# *Dynamic Programming and Optimal Control*

## *Volume I*

THIRD EDITION

Dimitri P. Bertsekas

Massachusetts Institute of Technology

WWW site for book information and

http://www.athenasc.com

Athena Scientific, Belmont, Massachusetts

Athena Scientific
Post Office Box 805
Nashua, NH 03061-0805
U.S.A.

ErnaH: info@athenasc.com
WWW: http://www.athenasc.co:m

Cover Design: Ann Gallager, www.gallagerdesign.com



Publisher's Cataloging-in-Publication Data

Bertsekas, Dimitri P.
Dynamic Programming and Optimal Control
Includes Bibliography and Index
1. Mathematical Optimization. 2. Dynamic Programming. L Title.
QA402.5 .13465 2005 519.703 00-91281

ISBN 1-886529-26-4

# ABOUT THE AUTHOR

Dimitri Bertsekas studied Mechanical and Electrical Engineering at the National Technical University of Athens, Greece, and obtained his Ph.D. in system science from the Massachusetts Institute of Technology. He has held faculty positions with the Engineering-Economic Systems Dept., Stanford University, and the Electrical Engineering Dept. of the University of Illinois, Urbana. Since 1979 he has been teaching at the Electrical Engineering and Computer Science Department of the Massachusetts Institute of Technology (M.LT.), where he is currently McAfee Professor of Engineering.

His research spans several fields, including optimization, control, la,rge-scale computation, and data communication networks, and is closely tied to his teaching and book authoring activities. He has written llUInerous research papers, and thirteen books, several of which are used as textbooks in MIT classes. He consults regularly with private industry and has held editorial positions in several journals.

Professor Bertsekas was awarded the INFORMS 1997 Prize for Research Excellence in the Interface Between Operations Research and Computer Science for his book "Neuro-Dynamic Programming" (co-authored with John Tsitsiklis), the 2000 Greek National Award for Operations Research, and the 2001 ACC John R. Ragazzini Education Award. In 2001, he was elected to the United States National Academy of Engineering.

ATHENA SCIENTIFIC

OPTIMIZATION AND CO\1PUTATION SERIES

1. Convex Analysis and Optimization, by Dimitri P. Bertsekas, with Angelia Nedić and Asuman E. Ozdaglar, 2003, ISBN 1-886529-45-0, 560 pages
2. Introduction to Probability, by Dimitri P. Bertsekas and John N. Tsitsiklis, 2002, ISBN 1-886529-40-X, 430 pages
3. Dynamic Programming and Optimal Control, Two-Volume Set, by Dimitri P. Bertsekas, 2005, ISBN 1-886529-08-6, 840 pages
4. Nonlinear Programming, 2nd Edition, by Dimitri P. Bertsekas, 1999, ISBN 1-886529-00-0, 791 pages
5. Network Optimization: Continuous and Discrete Models, by Dimitri P. Bertsekas, 1998, ISBN 1-886529-02-7, 608 pages
6. Network Flows and Monotropic Optimization, by R. Tyrrell RockareUar, 1998, ISBN 1-886529-06-X, 634 pages
7. Introduction to Linear Optimization, by Dimitris Bertsimas and John N. Tsitsiklis, 1997, ISBN 1-886529-19-1, 608 pages
8. Parallel and Distributed Computation: Numerical Methods, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1997, ISBN 1-886529-01-9, 718 pages
9. Neuro-Dynamic Programming, by Dimitri P. Bertsekas and John N. Tsitsiklis, 1996, ISBN 1-886529-10-8, 512 pages
10. Constra,ined Optimization and Lagrange Multiplier Methods, by Dimitri P. Bertsekas, 1996, ISBN 1-886529-04-3, 410 pages
11. Stochastic Optirnal Control: The Discrete-Time Case, by Dimitri P. Bertsekas and Steven E. Shreve, 1996, ISBN 1-886529-03-5, 330 pages

# *Contents*

# CONTENTS OF VOLUIVIE II

# *Preface*

This two-volume book is based on a first-year graduate course on dynamic programming and optimal control that I have taught for over twenty years at Stanford University, the University of Illinois, and the Massachusetts Institute of Technology. The course has been typically attended by students from engineering, operations research, economics, and applied mathematics. Accordingly, a principal objective of the book has been to provide a unified treatment of the subject, suitable for a broad audience. In particular, problems with a continuous character, such as stochastic control problems, popular in modern control theory, are simultaneously treated with problems with a discrete character, such as Markovian decision problems, popular in operations research. Furthermore, many applications and examples, drawn from a broad variety of fields, are discussed.

The book may be viewed as a greatly expanded and pedagogically improved version of my 1987 book "Dynamic Programming: Deterministic and Stochastic Models," published by Prentice-Hall. I have included much new material on deterministic and stochastic shortest path problems, as well as a new chapter on continuous-time optimal control problems and the Pontryagin Minimum Principle, developed from a dynamic programming viewpoint. I have also added a fairly extensive exposition of simulation-based approximation techniques for dynamic programming. These techniques, which are often referred to as "neuro-dynamic programming" or "reinforcement learning," represent a breakthrough in the practical application of dynamic programming to complex problems that involve the dual curse of large dimension and lack of an accurate mathematical model. Other material was also augmented, substantially modified, and updated.

With the new material, however, the book grew so much in size that it became necessary to divide it into two volumes: one on finite horizon, and the other on infinite horizon problems. This division was not only natural **in** terms of size, but also in terms of style and orientation. The first volume is more oriented towards modeling, and the second is more oriented towards mathematical analysis and computation. I have included in the first volume a final chapter that provides an introductory treatment of infinite horizon problems. The purpose is to make the first volume self-

# *The Dynamic Programming Algorithm*

Contents

Life can only be understood going backwards,
but it lllust be lived going forwards.
Kierkegaard

## 1.1 INTRODUCTION

This book deals with situations where decisions are made in stages. The outcome of each decision may not be fully predictable but can be anticipated to some extent before the next decision is made. The objective is to minimize a certain cost a mathematical expression of what is considered an undesirable outcome.

A key aspect of such situations is that decisions cannot be viewed in isolation since one must balance the desire for low present cost with the undesirability of high future costs. The dynamic programming technique captures this tradeoff. At each stage, it ranks decisions based on the sum of the present cost and the expected future cost, assuming optimal decision making for subsequent stages.

There is a very broad variety of practical problems that can be treated by dynamic programming. In this book, we try to keep the main ideas uncluttered by irrelevant assumptions on problem structure. To this end, we formulate in this section a broadly applicable model of optimal control of a dynamic system over a finite number of stages (a finite horizon). This model will occupy us for the first six chapters; its infinite horizon version will be the subject of the last chapter as well as Vol. II.

Our basic model has two principal features: (1) an underlying *discrete-time dynamic system*, and (2) a *cost function that is additive over time*. The dynamic system expresses the evolution of some variables, the system's "state", under the influence of decisions made at discrete instances of time. T'he system has the form

$$x_{k+1} = f_k(x_k, u_k, w_k), \qquad k = 0,1, \ldots, N-1,$$

where

$k$ indexes discrete time,

$x_k$ is the state of the system and summarizes past information that is relevant for future optimization,

$u_k$ is the control or decision variable to be selected at time $k$,

$w_k$ is a random parameter (also called disturbance or noise depending on the context),



$N$ is the horizon or number of times control is applied,

and $f_k$ is a function that describes the system and in particular the mechanism by which the state is updated.

The cost function is additive in the sense that the cost incurred at time $k$, denoted by $g_k(x_k, u_k, w_k)$, accumulates over time. The total cost is

$$g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k),$$

where $g_N(x_N)$ is a terminal cost incurred at the end of the process. However, because of the presence of $w_k$, the cost is generally a random variable and cannot be meaningfully optimized. We therefore formulate the problem as an optimization of the *expected cost*

$$E\left\{ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k) \right\},$$

where the expectation is with respect to the joint distribution of the random variables involved. The optimization is over the controls $u_0, u_1, \ldots, u_{N-1}$, but some qualification is needed here; each control $u_k$ is selected with some knowledge of the current state $x_k$ (either its exact value or some other related information).

A more precise definition of the terminology just used will be given shortly. We first provide some orientation by means of examples.

### Example 1.1.1 (Inventory Control)

Consider a problem of ordering a quantity of a certain item at each of $N$ periods so as to (roughly) meet a stochastic demand, while minimizing the incurred expected cost. Let us denote

$x_k$ stock available at the beginning of the $k$th period,

$u_k$ stock ordered (and immediately delivered) at the beginning of the kth period,

$w_k$ demand during the kth period with given probability distribution.

We assume that $w_0, w_1, \ldots, w_{N-1}$ are independent random variables, and that excess demand is backlogged and filled as soon as additional inventory becomes available. Thus, stock evolves according to the discrete-time equation

$$x_{k+1} = x_k + u_k - w_k,$$

where negative stock corresponds to backlogged demand (see Fig. 1.1.1).

The cost incurred in period $k$ consists of two components:

(a) A cost $r(x_k)$ representing a penalty for either positive stock $x_k$ (holding cost for excess inventory) or negative stock $x_k$ (shortage cost for unfilled demand).

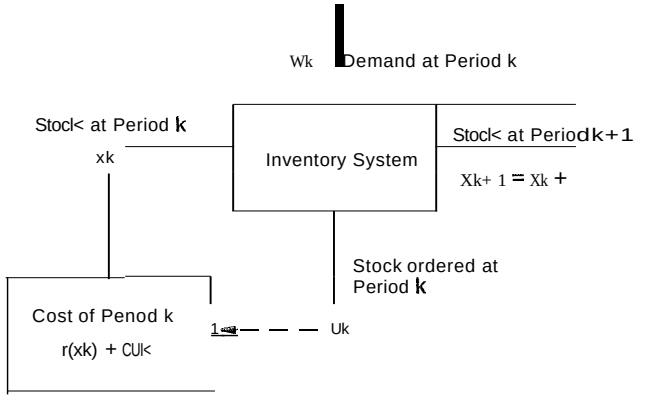



Figure 1.1.1 Inventory control example. At period $k$, the current stock (state) $x_k$, the stock ordered (control) $u_k$, and the demand (random disturbance) $w_k$ determine the cost $r(x_k)+cu_k$ and the stock $x_{k+1} = x_k + u_k - w_k$ at the next period.

(b) The purchasing cost $cu_k$, where c is cost per unit ordered.

There is also a terminal cost $R(x_N)$ for being left with inventory $x_N$ at the end of $N$ periods. Thus, the total cost over $N$ periods is

$$E\left\{R(x_N) + \sum_{k=0}^{N-1}\big(r(x_k) + cu_k\big)\right\}.$$

We want to minimize this cost by proper choice of the orders $u_0, \ldots, u_{N-1}$, subject to the natural constraint $u_k \geq 0$ for all $k$.

At this point we need to distinguish between *closed-loop* and *open-loop* minimization of the cost. In open-loop minimization we select all orders $u_0, \ldots, u_{N-1}$ at once at time 0, without waiting to see the subsequent demand levels. In closed-loop minimization we postpone placing the order $u_k$ until the last possible moment (time $k$) when the current stock $x_k$ will be known. The idea is that since there is no penalty for delaying the order $u_k$ up to time $k$, we can take advantage of information that becomes available between times 0 and $k$ (the demand and stock level in past periods).

Closed-loop optimization is of central importance in dynamic programming and is the type of optimization that we will consider almost exclusively in this book. Thus, in our basic formulation, decisions are made in stages while gathering information between stages that will be used to enhance the quality of the decisions. The effect of this on the structure of the resulting optimization problem is quite profound. In particular, in closed-loop inventory optimization we are not interested in finding optimal numerical values of the orders but rather we want to find an *optimal rule for selecting at each period k an order* $u_k$ *for each possible value of stock* $x_k$ *that can conceivably occur*. This is an "action versus strategy" distinction.

Mathematically, in closed-loop inventory optimization, we want to find a sequence of functions $\mu_k$, $k = 0, \ldots, N-1$, mapping stock $x_k$ into order $u_k$



so as to minimize the expected cost. The meaning of $\mu_k$ is that, for each $k$ and each possible value of $x_k$,

$\mu_k(x_k)$ = amount that should be ordered at time $k$ if the stock is $x_k$.

The sequence $\pi = \{\mu_0, \ldots, \mu_{N-1}\}$ will be referred to as a *policy* or *control law*. For each $\pi$, the corresponding cost for a fixed initial stock $x_0$ is

$$J_\pi(x_0) = E\left\{R(x_N) + \sum_{k=0}^{N-1}\big(r(x_k) + c\mu_k(x_k)\big)\right\},$$

and we want to minimize $J_\pi(x_0)$ for a given $x_0$ over all $\pi$ that satisfy the constraints of the problem. This is a typical dynamic programming problem. We will analyze this problem in various forms in subsequent sections. For example, we will show in Section 4.2 that for a reasonable choice of the cost function, the optimal ordering policy is of the form

$$\mu_k(x_k) = \begin{cases} S_k - x_k & \text{if } x_k < S_k, \\ 0 & \text{otherwise,} \end{cases}$$

where $S_k$ is a suitable threshold level determined by the data of the problem. In other words, when stock falls below the threshold $S_k$, order just enough to bring stock up to $S_k$.

The preceding example illustrates the main ingredients of the basic problem formulation:

(a) A *discrete-time system* of the form

$$x_{k+1} = f_k(x_k, u_k, w_k),$$

where $f_k$ is some function; for example in the inventory case, we have $f_k(x_k, u_k, w_k) = x_k + u_k - w_k$.

(b) *Independent random parameters* $w_k$. This will be generalized by allowing the probability distribution of $w_k$ to depend on $x_k$ and $u_k$; in the context of the inventory example, we can think of a situation where the level of demand $w_k$ is influenced by the current stock level $x_k$.

(c) A *control constraint;* in the example, we have $u_k \geq 0$. In general, the constraint set will depend on $x_k$ and the time index $k$, that is, $u_k \in U_k(x_k)$. To see how constraints dependent on $x_k$ can arise in the inventory context, think of a situation where there is an upper bound $B$ on the level of stock that can be accommodated, so $u_k \leq B - x_k$.

(d) An *additive cost* of the form

$$E\left\{g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)\right\},$$

where $g_k$ are some functions; in the inventory example, we have

$$g_N(x_N) \qquad g_k(x_k, u_k, w_k) = r(x_k) + cu_k.$$

(e) *Optimization over (closed-loop) policies*, that is, rules for choosing $u_k$ for each $k$ and each possible value of $x_k$.

## Discrete-State and Finite-State Problems

In the preceding example, the state $x_k$ was a continuous real variable, and it is easy to think of multidimensional generalizations where the state is an n-dimensional vector of real variables. It is also possible, however, that the state takes values from a discrete set, such as the integers.

A version of the inventory problem where a discrete viewpoint is more natural arises when stock is measured in whole units (such as cars), each of which is a significant fraction of $x_k$, $u_k$, or $w_k$. It is more appropriate then to take as state space the set of all integers rather than the set of real numbers. The form of the system equation and the cost per period will, of course, stay the same.

Generally, there are many situations where the state is naturally discrete and there is no continuous counterpart of the problem. Such situations are often conveniently specified in terms of the probabilities of transition between the states. What we need to know is $P_{ij}(u, k)$, which is the probability at time $k$ that the next state will be $j$, given that the current state is $i$, and the control selected is $u$, Le.,

$$p_{ij}(u, k) \quad P\{x_{k+1} = j \mid x_k = i, u_k = u\}.$$

This type of state transition can alternatively be described in terms of the discrete-time system equation

$$x_{k+1} = w_k,$$

where the probability distribution of the random parameter $w_k$ is

$$P\{w_k = j \mid x_k = i, u_k = u\} = p_{ij}(u, k).$$

Conversely, given a discrete-state system in the form

$$x_{k+1} = f_k(x_k, u_k, w_k),$$

together with the probability distribution $P_k(w_k \mid x_k, u_k)$ of $w_k$, we can provide an equivalent transition probability description. The corresponding transition probabilities are given by

$$p_{ij}(u, k) = P_k\{W_k(i, u, j) \mid x_k = i, u_k = u\},$$



where $W(i, u, j)$ is the set

$$W_k(i, u, j) = \{w \mid j = f_k(i, u, w)\}.$$

Thus a discrete-state system can equivalently be described in terms of a difference equation or in terms of transition probabilities. Depending on the given problem, it may be notationally or mathematically more convenient to use one description over the other.

The following examples illustrate discrete-state problems. The first example involves a *deterministic* problem, that is, a problem where there is no stochastic uncertainty. In such a problem, when a control is chosen at a given state, the next state is fully determined; that is, for any state i, control $u$, and time $k$, the transition probability $P_{ij}(u, k)$ is equal to 1 for a single state $j$, and it is 0 for all other candidate next states. The other three examples involve stochastic problems, where the next state resulting from a given choice of control at a given state cannot be determined a priori.

### Example 1.1.2 (A Deterministic Scheduling Problem)

Suppose that to produce a certain product, four operations must be performed on a certain machine. The operations are denoted by A, B, C, and D. We assume that operation B can be performed only after operation A has been performed, and operation D can be performed only after operation B has been performed. (Thus the sequence CDAB is allowable but the sequence CDBA is not.) The setup cost $C_{mn}$ for passing from any operation $m$ to any other operation $n$ is given. There is also an initial startup cost $S_A$ or $S_C$ for starting with operation A or C, respectively. The cost of a sequence is the sum of the setup costs associated with it; for example, the operation sequence ACDB has cost

$$S_A + C_{AC} + C_{CD} + C_{DB}.$$

We can view this problem as a sequence of three decisions, namely the choice of the first three operations to be performed (the last operation is determined from the preceding three). It is appropriate to consider as state the set of operations already performed, the initial state being an artificial state corresponding to the beginning of the decision process. The possible state transitions corresponding to the possible states and decisions for this problem is shown in Fig. 1.1.2. Here the problem is deterministic, Le., at a given state, each choice of control leads to a uniquely determined state. For example, at state AC the decision to perform operation D leads to state ACD with certainty, and has cost $C_{CD}$. Deterministic problems with a finite number of states can be conveniently represented in terms of transition graphs' such as the one of Fig. 1.1.2. The optimal solution corresponds to the path that starts at the initial state and ends at some state at the terminal time and has minimum sum of arc costs plus the terminal cost. We will study systematically problems of this type in Chapter 2.

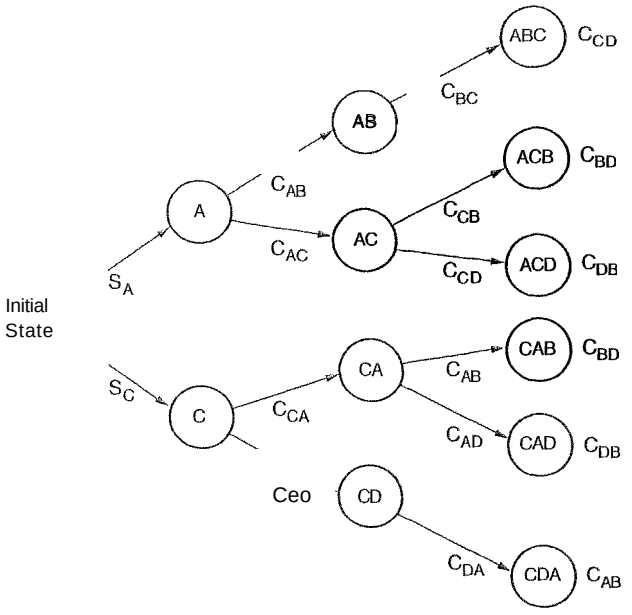



Figure 1.1.2 The transition graph of the deterministic scheduling problem of Exarnple 1.1.2. Each arc of the graph corresponds to a decision leading from some state (the start node of the arc) to some other state (the end node of the arc). The corresponding cost is shown next to the arc. The cost of the last operation is shown as a terminal cost next to the terminal nodes of the graph.

## Exarnple 1.1.3 (Machine Replacement)

Consider a problem of operating efficiently over $N$ time periods a machine that can be in anyone of $n$ states, denoted 1,2, ... , $n$. We denote by $g(i)$ the operating cost per period when the machine is in state i, and we assume that

$$g(l) \leq g(2) \leq \ldots \leq g(n).$$

The implication here is that state i is better than state i + 1, and state 1 corresponds to a machine in best condition.

During a period of operation, the state of the machine can become worse or it may stay unchanged. We thus assume that the transition probabilities

$$p_{ij} = P\{\text{next state will be j} \mid \text{current state is i}\}$$

satisfy

$$p_{ij} = 0 \qquad \text{if } j < \text{i}.$$

We assume that at the start of each period we know the state of the machine and we must choose one of the following two options:



(a) Let the machine operate one more period in the state it currently is.

(b) Repair the machine and bring it to the best state 1 at a cost $R$.

We assume that the machine, once repaired, is guaranteed to stay in state 1 for one period. In subsequent periods, it may deteriorate to states $j > 1$ according to the transition probabilities $p_{ij}$.

Thus the objective here is to decide on the level of deterioration (state) at which it is worth paying the cost of machine repair, thereby obtaining the benefit of smaller future operating costs. Note that the decision should also be affected by the period we are in. For example, we would be less inclined to repair the machine when there are few periods left.

The system evolution for this problem can be described by the graphs of Fig. 1.1.3. These graphs depict the transition probabilities between various pairs of states for each value of the control and are known as *transition probability graphs* or simply *transition graphs*. Note that there is a different graph for each control; in the present case there are two controls (repair or not repair).

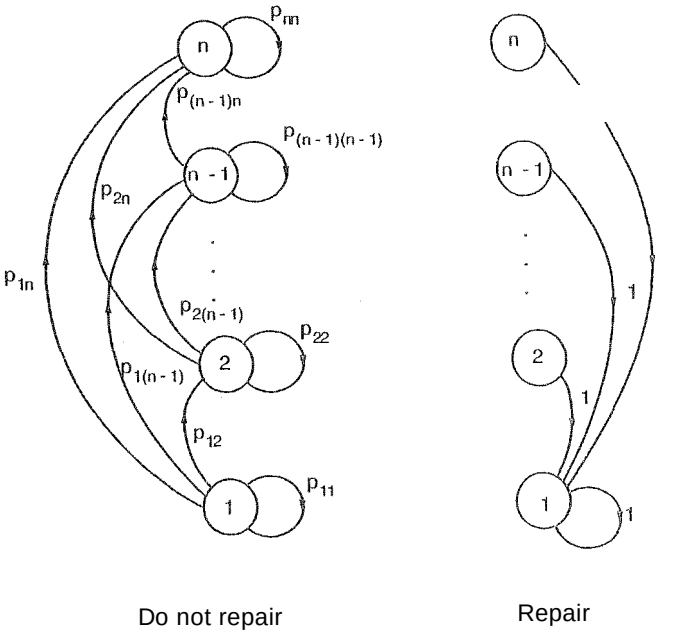



Figure 1.1.3 Machine replacement example. Transition probability graphs for each of the two possible controls (repair or not repair). At each stage and state i, the cost of repairing is $R+g(l)$, and the cost of not repairing is $g(i)$. The terminal cost is 0.

from the game is

$$\sum_{k=0}^{\infty} \frac{1}{2^{k+1}} \cdot 2^k = \infty,$$

so if his acceptance eriterion is based on maximization of expected profit, he is willing to pay any amount $x$ to enter the game. This, however, is in strong disagreement with observed behavior, due to the risk element involved in entering the game, and shows that a different formulation of the problem is needed. The formulation of problems of deeision under uncertainty so that risk is properly taken into aeeount is a deep subject with an interesting theory. An introduction to this theory is given in Appendix G. It is shown in particular that minimization of expected cost is appropriate under reasonable assumptions, provided the cost function is suitably chosen so that it properly eneodes the risk preferences of the deeision maker.

## 1.3 THE DYNAMIC PROGRAMMING ALGORITHM

The dynamie programming (DP) technique rests on a very simple idea, the *principle of optimality*. The name is due to Bellman, who contributed a great deal to the popularization of DP and to its transformation into a systematic tool. Roughly, the principle of optimality states the following rather obvious fact.

**Principle of Optimality**

Let $\pi^* = \{\mu_0^*, \mu_1^*, \ldots, \mu_{N-1}^*\}$ be an optimal policy for the basic problem, and assume that when using $\pi^*$, a given state $x_i$ occurs at time i with positive probability. Consider the subproblem whereby we are at $x_i$ at time i and wish to minimize the "cost-to-go" from time i to time $N$

$$E\left\{ g_N(x_N) + \sum_{k=i}^{N-1} g_k\big(x_k, \mu_k(x_k), w_k\big) \right\}.$$

Then the truncated poliey $\{\mu_i^*, \mu_{i+1}^*, \ldots, \mu_{N-1}^*\}$ is optimal for this subproblem.

The intuitive justification of the prineiple of optimality is very simple. If the truncated policy $\{\mu_i^*, \mu_{i+1}^*, \ldots, \mu_{N-1}^*\}$ were not optimal as stated, we would be able to reduce the cost further by switching to an optimal policy for the subproblem once we reach $x_i$. For an auto travel analogy, suppose that the fastest route from Los Angeles to Boston passes through Chicago. The principle of optimality translates to the obvious fact that the Chicago to Boston portion of the route is also the fastest route for a trip that starts from Chicago and ends in Boston.



The principle of optimality suggests that an optimal policy can be constructed in piecemeal fashion, first constructing an optimal policy for the "tail subproblem" involving the last stage, then extending the optimal policy to the "tail subproblem" involving the last two stages, and continuing in this manner until an optimal policy for the entire problem is constructed. The DP algorithm is based on this idea: it proceeds sequentially, by solving all the tail subproblems of a given time length, using the solution of the tail subproblems of shorter time length. We introduce the algorithm with two examples, one deterministic and one stochastic.

### The DP Algorithm for a Deterministic Scheduling Example

Let us consider the scheduling example of the preceding section, and let us apply the principle of optimality to calculate the optimal schedule. We have to schedule optimally the four operations A, B, C, and D. The transition and setup costs are shown in Fig. 1.3.1 next to the corresponding arcs.

According to the principle of optimality, the "tail" portion of an optimal schedule must be optimal. For example, suppose that the optimal schedule is CABD. Then, having scheduled first C and then A, it must be optimal to complete the schedule with BD rather than with DB. With this in mind, we solve all possible tail subproblems of length two, then all tail subproblems of length three, and finally the original problem that has length four (the subproblems of length one are of course trivial because there is only one operation that is as yet unscheduled). As we will see shortly, the tail subproblems of length $k+1$ are easily solved once we have solved the tail subproblems of lengt.h $k$, and this is the essence of the DP technique.

*Tail Subproblems of Length* 2: These subproblems are the ones that involve two unscheduled operations and correspond to the states AB, AC, CA, anel CD (see Fig. 1.3.1)

*State AB:* Here it is only possible to schedule operation C as the next operation, so the optimal cost of this subproblem is 9 (the cost of scheduling C after B, which is 3, plus the cost of scheduling Dafter C, which is 6).

*State AC:* Here the possibilities are to (a) schedule operation 13 and then D, which has cost 5, or (b) schedule operation D anel then B, which has cost 9. The first possibility is optimal, and the corresponding cost of the tail subproblem is 5, as shown next to node AC in Fig. 1.3.l.

*State CA:* Here the possibilities are to (a) schedule operation 13 and then D, which has cost 3, or (b) schedule operation D and then 13, which has cost 7. The first possibility is optimal, and the correspond-

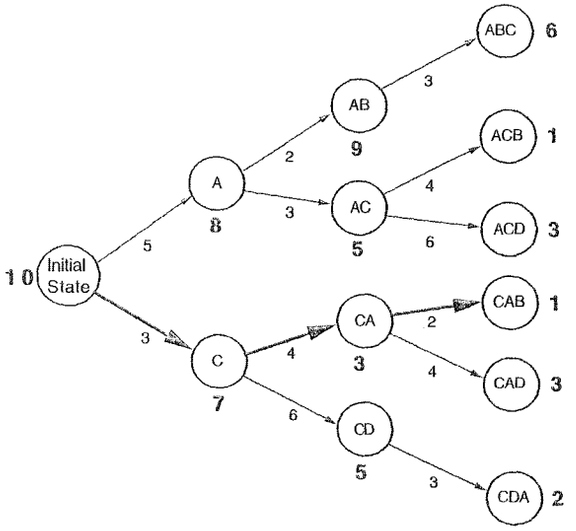



Figure 1.3.1 Transition graph of the deterministic scheduling problem, with the cost of each decision shown next to the corresponding arc. Next to each node/state we show the cost to optimally complete the schedule starting from that state. This is the optimal cost of the corresponding tail subproblem (ef. the principle of optimality). The optimal cost for the original problem is equal to 10, as shown next to the initial state. The optimal schedule corresponds to the thick-line arcs.

ing cost of the tail subproblem is 3, as shown next to node CA in Fig. 1.3.1.

*State CD:* Here it is only possible to schedule operation A as the next operation, so the optimal cost of this subproblem is 5.

*Tail Subproblems of Length* 3: These subproblems can now be solved using the optimal costs of the subproblems of length 2.

*State A:* Here the possibilities are to (a) schedule next operation B (cost 2) and then solve optimally the corresponding subproblem of length 2 (cost 9, as computed earlier), a total cost of 11, or (b) schedule next operation C (cost 3) and then solve optimally the corresponding subproblem of length 2 (cost 5, as computed earlier), a total cost of 8. The second possibility is optimal, and the corresponding cost of the tail subproblem is 8, as shown next to node A in Fig. 1.3.1.

*State C:* Here the possibilities are to (a) schedule next operation A (cost 4) and then solve optimally the corresponding subproblem of length 2 (cost 3, as computed earlier), a total cost of 7, or (b) schedule next operation D (cost 6) and then solve optimally the corresponding



subproblem of length 2 (cost 5, as computed earlier), a total cost of 11. The first possibility is optimal, and the corresponding cost of the tail subproblem is 7, as shown next to node A in Fig. 1.3.1.

*Original Problem of Length* 4: The possibilities here are (a) start with operation A (cost 5) and then solve optimally the corresponding subproblem of length 3 (cost 8, as computed earlier), a total cost of 13, or (b) start with operation C (cost 3) and then solve optimally the corresponding subproblem of length 3 (cost 7, as computed earlier), a total cost of 10. The second possibility is optimal, and the corresponding optimal cost is 10, as shown next to the initial state node in Fig. 1.3.1.

Note that having computed the optimal cost of the original problem through the solution of all the tail subproblems, we can construct the optimal schedule by starting at the initial node and proceeding forward, each time choosing the operation that starts the optimal schedule for the corresponding tail subproblem. In this way, by inspection of the graph and the computational results of Fig. 1.3.1, we determine that CABD is the optimal schedule.

## The DP Algorithm for the Inventory Control Example

Consider the inventory control example of the previous section. Similar to the solution of the preceding deterministic scheduling problem, we calculate sequentially the optimal costs of all the tail subproblems, going from shorter to longer problems. The only difference is that the optimal costs are computed as expected values, since the problem here is stochastic.

*Ta'il Subproblems of Length* 1: Assume that at the beginning of period $N - 1$ the stock is $x_{N-1}$. Clearly, no matter what happened in the past, the inventory manager should order the amount of inventory that minimizes over $u_{N-1} \geq 0$ the sum of the ordering cost and the expected terminal holding/shortage cost. Thus, he should minimize over $u_{N-1}$ the sum $c u_{N-1} + E\{R(x_N)\}$, which can be written as

$$c u_{N-1} + \mathop{E}_{w_{N-1}} \{R(x_{N-1} + u_{N-1} - w_{N-1})\}.$$

Adding the holding/shortage cost of period $N - 1$, we see that the optimal cost for the last period (plus the terminal cost) is given by

$$J_{N-1}(x_{N-1}) = r(x_{N-1}) + \min_{u_{N-1} \geq 0} \left[ c u_{N-1} + \mathop{E}_{w_{N-1}} \{R(x_{N-1} + u_{N-1} - w_{N-1})\} \right].$$

Naturally, $J_{N-1}$ is a function of the stock $x_{N-1}$. It is calculated either analytically or numerically (in which case a table is used for computer

storage ofthe function *IN-1)*. In the process of caleulating *IN-1*, we obtain the optimal inventory policy $\mu^*_{N-1}(x_{N-1})$ for the last period: $\mu^*_{N-1}(x_{N-1})$ is the value of $u_{N-1}$ that minimizes the right-hand side of the preceding equation for a given value of *XN-1*.

*Tail Subproblems of Length* 2: Assume that at the beginning of period $N$ 2 the stock is $x_{N-2}$. It is clear that the inventory manager should order the amount of inventory that minimizes not just the expected cost of period $N - 2$ but rather the

(expected cost of period $N - 2$) + (expected cost of period $N - 1$, given that an optimal policy will be used at period $N - 1$),

which is equal to

$$r(x_{N-2}) + cu_{N-2} + E\{J_{N-1}(x_{N-1})\}.$$

Using the system equation $x_{N-1} = XN\text{-}2 + UN\text{-}2 - WN\text{-}2$, the last term is also written as $IN\text{-}1(XN\text{-}2 + UN\text{-}2 \quad WN\text{-}2)$.

Thus the optimal cost for the last two periods given that we are at state $x_{N-2}$, denoted $IN\text{-}2(XN\text{-}2)$, is given by

$$= T(XN\text{-}2) + \min_{u_{N-2}\geq 0}\left[cu_{N-2} + E_{WN\text{-}2}\{IN\text{-}1(XN\text{-}2 + u_{N-2} - WN\text{-}2)\}\right]$$

Again $J_{N-2}(x_{N-2})$ is caleulated for every *XN-2*. At the same time, the optimal policy $\mu^*_{N-2}(x_{N-2})$ is also computed.

*Tail Subproblems of Length N - k:* Similarly, we have that at period $k$, when the stock is $x_k$, the inventory manager should order $uk$ to minimize

(expected cost of period $k$) + (expected cost of periods $k + 1, \ldots, N - 1$, given that an optimal policy will be used for these periods).

By denoting by *Jk(Xk)* the optimal cost, we have

$$J_k(x_k) = r(x_k) + \min_{u_k\geq 0}\left[cu_k + \underset{w_k}{E}\{J_{k+1}(x_k + u_k - w_k)\}\right], \qquad (1.4)$$

which is actually the dynamic programming equation for this problem.

The functions $J_k(x_k)$ denote the optimal expected cost for the tail subproblem that starts at period $k$ with initial inventory *Xk*. These functions are computed recursively backward in time, starting at period $N - 1$ and ending at period 0. The value $Jo(x_0)$ is the optimal expected cost when the initial stock at time 0 is $x_0$. During the caleulations, the optiInal policy is simultaneously computed from the minimization in the right-hand side of Eq. (1.4).



The example illustrates the main advantage offered by DP. While the original inventory problem requires an optimization over the set of policies, the DP algorithm of Eq. (1.4) decomposes this problem into a sequence of minimizations carried out over the set of controls. Each of these minimizations is much simpler than the original problem.

## The DP Algorithm

We now state the DP algorithm for the basic problem and show its optimality by translating into mathematical terms the heuristic argument given above for the inventory example.

**Proposition 1.3.1:** For every initial state *Xo*, the optimal cost *J*(xo)* of the basic problem is equal to *Jo(xo)*, given by the last step of the following algorithm, which proceeds backward in time from period $N - 1$ to period 0:

$$J_N(x_N) = g_N(x_N), \qquad (1.5)$$

$$Jk(Xk) = \min_{UkEUk(Xk)} \underset{Wk}{E}\{9k(Xk"Uk,Wk) + J_{k+1}\big(f_k(x_k, u_k, w_k)\big)\}, \quad k = 0,1, \ldots, N - 1, \qquad (1.6)$$

where the expectation is taken with respect to the probability distribution of $w_k$, which depends on *Xk* and *uk*. :Furthermore, if $u^*_k = \mu^*_k(x_k)$ minimizes the right side of Eq. '(1.6) for each *Xk* and *k*, the policy $\pi^* = \{\mu^*_0, \ldots, \mu^*_N\text{-}1\}$ is optimal.

Proof: t For any admissible policy $\pi = \{\mu_0, \mu_1, \ldots, \mu_{N-1}\}$ and each $k = 0,1, \ldots, N\text{-}1$, denote $\pi^k = \{\mu_k, \mu_{k+1}, \ldots, \mu_{N-1}\}$. For $k$ 0,1, ... ,*N-1*, let $J^*_k(x_k)$ be the optimal cost for the *(N* - k)-stage problem that starts at state *Xk* and time *k*, and ends at time *N*,

$$J^*_k(x_k) = \min_{\pi^k} \underset{w_k,\ldots,w_{N-1}}{E}\left\{g_N(x_N) + \sum_{i=k}^{N-1} g_i\big(x_i, \mu_i(x_i), w_i\big)\right\}.$$

t Our proof is somewhat informal and assumes that the functions $J_k$ are well-defined and finite. For a strictly rigorous proof, some technical mathematical issues must be addressed; see Section 1.5. These issues do not arise if the disturbance $w_k$ takes a finite or countable number of values and the expected values of all terms in the expression of the cost function (1.1) are well-defined and finite for every admissible policy $\pi$.

For $k$ IV, we define $J_N^*(x_N)$ = gN(XN). We will show by induction that the functions $J_k^*$ are equal to the functions Jk generated by the DP algorithm, so that for $k$ = 0, we will obtain the desired result.

Indeed, we have by definition $J_N^*$ = JN = gN. Assume that for some $k$ and all Xk+l, we have $J_{k+1}^*(x_{k+1})$ = Jk+1(Xk+l). Then, since $\pi^k$ $(\mu_k, \pi^{k+1})$, we have for all xk

$$J_k^*(x_k) = \min_{(\mu_k,\pi^{k+1})} \underset{w_k,\ldots,w_{N-1}}{E} \left\{ g_k\big(x_k,\mu_k(x_k),w_k\big) + g_N(x_N) + \sum_{i=k+1}^{N-1} g_i\big(x_i,\mu_i(x_i),w_i\big) \right\}$$

$$= \min_{\mu_k} \underset{\text{T1Jk}}{E} \{9k\,(Xk'\,\mu_k(x_k),w_k\big) + \min_{\pi^{k+1}} \left[ \underset{\text{wk+I,...,'I1JN-I}}{E} \{9N(XN) + \sum_{i=k+1}^{N-1} gi\,(Xi,\mu_i(x_i),'Wi)\} \right] \}$$

$$= \min_{\mu_k} \underset{\text{T1Jk}}{E} \{9k\,(Xk'\,\mu_k(x_k),w_k\big) + J_{k+1}^*\,Uk\,(Xk'\,\mu_k(x_k),Wk))\}$$

$$= \min_{\mu_k} \underset{w_k}{E} \left\{ g_k\big(x_k,\mu_k(x_k),w_k\big) + J_{k+1}\big(f_k\big(x_k,\mu_k(x_k),w_k\big)\big) \right\}$$

$$= \min_{u_k\in U_k(x_k)} \underset{w_k}{E} \{9k(Xk,'Uk,'Wk) + Jk+l(fk(Xk,Uk,'W_k))\}$$

$$= Jk(Xk),$$

cOInpleting the induction. In the second equation above, we moved the minimum over $\pi^{k+1}$ inside the braced expression, using a principle of optimalityargument: "the tail portion of an optimal policy is optimal for the tail subproblem" (a more rigorous justification of this step is given in Section 1.5). In the third equation, we used the definition of $J_{k+1}^*$, and in the fourth equation we used the induction hypothesis. In the fifth equation, we converted the minimization over $\mu_k$ to a minimization over $u_k$, using the fact that for any function F of $x$ and $u$, we have

$$\min_{\mu\in M} F(x,\mu(x)\big) = \min_{U\in U(x)} F(x,u),$$

where $M$ is the set of all functions $\mu(x)$ such that $\mu(x)$ ∈ U(x) for all $x$. Q.E.D.

The argument of the preceding proof provides an interpretation of $J_k(x_k)$ as the optimal cost for an (N ‒ k)-stage problem starting at state $x_k$ and time $k$, and ending at time $N$. We consequently call Jk(Xk) the *cost-to-go* at state Xk and time k, and refer to Jk as the *cost-to-go function* at time $k$.



Ideally, we would like to use the DP algorithm to obtain closed-form expressions for $J_k$ or an optimal policy. In this book, we will discuss a large number of models that admit analytical solution by DP. Even if such models rely on oversimplified assumptions, they are often very useful. They may provide valuable insights about the structure of the optimal solution of more complex models, and they may form the basis for suboptimal control schemes. Furthermore, the broad collection of analytically solvable models provides helpful guidelines for modeling: when faced with a new problem it is worth trying to pattern its model after one of the principal analytically tractable models.

Unfortunately, in many practical cases an analytical solution is not possible, and one has to resort to numerical execution of the DP algorithm. This may be quite time-consuming since the minimization in the DP Eq. (1.6) must be carried out for each value of Xk. The state space must be discretized in some way if it is not already a finite set. The computational requirements are proportional to the number of possible values of Xk, so for complex problems the computational burden may be excessive. Nonetheless, DP is the only general approach for sequential optimization under uncertainty, and even when it is computationally prohibitive, it can serve as the basis for more practical suboptimal approaches, which will be discussed in Chapter 6.

The following examples illustrate some of the analytical and computational aspects of DP.

### Example 1.3.1

A certain material is passed through a sequence of two ovens (see Fig. 1.3.2). Denote

xo: initial temperature of the material,

Xk, k = 1,2: temperature of the material at the exit of oven k,

$u_{k-1}$, k = 1,2: prevailing temperature in oven k.

We assume a model of the form

$$x_{k+1} = (1-a)x_k + au_k, \qquad k \quad 0,1,$$

where $a$ is a known scalar from the interval (0,1). The objective is to get the final temperature $x_2$ close to a given target T, while expending relatively little energy. This is expressed by a cost function of the form

$$r(x_2 - T')^2 + u_0^2 + u_1^2,$$

where r > 0 is a given scalar. We assume no constraints on $u_k$. (In reality, there are constraints, but if we can solve the unconstrained problem and verify that the solution satisfies the constraints, everything will be fine.) The problem is deterministic; that is, there is no stochastic uncertainty. However,

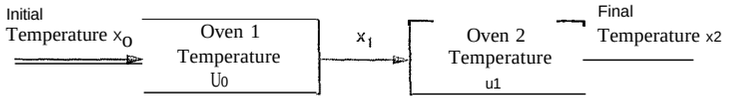



Figure 1.3.2 Problem of Example 1.3.1. The temperature of the material evolves according to $x_{k+1} = (1 - a)x_k + au_k$, where $a$ is some scalar with $0 < a < 1$.

such problems can be placed within the basic framework by introducing a fictitious disturbance taking a unique value with probability one.

We have $N = 2$ and a terminal cost $g_2(x_2) = r(x_2 - T)^2$, so the initial condition for the DP algorithm is [ef. Eq. (1.5)]

$$J_2(x_2) = r(x_2 - T)^2.$$

For the next-to-last stage, we have [cf. Eq. (1.6)]

$$J_1(x_1) = \min_{u_1} \left[u_1^2 + J_2(x_2)\right]$$
$$= \min_{u_1} \left[u_1^2 + J_2\left((1 - a)x_1 + au_1\right)\right].$$

Substituting the previous form of $J_2$, we obtain

$$J_1(x_1) = \min_{u_1} \left[u_1^2 + r\left((1 - a)x_1 + au_1 - T\right)^2\right]. \qquad (1.7)$$

This minimization will be done by setting to zero the derivative with respect to $u_1$. This yields

$$0 = 2u_1 + 2ra\left((1 - a)x_1 + au_1 - T\right),$$

and by collecting terms and solving for $u_1$, we obtain the optimal temperature for the last oven:

$$\mu_1^*(x_1) = \frac{ra(T - (1 - a)x_1)}{1 + ra^2}.$$

Note that this is not a single control but rather a control function, a rule that tells us the optimal oven temperature $u_1 = \mu_1^*(x_1)$ for each possible state $x_1$.

By substituting the optimal $u_1$ in the expression (1.7) for $J_1$, we obtain

$$J_1(x_1) = \frac{r^2a^2\left((1 - a)x_1 - T\right)^2}{(1 + ra^2)^2} + r\left((1 - a)x_1 + \frac{ra^2\left(T - (1 - a)x_1\right)}{1 + ra^2} - T\right)^2$$
$$= \frac{r^2a^2\left((1 - a)x_1 - T\right)^2}{(1 + ra^2)^2} + r\left(\frac{ra^2}{1 + ra^2} - 1\right)^2\left((1 - a)x_1 - T\right)^2$$
$$= \frac{r\left((1 - a)x_1 - T\right)^2}{1 + ra^2}.$$



We now go back one stage. We have [ef. Eq. (1.6)]

$$J_0(x_0) = \min_{u_0}\left[u_0^2 + J_1(x_1)\right] = \min_{u_0}\left[u_0^2 + J_1\left((1 - a)x_0 + au_0\right)\right],$$

and by substituting the expression already obtained for $J_1$, we have

$$J_0(x_0) = \min_{u_0}\left[u_0^2 + \frac{r((1 - a)^2x_0 + (1 - a)au_0 - T)^2}{1 + ra^2}\right].$$

We minimize with respect to $u_0$ by setting the corresponding derivative to zero. We obtain

$$0 = 2u_0 + \frac{2r(1 - a)a((1 - a)^2x_0 + (1 - a)au_0 - T)}{1 + ra^2}.$$

This yields, after some calculation, the optimal temperature of the first oven:

$$\mu_0^*(x_0) = \frac{r(1 - a)a\left(T - (1 - a)^2x_0\right)}{1 + ra^2\left(1 + (1 - a)^2\right)}.$$

The optimal cost is obtained by substituting this expression in the formula for $J_0$. This leads to a straightforward but lengthy calculation, which in the end yields the rather simple formula

$$J_0(x_0) = \frac{r\left((1 - a)^2x_0 - T\right)^2}{1 + ra^2\left(1 + (1 - a)^2\right)}$$

This completes the solution of the problem.

One noteworthy feature in the preceding example is the facility with which we obtained an analytical solution. A little thought while tracing the steps of the algorithm will convince the reader that what simplifies the solution is the quadratic nature of the cost and the linearity of the system equation. In Section 4.1 we will see that, generally, when the system is linear and the cost is quadratic, the optimal policy and cost-to-go function are given by closed-form expressions, regardless of the number of stages $N$.

Another noteworthy feature of the example is that the optimal policy remains unaffected when a zero-mean stochastic disturbance is added in the system equation. To see this, assume that the material's temperature evolves according to

$$x_{k+1} = (1 - a)x_k + au_k + w_k, \qquad k = 0,1,$$

where $w_0$, $w_1$ are independent random variables with given distribution, zero mean

$$E\{w_0\} = E\{w_1\} = 0,$$

and finite variance. Then the equation for $J_1$ [ref. Eq. (1.6)] becomes

$$J_1(x_1) = \min_{u_1} E_{w_1}\left\{u_1^2 + r((1-a)x_1 + au_1 + w_1 - T)^2\right\}$$
$$= \min_{u_1}\Big[u_1^2 + r((1-a)x_1 + au_1 - T)^2 + 2rE\{w_1\}((1-a)x_1 + au_1 - T) + rE\{w_1^2\}\Big].$$

Since $E\{w_1\} = 0$, we obtain

$$J_1(x_1) = \min_{u_1}\left[u_1^2 + r\big((1-a)x_1 + au_1 - T\big)^2\right] + rE\{w_1^2\}.$$

Comparing this equation with Eq. (1.7), we see that the presence of $w_1$ has resulted in an additional inconsequential term, $rE\{w_1^2\}$. Therefore, the optimal policy for the last stage remains unaffected by the presence of $w_1$, while $J_1(x_1)$ is increased by the constant term $rE\{w_1^2\}$. It can be seen that a similar situation also holds for the first stage. In particular, the optimal cost is given by the same expression as before except for an additive constant that depends on $E\{w_0^2\}$ and $E\{w_1^2\}$.

If the optimal policy is unaffected when the disturbances are replaced by their means, we say that *certainty equivalence* holds. We will derive certainty equivalence results for several types of problems involving a linear system and a quadratic cost (see Sections 4.1, 5.2, and 5.3).

### Example 1.3.2

To illustrate the computational aspects of DP, consider an inventory control problem that is slightly different from the one of Sections 1.1 and 1.2. In particular, we assume that inventory $x_k$ and the demand $w_k$ are nonnegative integers, and that the excess demand $(w_k - x_k - u_k)$ is lost. As a result, the stock equation takes the form

$$x_{k+1} = \max(0, x_k + u_k - w_k).$$

We also assume that there is an upper bound of 2 units on the stock that can be stored, i.e. there is a constraint $x_k + u_k \leq 2$. The holding/storage cost for the kth period is given by

$$(x_k + u_k - w_k)^2,$$

implying a penalty both for excess inventory and for unmet demand at the end of the *kth* period. The ordering cost is 1 per unit stock ordered. Thus the cost per period is

$$g_k(x_k, u_k, w_k) = u_k + (x_k + u_k - w_k)^2.$$



The terminal cost is assumed to be 0,

$$g_N(x_N) = 0.$$

The planning horizon $N$ is 3 periods, and the initial stock $x_0$ is 0. The demand $w_k$ has the same probability distribution for all periods, given by

$$p(w_k = 0) = 0.1, \qquad p(w_k = 1) = 0.7, \qquad p(w_k = 2) = 0.2.$$

The system can also be represented in terms of the transition probabilities $p_{ij}(u)$ between the three possible states, for the different values of the control (see Fig. 1.3.3).

The starting equation for the DP algorithm is

$$J_3(x_3) = 0,$$

since the terminal state cost is 0 [ef. Eq. (1.5)]. The algorithm takes the form [cf. Eq. (1.6)]

$$J_k(x_k) = \min_{\substack{0\leq u_k\leq 2-x_k \\ u_k=0,1,2}} E_{w_k}\left\{u_k + (x_k + u_k - w_k)^2 + J_{k+1}(\max(0, x_k + u_k - w_k))\right\},$$

where $k = 0,1,2$, and $x_k, u_k, w_k$ can take the values 0,1, and 2.

Period 2: We compute $J_2(x_2)$ for each of the three possible states. We have

$$J_2(0) = \min_{u_2=0,1,2} E_{w_2}\left\{u_2 + (u_2 - w_2)^2\right\}$$
$$= \min_{u_2=0,1,2}\left[u_2 + 0.1(u_2)^2 + 0.7(u_2 - 1)^2 + 0.2(u_2 - 2)^2\right].$$

We calculate the expectation of the right side for each of the three possible values of $u_2$:

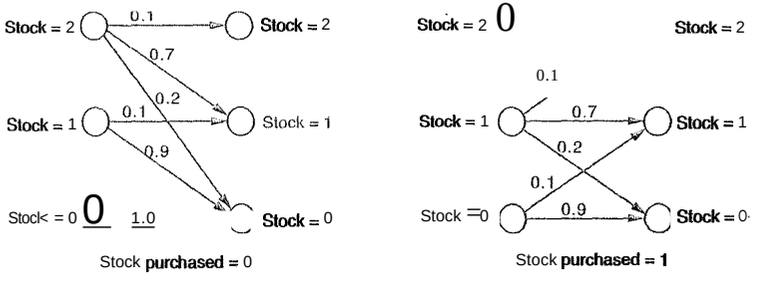
$$u_2 = 0 : E\{\cdot\} = 0.7\cdot 1 + 0.2\cdot 4 = 1.5,$$
$$u_2 = 1 : E\{\cdot\} = 1 + 0.1\cdot 1 + 0.2\cdot 1 = 1.3,$$
$$u_2 = 2 : E\{\cdot\} = 2 + 0.1\cdot 4 + 0.7\cdot 1 = 3.1.$$

Hence we have, by selecting the minimizing $u_2$,

$$J_2(0) = 1.3, \qquad \mu_2^*(0) = 1.$$

For $x_2 = 1$, we have

$$J_2(1) = \min_{u2=0,1} E_{w_2}\{u_2 + (1 + u_2 - w_2)^2\}$$
$$= \min_{u2=0,1}[u_2 + 0.1(1 + u_2)^2 + 0.7(u_2)^2 + 0.2(u_2 - 1)^2].$$




| Stock | Stage 0 Cost-to-go | Stage 0 Optimal stock to purchase | Stage 1 Cost-to-go | Stage 1 Optimal stock to purchase | Stage 2 Cost-to-go | Stage 2 Optimal stock to purchase |
|---|---|---|---|---|---|---|
| 0 | 3.7 | 1 | 2.5 | 1 | 1.3 | 1 |
| I | 2.7 | 0 | 1.5 | 0 | 0.3 | 0 |
| 2 | 2.818 | 0 | 1.68 | 0 | 1.1 | 0 |

Figure 1.3.3 System and DP results for Example 1.3.2. The transition probability diagrams for the different values of stock purchased (control) are shown. The numbers next to the arcs are the transition probabilities. The control $u = 1$ is not available at state 2 because of the limitation $x_k - u_k \leq 2$. Similarly, the control $u = 2$ is not available at states 1 and 2. The results of the DP algorithm are given in the table.

The expected value in the right side is

$$u_2 = 0\text{: } E\{\cdot\} = 0.1 \cdot 1 + 0.2 \cdot 1 = 0.3,$$
$$u_2 = 1 : E\{\cdot\} = 1 + 0.1 \cdot 4 + 0.7 \cdot 1 = 2.1.$$

Hence

$$J_2(1) = 0.3, \qquad \mu_2^*(1) = 0.$$



For $x_2$ 2, the only admissible control is $u_2 = 0$, so we have

$$J_2(2) = \underset{w_2}{E}\{(2 - w_2)^2\} = 0.1 \cdot 4 + 0.7 \cdot 1 = 1.1,$$

$$J_2(2) = 1.1, \qquad \mu_2^*(2) = 0.$$

**Period 1:** Again we compute $J_1(x_1)$ for each of the three possible states $x_1 = 0,1,2$, using the values $J_2(0)$, $J_2(1)$, $J_2(2)$ obtained in the previous period. For $x_1$ 0, we have

$$J_1(0) = \min_{u_1=0,1,2} \underset{w_1}{E}\left\{u_1 + (u_1 - w_1)^2 + J_2\big(\max(0, u_1 - w_1)\big)\right\},$$

$u_1 = 0 : E\{\cdot\} = 0.1 \cdot J_2(0) + 0.7(1 + J_2(0)) + 0.2(4 + J_2(0))$ 2.8,
$u_1 = 1 : E\{\cdot\} = 1 + 0.1(1 + J_2(1)) + 0.7 \cdot J_2(0) + 0.2(1 + J_2(0)) = 2.5,$
$u_1 = 2\text{: } E\{\cdot\} = 2 + 0.1(4 + J_2(2)) + 0.7(1 + J_2(1)) + 0.2 \cdot J_2(0) = 3.68,$

$$J_1(0) = 2.5, \qquad \mu_1^*(0) = 1.$$

For $x_1 = 1$, we have

$$J_1(1) = \min_{u_1=0,1} \underset{w_1}{E}\left\{u_1 + (1 + u_1 - w_1)^2 + J_2\big(\max(0, 1 + u_1 - w_1)\big)\right\},$$

$u_1 = 0\text{: } E\{\cdot\} = 0.1(1 + J_2(1)) + 0.7 \cdot J_2(0) + 0.2(1 + J_2(0)) = 1.5,$
$u_1$ 1: $E\{\cdot\} = 1 + 0.1(4 + J_2(2)) + 0.7(1 + J_2(1)) + 0.2 \cdot J_2(0)$ 2.68,

$$J_1(1) \quad 1.5, \qquad \mu_1^*(1) = 0.$$

For $x_1 = 2$, the only admissible control is $u_1 = 0$, so we have

$$J_1(2) = \underset{w_1}{E}\left\{(2 - w_1)^2 + J_2(\max(0, 2 \quad w_1))\right\}$$
$$= 0.1(4 + J_2(2)) + 0.7(1 + J_2(1)) + 0.2 \cdot J_2(0)$$
$$= 1.68,$$

$$J_1(2) \quad 1.68, \qquad \mu_1^*(2) \quad 0.$$

**Period 0:** Here we need to compute only Jo(O) since the initial state is known to be 0. We have

$$J_0(0) = \min_{u0=0,1,2} \underset{w_0}{E}\{u_0 + (u_0 \quad w_0)^2 + J_1\big(\max(0, u_0 \quad w_0)\big)\},$$

$u_0$ 0 : $E\{\cdot\} = 0.1 \cdot J_1(0) + 0.7(1 + J_1(0)) + 0.2(4 + J_1(0)) = 4.0,$
$u_0 = 1\text{: } E\{\cdot\} = 1 + 0.1(1 + J_1(1)) + 0.7 \cdot J_1(0) + 0.2(1 + J_1(0)) = 3.7,$
$u_0 = 2\text{: } E\{\cdot\} = 2 + 0.1(4 + J_1(2)) + 0.7(1 + J_1(1)) + 0.2 \cdot J_1(0) = 4.818,$

$$Jo(O) = 3.7, \qquad \mu_0^*(0) = 1.$$

If the initial state were not known a priori, we would have to compute in a similar manner $J_0(1)$ and $J_0(2)$, as well as the minimizing $u_0$. The reader may verify (Exercise 1.2) that these calculations yield

$$Jo(1) = 2.7, \qquad \mu_0^*(1) \quad 0,$$

$$Jo(2) = 2.818, \qquad \mu_0^*(2) = 0.$$

Thus the optimal ordering policy for each period is to order one unit if the current stock is zero and order nothing otherwise. The results of the DP algorithm are given in tabular form in Fig. 1.3.3.

### Example 1.3.3 (Optimizing a Chess Match Strategy)

Consider the chess match example of Section 1.1. There, a player can select timid play (probabilities $Pd$ and $1 - Pd$ for a draw or loss, respectively) or bold play (probabilities $p_w$ and $1 - p_w$ for a win or loss, respectively) in each game of the match. We want to formulate a DP algorithm for finding the policy that maximizes the player's probability of winning the match. Note that here we are dealing with a maximization problem. We can convert the problem to a minimization problem by changing the sign of the cost function, but a simpler alternative, which we will generally adopt, is to replace the minimization in the DP algorithm with maximization.

Let us consider the general case of an N-game match, and let the state be the *net score*, that is, the difference between the points of the player minus the points of the opponent (so a state of 0 corresponds to an even score). The optimal cost-to-go function at the start of the kth game is given by the dynamic programming recursion

$$J_k(x_k) = \max [PdJk+1(Xk) + (1 - Pd)Jk+1(Xk - 1), \; PwJk+1(Xk + 1) + (1 - p_w)J_{k+1}(x_k - 1)]. \tag{1.8}$$

The maximum above is taken over the two possible decisions:

(a) Timid play, which keeps the score at $Xk$ with probability $Pd$, and changes $x_k$ to $x_k$ 1 with probability 1 $Pd$.

(b) Bold play, which changes $Xk$ to $Xk + 1$ or to $Xk$ - 1 with probabilities $p_w$ or $(1 - Pw)$, respectively.

It is optimal to play bold when

$$p_w J_{k+1}(x_k + 1) + (1 - p_w)J_{k+1}(x_k - 1) \geq p_d J_{k+1}(x_k) + (1 - p_d)J_{k+1}(x_k - 1)$$

or equivalently, if

$$\frac{Pw}{Pd} > \frac{Jk+1(Xk) - Jk+1(Xk - 1)}{Jk+1(Xk + 1) \quad Jk+1(Xk - 1)}. \tag{1.9}$$



The dynamic programming recursion is started with

$$J_N(x_N) = \begin{cases} 1 & \text{if } xN > 0, \\ p_w & \text{if } xN = 0, \\ 0 & \text{if } xN < 0. \end{cases} \tag{1.10}$$

In this equation, we have $IN(O) = p_w$ because when the score is even after $N$ games ($xN = 0$), it is optimal to play bold in the first game of sudden death.

By executing the DP algorithm (1.8) starting with the terminal condition (1.10), and using the criterion (1.9) for optimality of bold play, we find the following, assuming that $Pd > p_w$:

$$J_{N-1}(x_{N-1}) = 1 \text{ for } x_{N-1} > 1; \quad \text{optirnal play: either}$$

$$IN\text{-}1(1) = \max[\text{pd} + (1 - Pd)Pw, \; p_w + (1 \quad p_w)p_w]$$

$$Pd + (1 - p_d)p_w; \quad \text{optimal play: timid}$$

$$IN\text{-}1(0) = p_w; \quad \text{optimal play: bold}$$

$$J_{N-1}(-1) = p_w^2; \quad \text{optimal play: bold}$$

$$IN\text{-}l(XN\text{-}1) = 0 \text{ for } XN\text{-}1 < -1; \quad \text{optimal play: either.}$$

Also, given $IN\text{-}l(XN\text{-}1)$, and Eqs. (1.8) and (1.9) we obtain

$$IN\text{-}2(0) = \max [PdPW + (1 - p_d)p_w^2, \; pw\left(p_d + (1 - Pd)Pw\right) + (1 - p_w)p_w^2\Big]$$

$$= pw\,(Pw + (p_w + Pd)(l \quad Pw))$$

and that if the score is even with 2 games remaining, it is optirnal to play bold. Thus for a 2-game match, the optimal policy for both periods is to play timid if and only if the player is ahead in the score. The region of pairs $(Pw,Pd)$ for which the player has a better than 50-50 chance to win a 2-game match is

$$R_2 = \left\{(p_w, p_d) \mid J_0(0) = p_w\left(p_w + (p_w + p_d)(1 - p_w)\right) > 1/2\right\},$$

and, as noted in the preceding section, it includes points where $pw < 1/2$.

### Example 1.3.4 (Finite-State Systems)

We mentioned earlier (d. the examples in Section 1.1) that systems with a finite number of states can be represented either in terms of a discrete-time system equation or in terms of the probabilities of transition between the states. Let us work out the DP algorithm corresponding to the latter caSe. We assume for the sake of the following discussion that the problem is stationary (i.e., the transition probabilities, the cost per stage, and the control constraint sets do not change from one stage to the next). Then, if

$$p_{ij}(u) = P\{x_{k+1} = j \mid x_k = i, u_k$$

(a) Show that the problem can be formulated as a shortest path problem, and write the corresponding DP algorithm.

(b) Suppose he is at location i on day $k$. Let

$$R_k^i = r_k^{\bar{i}} - r_k^i,$$

where $\bar{i}$ denotes the location that is not equal to i. Show that if $R_k^i \leq 0$ it is optimal to stay at location i, while if $R_k^i \geq 2c$, it is optimal to switch.

(c) Suppose that on each day there is a probability of rain $P_i$ at location $i$ independently of rain in the other location, and independently of whether it rained on other days. If he is at location i and it rains, his profit for the day is reduced by a factor $\beta_i$. Can the problem still be formulated as a shortest path problem? Write a DP algorithm.

(d) Suppose there is a possibility of rain as in part (c), but the businessman receives an accurate rain forecast just before making the decision to switch or not switch locations. Can the problem still be formulated as a shortest path problem? Write a DP algorithm.

# 3

# eterministic Continuous-Time Optimal ntrol

Contents

In this chapter, we provide an introduction to continuous-time deterministic optimal control. We derive the analog of the DP algorithm, which is the Hamilton-Jacobi-Bellman equation. Furthermore, we develop a celebrated theorem of optimal control, the Pontryagin Minimum Principle and its variations. We discuss two different derivations of this theorem, one of which is based on DP. We also illustrate the theorem by means of examples.

## 3.1 CONTINUOU8-TIl\!IE OPTIMAL CONTROL

We consider a continuous-time dynamic system

$$\dot{x}(t) = f(x(t),u(t)), \qquad 0 \le t \le T, \tag{3.1}$$

$$x(0) : \text{given},$$

where $x(t) \in \Re^n$ is the state vector at time t, $\dot{x}(t) \in \Re^n$ is the vector of first order time derivatives of the states at time $t$, $u(t) \in U \subset \Re^m$ is the control vector at time $t$, $U$ is the control constraint set, and $T$ is the terminal time. The components of $f$, $x$, $\dot{x}$, and $u$ will be denoted by $fi$, $Xi$, $\dot{x}_i$, and $Ui$, respectively. Thus, the system (3.1) represents the $n$ first order differential equations

$$\frac{dXi(t)}{dt} = f_i\big(x(t),u(t)\big), \qquad i = 1, \ldots, no$$

We view $\dot{x}(t)$, $x(t)$, and $u(t)$ as column vectors. We assume that the system function $Ii$ is continuously differentiable with respect to $x$ and is continuous with respect to $u$. The admissible control functions, also called *control trajectories*, are the piecewise continuous functions $\{u(t) \mid t \in [0, Tn$ with $u(t) \in U$ for all $t \in [0, T)$.

We should stress at the outset that the subject of this chapter is highly sophisticated, and it is beyond our scope to develop it according to high standards of mathematical rigor. In particular, we assume that, for any admissible control trajectory $\{u(t) \mid t \in [0, Tn$, the system of differential equations (3.1) has a unique solution, which is denoted $\{xu(t) \mid t \in [0, T)\}$ and is called the corresponding *state tmjectory*. In a more rigorous treatment, the issue of existence and uniqueness of this solution would have to be addressed more carefully.

We want to find an admissible control trajectory $\{u(t) \mid t \in [0, Tn$, which, together with its corresponding state trajectory $\{x(t) \mid t \in [0, Tn$, minimizes a cost function of the form

$$h(x(T)) + \int_0^T g(x(t),1t(t))dt,$$

where the functions $9$ and $h$ are continuously differentiable with respect to $x$, and $9$ is continuous with respect to $u$.



### Example 3.1.1 (Motion Control)

A unit mass moves on a line under the influence of a force $u$. Let $x1(t)$ and $X2(t)$ be the position and velocity of the mass at time $t$, respectively. From a given $(x1(0), X2(0))$ we want to bring the mass "near" a given final position-velocity pair $(\overline{x}_1, \overline{x}_2)$ at time $T$. In particular, we want to

$$\text{minimize } \big|x_1(T) - \overline{x}_1\big|^2 + \big|x_2(T) - \overline{x}_2\big|^2$$

subject to the control constraint

$$|u(t)| \le 1, \qquad \text{for allIt} \in [0, T].$$

The corresponding continuous-time system is

$$\dot{x}_1(t) = X2(t), \qquad \dot{x}_2(t) = u(t),$$

and the problem fits the general framework given earlier with cost functions given by

$$h(x(T)) = \big|x_1(T) - \overline{x}_1\big|^2 + IX2(T) - \overline{x}_2\big|^2,$$

$$g(x(t),u(t)) = 0, \qquad \text{for all } t \in [0,1'].$$

There are many variations of the problem; for example, the final position and/or velocity may be fixed. These variations can be handled by various reformulations of the general continuous-time optimal control problem, which will be given later.

### Example 3.1.2 (Resource Allocation)

A producer with production rate $x(t)$ at time $t$ may allocate a portion $u(t)$ of his/her production rate to reinvestment and $1 - u(t)$ to production of a storable good. Thus $x(t)$ evolves according to

$$\dot{x}(t) = \gamma u(t)x(t),$$

where $\gamma > 0$ is a given constant. The producer wants to maximize the total amount of product stored

$$\int_0^T (1 - n(t))x(t)dl$$

subject to

$$0 \le u(t) \le 1, \qquad \text{for all } t \in [0, T].$$

The initial production rate $x(O)$ is a given positivenurnber.

**Example 3.1.3 (Calculus of Variations Problems)**

Calculus of variations problems involve finding (possibly multidimensional) curves $x(t)$ with certain optimality properties. They are among the most celebrated problems of applied mathematics and have been worked on by many of the illustrious mathematicians of the past 300 years (Euler, Lagrange, Bernoulli, Gauss, etc.). We will see that calculus of variations problems can be reformulated as optimal control problems. We illustrate this reformulation by a simple example.

Suppose that we want to find a minimum length curve that starts at a given point and ends at a given line. The answer is of course evident, but we want to derive it by using a continuous-time optimal control formulation. Without loss of generality, we let $(0, \alpha)$ be the given point, and we let the given line be the vertical line that passes through $(T, 0)$, as shown in Fig. 3.1.1. Let also $(t, x(t))$ be the points of the curve $(0 \leq t \leq T)$. The portion of the curve joining the points $(t, x(t))$ and $(t + dt, x(t + dt))$ can be approximated, for small $dt$, by the hypotenuse of a right triangle with sides $dt$ and $x(t)dt$. Thus the length of this portion is

$$\sqrt{(dt)^2 + \big(\dot{x}(t)\big)^2 (dt)^2},$$

which is equal to

$$\sqrt{1 + (X(t))2}\ dt.$$

T'he length of the entire curve is the integral over $[0, T]$ of this expression, so the problem is to

$$\text{minimize} \quad \int_0^T \sqrt{1 + \big(\dot{x}(t)\big)^2}\ dl$$

$$\text{subject to} \quad x(O) = \alpha.$$

To refonnulate the problem as a continuous-time optimal control problem, we introduce a control $u$ and the system equation

$$\dot{x}(t) = u(t), \qquad x(O) = \alpha.$$

Our problem then becomes

$$\text{minimize} \quad \int_0^T \sqrt{1 + ("(i))2}\ dl,$$

This is a problem that fits our continuous-time optimal control framework.



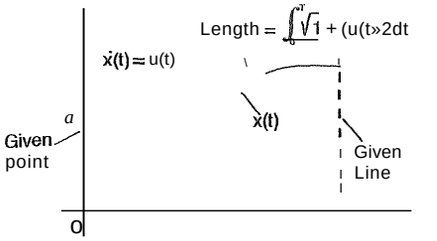



Figure 3.L1 Problem of finding a curve of minimum length from a given point to a given line, and its formulation as a calculus of variations problem.

## 3.2 THE HAMILTON-JACOBI-BELLMAN EQUATION

We will now derive informally a partial differential equation, which is satisfied by the optimal cost-to-go function, under certain assumptions. This equation is the continuous-time analog of the DP algorithm, and will be motivated by applying DP to a discrete-time approximation of the continuous-time optimal control problem.

Let us divide the time horizon $[0, T]$ into $N$ pieces using the discretization interval

$$\delta = \frac{T}{N'}$$

We denote

$$x_k \quad x(k\delta), \qquad k = 0,1, \quad ,N,$$

$$u_k = u(k\delta), \quad '. k = 0, 1, \quad ,N,$$

and we approximate the continuous-time system by

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

and the cost function by

$$h(x_N) + \sum_{k=()}^{N-J} g(x_k, u_k) \cdot \delta.$$

We now apply DP to the discrete-time approximation. Let

$J^*(t, x)$ : Optima.l cost-ta-go at time $t$ and state $x$ for the continuous-time problem,

$\tilde{J}^*(t, x)$ : Optimal cost-to-go at time $t$ and state $x$ for the discrete-time approximation.

The DP equations are

$$J^*(No,x) = h(x),$$

$$\tilde{J}^*(k\delta,x) = \min_{u\in U}\left[g(x,u)\cdot\delta+\tilde{J}^*\big((k+1)\cdot\delta, x+j(x,u)\cdot\delta\big)\right], \quad k = 0,\dots,N-1.$$

Assuming that $\tilde{J}^*$ has the required differentiability properties, we expand it into a first order Taylor series around $(k\delta, x)$, obtaining

$$\tilde{J}^*\big((k+1)\cdot\delta, x+f(x,u)\cdot\delta\big) = \tilde{J}^*(k\delta,x) + \nabla_t\tilde{J}^*(k\delta,x)\cdot\delta + \nabla_x\tilde{J}^*(k\delta,x)'j(x,u)\cdot\delta + o(\delta),$$

where $o(\delta)$ represents second order terms satisfying $\lim_{\delta\to 0} o(\delta)/\delta = 0$, $\nabla_t$ denotes partial derivative with respect to $t$, and $\nabla_x$ denotes the $n$-dilnensional (column) vector of partial derivatives with respect to $x$. Substituting in the DP equation, we obtain

$$\tilde{J}^*(k\delta,x) = \min_{u\in U}[g(x,u)\cdot\delta + \tilde{J}^*(k\delta,x) + \nabla_t\tilde{J}^*(k\delta,x)\cdot\delta + \nabla_x\tilde{J}^*(k\delta,x)'j(x,u)\cdot\delta + o(\delta)].$$

Canceling $\tilde{J}^*(k\delta,x)$ from both sides, dividing by $\delta$, and taking the limit as $\delta\to 0$, while assuming that the discrete-time cost-to-go function yields in the limit its continuous-time counterpart,

$$\lim_{k\to\infty,\,\delta\to 0,\,k\delta=t}\tilde{J}^*(k\delta,x) = J^*(t,x), \qquad \text{for all } t, x,$$

we obtain the following equation for the cost-to-go function $J^*(t,x)$:

$$0 = \min_{u\in U}[g(x,u) + \nabla_t J^*(t,x) + \nabla_x J^*(t,x)'j(x,u)], \qquad \text{for all } t, x,$$

with the boundary condition $J^*(T,x) = h(x)$.

This is the *Hamilton-Jacobi-Bellman (HJB) equation.* It is a *partial* differential equation, which should be satisfied for all time-state pairs $(t,x)$ by the cost-to-go function $J^*(t,x)$, based on the preceding informal derivation, which assumed among other things, differentiability of $J^*(t,x)$. In fact we do not know a priori that $J^*(t,x)$ is differentiable, so we do not know if $J^*(t,x)$ solves this equation. However, it turns out that if we can solve the HJB equation analytically or computationally, then we can obtain an optimal control policy by minimizing its right-hand-side. This is shown in the following proposition, whose statement is reminiscent of a corresponding statement for discrete-time DP: if we can execute the DP algorithm, which may not be possible due to excessive computational requirements, we can find an optimal policy by minimization of the right-hand side.



**Proposition 3.2.1:** (Sufficiency Theorem) Suppose $V(t,x)$ is a solution to the HJB equation; that is, $V$ is continuously differentiable in t and $x$, and is such that

$$0 = \min_{u\in U}[g(x,u) + \nabla_t V(t,x) + \nabla_x V(t,x)'f(x,u)], \qquad \text{for all } t, x, \tag{3.2}$$

$$V(T,x) = h(x), \qquad \text{for all } x. \tag{3.3}$$

Suppose also that $\mu^*(t,x)$ attains the minimum in Eq. (3.2) for all $t$ and $x$. Let $\{x^*(t) \mid t \in [0,T]\}$ be the state trajectory obtained from the given initial condition $x(0)$ when the control trajectory $u^*(t) = \mu^*(t,x^*(t))$, $t \in [0,T]$ is used [that is, $x^*(0) = x(0)$ and for all $t \in [0,T]$, $\dot{x}^*(t) = f(x^*(t),\mu^*(t,x^*(t)))$; we assume that this differential equation has a unique solution starting at any pair $(t,x)$ and that the control trajectory $\{\mu^*(t,x^*(t)) \mid t \in [0,T]\}$ is piecewise continuous as a function of $t$]. Then $V$ is equal to the optimal cost-to-go function, I.e.,

$$V(t,x) = J^*(t,x), \qquad \text{for all } t, x.$$

Furthermore, the control trajectory $\{u^*(t) \mid t \in [0,T]\}$ is optimal.

**Proof:** Let $\{\hat{u}(t) \mid t \in [0,T]\}$ be any admissible control trajectory and let $\{\hat{x}(t) \mid t \in [0,T]\}$ be the corresponding state trajectory. From Eq. (3.2) we have for all $t \in [0,T]$

$$0 \le g\big(\hat{x}(t),\hat{u}(t)\big) + \nabla_t V\big(t,\hat{x}(t)\big) + \nabla_x V\big(t,\hat{x}(t)\big)'f\big(\hat{x}(t),\hat{u}(t)\big).$$

Using the system equation $\dot{\hat{x}}(t) = f(\hat{x}(t),\hat{u}(t))$, the right-hand side of the above inequality is equal to the expression

$$g\big(\hat{x}(t),\hat{u}(t)\big) + \frac{d}{dt}\big(V(t,\hat{x}(t))\big),$$

where $d/dt(\cdot)$ denotes total derivative with respect to $t$. Integrating this expression over $t \in [0,T]$, and using the preceding inequality, we obtain

$$0 \le \int_0^T g\big(\hat{x}(t),\hat{u}(t)\big)dt + V(T,\hat{x}(T)) - V(0,x(0)).$$

Thus by using the terminal condition $V(T,x) = h(x)$ of Eq. (3.3) and the initial condition $\hat{x}(0) = x(0)$, we have

$$V(0,x(0)) \le h\big(\hat{x}(T)\big) + \int_0^T g\big(\hat{x}(t),\hat{u}(t)\big)dt.$$

If we use $u^*(t)$ and $x^*(t)$ in place of $\hat{u}(t)$ and $\hat{x}(t)$, respectively, the preceding inequalities becomes equalities, and we obtain

$$V(0, x(O)) = h(x'(T)) + \int_0^T g(x''(tj, u^*(t))\, dt.$$

Therefore the cost corresponding to $\{u^*(t)$ $It$ E $[O,T]\}$ is $V(O,x(O))$ and is no larger than the cost corresponding to any other admissible.control trajectory $\{\hat{u}(t)$ $It$ E $[0, Tn\cdot$ It follows that $\{u^*(t)$ $It$ E $[0, T]\}$ is optimal and that

$$V(O,x(O)) = J^*(O,x(O)).$$

We now note that the preceding argument can be repeated with any initial time $t$ E $[0, TJ$ and any initial state $x$. We thus obtain

$$V(t,x) = J^*(t,x), \qquad \text{for all } t, x.$$

Q.E.D.

### Example 3.2.1

To illustrate the HJB equation, let us consider a simple example involving the scalar system

$$x(t) = u(t),$$

with the constraint $|u(t)| \leq 1$ for all $t$ E [0, TJ. The cost is

$$\frac{1}{2}\big(x(T)\big)^2.$$

The HJB equation here is

$$0 = \min_{|u|\leq 1}\big[\nabla_t V(t, x) + \nabla_x V(t,x)u\big], \qquad \text{for all } t, x, \tag{3.4}$$

with the terminal condition

$$V(T,x) = \frac{1}{2}x^2. \tag{3.5}$$

There is an evident candidate for optimality, namely moving the state towards 0 as quickly as possible, and keeping it at 0 once it is at 0. The corresponding control policy is

$$\mu^*(t,x) = -\text{sgn}(x) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x > 0. \end{cases} \tag{3.6}$$

For a given initial time $t$ and initial state $x$, the cost associated with this policy can be calculated to be

$$J^*(t, x) = \frac{1}{2}\Big(\max\{0, |x| - (T-t)\}\Big)^2. \tag{3.7}$$



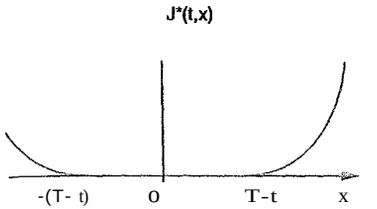



Figure 3.2.1 Optimal cost-to-go function $J^*(t,x)$ for Example 3.2.1.

This function, which is illustrated in Fig. 3.2.1, satisfies the terrninal condition (3.5), since $J^*(T,x) = (1/2)x^2$. Let us verify that this function also satisfies the HJB Eq. (3.4), and that $u$ = -sgn(x) attains the minimum in the right-hand side ofthe equation for all $t$ and $x$. Proposition 3.2.1 will then guarantee that the state and control trajectories corresponding to the policy $\mu^*(t,x)$ are optimal.

Indeed, we have

$$\nabla_t J^*(t,x) = \max\{0, |x| - (T-t)\},$$

$$\nabla_x J^*(t,x) = \text{sgn}(x)\cdot\max\{0,\ |x| - (T-i)\}.$$

Substituting these expressions, the HJB Eq. (3.4) becomes

$$0 = \min_{|u|\leq 1}[1 + \text{sgn}(x)\cdot u]\max\{0, |x| - (T-t)\},$$

which can be seen to hold as an identity for all $(i,x)$. Purthermore, the minimum is attained for $u = -\text{sgn}(x)$. We therefore conclude based on Prop. 3.2.1 that $J^*(t,x)$ as given by Eq. (3.7) is indeed the optimal cost-to-go function, and that the policy defined by Eq. (3.6) is optimal. Note, however, that the optimal policy is not unique. Based on Prop. 3.2.1, any policy for which the minimum is attained in Eq. (3.8) is optimal. In particular, when $IX(i)1 \leq T - i$, applying any control from the range [-1,1) is optimal.

The preceding derivation generalizes to the case of the cost

$$h(x(T»),$$

where $h$ is a nonnegative differentiable convex function with $h(O)$ 0. The corresponding optimal cost-to-go function is

$$J^*(t,x) = \begin{cases} fh(x-(T-t)) & \text{if } x > T - i, \\ h\big(x + (T-t)\big) & \text{if } x < -(T-i), \\ 0 & \text{if } |x| \leq T - t, \end{cases}$$

and can be similarly verified to be a solution of the HJB equation.

**Example 3.2.2 (Linear-Quadratic Problelns)**

Consider the n-dimensional linear system

$$\dot{x}(t) = Ax(t) + Bu(t),$$

where $A$ and $B$ are given matrices, and the quadratic cost

$$x(T)'Ql'x(T) + \int_0^T \big(x(t)'Qx(t) + u(t)'Ru(t)\big)dt,$$

where the matrices QT and $Q$ are symmetric positive semidefinite, and the rnatrix $R$ is symmetric positive definite (Appendix A defines positive definite and semidefinite matrices). The HJB equation is

$$0 = \min_{u\in\Re^m}\big[x'Qx + u'Ru + \nabla_t V(t,x) + \nabla_x V(t,x)'(Ax+Bu)\big], \qquad (3.9)$$

$$V(T,x) = x'Q_T x.$$

Let us try a solution of the form

$$Vet, x) = x_I K(t)x, \qquad K(t) : n \times n \text{ symmetric},$$

and see if we can solve the HJB equation. We have $\nabla_x V(t, x) = 2K(t)x$ and $\nabla_t V(t,x) = x_I \dot{K}(t)x$, where $\dot{K}(t)$ is the matrix with elements the first order derivatives of the elements of $K(t)$ with respect to time. By substituting these expressions in Eq. (3.9), we obtain

$$0 = \min_u [x_I Qx + u'Ru + x_I \dot{K}(t)x + 2x^I K(t)Ax + 2x^I K(t)Bu]. \qquad (3.10)$$

The minimum is attained at a $u$ for which the gradient with respect to $u$ is zero, that is,

$$2B^I K(t)x + 2Ru = 0$$

or

$$u = \_R\text{-}^1 B^I K(t)x. \qquad (3.11)$$

Substituting the minimizing value of $u$ in Eq. (3.10), we obtain

$$0 \quad x'\big(\dot{K}(t) + K(t)A + AIK(t) \text{--} K(t)BR\text{-}^1 B^I K(t) + Q\big)x, \quad \text{for all } (t, x).$$

Therefore, in order for $Vet, x) = x_I K(t)x$ to solve the HJB equation, $K(t)$ must satisfy the following matrix differential equation (known as the *continuous-time Riccati equation)*

$$\dot{K}(t) = \text{-}K(t)A \quad AIK(t) + K(t)BR\text{-}^1 B^I K(t) \quad Q \qquad (3.12)$$

with the terminal condition

$$K(T) = QT. \qquad (3.13)$$

Reversing the argument, we see that if $K(t)$ is a solution of the Riccati equation (3.12) with the boundary condition (3.13), then $Vet, x) = x_I K(t)x$ is a solution of the HJB equation. Thus, by using Prop. 3.2.1, we conclude that the optimal cost-to-go function is

$$J^*(t,x) = xIK(t)x.$$

F'urthermore, in view of the expression derived for the control that minimizes in the right-hand side of the HJB equation [ef. Eq. (3.11)], an optimal policy is

$$\mu^*(t,x) = -R^{-1}B'K(t)x.$$



## 3.3 THE PONTRYAGIN MINIMUM PRINCIPLE

In this section we discuss the continuous-time and the discrete-time versions of the Minimum Principle, starting with a DP-based informal argument.

### 3.3.1 An Informal Derivation Using the HJB Equation

Recall the HJB equation

$$0 = \min_{uEU}\big[g(x,u) + \nabla_t J^*(t,x) + \nabla_x J^*(t,x)'f(x,u)\big], \quad \text{for all } t,x, \qquad (3.14)$$

$$J^*(T,x) = hex), \qquad \text{for all } x. \qquad (3.15)$$

We argued that the optimal cost-to-go function $J^*(t, x)$ satisfies this equation under some conditions. Furthermore, the sufficiency theorem of the preceding section suggests that if for a given initial state $x(0)$, the control trajectory $\{u^*(t) \text{ It } \in [0, Tn$ is optimal with corresponding state trajectory $\{x^*(t) \text{ It } \in [O,T]\}$, then for all $t \in [O,T]$,

$$u^*(t) \quad \operatorname{argmin}_{uEU}\Big[g\big(x^*(t),u\big) + \nabla_x J^*\big(t,x^*(t)\big)'f\big(x^*(t),u\big)\Big].$$

Note that to obtain the optimal control trajectory via this equation, we do not need to know $\nabla xJ^*$ at *all* values of $x$ and $t$; it is sufficient to know $\nabla xJ^*$ at only *one* value of $x$ for each $t$, that is, to know only $\nabla xJ^*(t, x^*(t))$.

The Minimum Principle is basically the preceding Eq. (3.16). Its application is facilitated by streamlining the computation of $\nabla_x J^*(t, x^*(t))$. It turns out that we can often calculate $\nabla_x J^*\big(t, x^*(t)\big)$ along the optimal state trajectory far more easily than we can solve the HJB equation. In particular, $\nabla xJ^*(t, x^*(t))$ satisfies a certain differential equation, called the *adjo'int equation.* We will derive this equation informally by differentiating the HJB equation (3.14). We first need the following lemma, which indicates how to differentiate functions involving minima.

Lernma 3.3.1: Let $F(t, x, u)$ be a continuously differentiable function of $t \in \Re$, $x \in \Re^n$, and $u \in \Re^m$, and let $U$ be a convex subset of $\Re^m$. Assume that $JL^*(t, x)$ is a continuously differentiable function such that

$$J1^*(t, x) = \arg\min_{uEU} F(t, x, u), \qquad \text{for all } t, x.$$

Then

$$\nabla_t \left\{ \min_{u \in U} F(t,x,u) \right\} = \nabla_t F(t,x,\mu^*(t,x)), \qquad \text{for all } t, x,$$

$$\nabla_x \left\{ \min_{u \in U} F(t,x,u) \right\} = \nabla_x F(t,x,\mu^*(t,x)), \qquad \text{for all } t, x.$$

[Note: On the left-hand side, $\nabla_t\{\cdot\}$ and $\nabla_x\{\cdot\}$ denote the gradients of the function $G(t,x) = \min_{u \in U} F(t,x,u)$ with respect to $t$ and $x$, respectively. On the right-hand side, $\nabla_t$ and $\nabla_x$ denote the vectors of partial derivatives of $F$ with respect to $t$ and $x$, respectively, evaluated at $(t, x, \mu^*(t,x))$.]

Proof: For notational simplicity, denote $y = (t,x)$, $F(y,u) = F(t,x,u)$, and $\mu^*(y) = \mu^*(t,x)$. Since $\min_{u \in U} F(y,u) = F(y,L^*(y))$,

$$\nabla \{\min_{u \in U} F(y,u)\} = \nabla_y F(y,\mu^*(y)) + \nabla \mu^*(y) \nabla_u F(y,\mu^*(y)).$$

We will prove the result by showing that the second term in the right-hand side above is zero. This is true when $U = \Re^m$, because then $\mu^*(y)$ is an unconstrained minimum of $F(y,u)$ and $\nabla_u F(y,\mu^*(y)) = 0$. More generally, for every fixed $y$, we have

$$(u - \mu^*(y))' \nabla_u F(y,\mu^*(y)) \geq 0, \qquad \text{for all } u \in U,$$

[see Eq. (B.2) in Appendix B]. Now by Taylor's Theorem, we have that when $y$ changes to $y + \Delta y$, the minimizing $\mu^*(y)$ changes from $\mu^*(y)$ to some vector $\mu^*(y + \Delta y) = \mu^*(y) + \nabla \mu^*(y)' \Delta y + o(\|\Delta y\|)$ of $U$, so

$$(\nabla \mu^*(y)' \Delta y + o(\|\Delta y\|))' \nabla_u F(y,\mu^*(y)) \geq 0, \qquad \text{for all } \Delta y,$$

implying that

$$\nabla \mu^*(y) \nabla_u F(y,\mu^*(y)) = 0.$$

Q.E.D.

Consider the HJB equation (3.14), and for any $(t,x)$, suppose that $\mu^*(t,x)$ is a control attaining the minimum in the right-hand side. We make the restrictive assumptions that $U$ is a convex set, and that $\mu^*(t,x)$ is continuously differentiable in $(t,x)$, so that we can use Lemma 3.3.1. (We note, however, that alternative derivations of the Minimum Principle do not require these assumptions; see Section 3.3.2.)



We differentiate both sides of the fIJB equation with respect to $x$ and with respect to $t$. In particular, we set to zero the gradient with respect to $x$ and t of the function

$$g(x, \mu^*(t,x)) + \nabla_t J^*(t,x) + \nabla_x J^*(t,x)' f(x, \mu^*(t,x)),$$

and we rely on Lemma 3.3.1 to disregard the terms involving the derivatives of $\mu^*(t,x)$ with respect to t and $x$. We obtain for all $(t,x)$,

$$0 = \nabla_x g(x, \mu^*(t,x)) + \nabla^2_{xt} J^*(t,x) + \nabla^2_{xx} J^*(t,x) f(x,\mu^*(t,x)) + \nabla_x f(x, \mu^*(t,x)) \nabla_x J^*(t,x), \tag{3.17}$$

$$0 = \nabla^2_{tt} J^*(t,x) + \nabla^2_{xt} J^*(t,x)' f(x,\mu^*(t,x)), \tag{3.18}$$

where $\nabla_x f(x, \mu^*(t,x))$ is the matrix

$$\nabla_x f = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots \\ \vdots & \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_n}{\partial x_n} \end{pmatrix}$$

with the partial derivatives evaluated at the argument $(x, \mu^*(t,x))$.

The above equations hold for all $(t,x)$. Let us specialize them along an optimal state and control trajectory $\{(x^*(t), u^*(t)) \mid t \in [0,T]\}$, where $u^*(t) = \mu^*(t, x^*(t))$ for all $t \in [0,T]$. We have for all t,

$$\dot{x}^*(t) = f(x^*(t), u^*(t)),$$

so that the term

$$\nabla^2_{xt} J^*(t, x^*(t)) + \nabla^2_{xx} J^*(t, x^*(t)) f(x^*(t), u^*(t))$$

in Eq. (3.17) is equal to the following total derivative with respect to t

$$\frac{d}{dt} \left( \nabla_x J^*(t, x^*(t)) \right).$$

Similarly, the term

$$\nabla^2_{tt} J^*(t, x^*(t)) + \nabla^2_{xt} J^*(t, x^*(t))' f(x^*(t), u^*(t))$$

in Eq. (3.18) is equal to the total derivative

$$\frac{d}{dt} \left( \nabla_t J^*(t, x^*(t)) \right).$$

Thus, by denoting

$$p(t) = \nabla_x J^*(t, x^*(t)), \tag{3.19}$$

$$p_0(t) = \nabla_t J^*(t, x^*(t)), \tag{3.20}$$

Eq. (3.17) becomes

$$\dot{p}(t) = -\nabla_x f(x^*(t), u^*(t))p(t) - \nabla_x g\big(x^*(t), u^*(t)\big) \tag{3.21}$$

and Eq. (3.18) becomes

$$\dot{p}_0(t) = 0$$

or equivalently,

$$p_0(t) = \text{constant}, \qquad \text{for all } t \in [0, T]. \tag{3.22}$$

Equation (3.21) is a system of $n$ first order differential equations known as the *adjoint equation*. From the boundary condition

$$J^*(T,x) = h(x), \qquad \text{for all } x,$$

we have, by differentiation with respect to $x$, the relation $\nabla_x J^*(T,x) = \nabla h(x)$, and by using the definition $\nabla_x J^*(t,x^*(t)) = p(t)$, we obtain

$$p(T) = \nabla h(x^*(T)). \tag{3.23}$$

Thus, we have a terminal boundary condition for the adjoint equation (3.21).

To summarize, along optimal state and control trajectories $x^*(t)$, $u^*(t)$, $t \in [0, T]$, the adjoint equation (3.21) holds together with the boundary condition (3.23), while Eq. (3.16) and the definition of $p(t)$ imply that $u^*(t)$ satisfies

$$u^*(t) = \arg\min_{u \in U} \big[g(x^*(t), u) + p(t)' f(x^*(t), u)\big], \qquad \text{for all } t \in [0, T]. \tag{3.24}$$

### Harniltonian Formulation

Motivated by the condition (3.24), we introduce the Hamiltonian function mapping triplets $(x, u, p) \in \Re^n \times \Re^m \times \Re^n$ to real numbers and given by

$$H(x, u, p) = g(x,u) + p' f(x, u).$$

Note that both the system and the adjoint equations can be compactly written in terms of the Hamiltonian as

$$\dot{x}^*(t) \quad \nabla_p H\big(x^*(t), u^*(t), p(t)\big), \qquad \dot{p}(t) = -\nabla_x H(x^*(t), u^*(t), p(t)).$$

We state the Minimum Principle in terms of the Hamiltonian function.



**Proposition 3.3.1: (Minimum Principle)** Let $\{u^*(t) \mid t \in [0, T]\}$ be an optimal control trajectory and let $\{x^*(t) \mid t \in [0, T]\}$ be the corresponding state trajectory, i.e.,

$$\dot{x}^*(t) = f(x^*(t), u^*(t)), \qquad x^*(0) = x(0) : \text{given}.$$

Let also $p(t)$ be the solution of the adjoint equation

$$\dot{p}(t) = -\nabla_x H\big(x^*(t), u^*(t), p(t)\big),$$

with the boundary condition

$$p(T) = \nabla h(x^*(T)),$$

where $h(\cdot)$ is the terminal cost function. Then, for all $t \in [0, T]$,

$$u^*(t) = \arg\min_{u \in U} H(x^*(t), u, p(t)).$$

Furthermore, there is a constant 0 such that

$$H(x^*(t), u^*(t), p(t)) = 0, \qquad \text{for all } t \in [0, T].$$

All the assertions of the Minimum Principle have been (informally) derived earlier except for the last assertion. To see why the Hamiltonian function is constant for $t \in [0, T]$ along the optimal state and control trajectories, note that by Eqs. (3.14), (3.19), and (3.20), we have for all $t \in [0, T]$

$$H(x^*(t), u^*(t), p(t)) = -\nabla_t J^*(t, x^*(t)) = -p_0(t),$$

and $p_0(t)$ is constant by Eq. (3.22). We should note here that the Hamiltonian function need not be constant along the optimal trajectory if the system and cost are not time-independent, contrary to our assumption thus far (see Section 3.4.4).

It is important to note that the Minimum Principle provides *necessary* optimality conditions, so all optimal control trajectories satisfy these conditions, but if a control trajectory satisfies these conditions, it is not necessarily optimal. Further analysis is needed to guarantee optimality. One method that often works is to prove that an optirnal control trajectory exists, and to verify that there is only one control trajectory satisfying the conditions of the Minimum Principle (or that all control trajectories satisfying these conditions have equal cost). Another possibility to conclude optimality arises when the system function $f$ is linear in $(x, u)$, the

constraint set $U$ is convex, and the cost functions $h$ and $g$ are convex. Then it can be shown that the conditions of the Minimum Principle are both necessary and sufficient for optimality.

The Minimum Principle can often be used as the basis of a numerical solution. One possibility is the *two-point boundary problem method.* In this method, we use the minimum condition

$$u^*(t) = \arg\min_{u \in U} H(x^*(t), u, p(t)),$$

to express $u^*(t)$ in terms of $x^*(t)$ and $p(t)$. We then substitute the result into the system and the adjoint equations, to obtain a set of $2n$ first order differential equations in the components of $x^*(t)$ and $p(t)$. These equations can be solved using the split boundary conditions

$$x^*(0) = x(0), \qquad p(T) = \nabla h(x^*(T)).$$

The number of boundary conditions (which is $2n$) is equal to the number of differential equations, so that we generally expect to be able to solve these differential equations numerically (although in practice this may not be simple).

Using the Minimum Principle to obtain an analytical solution is possible in many interesting problems, but typically requires considerable creativity. We give some simple examples.

### Example 3.3.1 (Calculus of Variations Continued)

Consider the problem of finding the curve of minimum length from a point $(0, \alpha)$ to the line $\{(T, y) \mid y \in \Re\}$. In Section 3.1, we formulated this problem as the problem of finding an optimal control trajectory $\{u(t) \mid t \in [0, T]\}$ that minimizes

$$\int_0^T \sqrt{1 + \big(u(t)\big)^2}\, dt$$

subject to

$$\dot{x}(t) = u(t), \qquad x(0) = \alpha.$$

Let us apply the preceding necessary conditions. The Hamiltonian is

$$H(x, u, p) = \sqrt{1 + u^2} + pu,$$

and the adjoint equation is

$$\dot{p}(t) = 0, \qquad p(T) = 0.$$

It follows that

$$p(t) = 0, \qquad \text{for all } t \in [0, T],$$



so minimization of the Hamiltonian gives

$$u^*(t) = \arg\min_{u \in \Re} \sqrt{1 + u^2} = 0, \qquad \text{for all } t \in [0, T].$$

Therefore we have $\dot{x}^*(t) = 0$ for all $t$, which implies that $x^*(t)$ is constant. Using the initial condition $x^*(0) = \alpha$, it follows that

$$x^*(t) = \alpha, \qquad \text{for all } t \in [0, T].$$

We thus obtain the (a priori obvious) optimal solution, which is the horizontal line passing through $(0, \alpha)$. Note that since the Minimum Principle is only a necessary condition for optimality, it does not guarantee that the horizontal line solution is optimal. For such a guarantee, we should invoke the linearity of the system function, and the convexity of the cost function. As mentioned (but not proved) earlier, under these conditions, the Minimum Principle is both necessary and sufficient for optimality.

### Example 3.3.2 (Resource Allocation Continued)

Consider the optimal production problem (Example 3.1.2). We want to maximize

$$\int_0^T \big(1 - u(t)\big)x(t)\, dt$$

subject to

$$0 \le u(t) \le 1, \qquad \text{for all } t \in [0, T],$$

$$\dot{x}(t) = \gamma u(t) x(t), \qquad x(0) > 0 \text{ : given.}$$

The Hamiltonian is

$$H(x, u, p) = (1 - u)x + p\gamma u x.$$

The adjoint equation is

$$\dot{p}(t) = -\gamma u^*(t) p(t) - 1 + u^*(t),$$
$$p(T) = 0.$$

Maximization of the Hamiltonian over $u \in [0, 1]$ yields

$$u^*(t) = \begin{cases} 0 & \text{if } p(t) < \frac{1}{\gamma}, \\ 1 & \text{if } p(t) \ge \frac{1}{\gamma}. \end{cases}$$

Since $p(T) = 0$, for $t$ close to $T$ we will have $p(t) < 1/\gamma$ and $u^*(t) = 0$. Therefore, for $t$ near $T$ the adjoint equation has the form $\dot{p}(t) = -1$ and $p(t)$ has the form shown in Fig. 3.3.1.

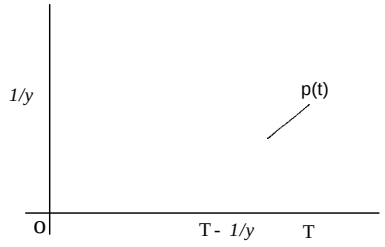



Figure 3.3.1 Form of the adjoint variable $p(t)$ for $t$ near $T$ in the resource allocation example.

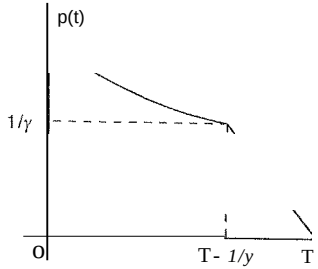
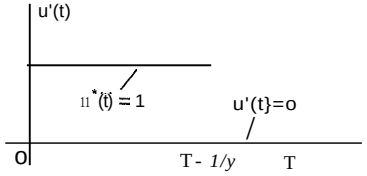



Figure 3.3.2 Form of the adjoint variable $p(t)$ and the optimal control in the resource allocation example.




Thus, near $t = T$, $p(t)$ decreases with slope -1. For $t = T - 1/\gamma$, $p(t)$ is equal to $1/\gamma$, so $u^*(t)$ changes to $u^*(t) = 1$. It follows that for $t < T - 1/\gamma$, the adjoint equation is

$$\dot{p}(t) = -\gamma p(t)$$

or

$$p(t) \quad e^{-\gamma t} \cdot \text{constant}.$$

Piecing together $p(t)$ for $t$ greater and less than $T - 1/\gamma$, we obtain the form shown in Fig. 3.3.2 for $p(t)$ and $u^*(t)$. Note that if $T < 1/\gamma$, the optimal control is $u^*(t) = 0$ for all $t \in [0, T]$; that is, for a short enough horizon, it does not pay to reinvest at any time.

### Exarnple 3.3.3 (A Linear-Quadratic Problem)

Consider the one-dimensional linear system

$$\dot{x}(t) = ax(t) + bu(t),$$



where $a$ and $b$ are given scalars. We want to find an optimal control over a given interval $[0, T]$ that minimizes the quadratic cost

$$\frac{1}{2}q \cdot \big(x(T)\big)^2 + \frac{1}{2}\int_0^T \big(u(t)\big)^2 dt,$$

where $q$ is a given positive scalar. There are no constraints on the control, so we have a special case of the linear-quadratic problem of Example 3.2.2. We will solve this problem via the Minimum Principle.

The Hamiltonian here is

$$H(x, u, p) = \frac{1}{2}u^2 + p(ax + bu),$$

and the adjoint equation is

$$\dot{p}(t) = -ap(t),$$

with the terminal condition

$$p(T) \quad qx^*(T).$$

The optimal control is obtained by minimizing the Hamiltonian with respect to $u$, yielding

$$u^*(t) = \arg\min_u \left[\frac{1}{2}u^2 + p(t)\big(ax^*(t) + bu\big)\right] \quad -bp(t). \tag{3.25}$$

We will extract the optimal solution from these conditions using two different approaches.

In the first approach, we solve the two-point boundary value problem discussed following Prop. 3.3.1. In particular, by eliminating the control from the system equation using Eq. (3.25), we obtain

$$\dot{x}^*(t) = ax^*(t) - b^2 p(t).$$

Also, from the adjoint equation, we see that

$$p(t) = e^{-at}\xi, \qquad \text{for all } t \in [0, T]'$$

where $\xi = p(0)$ is an unknown parameter. The last two equations yield

$$\dot{x}^*(t) = ax^*(t) - b^2 e^{-at}\xi. \tag{3.26}$$

This differential equation, together with the given initial condition $x^*(0)$ $x(0)$ and the terminal condition

$$x^*(T) = \frac{e^{-aT}\xi}{q},$$

(which is the terminal condition for the adjoint equation) can be solved for the unknown variable $\xi$. In particular, it can be verified that the solution of the differential equation (3.26) is given by

$$x^*(t) = x(0)e^{at} + \frac{b^2\xi}{2a}(e^{at} - e^{at}),$$

and $\xi$ can be obtained from the last two relations. Given $\xi$, we obtain $p(t) = e^{-at}\xi$ and from $p(t)$, we can then determine the optimal control trajectory as $u^*(t) = -bp(t)$, $t \in [0, T]$ [ef. Eq. (3.25)].

In the second approach, we basically derive the Riccati equation encountered in Example 3.2.2. In particular, we hypothesize a linear relation between $x^*(t)$ and $p(t)$, that is,

$$K(t)x^*(t) = p(t), \qquad \text{for all } t \in [0, T]'$$

and we show that $K(t)$ can be obtained by solving the Riccati equation. Indeed, from Eq. (3.25) we have

$$u^*(t) = -bK(t)x^*(t),$$

which by substitution in the system equation, yields

$$\dot{x}^*(t) = (a - b^2K(t))x^*(t).$$

By differentiating the equation $K(t)x^*(t) = p(t)$ and by also using the adjoint equation, we obtain

$$\dot{K}(t)x^*(t) + K(t)\dot{x}^*(t) = \dot{p}(t) = -ap(t) = -aK(t)x^*(t).$$

By combining the last two relations, we have

$$k(t)x^*(t) + K(t)(a - b^2K(t))x^*(t) = -aK(t)x^*(t),$$

from which we see that $K(t)$ should satisfy

$$\dot{K}(t) = -2aK(t) + b^2\big(K(t)\big)^2.$$

This is the Riccati equation of Example 3.2.2, specialized to the problem of the present example. This equation can be solved using the terminal condition

$$K(T) = q,$$

which is implied by the terminal condition $p(T) = qx^*(T)$ for the adjoint equation. Once $K(t)$ is known, the optimal control is obtained in the closed-loop form $u^*(t) = -bK(t)x^*(t)$. By reversing the preceding arguments, this control can then be shown to satisfy all the conditions of the Minimum Principle.



### 3.3.2 A Derivation Based on Variational ]Ideas

In this subsection we outline an alternative and more rigorous proof of the Minimum Principle. This proof is primarily directed towards the advanced reader, and is based on making small variations in the optimal trajectory and comparing it with neighboring trajectories.

For convenience, we restrict attention to the case where the cost is

$$h(x(T)).$$

The more general cost

$$h(x(T)) + \int_0^T g(x(t), u(/))dt \tag{3.27}$$

can be reformulated as a terminal cost by introducing a new state variable $y$ and the additional differential equation

$$\dot{y}(t) = g(x(t), u(t)). \tag{3.28}$$

The cost then becomes

$$h(x(T)) + y(T), \tag{3.29}$$

and the Minimum Principle corresponding to this terminal cost yields the Minimum Principle for the general cost (3.27).

We introduce some assumptions:

Convexity Assumption: For every state $x$ the set

$$D = \{f(x, u) \mid u \in U\}$$

is convex.

The convexity assumption is satisfied if $U$ is a convex set and $f$ is linear in $u$ [and $9$ is linear in $u$ in the case where there is an integral cost of the form (3.27), which is reformulated as a terminal cost by using the additional state variable $y$ of Eq. (3.28)]. Thus the convexity assumption is quite restrictive. However, the Minimum Principle typically holds without the convexity assumption, because even when the set $D = \{f(x, u) \mid u \in U\}$ is nonconvex, any vector in the convex hull of $D$ can be generated by quick alternation between vectors from $D$ (for an example, see Exercise 3.10). This involves the complicated mathematieal concept of *randomized* or *relaxed* controls and will not be discussed further.

**Regularity Assumption:** Let $u(t)$ and $u^*(t)$, $t \in [0,T]$, be any two admissible control trajectories and let $\{x^*(t) \mid t \in [0,T]\}$ be the state trajectory corresponding to $u^*(t)$. For any $\epsilon \in [0,1]$ the solution $\{x_\epsilon(t) \mid t \in [0,T]\}$ of the system

$$\dot{x}_\epsilon(t) = (1-\epsilon)f\big(x_\epsilon(t), u^*(t)\big) + \epsilon f\big(x_\epsilon(t), u(t)\big), \tag{3.30}$$

with $x_\epsilon(0) = x^*(0)$, satisfies

$$x_\epsilon(t) = x^*(t) + \epsilon\xi(t) + o(\epsilon), \tag{3.31}$$

where $\{\xi(t) \mid t \in [0,T]\}$ is the solution of the linear differential system

$$\dot{\xi}(t) = \nabla_x f\big(x^*(t), u^*(t)\big)\xi(t) + f(x^*(t), u(t)) - f\big(x^*(t), u^*(t)\big), \tag{3.32}$$

with initial condition $\xi(0) = 0$.

The regularity assumption "typically" holds because from Eq. (3.30) we have

$$\dot{x}_\epsilon(t) - \dot{x}^*(t) = f(x_\epsilon(t), u^*(t)) - f(x^*(t), u^*(t)) + \epsilon\Big(f\big(x_\epsilon(t), u(t)\big) - f\big(x_\epsilon(t), u^*(t)\big)\Big),$$

so from a first order Taylor series expansion we obtain

$$\delta\dot{x}(t) = \nabla f\big(x^*(t), u^*(t)\big)'\delta x(t) + o(\|\delta x(t)\|) + \epsilon\Big(f\big(x_\epsilon(t), u(t)\big) - f\big(x_\epsilon(t), u^*(t)\big)\Big),$$

where

$$\delta x(t) = x_\epsilon(t) - x^*(t).$$

Dividing by $\epsilon$ and taking the limit as $\epsilon \to 0$, we see that the function

$$\xi(t) = \lim_{\epsilon\to 0} \delta x(t)/\epsilon, \qquad t \in [0,T],$$

should "typically" solve the linear system of differential equations (3.32), while satisfying Eq. (3.31).

In fact, if the system is linear of the form

$$\dot{x}(t) = Ax(t) + Bu(t),$$

where $A$ and $B$ are given matrices, it can be shown that the regularity assumption is satisfied. To see this, note that Eqs. (3.30) and (3.32) take the forms

$$\dot{x}_\epsilon(t) = Ax_\epsilon(t) + Bu^*(t) + \epsilon B\big(u(t) - u^*(t)\big),$$



and

$$\dot{\xi}(t) = A\xi(t) + B\big(u(t) - u^*(t)\big),$$

respectively. Thus, taking into account the initial conditions $x_\epsilon(0) = x^*(0)$ and $\xi(0) = 0$, we see that

$$x_\epsilon(t) = x^*(t) + \epsilon\xi(t), \qquad t \in [0,T],$$

so the regularity condition (3.31) is satisfied.

We now prove the Minimum Principle assuming the convexity and regularity assumptions above. Suppose that $\{u^*(t) \mid t \in [0,T]\}$ is an optimal control trajectory, and let $\{x^*(t) \mid t \in [0,T]\}$ be the corresponding state trajectory. Then for any other admissible control trajectory $\{u(t) \mid t \in [0,T]\}$ and any $\epsilon \in [0,1]$, the convexity assumption guarantees that for each $t$, there exists a control $\overline{u}(t) \in U$ such that

$$f(x_\epsilon(t), u(t)) = (1-\epsilon)f\big(x_\epsilon(t), u^*(t)\big) + \epsilon f\big(x_\epsilon(t), u(t)\big).$$

Thus, the state trajectory $\{x_\epsilon(t) \mid t \in [0,T]\}$ of Eq. (3.30) corresponds to the admissible control trajectory $\{\overline{u}(t) \mid t \in [0,T]\}$. Hence, using the optimality of $\{x^*(t) \mid t \in [0,T]\}$ and the regularity assumption, we have

$$\begin{aligned} h(x^*(T)) &\le h\big(x_\epsilon^*(T)\big) \\ &= h\big(x^*(T) + \epsilon\xi(T) + o(\epsilon)\big) \\ &= h(x^*(T)) + \epsilon\nabla h\big(x^*(T)\big)'\xi(T) + o(\epsilon), \end{aligned}$$

which implies that

$$\nabla h\big(x^*(T)\big)'\xi(T) \ge 0. \tag{3.33}$$

Using a standard result in the theory of linear differential equations (see e.g. [CoL65]), the solution of the linear differential system (3.32) can be written in closed form as

$$\xi(t) = \Phi(t,\tau)\xi(\tau) + \int_\tau^t \Phi(t,\tau)\Big(f\big(x^*(\tau), u(\tau)\big) - f\big(x^*(\tau), u^*(\tau)\big)\Big)d\tau, \tag{3.34}$$

where the square matrix $\Phi$ satisfies for all $t$ and $\tau$,

$$\frac{\partial\Phi(t,\tau)}{\partial\tau} = -\Phi(t,\tau)\nabla_x f\big(x^*(\tau), u^*(\tau)\big)', \tag{3.35}$$

$$\Phi(t,t) = I.$$

Since $\xi(0) = 0$, we have from Eq. (3.34),

$$\xi(T) = \int_0^T \Phi(T,t)\Big(f(x'(t), u(t)) - f\big(x^*(t), u^*(t)\big)\Big)\,dt. \tag{3.36}$$

Define

$$p(T) = \nabla h\big(x^*(T)\big), \qquad p(t) = \Phi(T,t)'p(T), \quad t \in [0,T]. \tag{3.37}$$

By differentiating with respect to $t$, we obtain

$$\dot{p}(t) = \frac{\partial \Phi(T,t)}{\partial t}' p(T).$$

Combining this equation with Eqs. (3.35) and (3.37), we see that $p(t)$ is generated by the differential equation

$$\dot{p}(t) = -\nabla_x f\big(x^*(t), u^*(t)\big)p(t),$$

with the terminal condition

$$p(T) = \nabla h\big(x^*(T)\big).$$

This is the adjoint equation corresponding to $\{(x^*(t), u^*(t)) \mid t \in [0,T]\}$. Now, to obtain the Minimum Principle, we note that from Eqs. (3.33), (3.36), and (3.37) we have

$$\begin{aligned} 0 &\le p(T)'\xi(T) \\ &= p(T)' \int_0^T \Phi(T,t)\Big(f\big(x^*(t),u(t)\big) - f(x^*(t), u^*(t))\Big)\,dt \\ &= \int_0^T p(t)'\Big(f\big(x^*(t), u(t)\big) - f\big(x^*(t), u^*(t)\big)\Big)\,dt, \end{aligned} \tag{3.38}$$

from which it can be shown that for all $t$ at which $u^*(\cdot)$ is continuous, we have

$$p(t)'f\big(x^*(t),u^*(t)\big) \le p(t)'f(x^*(t),u), \qquad \text{for all } u \in U. \tag{3.39}$$

Indeed, if for some $\hat{u} \in U$ and $t_0 \in [0,T)$, we have

$$p(t_0)'f\big(x^*(t_0),u^*(t_0)\big) > p(t_0)'f(x^*(t_0),\hat{u}),$$

while $\{u^*(t) \mid t \in [0,T]\}$ is continuous at $t_0$, we would also have

$$p(t)'f(x^*(t),u^*(t)) > p(t)'f\big(x^*(t),\hat{u}\big),$$

for all $t$ in some nontrivial interval $I$ containing $t_0$. By taking

$$u(t) = \begin{cases} \hat{u} & \text{for } t \in I, \\ u^*(t) & \text{for } t \notin I, \end{cases}$$



we would then obtain a contradiction of Eq. (3.38).

We have thus proved the Minimum Principle (3.39) under the convexity and regularity assumptions, and the assumption that there is only a terminal cost $h(x(T))$. We have also seen that in the case where the constraint set $U$ is convex and the system is linear, the convexity and regularity assumptions are satisfied. To prove the Minimum Principle for the more general integral cost function (3.27), we can apply the preceding development to the system of differential equations $\dot{x} = f(x,u)$ augmented by the additional Eq. (3.28) and the equivalent terminal cost (3.29). The corresponding convexity and regularity assumptions are automatically satisfied if the constraint set $U$ is convex and the system function $f(x,u)$ as well as the cost function $g(x,u)$ are linear. This is necessary in order to maintain the linearity of the augmented system, thereby maintaining the validity of the regularity assumption.

### 3.3.3 Minimum Principle for Discrete-Time Problems

In this subsection we briefly derive a version of the Minimum Principle for discrete-time deterministic optimal control problems. Interestingly, it is essential to make some convexity assumptions in order for the Minimum Principle to hold. For continuous-time problems these convexity assumptions are typically not needed, because, as mentioned earlier, the differential system can generate any $\dot{x}(t)$ in the convex hull of the set of possible vectors $f(x(t), u(t))$ by quick alternation between different controls (see for example Exercise 3.10).

Suppose that we want to find a control sequence $(u_0, u_1, \ldots, u_{N-1})$ and a corresponding state sequence $(x_0, x_1, \ldots, x_N)$, which minimize

$$J(u) = g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k),$$

subject to the discrete-time system constraints

$$x_{k+1} = f_k(x_k, u_k), \quad k = 0, \ldots, N-1, \qquad x_0 \text{ given},$$

and the control constraints

$$u_k \in U_k \subset \Re^m, \qquad k = 0, \ldots, N-1.$$

We first develop an expression for the gradient $\nabla J(u_0, \ldots, u_{N-1})$. We have, using the chain rule,

$$\begin{aligned} \nabla_{u_{N-1}} J(u_0, \ldots, u_{N-1}) &= \nabla_{u_{N-1}} \big(g_N(f_{N-1}(x_{N-1}, u_{N-1})) \\ &\qquad + g_{N-1}(x_{N-1}, u_{N-1})\big) \\ &= \nabla_{u_{N-1}} f_{N-1} \cdot \nabla g_N + \nabla_{u_{N-1}} g_{N-1}, \end{aligned}$$

where all gradients are evaluated along the control trajectory *(uo, ... ,UN-1)* and the corresponding state trajectory. Similarly, for all *k,*

$$\begin{aligned}\nabla_{u_k} J(u_0, \ldots, u_{N-1}) = {} & \nabla_{u_k} f_k \cdot \nabla_{x_{k+1}} f_{k+1} \cdot \cdots \nabla_{x_{N-1}} f_{N-1} \cdot \nabla g_N \\ & + \nabla_{u_k} f_k \cdot \nabla_{x_{k+1}} f_{k+1} \cdot \cdots \nabla_{x_{N-2}} f_{N-2} \cdot \nabla_{x_{N-1}} g_{N-1} \\ & + \nabla_{u_k} f_k \cdot \nabla_{x_{k+1}} g_{k+1} \\ & + \nabla_{u_k} g_k,\end{aligned} \tag{3.40}$$

which can be written in the form

$$\nabla_{u_k} J(u_0, \ldots, u_{N-1}) = \nabla_{u_k} f_k \cdot p_{k+1} + \nabla_{u_k} g_k,$$

for an appropriate vector *Pk+l,* or

$$\nabla_{u_k} J(u_0, \ldots, u_{N-1}) = \nabla_{u_k} H_k(x_k, u_k, p_{k+1}), \tag{3.41}$$

where *11k* is the Hamiltonian function defined by

$$H_k(x_k, u_k, p_{k+1}) = g_k(x_k, u_k) + p'_{k+1} f_k(x_k, u_k).$$

It can be seen from Eq. (3.40) that the vectors *Pk+l* are generated backwards by the *discrete-time adjoint equation*

$$p_k = \nabla_{x_k} f_k \cdot p_{k+1} + \nabla_{x_k} g_k, \qquad k = 1, \ldots, N-1,$$

with terminal condition

$$p_N = \nabla g_N.$$

We will assume that the constraint sets *Uk* are convex, so that we can apply the optimality condition

$$\sum_{k=0}^{N-1} \nabla_{u_k} J(u_0^*, \ldots u_{N-1}^*)'(u_k - u_k^*) \geq 0,$$

for all feasible $(u_0, \ldots, u_{N-1})$ (see Appendix B). This condition can be decomposed into the N conditions

$$\nabla_{u_k} J(u_0^*, \ldots, u_{N-1}^*)'(u_k - u_k^*) \geq 0, \qquad \text{for all } u_k \in U_k,\ k = 0, \ldots, N-1. \tag{3.42}$$

We thus obtain:



**Proposition** 3.3.2: (Discrete-Time Minimllm Suppose that $(u_0^*, u_1^* \ldots, u_{N-1}^*)$ is an optimal control trajectory and that $(x_0^*, x_1^*, \ldots, x_N^*)$ is the corresponding state trajectory. Assume also that the constraint sets *Uk* are convex. Then for all $k = 0, \ldots, N-1$, we have

$$\nabla_{u_k} H_k(x_k^*, u_k^*, p_{k+1})'(u_k - u_k^*) \geq 0, \qquad \text{for all } u_k \in U_k, \tag{3.43}$$

where the vectors *PI, ... ,PN* are obtained from the adjoint equation

$$p_k = \nabla_{x_k} f_k \cdot p_{k+1} + \nabla_{x_k} g_k, \qquad k = 1, \ldots, N-1,$$

with the terminal condition

$$p_N = \nabla g_N(x_N^*).$$

The partial derivatives above are evaluated along the optimal state and control trajectories. If, in addition, the Hamiltonian $H_k$ is a convex function of $u_k$ for any fixed *Xk* and $p_{k+1}$, we have

$$u_k^* = \arg \min_{u_k \in U_k} H_k(x_k^*, u_k, p_{k+1}), \qquad \text{for all } k = 0, \ldots, N-1. \tag{3.44}$$

Proof: Equation (3.43) is a restatement of the necessary condition (3.42) using the expression (3.41) for the gradient of J. If *11k* is convex with respect to *Uk,* Eq. (3.42) is a sufficient condition for the minimum condition (3.44) to hold (see Appendix B). Q.E.D.

## 3.4 EXTENSIONS OF THE MINIlVfUM PRINCIPLE

We now consider some variations of the continuous-time optimal control problem and derive corresponding variations of the Minimum Principle.

### 3.4.1 Fixed Terminal State

Suppose that in addition to the initial state *x(O),* the final state $x(T)$ is given. Then the preceding informal derivations still hold except that the terminal condition *J*(T, x) = h(x)* is not true anynlOre. In effect, here we have

$$J^*(T, x) = \begin{cases} 0 & \text{if } x = x(T), \\ \infty & \text{otherwise.} \end{cases}$$

Thus $J^*(T, x)$ cannot be differentiated with respect to $x$, and the terminal boundary condition $p(T) = \nabla h(x^*(T))$ for the adjoint equation does not hold. However, as compensation, we have the extra condition

$$x(T) \text{ : given,}$$

thus maintaining the balance between boundary conditions and unknowns.

If only *some* of the terminal states are fixed, that is,

$$x_i(T) \text{ : given,} \qquad \text{for all } i \in I,$$

where $I$ is some index set, we have the partial boundary condition

$$p_j(T) = \frac{\partial h(x^*(T))}{\partial x_j}, \qquad \text{for all } j \notin I,$$

for the adjoint equation.

### Example 3.4.1

Consider the problem of finding the curve of minimum length connecting two points $(0, \alpha)$ and $(T, \beta)$. This is a fixed endpoint variation of Example 3.3.1 in the preceding section. We have

$$\dot{x}(t) = u(t),$$

$$x(0) = \alpha, \qquad x(T) = \beta,$$

and the cost is

$$\int_0^T \sqrt{1 + (u(t))^2}\, dt.$$

The adjoint equation is

$$\dot{p}(t) = 0,$$

implying that

$$p(t) = \text{constant}, \qquad \text{for all } t \in [0, T].$$

Minimization of the Hamiltonian,

$$\min_{u \in \Re} \left[ \sqrt{1 + u^2} + p(t)u \right],$$

yields

$$u^*(t) = \text{constant}, \qquad \text{for all } t \in [0, T].$$

Thus the optimal trajectory $\{x^*(t) \mid t \in [0, T]\}$ is a straight line. Since this trajectory must pass through $(0, \alpha)$ and $(T, \beta)$, we obtain the (a priori obvious) optimal solution shown in Fig. 3.4.1.



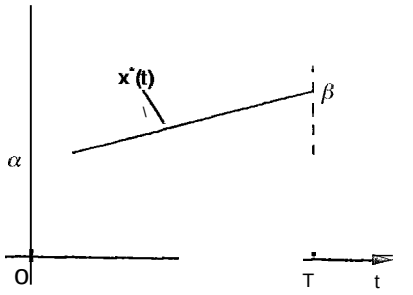



Figure 3.4.1 Optimal solution of the problem of connecting the two points $(0, \alpha)$ and $(T, \beta)$ with a minimum length curve (cf. Example 3.4.1).

### Example 3.4.2 (The Brachistochrone Problem)

In 1696 Johann Bernoulli challenged the mathematical world of his time with a problem that played an instrumental role in the development of the calculus of variations: Given two points A and B, find a curve connecting A and B such that a body moving along the curve under the force of gravity reaches B in minimum time (see Fig. 3.4.2). Let A he (0,0) and B be $(T, -b)$ with $b > 0$. Then it can he seen that the problem is to find $\{x(t) \mid t \in [0, T]\}$ with $x(0) = 0$ and $x(T) = b$, which minimizes

$$\int_0^T \frac{\sqrt{1 + (\dot{x}(t))^2}}{\sqrt{2\gamma x(t)}}\, dt,$$

where $\gamma$ is the acceleration due to gravity. Here $\{(t, -x(t)) \mid t \in [0,T]\}$ is the desired curve, the term $\sqrt{1 + (\dot{x}(t))^2}\, dt$ is the length of the curve from $x(t)$ to $x(t + dt)$, and the term $\sqrt{2\gamma x(t)}$ is the velocity of the body upon reaching the level $x(t)$ [if $m$ and $v$ denote the mass and the velocity of the body, the kinetic energy is $mv^2/2$, which at level $x(t)$ must be equal to the change in potential energy, which is $mgx(t)$; this yields $v = \sqrt{2gx(t)}$ ].

We introduce the system $\dot{x} = u$, and we obtain a fixed terminal state problem $[x(0) = 0$ and $x(T) = b]$. Letting

$$g(x, u) = \frac{\sqrt{1 + u^2}}{\sqrt{2\gamma x}},$$

the Hamiltonian is

$$H(x, u, p) = g(x, u) + pu.$$

We minimize the Hamiltonian by setting to zero its derivative with respect to $u$:

$$p(t) = -\nabla_u g\big(x^*(t), u^*(t)\big).$$

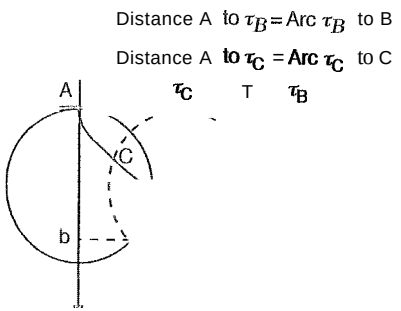



Figure 3.4.2 Formulation and optimal solution of the brachistochrone problem.

We know from the Minimum Principle that the Hamiltonian is constant along an optimal trajectory, Le.,

$$g\big(x^*(t),u^*(t)\big) - \nabla_u g\big(x^*(t),u^*(t)\big)u^*(t) = \text{constant}, \qquad \text{for all } t \in [0,T].$$

Using the expression for $g$, this can be written as

$$\frac{\sqrt{1+\big(u^*(t)\big)^2}}{\sqrt{2\gamma x^*(t)}} - \frac{\big(u^*(t)\big)^2}{\sqrt{1+\big(u^*(t)\big)^2}\sqrt{2\gamma x^*(t)}} = \text{constant}, \qquad \text{for all } t \in [0,T],$$

or equivalently

$$\frac{1}{\sqrt{1+\big(u^*(t)\big)^2}\sqrt{2\gamma x^*(t)}} = \text{constant}, \qquad \text{for all } t \in [0,T].$$

Using the relation $\dot{x}^*(t) = u^*(t)$, this yields

$$x^*(t)(1+x^*(t)^2) = C, \qquad \text{for all } t \in [0,T],$$

for some constant C. Thus an optimal trajectory satisfies the differential equation

$$\dot{x}^*(t) = \sqrt{\frac{C - x^*(t)}{x^*(t)}} \qquad \text{for all } t \in [0,T].$$

The solution of this differential equation was known at Bernoulli's time to be a *cycloid;* see Fig. 3.4.2. The unknown parameters of the cycloid are determined by the boundary conditions $x^*(0) = 0$ and $x^*(T) = b$.



### 3.4.2 Free Initial State

If the initial state $x(0)$ is not fixed but is subject to optimization, we have

$$J^*(0,x^*(0)) \le J^*(0,x), \qquad \text{for all } x \in \Re^n,$$

yielding

$$\nabla_x J^*\big(0,x^*(0)\big) = 0$$

and the extra boundary condition for the adjoint equation

$$p(0) = 0.$$

Also if there is a cost $\ell\big(x(0)\big)$ on the initial state, i.e., the cost is

$$\ell\big(x(0)\big) + \int_0^T g\big(x(t),u(t)\big)dt + h\big(x(T)\big),$$

the boundary condition becomes

$$p(0) = -\nabla\ell(x^*(0)).$$

This follows by setting to zero the gradient with respect to $x$ of $\ell(x) + J(0,x)$, Le.,

$$\nabla_x\big\{\ell(x) + J(0,x)\big\}|_{x=x^*(0)} = 0.$$

### 3.4.3 Free Terminal Time

Suppose the initial state and/or the terminal state are given, but the terminal time $T$ is subject to optimization.

Let $\{(x^*(t),u^*(t)) \mid t \in (0,T)\}$ be an optimal state-control trajectory pair and let $T^*$ be the optimal terminal time. Then if the terminal time were fixed at $T^*$, the pair $\big\{\big(u^*(t),x^*(t)\big) \mid t \in [0,T^*]\big\}$ would satisfy the conditions of the Minimum Principle. In particular,

$$u^*(t) = \arg\min_{u\in U} H\big(x^*(t),u,p(t)\big), \qquad \text{for all } t \in [0,T^*],$$

where $p(t)$ is the solution of the adjoint equation. What we lose with the terminal time being free, we gain with an extra condition derived as follows.

We argue that if the terminal time were fixed at $T^*$ and the initial state were fixed at the given $x(0)$, but instead the initial time were subject to optimization, it would be optimal to start at t = 0. This means that the first order variation of the optimal cost with respect to the initial time must be zero;

$$\nabla_t J^*(t,x^*(t))|_{t=0} = 0.$$

The I-IJB equation can be written along the optimal trajectory as

$$\nabla_t J^*\big(t, x^*(t)\big) = -H(X^*(t), 1t^*(t), p(t)), \qquad \text{for all t E } [O, T^*)$$

[cf. Eqs. (3.14) and (3.19)), so the preceding two equations yield

$$H(x^*(O), u^*(O), p(O)) = 0.$$

Since the Hamiltonian was shown earlier to be constant along the optimal trajectory, we obtain for the case of a free terminal time

$$H\big(x^*(t), u^*(t), p(t)\big) = 0, \qquad \text{for all t E } [0, T^*).$$

### Example 3.4,3 (lVIinimum-Time Problem)

A unit mass object moves horizontally under the influence of a force $u(t)$, so that

$$\ddot{y}(t) = u(t),$$

where *yet)* is the position of the object at time *t.* Given the object's initial position *y(O)* and initial velocity $\dot{y}(0)$, it is required to bring the object to rest (zero velocity) at a given position, say zero, while using at most unit magnitude force,

$$-1 \leq u_(t) \leq 1, \qquad \text{for all } t.$$

We want to accomplish this transfer in minimum time. Thus, we want to

$$\text{minimize } T = \int_0^T 1dt.$$

Note that the integral cost, $g(x(t),u(t)) \equiv 1$, is unusual here; it does not depend on the state or the control. However, the theory does not preclude this possibility, and the problem is still meaningful because the terminal time $T$ is free and subject to optimization.

Let the state variables be

$$\text{x1}(t) = yet), \qquad x_2(t) = \dot{y}(t),$$

so the system equation is

$$\dot{x}_1(t) = x_2(t), \qquad \dot{x}_2(t) = u(t).$$

The initial state *(XI(0),X2(0))* is given and the terminal state is also given

$$xI(T) = 0, \qquad x2(T) = 0.$$

If $\{u^*(t)$ It E [0, Tn is an optimal control trajectory, $u^*(t)$ must minimize the Hamiltonian for each *t,* i.e.,

$$u^*(t) = \arg \min_{-1 \leq u \leq 1} \big[1 + p_1(t)x_2^*(t) + p_2(t)u\big].$$



Therefore

$$u^*(t) = \begin{cases} 1 & \text{if } p2(t) < 0, \\ -1 & \text{if } P2(t) \geq 0. \end{cases}$$

The adjoint equation is

$$\dot{p}_1(t) = 0, \qquad \dot{p}_2(t) = -p_1(t),$$

so

$$p_1(t) = c_1, \qquad p_2(t) = c_2 - c_1 t,$$

where c1 and c2 are constants. It follows that $\{p2(t)$ I $t$ E [0, Tn has one of the four forms shown in Fig. 3.4.3(a); that is, $\{p2(t)$ I $t$ E [0, Tn switches at most once in going from negative to positive or reversely. [Note that it is not possible for *P2(t)* to be equal to 0 for all *t* because this implies that Pl *(t)* is also equal to 0 for all *t,* so that the Hamiltonian is equal to 1 for all *t;* the necessary conditions require that the Hamiltonian be 0 along the optimal trajectory.] The corresponding control trajectories are shown in Fig. 3.4.3(b). The conclusion is that, for each *t,* $u^*(t)$ is either +1 or -1, and $\{u^*(t)$ It E [0, Tn has at most one switching point in the interval [0, *T].*

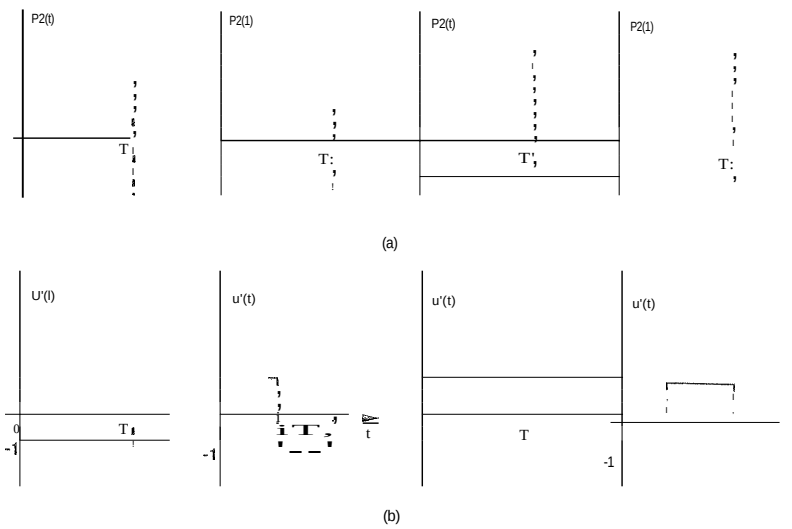



Figure 3.4.3 (a) Possible forms of the adjoint variable $p_2(t)$. (b) Corresponding forms of the optimal control trajectory.

To determine the precise form of the optimal control trajectory, we use the given initial and final states. For $u(t) \equiv \zeta$, where $\zeta = \pm 1$, the system evolves according to

$$x_1(t) = x_1(0) + x_2(0)t + \frac{\zeta}{2}t^2, \qquad x_2(t) = x_2(0) +$$

By eliminating the time $t$ in these two equations, we see that for all $t$

$$x_1(t) - \frac{1}{2\zeta}(x_2(t))^2 = x_1(0) - \frac{1}{2\zeta}(x_2(0))^2.$$

Thus for intervals where $u(t) \equiv 1$, the system moves along the curves where $x_1(t) - \frac{1}{2}(x_2(t))^2$ is constant, shown in Fig. 3.4.4(a). For intervals where $u(t) \equiv -1$, the system moves along the curves where $x_1(t) + \frac{1}{2}(x_2(t))^2$ is constant, shown in Fig. 3.4.4(b).

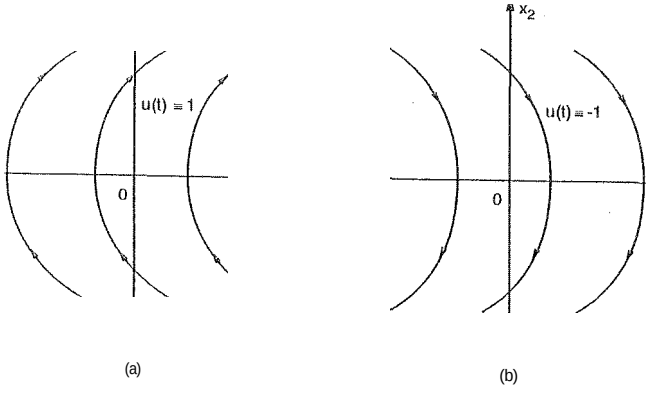



Figure 3.4.4 State trajectories when the control is $u(t) \equiv 1$ [Fig. (a)] and when the control is $u(t) \equiv -1$ [Fig. (b)].

To bring the system from the initial state $(x_1(0), x_2(0))$ to the origin with at most one switch in the value of control, we must apply control according to the following rules involving the *switching curve* shown in Fig. 3.4.5.

(a) If the initial state lies *above* the switching curve, use $u^*(t) \equiv -1$ until the state hits the switching curve; then use $u^*(t) \equiv 1$ until reaching the origin.

(b) If the initial state lies *below* the switching curve, use $u^*(t) \equiv 1$ until the state hits the switching curve; then use $u^*(t) \equiv -1$ until reaching the origin.

(c) If the initial state lies on the top (bottom) part of the switching curve, use $u^*(t) \equiv -1$ [$u^*(t) \equiv 1$, respectively] until reaching the origin.

### 3.4.4 Thne-Varying System and Cost

If the system equation and the integral cost depend on the time $t$, Le.,

$$\dot{x}(t) = f(x(t), u(t), t),$$



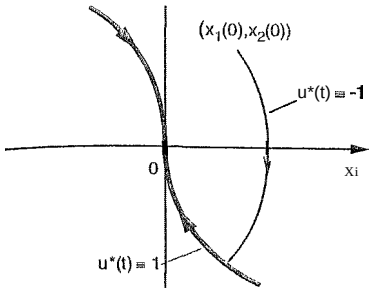



Figure 3.4.5 Switching curve (shown with a thick line) and closed-loop optimal control for the minimmn time example.

$$\text{cost} = h(x(T)) + \int_0^T g(x(t), u(t), t)dt,$$

we can convert the problem to one involving a time-independent system and cost by introducing an extra state variable $y(t)$ representing time:

$$\dot{y}(t) = 1, \qquad y(0) = 0,$$

$$\dot{x}(t) = f(x(t), u(t), y(t)), \qquad x(0) : \text{given},$$

$$\text{cost} = h(x(T)) + \int_0^T g(x(t), u(t), y(t))dt.$$

After working out the corresponding optimality conditions, we see that they are the same as when the system and cost are time-independent. The only difference is that the Hamiltonian need not be constant along the optimal trajectory.

### 3.4.5 Singular Problelms

In some cases, the minimum condition

$$u^*(t) = \arg\min_{u \in U} H(x^*(t), u, p(t), t) \tag{3.45}$$

is insufficient to determine $u^*(t)$ for all $t$, because the values of $x^*(t)$ and $p(t)$ are such that $H(x^*(t), u, p(t), t)$ is independent of $u$ over a nontrivial interval of time. Such problems are called *singular*. Their optimal trajectories consist of portions, called *regular arcs*, where $u^*(t)$ can be determined from the minimum condition (3.45), and other portions, called *singular arcs*, which can be determined from the condition that the Hamiltonian is independent of $u$.

### Example 3.4.4 (Road Construction)

Suppose that we want to construct a road over a one-dimensional terrain whose ground elevation (altitude measured from some reference point) is known and is given by $z(t)$, $t \in [0, T]$. The elevation of the road is denoted by $x(t)$, $t \in [0,1']$, and the difference $x(t) - z(t)$ must be made up by fill-in or excavation. It is desired to minimize

$$\frac{1}{2}\int_0^T \big(x(t) - z(t)\big)^2 dt,$$

subject to the constraint that the gradient of the road $\dot{x}(t)$ lies between $-a$ and $a$, where $a$ is a specified maximum allowed slope. Thus we have the constraint

$$|u(t)| \leq a, \qquad t \in [0,1'],$$

where

$$\dot{x}(t) = u(t), \qquad t \in [0,1'].$$

The adjoint equation here is

$$\dot{p}(t) = -x^*(t) + z(t),$$

with the terminal condition

$$p(T) = 0.$$

Minimization of the Hamiltonian

$$H(x^*(t),u,p(t),t) = \frac{1}{2}\big(x^*(t) - Z(t)\big)2 + p(t)u$$

with respect to $u$ yields

$$u^*(t) = \arg\min_{|u|\leq a} p(t)u,$$

for all $t$, and shows that optimal trajectories are obtained by concatenation of three types of arcs:

(a) Regular arcs where $p(t) > 0$ and $u^*(t) = -a$ (maximum downhill slope arcs).

(b) H.egular arcs where $p(t) < 0$ and $u^*(t) = a$ (maximum uphill slope arcs).

(c) Singular arcs where $p(t) = 0$ and $u^*(t)$ can take any value in $[-a, a]$ that maintains the condition $p(t) = 0$. From the adjoint equation we see that singular arcs are those along which $p(t) = 0$ and $x^*(t) = z(t)$, Le., the road follows the ground elevation (no fill-in or excavation). Along such arcs we must have

$$\dot{z}(t) = u^*(t) \in [-a, a].$$



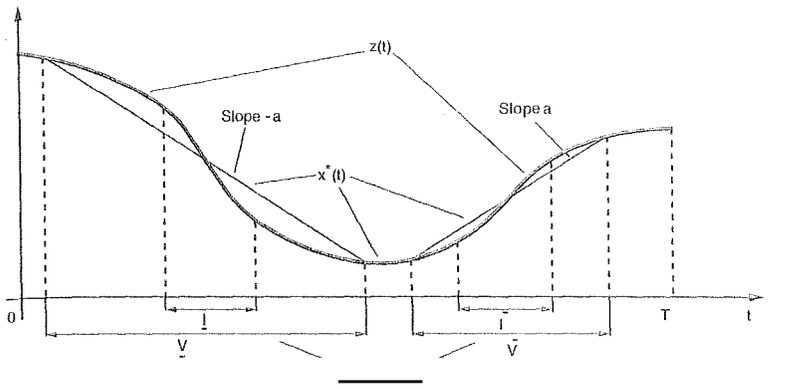



Figure 3.4.6 Graphical method for solving the road construction example. The sharply uphill (downhill) intervals $\overline{I}$ (respectively, $\underline{I}$) are first identified, and are then embedded within maximum uphill (respectively, downhill) slope regular arcs $\overline{V}$ (respectively, $\underline{V}$) within which the total fill-in is equal to the total excavation. The regular arcs are joined by singular arcs where there is no fill-in or excavation. The graphical process is started at the endpoint $t = T$.

Optimal solutions can be obtained by a graphical method using the above observations. Consider the *sharply uphill intervals* $\overline{I}$ such that $\dot{z}(t) \geq a$ for all $t \in \overline{I}$, and the *sharply downhill intervals* $\underline{I}$ such that $\dot{z}(t) \leq -a$ for all $t \in \underline{I}$. Clearly, within each sharply uphill interval $\overline{I}$ the optimal slope is $u^*(t) = a$, but the optimal slope is also equal to $a$ within a larger maximum uphill slope interval $\overline{V} \supset \overline{I}$, which is such that $p(t) < 0$ within $\overline{V}$ and

$$p(t_1) = p(t_2) = 0$$

at the endpoints t1 and $t2$ of $\overline{V}$. In view of the forn1 of the adjoint equation, we see that the endpoints t1 and $t2$ of $\overline{V}$ should be such that

$$\int_{t_1}^{t_2} \big(z(t) - x^*(t)\big)dt = 0;$$

that is, the *total fill-in should be equal to the total excavation within* (see Fig. 3.4.6). Similarly, each sharply downhill interval $\underline{I}$ should be contained within a larger maximum downhill slope interval $\underline{V} \supset \underline{I}$, which is such that $p(t) > 0$ within $\underline{V}$, while the *total fill-in should be equal to the total excavation within* $\underline{V}$, (see Fig. 3.4.6). Thus the regular arcs consist of the intervals $\overline{V}$ and $\underline{V}$ described above. Between the regular arcs there can be one or IllOre singular arcs where $x^*(t) = z(t)$. The optimal solution can be pieced together starting at the endpoint $t = 1'$ [where we know that $p(T) = 0$], and proceeding backwards.

## NOTES, SOURCES, AND EXERCISES

The calculus of variations is a classical subject that originated with the works of the great mathematicians of the 17th and 18th centuries. Its rigorous developrnent (by modern mathematical standards) took place in the 1930s and 19/10s, with the work of a group of mathematicians that originated mostly from the University of Chicago; Bliss, McShane, and Hestenes are some of the most prominent members of this group.· Curiously, this development preceded the development of nonlinear programming by many years.† The modern theory of deterministic optimal control has its roots primarily in the work of Pontryagin, Boltyanski, Gamkrelidze, and Mishchenko in the 1950s [PBG65]. A highly personal but controversial historical account of this work is given by Boltyanski in [BMS96]. The theoretical and applications literature on the subject is very extensive. We give three representative references: the book by Athans and Falb [AtF66] (a classical extensive text that includes engineering applications), the book by Hestenes [Hes66] (a rigorous mathematical treatment, containing important work that predates the work of Pontryagin et a1.), and the book by Luenberger [LlIe69] (which deals with optimal control within a broader infinite dimensional context). The author's nonlinear programming book [13er99] gives a detailed treatment of optimality conditions and computational methods for discrete-time optimal control.

## EXERCISES

### 3.1

Solve the problem of Example 3.2.1 for the case where the cost function is

$$(X(T))^2 + \int_0^T \big(u(t)\big)^2 dt.$$

Also, calculate the cost-to-go function $J^*(t, x)$ and verify that it satisfies the HJB equation.

† In the 30s and 40s journal space was at a premium, and finite-dimensional optimization research was thought to be a simple special case of the calculus of variations, thus insufficiently challenging or novel for publication. Indeed the modern optimality conditions of finite-dimensional optimization subject to equality and inequality constraints were first developed in the 1939 Master's thesis by Karush, but first appeared in a journal quite a few years later under the names of other researchers.



### 3.2

A young investor has earned in the stock market a large amount of money S and plans to spend it so as to maximize his enjoyment through the rest of his life without working. He estimates that he will live exactly T more years and that his capital $x(t)$ should be reduced to zero at time T, i.e., $x(T) = 0$. Also he models the evolution of his capital by the differential equation

$$\frac{dx(t)}{dt} = \alpha x(t) - u(t),$$

where $x(0) = S$ is his initial capital, $\alpha > 0$ is a given interest rate, and $u(t) \geq 0$ is his rate of expenditure. The total enjoyment he will obtain is given by

$$\int_0^T e^{-\beta t} \sqrt{u(t)}\, dt.$$

Here $\beta$ is some positive scalar, which serves to discount future enjoyment. Find the optimal $\{u(t) \mid t \in [0, T)\}$.

### 3.3

Consider the system of reservoirs shown in Fig. 3.5.1. The system equations are

$$\dot{x}_1(t) = -x_1(t) + u(t),$$
$$\dot{x}_2(t) = x_1(t),$$

and the control constraint is $0 \leq u(t) \leq 1$ for all $t$. Initially

$$x_1(0) = x_2(0) = 0.$$

We want to maximize $x_2(1)$ subject to the constraint $x_1(1) \geq 0.5$. Solve the problem.

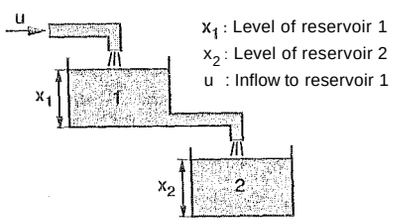



Figure 3.5.1 Reservoir system for Exercise 3.3.

**3.4**

Work out the minimum-time problem (Example 3.4.3) for the case where there is friction and the object's position moves according to

$$\ddot{y}(t) = -a\dot{y}(t) + u(t),$$

where $a > 0$ is given. *Hint:* The solution of the system

$$\dot{p}_1(t) = 0,$$
$$\dot{p}_2(t) = -p_1(t) + ap_2(t),$$

is

$$p_1(t) = p_1(0),$$
$$p_2(t) = \frac{1}{a}(1 - e^{at})p_1(0) + e^{at}p_2(0).$$

The trajectories of the system for $u(t) \equiv -1$ and $u(t) \equiv 1$ are sketched in Fig. 3.5.2.

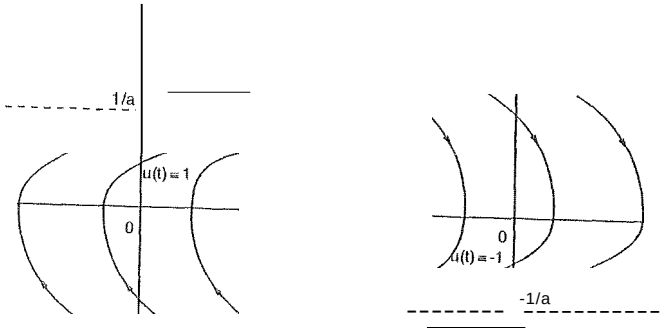



Figure 3.5.2 State trajectories of the system of Exercise 3.4 for $u(t) \equiv -1$ and $u(t) \equiv 1$.

## 3.5 (Isoperirnetric Problem)

Analyze the problem of finding a curve $\{x(t) \mid t \in [0, T]\}$ that maximizes the area under $x$,

$$\int_0^T x(t)dt,$$



subject to the constraints

$$x(0) = a, \qquad x(T) = b, \qquad \int_0^T \sqrt{1 + \left(\dot{x}(t)\right)^2}\, dt = L,$$

where $a$, $b$, and $L$ are given positive scalars. The last constraint is known as an isoperimetric constraint; it requires that the length of the curve be $L$. *Hint*: Introduce the system $\dot{x}_1 = u$, $\dot{x}_2 = \sqrt{1 + u^2}$, and view the problem as a fixed terminal state problem. Show that the sine of the optimal $u(t)$ depends linearly on $t$. Under some assumptions on $a$, $b$, and $L$, the optimal curve is a circular arc.

## 3.6 (L'Hôpital's Problern)

Let $a$, $b$, and $T$ be positive scalars, and let $A = (0, a)$ and $B = (T, b)$ be two points in a medium within which the velocity of propagation of light is proportional to the vertical coordinate. Thus the time it takes for light to propagate from $A$ to $B$ along a curve $\{x(t) \mid t \in [0, T]\}$ is

$$\int_0^T \frac{\sqrt{1 + (\dot{x}(t))^2}}{Cx(t)}\, dt,$$

where $C$ is a given positive constant. Find the curve of minimum travel time of light from $A$ to $B$, and show that it is an arc of a circle of the form

$$(x(t) - d)^2 + t^2 = D,$$

where $d$ and $D$ are some constants.

## 3.7

A boat moves with constant unit velocity in a stream moving at constant velocity $s$. The problem is to find the steering angle $u(t)$, $0 \le t \le T$, which minimizes the time $T$ required for the boat to move between the point (0,0) to a given point $(a, b)$. The equations of motion are

$$\dot{x}_1(t) = s + \cos u(t), \qquad \dot{x}_2(t) = \sin u(t),$$

where $x_1(t)$ and $x_2(t)$ are the positions of the boat parallel and perpendicular to the stream velocity, respectively. Show that the optimal solution is to steer at a constant angle.

A unit mass object moves on a straight line from a given initial position $x_1(0)$ and velocity $x_2(0)$. Find the force $\{u(t) \mid t \in [0,1]\}$ that brings the object at time 1 to rest $[x_2(1) = 0]$ at position $x_1(1) = 0$, and minimizes

$$\int_0^1 \left(u(t)\right)^2 dt.$$

### 3.9

Use the Minimum Principle to solve the linear-quadratic problem of Example 3.2.2. *Hint:* Follow the lines of Example 3.3.3.

### 3.10 (On the Need for Convexity Assumptions)

Solve the continuous-time problem involving the system $\dot{x}(t) = u(t)$, the terminal cost $\big(x(T)\big)^2$, and the control constraint $u(t) = -1 \text{ or } 1$ for all $t$, and show that the solution satisfies the Minimum Principle. Show that, depending on the initial state $x_0$, this may not be true for the discrete-time version involving the system $x_{k+1} = x_k + u_k$, the terminal cost $x_N^2$, and the control constraint $u_k = -1 \text{ or } 1$ for all $k$.

### 3.11

Use the discrete-time Minimum Principle to solve Exercise 1.14 of Chapter 1, assuming that each $w_k$ is fixed at a known deterministic value.

### 3.12

Use the discrete-time Minimum Principle to solve Exercise 1.15 of Chapter 1, assuming that $\gamma_k$ and $\delta_k$ are fixed at known deterministic values.

### 3.13 (Lagrange Multipliers and the Minimum Principle)

Consider the discrete-time optimal control problem of Section 3.3.3, where there are no control constraints ($U = \Re^m$). Introduce a Lagrange multiplier vector $p_{k+1}$ for each of the constraints

$$f_k(x_k, u_k) - x_{k+1} = 0, \qquad k = 0, \ldots, N-1,$$

and form the Lagrangian function

$$g_N(x_N) + \sum_{k=0}^{N-1} \big(g_k(x_k, u_k) + p_{k+1}'\big(f_k(x_k, u_k) - x_{k+1}\big)\big)$$

(cL Appendix B). View both the state and the control vectors as the optimization variables of the problem, and show that by differentiation of the Lagrangian function with respect to $x_k$ and $u_k$, we obtain the discrete-time Minimum Principle.

4

# *Problems with Perfect State Information*

Contents

In this chapter we consider a number of applications of discrete-time stochastic optimal control with perfect state information. These applications are special cases of the basic problem of Section 1.2 and can be addressed via the DP algorithm. In all these applications the stochastic nature of the disturbances is significant. For this reason, in contrast with the deterministic problems of the preceding two chapters, the use of closed-loop control is essential to achieve optimal performance.

## 4.1 LINEAR SYSTEMS AND QUADRATIC COST

In this section we consider the special case of a linear system

$$x_{k+1} = A_k x_k + B_k u_k + w_k, \qquad k = 0,1,\dots,N-1,$$

and the quadratic cost

$$\underset{\substack{w_k \\ k=0,1,\dots,N-1}}{E} \left\{ x_N' Q_N x_N + \sum_{k=0}^{N-1} (x_k' Q_k x_k + u_k' R_k u_k) \right\}.$$

In these expressions, $x_k$ and $u_k$ are vectors of dimension n and m, respectively, and the matrices $A_k$, $B_k$, $Q_k$, $R_k$ are given and have appropriate dimension. We assume that the matrices $Q_k$ are positive semidefinite symmetric, and the matrices $R_k$ are positive definite symmetric. The controls $u_k$ are unconstrained. The disturbances $w_k$ are independent random vectors with given probability distributions that do not depend on $x_k$ and $u_k$. Furthermore, each $w_k$ has zero mean and finite second moment.

The problem described above is a popular formulation of a regulation problem whereby we want to keep the state of the system close to the origin. Such problems are common in the theory of automatic control of a motion or a process. The quadratic cost function is often reasonable because it induces a high penalty for large deviations of the state from the origin but a relatively small penalty for small deviations. Also, the quadratic cost is frequently used, even when it is not entirely justified, because it leads to a nice analytical solution. A number of variations and generalizations have similar solutions. For example, the disturbances $w_k$ could have nonzero means and the quadratic cost could have the form

$$E\left\{ (x_N - \bar{x}_N)' Q_N (x_N - \bar{x}_N) + \sum_{k=0}^{N-1} \left( (x_k - \bar{x}_k)' Q_k (x_k - \bar{x}_k) + u_k' R_k u_k \right) \right\},$$

which expresses a desire to keep the state of the system close to a given trajectory $(\bar{x}_0, \bar{x}_1, \dots, \bar{x}_N)$ rather than close to the origin. Another generalized version of the problem arises when $A_k$, $B_k$ are independent random



matrices, rather than being known. This case is considered at the end of this section.

Applying now the DP algorithm, we have

$$J_N(x_N) = x_N' Q_N x_N,$$

$$J_k(x_k) = \min_{u_k} E\left\{ x_k' Q_k x_k + u_k' R_k u_k + J_{k+1}(A_k x_k + B_k u_k + w_k) \right\}. \quad (4.1)$$

It turns out that the cost-to-go functions $J_k$ are quadratic and as a result the optimal control law is a linear function of the state. These facts can be verified by straightforward induction. We write Eq. (4.1) for $k = N - 1$,

$$J_{N-1}(x_{N-1}) = \min_{u_{N-1}} E\{ x_{N-1}' Q_{N-1} x_{N-1} + u_{N-1}' R_{N-1} u_{N-1} + (A_{N-1} x_{N-1} + B_{N-1} u_{N-1} + w_{N-1})' Q_N (A_{N-1} x_{N-1} + B_{N-1} u_{N-1} + w_{N-1}) \},$$

and we expand the last quadratic form in the right-hand side. We then use the fact $E\{w_{N-1}\} = 0$ to eliminate the term $E\{w_{N-1}' Q_N (A_{N-1} x_{N-1} + B_{N-1} u_{N-1})\}$, and we obtain

$$J_{N-1}(x_{N-1}) = x_{N-1}' Q_{N-1} x_{N-1} + \min_{u_{N-1}} \left[ u_{N-1}' R_{N-1} u_{N-1} + u_{N-1}' B_{N-1}' Q_N B_{N-1} u_{N-1} + 2 x_{N-1}' A_{N-1}' Q_N B_{N-1} u_{N-1} \right] + x_{N-1}' A_{N-1}' Q_N A_{N-1} x_{N-1} + E\{ w_{N-1}' Q_N w_{N-1} \}.$$

By differentiating with respect to $u_{N-1}$ and by setting the derivative equal to zero, we obtain

$$(R_{N-1} + B_{N-1}' Q_N B_{N-1}) u_{N-1} = -B_{N-1}' Q_N A_{N-1} x_{N-1}.$$

The matrix multiplying $u_{N-1}$ on the left is positive definite (and hence invertible), since $R_{N-1}$ is positive definite and $B_{N-1}' Q_N B_{N-1}$ is positive semidefinite. As a result, the minimizing control vector is given by

$$u_{N-1}^* = -(R_{N-1} + B_{N-1}' Q_N B_{N-1})^{-1} B_{N-1}' Q_N A_{N-1} x_{N-1}.$$

By substitution into the expression for $J_{N-1}$, we have

$$J_{N-1}(x_{N-1}) = x_{N-1}' K_{N-1} x_{N-1} + E\{ w_{N-1}' Q_N w_{N-1} \},$$

where by straightforward calculation, the matrix $K_{N-1}$ is verified to be

$$K_{N-1} = A_{N-1}' (Q_N - Q_N B_{N-1} (B_{N-1}' Q_N B_{N-1} + R_{N-1})^{-1} B_{N-1}' Q_N) A_{N-1} + Q_{N-1}.$$

The matrix $K_{N-1}$ is clearly symmetric. It is also positive semidefinite. To see this, note that from the preceding calculation we have for $x \in \Re^n$

$$x'K_{N-1}x = \min_u [x'Q_{N-1}x + u'R_{N-1}u + (A_{N-1}x + B_{N-1}u)'Q_N(A_{N-1}x + B_{N-1}u)].$$

Since $Q_{N-1}$, $R_{N-1}$, and $Q_N$ are positive semidefinite, the expression within brackets is nonnegative. Minimization over $u$ preserves nonnegativity, so it follows that $x'K_{N-1}x \geq 0$ for all $x \in \Re^n$. Hence $K_{N-1}$ is positive semidefinite.

Since $J_{N-1}$ is a positive semidefinite quadratic function (plus an inconsequential constant term), we may proceed similarly and obtain from the DP equation (4.1) the optimal control law for stage $N - 2$. As earlier, we show that $J_{N-2}$ is a positive semidefinite quadratic function, and by proceeding sequentially, we obtain the optimal control law for every $k$. It has the form

$$\mu_k^*(x_k) = L_k x_k, \tag{4.2}$$

where the gain matrices $L_k$ are given by the equation

$$L_k = -(B_k'K_{k+1}B_k + R_k)^{-1}B_k'K_{k+1}A_k,$$

and where the symmetric positive semidefinite matrices $K_k$ are given recursively by the algorithm

$$K_N = Q_N, \tag{4.3}$$

$$K_k = A_k'(K_{k+1} - K_{k+1}B_k(B_k'K_{k+1}B_k + R_k)^{-1}B_k'K_{k+1})A_k + Q_k. \tag{4.4}$$

Just like DP, this algorithm starts at the terminal time $N$ and proceeds backwards. The optimal cost is given by

$$J_0(x_0) = x_0'K_0x_0 + \sum_{k=0}^{N-1} E\{w_k'K_{k+1}w_k\}.$$

The control law (4.2) is simple and attractive for implementation in engineering applications: the current state $x_k$ is being fed back as input through the linear feedback gain matrix $L_k$ as shown in Fig. 4.1.1. This accounts in part for the popularity of the linear-quadratic formulation. As we will see in Chapter 5, the linearity of the control law is still maintained even for problems where the state $x_k$ is not completely observable (imperfect state information).



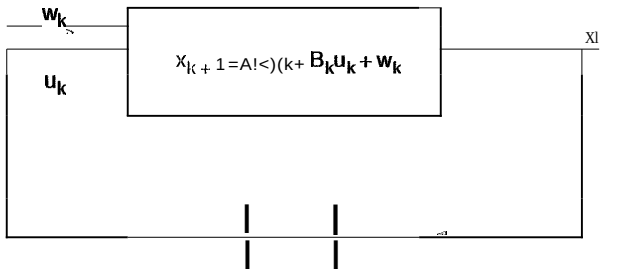



Figure 4.1.1 Linear feedback structure of the optimal controller for the linear-quadratic problem.

## The Riccati Equation and Its Asymptotic Behavior

Equation (4.4) is called the *discrete-time Riccati equation*. It plays an important role in control theory. Its properties have been studied extensively and exhaustively. One interesting property of the Riccati equation is that if the matrices $A_k$, $B_k$, $Q_k$, $R_k$ are constant and equal to $A$, $B$, $Q$, $R$, respectively, then the solution $K_k$ converges as $k \to -\infty$ (under mild assumptions) to a steady-state solution $K$ satisfying the *algebraic Riccati equation*

$$K = A'(K - KB(B'KB + R)^{-1}B'K)A + Q \tag{4.5}$$

This property, to be proved shortly, indicates that for the system

$$x_{k+1} = Ax_k + Bu_k + w_k, \qquad k = 0, 1, \ldots, N-1,$$

and a large number of stages $N$, one can reasonably approximate the control law (4.2) by the control law $\{\mu^*, \mu^*, \ldots, \mu^*\}$, where

$$\mu^*(x) = Lx, \tag{4.6}$$

$$L = -(B'KB + R)^{-1}B'KA,$$

and $K$ solves the algebraic Riccati equation (4.5). This control law is *stationary;* that is, it does not change over time.

We now turn to proving convergence of the sequence of matrices $\{K_k\}$ generated by the Riccati equation (4.4). We first introduce the notions of controllability and observability, which are very important in control theory.

Definition 4.1.1: A pair $(A, B)$, where $A$ is an $n$ x $n$ matrix and $B$ is an $n$ x $m$ matrix, is said to be *controllable* if the $n$ x $nm$ matrix

$$[B, AB, A^2B, \ldots, A^{n-1}B]$$

has full rank (Le., has linearly independent rows). A pair $(A, C)$, where $A$ is an $n$ x $n$ matrix and $C$ an m x $n$ matrix, is said to be *observable* if the pair $(A', C')$ is controllable, where $A'$ and $C'$ denote the transposes of $A$ and $C$, respectively.

One may show that if the pair $(A, B)$ is controllable, then for any initial state $x_0$, there exists a sequence of control vectors $u_0, u_1, \ldots, u_{n-1}$ that force the state $x_n$ of the system

$$x_{k+1} = Ax_k + Bu_k$$

to be equal to zero at time $n$. Indeed, by successively applying the above equation for $k = n-1, n-2, \ldots, 0$, we obtain

$$x_n = A^n x_0 + Bu_{n-1} + ABu_{n-2} + \ldots + A^{n-1}Bu_0$$

or equivalently

$$x_n - A^n x_0 = (B, AB, \ldots, A^{n-1}B) \begin{pmatrix} u_{n-1} \\ u_{n-2} \\ \vdots \\ u_0 \end{pmatrix}. \tag{4.7}$$

If $(A, B)$ is controllable, the matrix $(B, AB, \ldots, A^{n-1}B)$ has full rank and as a result the right-hand side of Eq. (4.7) can be made equal to any vector in $\Re^n$ by appropriate selection of $(u_0, u_1, \ldots, u_{n-1})$. In particular, one can choose $(u_0, u_1, \ldots, u_{n-1})$ so that the right-hand side of Eq. (4.7) is equal to $-A^n x_0$, which implies $x_n = 0$. This property explains the name "controllable pair" and in fact is often used to define controllability.

The notion of observability has an analogous interpretation in the context of estimation problems; that is, given measurements $z_0, z_1, \ldots, z_{n-1}$ of the form $z_k = Cx_k$, it is possible to infer the initial state $x_0$ of the system $x_{k+1} = Ax_k$ in view of the relation

$$\begin{pmatrix} z_{n-1} \\ \vdots \\ z_1 \\ z_0 \end{pmatrix} = \begin{pmatrix} CA^{n-1} \\ \vdots \\ CA \\ C \end{pmatrix} x_0.$$

Alternatively, it can be seen that observability is equivalent to the property that, in the absence of control, if $CA^k \to 0$ then $x_k \to 0$.



The notion of stability is of paramount importance in control theory. In the context of our problem it is important that the stationary control law (4.6) results in a stable closed-loop system; that is, in the absence of input disturbance, the state of the system

$$x_{k+1} = (A + BL)x_k, \qquad k = 0, 1, \ldots,$$

tends to zero as $k \to \infty$. Since $x_k = (A + BL)^k x_0$, it follows that the closed-loop system is stable if and only if $(A + BL)^k \to 0$, or equivalently (see Appendix A), if and only if the eigenvalues of the matrix $(A + BL)$ are strictly within the unit circle.

The following proposition shows that for a stationary controllable system and constant matrices $Q$ and $R$, the solution of the Riccati equation (4.4) converges to a positive definite symmetric matrix $K$ for an arbitrary positive semidefinite symmetric initial matrix. In addition, the proposition shows that the corresponding closed-loop system is stable. The proposition also requires an observability assumption, namely, that $Q$ can be written as $C'C$, where the pair $(A, C)$ is observable. Note that if $r$ is the rank of $Q$, there exists an $r$ x $n$ matrix $C$ of rank $r$ such that $Q = C'C$ (see Appendix A). The implication of the observability assumption is that in the absence of control, if the state cost per stage $x_k'Qx_k$ tends to zero or equivalently $Cx_k \to 0$, then also $x_k \to 0$.

To simplify notation, we reverse the time indexing of the Riccati equation. Thus, $P_k$ in the following proposition corresponds to $K_{N-k}$ in Eq. (4.4). A graphical proof of the proposition for the case of a scalar system is given in Fig. 4.1.2.

Proposition 4.4.1: Let $A$ be an $n$ x $n$ niatrix, $B$ be an $n$ x $m$ matrix, $Q$ be an $n$ x $n$ positive semidefinite symmetric matrix, and $R$ be an m x m positive definite symmetric matrix. Consider the discrete-time Riccati equation

$$P_{k+1} = A'\big(P_k - P_k B(B'P_k B + R)^{-1}B'P_k\big)A + Q, \qquad k = 0, 1, \ldots, \tag{4.8}$$

where the initial matrix $P_0$ is an arbitrary positive semidefinite symmetric matrix. Assume that the pair $(A, B)$ is controllable. Assume also that $Q$ may be written as $C'C$, where the pair $(A, C)$ is observable. Then:

(a) There exists a positive definite symmetric matrix $P$ such that for every positive semidefinite symmetric initial matrix $P_0$ we have

$$\lim_{k\to\infty} P_k = P.$$

Furthermore, P is the unique solution of the algebraic matrix equation

$$P = A'(P - PB(B'PB + R)^{-1}B'P)A + Q \qquad (4.9)$$

within the class of positive semidefinite symmetric matrices.

(b) The corresponding closed-loop system is stable; that is, the eigenvalues of the matrix

$$D \quad A+BL, \qquad (4.10)$$

where

$$L = -(B'PB + R)^{-1}B'PA, \qquad (4.11)$$

are strictly within the unit circle.

Proof: The proof proceeds in several steps. First we show convergence of the sequence generated by Eq. (4.8) when the initial matrix *Po* is equal to zero. Next we show that the corresponding matrix *D* of Eq. (4.10) satisfies $Dk \to 0$. Then we show the convergence of the sequence generated by Eq. (4.8) when *Po* is any positive semidefinite symmetric matrix, and finally we show uniqueness of the solution of Eq. (4.9).

*Initial MatTix Po* = 0. Consider the optimal control problem of finding $u_0, u_1, \ldots, u_{k-1}$ that minimize

$$\sum_{i=0}^{k-1}(x_i'Qx_i + u_i'Ru_i)$$

subject to

$$x_{i+1} = Ax_i + Bu_i, \qquad i = 0,1, \ldots, k \quad 1,$$

where $x_0$ is given. The optimal value of this problem, according to the theory of this section, is $x_0'P_k(0)x_0$,

where *Pk(O)* is given by the Riccati equation (4.8) with *Po* = 0. For any control sequence $(u_0, u_1, \ldots, u_k)$ we have

$$\sum_{i=0}^{k-1}(x_i'Qx_i + u_i'Ru_i) \le \sum_{i=0}^{k}(x_i'Qx_i + u_i'Ru_i)$$

and hence

$$\begin{aligned} x_0'P_k(0)x_0 &= \min_{u_i,\ i=0,\ldots,k-1} \sum_{i=0}^{k-1}(x_i'Qx_i + u_i'Ru_i) \\ &\le \min_{u_i,\ i=0,\ldots,k} \sum_{i=0}^{k}(x_i'Qx_i + u_i'Ru_i) \\ &= x_0'P_{k+1}(0)x_0, \end{aligned}$$



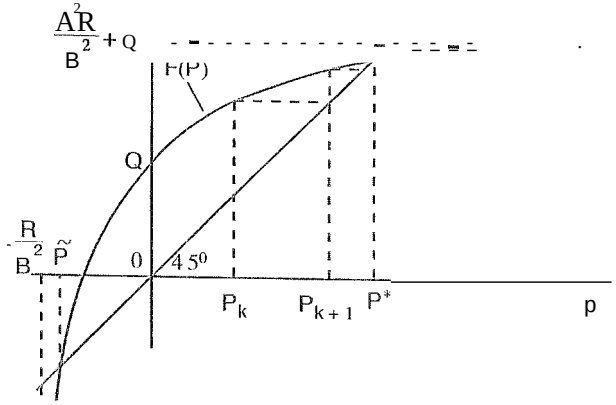



**Figure** 4.1.2 Graphical proof of Prop. 4.4.1 for the case of a scalar stationary system (one-dimensional state and control), assuming that $A \neq 0$, $B \neq 0$, $Q > 0$, and $R > 0$. The Riccati equation (4.8) is given by

$$P_{k+1} = A^2\left(P_k - \frac{B^2P_k^2}{B^2P_k + R}\right) + Q,$$

which can be equivalently written as

$$P_{k+1} = F(P_k),$$

where the function $F$ is given by

$$F(P) = \frac{A^2RP}{B^2P + R} + Q.$$

Because $F$ is concave and monotonically increasing in the interval $(-R/B^2, \infty)$, as shown in the figure, the equation $P = F(P)$ has one positive solution $P^*$ and one negative solution $\tilde{P}$. The Riccati iteration $P_{k+1} = F(P_k)$ converges to $P^*$ starting anywhere in the interval $(P, \infty)$ as shown in the figure.

where both minimizations are subject to the system equation constraint $x_{i+1} = Ax_i + Bu_i$. Furthermore, for a fixed $x_0$ and for every $k$, $x_0'P_k(0)x_0$ is bounded from above by the cost corresponding to a control sequence that forces *Xo* to the origin in *n* steps and applies zero control after that. Such a sequence exists by the controllability assumption. Thus the sequence $\{x_0'P_k(O)xo\}$ is nonc1ecreasing with respect to $k$ and bounded from above, and therefore converges to some real number for every $x_0 \in \Re^n$. It follows that the sequence $\{Ph(O)\}$ converges to some matrix P in the sense that each of the sequences of the elements of *Pk(O)* converges to the corresponcl-

ing elements of $P$. To see this, take $XQ = (1,0, \ldots ,0)$. Then $x_0' P_k(0) x_0$ is equal to the first diagonal element of $Pk(O)$, so it follows that the sequence offlrst diagonal elements of $Pk(O)$ converges; the limit of this sequence is the first diagonal element of $P$. Similarly, by taking $XQ = (0, \ldots ,0,1,0, \ldots ,0)$ with the 1 in the ith coordinate, for i = 2, ... ,n, it follows that all the diagonal elements of $Pk(0)$ converge to the corresponding diagonal elements of $P$. Next take $XQ = (1, 1, 0, \ldots ,0)$ to show that the second elements of the first row converge. Continuing similarly, we obtain

$$\lim_{k \to \infty} Pk(O) = P,$$

where $Pk(O)$ are generated by Eq. (4.8) with $PQ = 0$. Furthermore, since $Pk(O)$ is positive semidefinite and symmetric, so is the limit matrix $P$. Now by taking the limit in Eq. (4.8) it follows that $P$ satisfies

$$P = A'(P - PB(B'PB + R)\text{-}lB'P)A + Q.$$

In addition, by direct calculation we can verify the following useful equality

$$P \quad D'PD + Q + L'RL, \tag{4.12}$$

where $D$ and $L$ are given by Eqs. (4.10) and (4.11). An alternative way to derive this equality is to observe that from the DP algorithm corresponding to a finite horizon N we have for all states $xN\text{-}k$

$$x'_{N-k} P_{k+1}(0) x_{N-k} = x'_{N-k} Q x_{N-k} + \mu^*_{N-k}(x_{N-k})' R \mu^*_{N-k}(x_{N-k}) + x'_{N-k+1} Pk(O) XN\text{-}k+l.$$

By using the optimal controller expression $\mu^*_{N-k}(x_{N-k}) = LN\text{-}kXN\text{-}k$ and the closed-loop system equation $XN\text{-}k+l = (A + BLN\text{-}k)XN\text{-}k$, we thus obtain

$$P_{k+1}(0) = Q + L'_{N-k} R L_{N-k} + (A + BL_{N-k})' P_k(0)(A + BL_{N-k}). \tag{4.13}$$

Equation (4.12) then follows by taking the limit as $k \to$ oo in Eq. (4.13).

*Stability of the Closed-Loop System.* Consider the system

$$x_{k+1} = (A + BL)x_k = Dx_k \tag{4.14}$$

for an arbitrary initial state $XQ$. We will show that $Xk \to$ as $k \to$ oo. We have for all $k$, by using Eq. (4.12),

$$x'_{k+1} P x_{k+1} - x'_k P x_k = x'_k (D'PD - P) Xk = -x'_k (Q + L'RL) x_k.$$

Hence

$$x'_{k+1} P x_{k+1} = x'_0 P x_0 - \sum_{i=Q}^{k} x'_i (Q + L'RL) Xi' \tag{4.15}$$



rrhe left-hand side of this equation is bounded below by zero, so it follows that

$$\lim_{k \to \infty} x'_k (Q + L'RL) Xk = 0.$$

Since $R$ is positive definite and $Q$ may be written as $C'C$, we obtain

$$\lim_{k \to \infty} OXk = 0, \qquad \lim_{k \to \infty} LXk = \lim_{k \to \infty} \mu^*(x_k) = 0. \tag{4.16}$$

The preceding relations imply that as the control asymptotically becomes negligible, we have $\lim_{k \to \infty} OXk = 0$, and in view of the observability assumption, this implies that $Xk \to 0$. To express this argument more precisely, let us use the relation $Xk+l = (A + BL)x_k$ [d. Eq. (4.14)], to write

$$\begin{vmatrix} C\left(x_{k+n-1} - \sum_{i=1}^{n-1} A^{i-1} BLXk+n\text{-}i\text{-}l\right) \\ C\,(Xk+n\text{-}2 \quad \sum_{i=1}^{n-2} A^{i-1} BLXk+n\text{-}i\text{-}2) \\ \\ C(Xk+l - BLxk) \\ OXk \end{vmatrix} \begin{pmatrix} CA^{n-1} \\ CA^{n-2} \\ \vdots \\ CA \\ C \end{pmatrix} x_k. \tag{4.17}$$

Since $LXk \to$ by Eq. (4.16), the left-hand side tends to zero and hence the right-hand side tends to zero also. By the observability assumption, however, the matrix multiplying $Xk$ on the right side of (4.17) has full rank. It follows that $x_k \to 0$.

*Positive Definiteness of P.* Assume the contrary, i.e., there exists some $XQ \neq$ such that $x'_0 P x_0 = 0$. Since $P$ is positive semidefinite, from Eq. (4.15) we obtain

$$x'_k (Q + L'RL) x_k = 0, \qquad \text{k=O,1, ...}$$

Since $Xk \to 0$, we obtain $x'_k Q x_k = x'_k C'C x_k$ and $x'_k L'RL x_k = 0$, or

$$Cx_k = 0, \qquad Lx_k = 0, \qquad k = 0,1, \ldots$$

Thus all the controls $\mu^*(Xk) = LXk$ of the closed-loop system are zero while we have $CXk =$ for all $k$. Based on the observability assumption, we will show that this implies $XQ = 0$, thereby reaching a contradiction. Indeed, consider Eq. (4.17) for $k = 0$. By the preceding equalities, the left-ha,ncl side is zero and hence

$$0 = \begin{pmatrix} CA^{n-1} \\ \vdots \\ OA \\ C \end{pmatrix} xo.$$

Since the matrix multiplying $XQ$ above has full rank by the observability assumption, we obtain $XQ = 0$, which contradicts the hypothesis $XQ \neq 0$ and proves that $P$ is positive definite.

*Arbitrary Initial Matrix Po.* Next we show that the sequence of matrices $\{P_k(P_0)\}$, defined by Eq. (4.8) when the starting matrix is an arbitrary positive semidefinite symmetric matrix $P_0$, converges to $P$ $\lim_{k\to\infty} P_k(0)$. Indeed, the optimal cost of the problem of minimizing

$$x_k' P_0 x_k + \sum_{i=0}^{k-I} (x_i' Q x_i + u_i' R u_i) \tag{4.18}$$

subject to the system equation $Xi+I = AX'i + BUi$ is equal to $x_0' P_k(P_0) x_0$. Hence we ha,ve for every $Xo \in \Re^n$

$$x_0' P_k(0) x_0 \leq x_0' P_k(P_0) x_0.$$

Consider now the cost (4.18) corresponding to the controller $\mu(x_k) = Uk = Lx_k$, where $L$ is defined by Eq. (4.11). This cost is

$$x_0' \left( D^{k'} P_0 D^k + \sum_{i=0}^{k-1} D^{i'} (Q + L'RL) D^i \right) x_0$$

and is greater or equal to $x_0' P_k(P_0) x_0$, which is the optimal value of the cost (4.18). Hence we have for all $k$ and $x \in \Re^n$

$$x' Pk(O) x \leq x' P_k(P_0) x \leq x' \left( D^{k'} P_0 D^k + \sum_{i=0}^{k-1} Di'(Q + L'RL) Di \right) x.$$

We have proved that

$$\lim_{k\to\infty} P_k(0) = P,$$

and we also have, using the fact $\lim_{k\to\infty} Dk' PoDk = 0$, and the relation $Q + L'RL \quad P - D'PD$ fcf. Eq. (4.12)),

$$\begin{aligned} \lim_{k\to\infty} \Big\{ D^{k'} P_0 D_k + \sum_{i=0}^{k-1} D^{i'} (Q + L'RL) Di \Big\} & \\ = \lim_{k\to\infty} \Big\{ \sum_{i=0}^{k-1} Di'(Q + L'RL) Di \Big\} & \\ = \lim_{k\to\infty} \Big\{ \sum_{i=0}^{k-1} D^{i'} (\mathrm{P} - \mathrm{D'PD}) Di \Big\} & \\ p. & \end{aligned} \tag{4.19}$$

Combining the preceding three equations, we obtain

$$\lim_{k\to\infty} Pk(P_O) = P,$$



for an arbitrary positive semidefinite symmetric initial matrix Po.

*Uniqueness of Solution.* If $\tilde{P}$ is another positive semidefinite symmetric solution of the algebraic Riccati equation (4.9), we have $P_k(\tilde{P}) = \tilde{P}$ for all $k = 0,1, \ldots$ From the convergence result.just proved, we then obtain

$$\lim_{k\to\infty} P_k(\tilde{P}) = P,$$

implying that $\tilde{P} \quad P$. **Q.E.D.**

The assumptions of the preceding proposition can be relaxed somewhat. Suppose that, instead of controllability of the pair $B)$, we assume that the system is *stabilizable* in the sense that there exists an $m$ x $n$ feedback gain matrix $G$ such that the closed-loop system $x_{k+1} \quad (A + BG)x_k$ is stable. Then the proof of convergence of $Pk(O)$ to some positive semidefinite $P$ given previously carries through. [We use the stationary control law $\mu(x) = Gx$ for which the closed-loop system is stable to ensure that $x_0' P_k(0) x_0$ is bounded.] Suppose that, instead of observability of the pair $(A, 0)$, the system is assumed *detectable* in the sense that $A$ is such that if $u_k \to 0$ and $OXk \to 0$ then it follows that $Xk \to 0$. (This essentially means that instability of the system can be detected by looking at the measurement sequence $\{z_k\}$ with $zk = Cx_k$.) Then Eq. (4.16) implies that $Xk \to 0$ and that the system $x_{k+1} = (A + BL)Xk$ is stable. The other parts of the proof of the proposition follow similarly, with the exception of positive definiteness of $P$, which cannot be guaranteed anymore. (As an example, take $A = 0$, $B \quad 0$, $0 = 0$, $R > 0$. Then both the stabilizability and the detectability assumptions are satisfied, but $P = 0$.)

To summarize, if the controllability and observability assumptions of the proposition are replaced by the preceding stabilizability and detectability assumptions, the conclusions of the proposition hold with the exception of positive definiteness of the limit matrix $P$, which can now only be guaranteed to be positive semidefinite.

## Random System Matrices

We consider now the case where $\{A_O, B_O\}, \ldots, \{AN\text{-}I, B_{N-1}\}$ are not known but rather are independent random matrices that are also independent of $'WO, 'WI, \ldots, 'WN-1$. Their probability distributions are given, and they are assumed to have finite second moments. This problem falls again within the framework of the basic problem by considering as disturbance at each time $k$ the triplet $(A_k, B_k, w_k)$. The DP algorithm is written as

$$J_N(x_N) = x_N' Q_N x_N,$$

$$J_k(x_k) = \min_{u_k} \underset{w_k, A_k, B_k}{E} \left\{ x_k' Q_k x_k + u_k' R_k u_k + J_{k+1}(A_k x_k + B_k u_k + w_k) \right\}.$$