ be the value function arising from using $u(.,.)$ all the time. In the current case, if $u(x,v)$ yields the maximum in the Bellman equation when $J^u(x,v)$ is inserted, then $u(.,.)$ is optimal. (In other words, if we have been able to find a control $u(x,v)$ such that the pair $(u(x,v), J^u(x,v))$ behaves in this way, then $u(x.,.)$ is optimal.)

Most often, it can be imagined that first the Bellman equation were solved and a pair $(u(x,v), \hat{J}(x,v))$ satisfying it were found (in particular, then, $u(x,v)$ yields the maximum in the equation). Next, if we are lucky enough to be able to prove that $J^u(x,v) = \hat{J}(x,v)$, then all is well.

Sometimes it is useful to know the fact that if $J^u(s,x,v,T)$ is the value function arising from using $u = u(x,v)$ all the time from s until $t = T$ when starting at (s,x,v), then $J^u(0,x,v,T) \to J^u(x,v)$ as $T \to \infty$. Also $J(0,x,v,T) \to J(x,v)$ as $T \to \infty$, where $J(s,x,v,T)$ is the optimal value function in the problem with finite horizon T, and where we start at (s,x,v).

Note that, in the current case, the optimal value function J is $\leq \hat{J}$ for any other solution $\hat{J}$ of the Bellman equation for which $\hat{J} \geq \hat{M}(1-\alpha)$ for some $\hat{M} \leq 0$.

N. *Either* $g(x,u) \leq 0$ *for all* $(x,u) \in \mathbb{R}^n \times U$, *and* $\alpha = 1$, *or for some positive number* γ, $g(x,u) \leq \gamma$ *for all* $(x,u) \in \mathbb{R}^n \times U$ *and* $\alpha \in (0,1)$.

In this case, it is known that if $u(.,.)$ satisfies the Bellman equation with the optimal value function inserted, then $u(.,.)$ is optimal. It is also known that if U is compact, then $J(0,x,v,T) \to J(x,v)$ as $T \to \infty$. (In fact, for this result, we now do need that f and g are continuous and $\mathcal{V}$ is finite. For other assumptions on $\mathcal{V}$, see Section 1.6 below.)

How can this information be used? Assume that we have found a function $\hat{J}(x,v)$ satisfying $\hat{J} \leq \hat{M}(1-\alpha)$ for some $\hat{M} \geq 0$, together with a function $u(x,v)$ satisfying the Bellman equation. If we are able to prove that the Bellman equation has only one such solution $\hat{J}(x,v)$, then this is the optimal value function $J(x,v)$ (because $J(x,v)$ is known to satisfy the Bellman equation, both in case P. and N.), and then $u(x,v)$ is optimal. Another possibility is the following: Suppose that U is compact and that we can apply the limit result $\lim_{T\to\infty} J(0,x,v,T) = J(x,v)$ mentioned above. If we then find that $\lim\limits_{T\to\infty} J(0,x,v,T) = \hat{J}(x,v)$, then $\hat{J}(x,v)$ is the optimal value function and $u(x)$ is optimal.

Note that in case N., the optimal value function J is $\geq \hat{J}$ for any other solution $\hat{J}$ of the Bellman equation for which $\hat{J} \leq \hat{M}(1-\alpha)$ for some $\hat{M} \geq 0$. (This fact lies behind the uniqueness argument in the last paragraph.) □

Remark 1.7 (Modified boundedness conditions).* The boundedness condition (1.11) and the conditions in P. and N. need only hold for x in $\mathcal{X}(x_0) := \cup_s \mathcal{X}_s(x_0)$, where

$\mathcal{X}_s(x_0)$ is the set of states that can be reached at time s, when starting at x_0 at time 0, considering all outcomes and all controls.

The conclusions drawn in the case where (1.11) is satisfied also hold if the following alternative condition holds: There exist positive constants M, M^*, β, and δ such that for all $x \in \mathcal{X}(x_0)$, all $u \in U$ and all V, $|f(x,u,V)| \le M + \delta|x|$ and $|g(x,u)| \le M^*(1+|x|^\beta)$, with $\alpha\delta^\beta < 1$, and $\alpha \in (0,1)$. Moreover, the conclusions in case P. (respectively, case N.) hold if the next to last inequality is replaced by $g(x,u) \ge -M^*(1+|x|^\beta)$ (respectively, $g(x,u) \le M^*(1+|x|^\beta)$). Note that J needs to be defined only for x in $\mathcal{X}(x_0)$, this set having the property that if x belongs to the set, also $f(x,u,v)$ belongs to it. □

Example 1.8. Consider the problem

$$\max_{u_t \in (0,1)} E\left[\sum_{t=0}^{\infty} \beta^t x_t^{1-\gamma} u_t^{1-\gamma}\right] \tag{i}$$

$$x_{t+1} = V_{t+1}(1-u_t)x_t, \quad x_0 \text{ is a positive constant.} \tag{ii}$$

Here, $V_1, V_2, \ldots$ are identically and independently distributed non-negative stochastic variables, with $D = EV^{1-\gamma} < \infty$, where V is any of the V_t's. We may think of x_t as the assets of, say, some timeless institution. At each point in time an amount $u_t x_t$ is spent on some useful purpose, and the total effect is measured by the expectation in (i). (For a comment on (ii), see Example 1.1.) It is assumed that

$$\rho = (\beta D)^{1/\gamma} < 1, \quad \beta \in (0,1), \quad \gamma \in (0,1) \tag{iii}$$

Solution. In the notation of problem (1.7), (1.8), $g(x,u) = x^{1-\gamma}u^{1-\gamma}$ and $f(x,u,V) = V(1-u)x$. The equilibrium optimality equation (1.10) yields

$$J(x) = \max_{u \in (0,1)} \left[x^{1-\gamma}u^{1-\gamma} + \beta E J(V(1-u)x)\right] \tag{iv}$$

We guess that $J(x)$ had the form $J(x) = kx^{1-\gamma}$ for some constant k (the optimal value function had a similar form in the finite horizon version of this problem discussed in the previous section). Then, canceling the factor $x^{1-\gamma}$, (iv) reduces to

$$k = \max_{u \in (0,1)} \left[u^{1-\gamma} + \beta k D (1-u)^{1-\gamma}\right], \tag{v}$$

where $D = EV^{1-\gamma}$. Using the result from Example 1.5 (the maximization of φ) gives that the maximum in (v) is obtained for $u =$

$$u_* = \frac{1}{1+\rho k^{1/\gamma}}, \quad \rho = (\beta D)^{1/\gamma} \tag{vi}$$

and the maximum value in (v) equals $(1+\rho k^{1/\gamma})^\gamma$, so k is determined by the equation

$$k = (1+\rho k^{1/\gamma})^\gamma.$$

Raise each side to the power $1/\gamma$, and solve for $k^{1/\gamma}$ to obtain $k^{1/\gamma} = 1/(1-\rho)$, or $k = (1-\rho)^{-\gamma}$. Hence, the solution is $J(x) = (1-\rho)^{-\gamma}x^{1-\gamma}$, with $u = 1-\rho$.

In this example, the boundedness condition (1.11) is not satisfied for $x \in \mathcal{X}(x_0)$. One method out is to use the transformation $y_t = x_t/z_t$, $z_{t+1} = V_{t+1}z_t, z_0 = 1$, which gives that $y_{t+1} = (1-u)y_t, y_0 = x_0$. Replacing x_t by $y_t z_t$, as $Z_t = V_1 \cdot \ldots \cdot V_t$, taking the expectation inside the sum in the criterion (using actually what is called the monotone convergence theorem), the problem can be transformed into a deterministic one. The deterministic difference equation $y_{t+1} = (1-u)y_t, y_0 = x_0$ is the state equation, we have a new discount factor $\hat{\beta} = \beta EV^{1-\gamma}$ and a new g-function equal to $y^{1-\gamma}u^{1-\gamma} \in [0, x_0^{1-\gamma}]$ for all $y \in \mathcal{X}(y_0) \subset [0, x_0]$. In this problem, the modified boundedness condition in Remark 1.7 is satisfied. Another way out is the following: Let us use P. in Remark 1.6: Then we need to know that $J^{u_*}(x) = J(x)$. It is fairly easy to carry out the explicit calculation of $J^{u_*}(x)$, by taking the expectation inside the sum and summing the arising geometric series. But we don't need to do that. Noting that $x_t = x_0\rho^t V_1 \cdot \ldots \cdot V_t$, evidently, we must have that $J^{u_*}(x_0) = kx_0^{1-\gamma}$, for some k. We must also have that $J^{u_*}(x_0)$ satisfies the equilibrium optimality equation with $u = u_*$ and the maximization deleted, (in the problem where $U = u_*$, u_* is optimal!). But the only value of k for which this equation is satisfied we found above. Thus the test in P. works and u_* as specified in (vi) is optimal. □

1.3 State and Control-Dependent Probabilities

Suppose that the state equation is still of the form

$$X_{t+1} = f(t, X_t, u_t, V_{t+1}), \quad x_0, v_0 \text{ are given} \tag{1.12}$$

where V_{t+1} takes values in a finite set $\mathcal{V} = \{\bar{v}_0, \ldots, \bar{v}_m\}$, whose elements have probabilities $Pr[V_{t+1} = \bar{v}_0] = P^{(0)}(t, x_t, u_t, v_t), \ldots, Pr[V_{t+1} = \bar{v}_m] = P^{(m)}(t, x_t, u_t, v_t)$, respectively, hence these probabilities are conditional ones, also written $P_t(v|x_t, u_t, v_t)$, $v \in \mathcal{V}$. Thus, the probability $\Pr[V_{t+1} = v] = P_t(v|x_t, u_t, v_t)$ of the event $V_{t+1} = v$, is supposed to depend on the time t, the outcome v_t, the state x_t, and the control u_t we select at time t. We may allow $\mathcal{V}$ instead to be all $\mathbb{R}^{\hat{n}}$, for some $\hat{n}$, thus allowing the V_t's to be continuous stochastic variables. Then the distribution of V_{t+1} is often given by a density $p_t(v|x_t, u_t, v_t)$, separately piecewise continuous in each component of v, x_t, u_t and v_t. In the main theoretical discussions, we mostly stick to discrete random variables. However, the solution tools presented can also be used for continuous stochastic variables. Again it is assumed that x_t belongs to $\mathbb{R}^n$, that u_t belongs to a given subset U of $\mathbb{R}^r$, and that $t = 0, \ldots, T$.

Example 1.9. A machine is supposed to be in one of three states. Either "as good as new," denoted by (2), or "functioning" (1), or "broken" (0). After having been used all day, the machine is checked in the evening and its state is determined. The following table describes the "transition probabilities" of the state from one evening to the next one (it is an example of a so-called Markov process).

(1)

The state when the machine is checked	The state next evening: 0	1	2
0	1	0	0
1	0.4	0.6	0
2	0.2	0.4	0.4

The table should be understood as follows. The first column lists the three possible states of the machine, when it is checked in the evening. The uppermost row shows the possible states of the machine after it has run for one day. If the machine is, say, "as good as new," i.e., in state 2, then the last row says that upon checking the machine the next evening there is a probability of 0.2 of finding that it is "broken" (state 0), a probability of 0.4 of finding that it is functioning (state 1), and a probability of 0.4 of finding that it is as good as new (state 2). The other two rows below the bar are read similarly. If we use the symbols above, we let $\bar{v}_0 = 0$, $\bar{v}_1 = 1$, $\bar{v}_2 = 2$, $X_{t+1} = f(t, x_t, u_t, V_{t+1}) = V_{t+1}$, where $x_{t+1} \in \{0,1,2\}$ and $v_t \in \{0,1,2\}$. Moreover, $u_t \in \{0,1\}$, $u_t = 0$ means that we do not repair the machine after the evening check, whereas $u_t = 1$ means that we repair it. The above table describes the situation the next evening if we do not repair the machine. The elements in the matrix in (1) are hence the probabilities $P^{(i)}(t, x_t, 0) = P_t^{(i)}(t, x_t, u_t = 0)$, $i = 0$, 1, 2, $x_t = 0$, 1, 2, where i gives the column number and x_t the row number.

If the machine is repaired one evening, then it is simply assumed that it is as good as new (in state 2) the next evening. Thus, $P_t^{(2)}(t, x_t, 1) = 1$ and $P_t^{(i)}(t, x_t, 1) = 0$ for $i = 0$, 1, regardless of x_t. □

Let us return to the general problem. The process determined by (1.12) and the random events V_1, V_2, ..., is to be controlled in the best possible manner by appropriate choices of the variables u_t. The criterion to be maximized is the expectation

$$E\left[\sum_{t=0}^{T} f_0(t, X_t, u_t(X_t, V_t), V_t)\right]. \tag{1.13}$$

Again, each control u_t, $t = 0, 1, 2, \ldots, T$ should be a function, $u_t(x_t, v_t)$, $t = 0$, ..., T, of the current state x_t and the current outcome v_t. To compute the expectation in (1.13), i.e., to calculate $E[f_0(t, X_t, u_t(X_t, V_t), V_t)]$ for any given t, requires specifying the probabilities that lie behind the calculation of this expectation. Let us consider the case where $\mathcal{V}$ is discrete. Given that the policies $u_0(x_0, x_0)$, ..., $u_T(x_T, v_T)$ are used, note first that $X_s = X_s(V_1, \ldots, V_s)$ in other words, X_s depends on the outcomes of $V_1, \ldots, V_s$. The probability of the joint event $V_1 = v_1, V_2 = v_2, \ldots, V_t = v_t$, is given by $p^*(v_1, \ldots, v_t) =$

$$P_0(v_1|x_0, u_0, v_0) \cdot P_1(v_2|x_1, u_1, v_1) \cdot \ldots \cdot P_{t-1}(v_t|x_{t-1}, u_{t-1}, v_{t-1}) \tag{1.14}$$

where $u_0 = u_0(x_0, v_0)$, $u_1 = u_1(x_1, v_1)$,..., $u_{T-1} = u_{T-1}(x_{T-1}, v_{T-1})$ and where each $x_s = x_s(v_1, \ldots, v_s)$ (the x_s's forming a solution sequence of the state equation for the specified control sequence), so the expression in (1.14) is a function (only) of

$(v_1,\dots,v_t)$. Similarly, when inserting $X_t = X_t(V_1,..,V_t)$ in $f_0(t,X_t,u_t(X_t,V_t),V_t)$, this function becomes a function only of $(V_1,\dots,V_t)$ and the probabilities for the various outcomes $(v_1,\dots,v_t)$ we have already specified, so $E[f_0(t,X_t,u_t(X_t,V_t),V_t)]$ can be calculated. Thus, the expression in (1.13) is equal to

$$\left(\sum_{t=0}^{T}\sum_{v_1,\dots,v_t} f_0(t,x_t,u_t(x_t,v_t),v_t)\right) p^*(v_1,\dots,v_t), \tag{1.15}$$

where the inner sum is taken over all combinations of values $(v_1,\dots,v_t)$. The probabilities $p^*(v_1,\dots,v_t)$, and hence the expected value, depend on the policies chosen, so sometimes we write $E_{u_0,\dots,u_T}$ instead of E in (1.13).

Though not always necessary, we shall assume that f_0 and f are continuous in (x,u), even in (x,u,v) if $\mathcal{V}$ is nondiscrete.

The optimization problem is to find a sequence of policies $u_0^*(x_0,v_0),\dots,$ $u_T^*(x_T,v_T)$, which gives the expression in (1.13) the largest possible value, subject to the difference equation (1.12).

We now define

$$J(t,x_t,v_t) = \sup E_{u_t,\dots,u_T}\left[\sum_{s=t}^{T} f_0(s,X_s,u_s(X_s,V_s),V_s) \;\middle|\; x_t,v_t\right], \tag{1.16}$$

where the supremum is taken over all policy sequences $u_s = u_s(x_s,v_s), s = t,\dots,T$, given v_t and given that we start at the state x_t at time t, as indicated by "$|\, x_t,v_t$." The computation of the expectation is now based on conditional probabilities of the form

$$P_t(v_{t+1}|x_t,u_t(x_t,v_t),v_t)\cdot\ldots\cdot P_{T-1}(v_T|x_{T-1},u_{T-1}(x_{T-1},v_{T-1}),v_{T-1}).$$

In (1.16), and in these probabilities, given $u_t(.,.),\dots,u_{T-1}(.,.)$, for $s = t+1,\dots,T$, x_s is a function of $(v_{t+1},\dots,v_s)$ (and the given v_t), as well as of the given start value x_t, again determined by the difference equation (1.12).

(We seek a maximum in (1.16), and in a similar definition in Section 1.1, we wrote max and not sup. When we write max we indirectly say that a maximum exists, and being a little more formal in this section, we don't want to include such an assumption in the definition. A similar remark pertains to the optimality equation (1.17), (1.18) below.)

Again, the central tool in solving optimization problems of the type (1.12)–(1.13) is the following *optimality* equation (we write also here sup instead of max). For $t < T$

$$\begin{aligned} J(t-1,x_{t-1},v_{t-1}) = \sup_{u_{t-1}}\Big\{ & f_0(t-1,x_{t-1},u_{t-1},v_{t-1}) \\ & + \sum_{v_t\in\mathcal{V}} P_{t-1}(v|x_{t-1},u_{t-1},v_{t-1}) J(t,f(t-1,x_{t-1},u_{t-1},v_t),v_t)\Big\}. \end{aligned} \tag{1.17}$$

Of course also here, if possible, we want to maximize, and in the maximization, the vector u_{t-1} is constrained to lie in U. The equation can be written more concisely as

$$J(t-1,x_{t-1},v_{t-1}) = \sup_{u_{t-1}} \big\{ f_0(t-1,x_{t-1},u_{t-1},v_{t-1}) + E_{u_{t-1}}\left[J(t,X_t,V_t) \mid x_{t-1},v_{t-1}\right] \big\}. \tag{1.18}$$

Of course, this version is also valid for continuous stochastic variables. Moreover, when $t = T$, we must have

$$J(T,x_T,v_T) = \sup_{u_T} f_0(T,x_T,u_T,v_T). \tag{1.19}$$

The intuitive argument for (1.17) is exactly as before: Suppose the system is in state x_{t-1}. For a given u_{t-1}, the "instantaneous" reward is equal to $f_0(t,x_{t-1},u_{t-1},v_{t-1})$. In addition, the sum of rewards at all later times is at most $J(t,x_t)$ if $x_t = f(t-1,x_{t-1},u_{t-1},v)$, and the probability of this event is $P_{t-1}(v|x_{t-1},u_{t-1},v_{t-1})$. When using u_{t-1}, the total expected maximum value gained over all future time points (now including even $t-1$) is the sum in (1.17). The largest expected gain comes from choosing u_{t-1} to maximize this sum.

A formal proof is presented later on. In connection with the proof, certain theoretical questions are discussed. In particular, it can be shown that the maximal value of the criterion cannot be increased by allowing policies that depend on past states as well as on the present state.

Remark 1.10 (Criterion to be minimized). Suppose that we want to minimize the value of the criterion. Then, to obtain the optimal value functions $J(t,x_t,v_t)$ a minimization is carried out instead of a maximization. In the optimality equation, "max" (or "sup") must then be replaced by "min" ("inf").

To see that this is correct, recall that to minimize a criterion is the same as maximizing (-1) times the criterion. Thus, we can apply the above "maximization theory" to a problem where f_0 is replaced by $-f_0$. From this it is easy to see that the "min"-version of the optimality equation follows. □

Example 1.11. Consider Example 1.9 again. In this example, the values of f_0 will be costs, rather than rewards. Let the values of the function f_0 for all t be given by the table

(2)

		u	
		0	1
	0	2	5
x	1	0	1
	2	0	1/2

From the table, we see for instance that $f_0(t,x_t,u_t) = f_0(x_t,u_t) = 5$ when $x_t = 0$, $u_t = 1$. The costs in the table may be interpreted as follows: A broken machine leads to lost sales. But, if it is repaired, then that will add to the costs (see the numbers

2 and 5 in the table). Repair carried out on a machine in better shape costs less, as indicated by the last column. We are going to use the machine in a production run over a period of three days. Before we start (i.e., at time $t = 0$), the machine is in state 1. Because we are going to minimize costs, we replace sup with inf, (or min) in (1.18) and (1.19), see Remark 1.10. For $J(3, X_3)$ we get:

$$\begin{array}{ccc} J(3,0) = 2 & J(3,1) = 0 & J(3,2) = 0 \\ u_3^* = 0 & u_3^* = 0 & u_3^* = 0 \end{array} \tag{3}$$

We naturally choose $u_3^* = 0$ because we shall not produce anything the next day. Let us compute $J(2, x_2)$ for $x_2 = 0$, 1, 2.

First let $x_2 = 0$. If $u = 0$ is chosen, then the expected cost is $f_0(0,0) + 1 \cdot J(3,0) + 0 \cdot J(3,1) + 0 \cdot J(3,2) = 2 + 1 \cdot 2 + 0 \cdot 0 + 0 \cdot 0 = 4$, where the factors 1, 0, 0 make up the first row in the matrix (1) in Example 1.9. If $u = 1$ is chosen, the expected cost is $f_0(0,1) + 1 \cdot J(3,2) = 5 + 1 \cdot 0 = 5$. (Recall that a newly repaired machine is still as good as new $(x = 2)$ after one day's use.) The minimum of the numbers 4 and 5 is 4, attained by $u = 0$, so $J(2,0) = 4$.

Next, let $x_2 = 1$. If $u = 0$ is chosen, the expected cost is $f_0(1,0) + 0.4 \cdot J(3,0) + 0.6 \cdot J(3,1) + 0 \cdot J(3,2) = 0 + 0.4 \cdot 2 + 0.6 \cdot 0 + 0 \cdot 0 = 0.8$, where the factors 0.4, 0.6, and 0 make up the second row in table (1) in Example 1.9. If $u = 1$ is chosen, the expected cost is $f_0(1,1) + 1 \cdot J(3,2) = 1$. The minimum of the numbers 0.8 and 1 is 0.8, attained for $u = 0$.

Finally, put $x_2 = 2$. If $u = 0$ is chosen, we get $f_0(2,0) + 0.2 \cdot J(3,0) + 0.4 \cdot J(3,1) + 0.4 \cdot J(3,2) = 0 + 0.2 \cdot 2 + 0.4 \cdot 0 + 0.4 \cdot 0 = 0{,}4$. If $u = 1$ is chosen, we get $f_0(2,1) + 1 \cdot J(3,2) = 0{,}5 + 1 \cdot 0 = 0.5$. The minimum of the numbers 0.4 and 0.5 is 0.4, attained for $u = 0$. We summarize our calculations thus:

$$\begin{array}{ccc} J(2,0) = 4 & J(2,1) = 0.8 & J(2,2) = 0.4 \\ u_2^* = 0 & u_2^* = 0 & u_2^* = 0 \end{array} \tag{4}$$

Let us compute $J(1, x_1)$, $x_1 = 0$, 1, 2, in the same way.

Let $x_1 = 0$. If $u = 0$ is chosen, we get $f_0(0,0) + 1 \cdot J(2,0) + 0 \cdot J(2,1) + 0 \cdot J(2,2) = 2 + 1 \cdot 4 + 0 \cdot 0.8 + 0 \cdot 0.4 = 6$. If $u = 1$ is chosen, we get $f_0(0,1) + 1 \cdot J(2,2) = 5 + 1 \cdot 0.4 = 5.4$. The minimum, 5.4, is attained for $u = 1$.

Next, let $x_1 = 1$. If $u = 0$ is chosen, the expected cost is $f_0(1,0) + 0.4 \cdot J(2,0) + 0.6 \cdot J(2,1) + 0 \cdot J(2,2) = 0 + 0.4 \cdot 4 + 0.6 \cdot 0.8 + 0 \cdot 0.4 = 2.08$. If $u = 1$ is chosen, the expected cost is $f_0(1,1) + 1 \cdot J(2,2) = 1 + 0.4 = 1.4$. The minimum, 1.4, is attained for $u = 1$.

Finally, let $x_1 = 2$. If $u = 0$ is chosen, we obtain $f_0(2,0) + 0.2 \cdot J(2,0) + 0.4 \cdot J(2,1) + 0.4 \cdot J(2,2) = 0 + 0.2 \cdot 4 + 0.4 \cdot 0.8 + 0.4 \cdot 0.4 = 1.28$. If $u = 1$ is chosen, we get $f_0(2,1) + 1 \cdot J(2,2) = 0.5 + 1 \cdot 0.4 = 0.9$. The minimum, 0.9, is attained for $u = 1$.

This gives the following table:

$$\begin{array}{ccc} J(1,0) = 5.4 & J(1,1) = 1.4 & J(1,2) = 0.9 \\ u_1^* = 1 & u_1^* = 1 & u_1^* = 1 \end{array} \tag{5}$$

From (5), we now conclude that if two production days remain, we always repair the machine, whatever its state. If only one production day is left, then it is too expensive to repair the machine for such a short spell of time. □

In the next example, we go back to a very simple probability structure. (Recall that any minimization problem can be rewritten as a maximization problem by changing the sign of the criterion function, and in case of minimization, we get minimization also in the optimality equation.)

Example 1.12 (Linear quadratic multidimensional problem). Let H' be the transpose of the matrix H, and call a symmetric $n \times n$ matrix positive definite if, for all $x \in \mathbb{R}^n$, $x \neq 0$, $x'Hx > 0$, and positive semidefinite if $x'Hx \geq 0$. Consider the following problem with n state variables and r control variables:

$$\min_{u_0,\ldots,u_T} E\left[\sum_{0\leq t\leq T} x_t' R_t x_t + u_t' Q_t u_t\right], \tag{1.20}$$

where R_t and Q_t are given symmetric positive definite square matrices. The minimization is subject to the condition (equation)

$$x_{t+1} = A_t x_t + B_t u_t + \varepsilon_t, \quad u_t \in \mathbb{R}^r, \quad x_0 \text{ given in } \mathbb{R}^n, \tag{1.21}$$

where A_t and B_t are given $n \times n$ and $n \times r$ matrices, respectively, and where the random variables ε_t are independently distributed with mean zero and finite covariance matrices, their distributions being independent of history.

Solution. We will need the following result: Let Q be a symmetric and positive definite $r \times r$–matrix, let C be a symmetric and positive semidefinite $n \times n$–matrix, let A be a $n \times n$–matrix, and let B be an $n \times r$–matrix. The following equality is obtained by a completing-the-square argument presented below:

$$h(u) := u'Qu + (Ax + Bu)'C(Ax + Bu) = (u' + x'H')K(u + Hx) + x'Jx, \quad (*)$$

where $K = Q + B'CB$, $H = K^{-1}B'CA$, $J = A'CA - H'KH = A'CA - A'CB(Q + B'CB)^{-1}B'CA$ (K is symmetric and positive definite).
The equality (*) follows from $h(u) =$

$$\begin{aligned}
&u'Qu + u'B'CBu + x'A'CAx + x'A'CBu + u'B'CAx \\
&= u'Ku + x'A'CAx + x'A'CBK'^{-1}K'u + u'KK^{-1}B'CAx \\
&= u'Ku + u'KHx + x'H'Ku + x'H'KHx + x'Jx \\
&= (u' + x'H')K(u + Hx) + x'Jx.
\end{aligned}$$

The minimum point and minimal value of $h(u)$ are evidently given by

$$u = -Hx, \quad \min_u h(u) = x'Jx. \tag{**}$$

Define the symmetric, positive definite matrix C_t by the (backwards) Riccati equation

$$C_t = R_t + A_t' C_{t+1} A_t - (A_t' C_{t+1} B_t)(Q_t + B_t' C_{t+1} B_t)^{-1}(B_t' C_{t+1} A_t), \tag{1.22}$$

$C_{T+1} = 0$. As a backwards induction hypothesis, assume that $J(t,x)$ is of the form $x'C_t x + d_t$ for t replaced by $t+1$ and let us prove that then the formula is also correct for t (it *is* correct for $t = T$, for $C_T = R_T, d_T = 0$). Using the induction hypothesis, the optimality equation is:

$$J(t,x) = \min_u \{x' R_t x + u' Q_t u + E(A_t x + B_t u + \varepsilon_t)' C_{t+1}(A_t x + B_t u + \varepsilon_t)\} + d_{t+1}$$

Now,

$$\begin{aligned} E[(A_t x + B_t u + \varepsilon_t)' C_{t+1}(A_t x + B_t u + \varepsilon_t)] &= (A_t x + B_t u)' C_{t+1}(A_t x + B_t u) \\ &+ E[(A_t x + B_t u)' C_{t+1}\varepsilon_t + \varepsilon_t' C_{t+1}(A_t x + B_t u)] + E(\varepsilon_t' C_{t+1}\varepsilon_t), \end{aligned} \tag{1.23}$$

where the second term on the right-hand side vanishes. Only the first of the three terms is relevant to the minimization, because the third one is independent of u, so

$$J(t,x) = \min_u \{x' R_t x + u' Q_t u + (A_t x + B_t u)' C_{t+1}(A_t x + B_t u)\} + d_t,$$

where $d_t = E(\varepsilon_t' C_{t+1}\varepsilon_t) + d_{t+1}$. Using $(**)$, we have that the optimal control $u = u_t$ satisfies $u_t = -D_t x$, where $D_t = (Q_t + B_t' C_{t+1} B_t)^{-1} B_t' C_{t+1} A_t$. Moreover, using $(**)$ and (1.22), we get $J(t,x) = x' C_t x + d_t$, where d_t satisfies the backwards recursion

$$d_t = d_{t+1} + \Sigma_{i,j} N_t^{ij} C_{t+1}^{ij}, N_t = Cov(\varepsilon_t), d_T = 0,$$

(the top indices ij indicating elements in the matrices). We have obtained results in conformity with the so-called "certainty equivalence principle," namely that the control is the same as that obtained by taking expectation on the right-hand side of the state equation, i.e., by putting $\varepsilon_t = 0$, as if there were no uncertainty. This is a rather exceptional result, completely dependent on the particular structure of the problem. □

Proof of the optimality equation (1.18), (1.19)

The proof is provided for the specially interested reader and we assume that $\mathcal{V}$ is finite. Write $z_t = (x_t, v_t)$. For simplicity, $f_0(s,.,.,.)$ is assumed to be independent of v_s. We define, as before,

$$J(t,z_t) = \sup_{u_t,\ldots,u_T} E_{u_t,\ldots,u_T}\left[\sum_{s=t}^{T} f_0(s, X_s, u_s(Z_s)) \,\middle|\, z_t\right], \tag{1.24}$$

for $t < T$, with

$$J(T,z_T) = \sup_{u_T} f_0(T, x_T, u_T). \tag{1.25}$$

The optimality equation to be proved is

$$J(t-1,z_{t-1}) = \sup_{u_{t-1}} \big\{ f_0(t-1,x_{t-1},u_{t-1}) + E_{u_{t-1}} [J(t,Z_t) \mid z_{t-1}] \big\}. \tag{1.26}$$

In the proof, we shall consider a larger class of control policies, namely the general history-dependent controls $u_t(z_1,\ldots,z_t)$. Thus the controls are allowed to depend on all previous events v and states x. The proof to be presented makes it possible to answer the following question: Is it possible to achieve even better results if we are allowed to select policies from this larger collection of policies?

The argument below uses the following iterated expectation rule that can be found in standard texts on probability theory (see also the Appendix):

$$E[Y \mid X_1,\ldots,X_m] = E[E[Y \mid X_1,\ldots,X_n] \mid X_1,\ldots,X_m], m < n.$$

Let us write $J(t,z_{\to t})$ for the value that results when the policies in (1.24) are chosen from the class of policies $u_s(z_{\to s}) := u_s(z_1,\ldots,z_s)$, where the symbol $z_{\to s}$ means the sequence $(z_1,\ldots,z_s)$, and where we condition on $z_{\to t}$ rather than on just z_t.

Write $E^{s-1} := E_{u_{s-1},\ldots,u_T}$ (if the probabilities P_t do not depend on the controls, drop the superscripts $s-1$ and s on E below, the reader may want to concentrate on this slightly simpler case). The following sequence of equalities will be explained shortly:

$$\begin{aligned}
J(s-1,z_{\to s-1}) &= \sup_{t\geq s-1} E^{s-1}\left[\sum_{\tau=s-1}^{T} f_0(\tau,X_\tau,u_\tau(Z_{\to\tau})) \mid z_{\to s-1}\right] \\
&= \sup_{u_{s-1}(\cdot)}\Bigg[f_0(s-1,x_{s-1},u_{s-1}(z_{\to s-1})) \\
&\qquad + \sup_{t\geq s} E^{s-1}\left\{\sum_{\tau=s}^{T} f_0(\tau,X_\tau,u_\tau(Z_{\to\tau})) \mid z_{\to s-1}\right\}\Bigg] \\
&= \sup_{u_{s-1}(\cdot)}\Bigg[f_0(s-1,x_{s-1},u_{s-1}(z_{\to s-1})) \\
&\qquad + \sup_{t\geq s} E^{s-1}\left[E^{s}\left\{\sum_{\tau=s}^{T} f_0(\tau,X_\tau,u_\tau(Z_{\to\tau})) \mid Z_{\to s}\right\} \mid z_{\to s-1}\right]\Bigg] \\
&= \sup_{u_{s-1}(\cdot)}\Bigg[f_0(s-1,x_{s-1},u_{s-1}(z_{\to s-1})) \\
&\qquad + E^{s-1}\left\{\sup_{t\geq s} E^{s}\left[\sum_{t=s}^{T} f_0(\tau,X_\tau,u_\tau(Z_{\to\tau})) \mid Z_{\to s}\right] \mid z_{\to s-1}\right\}\Bigg] \\
&= \sup_{u_{s-1}} \big[f_0(s-1,x_{s-1},u_{s-1}) + E^{s-1}[J(s,Z_{\to s}) \mid z_{\to s-1}] \big].
\end{aligned} \tag{1.27}$$