Mathematical Aspects of Scheduling and Applications

by

R. BELLMAN

A. O. ESOGBUE

and

I. NABESHIMA

Mathematical Aspects
of
Scheduling and Applications

by

R. BELLMAN
University of Southern California
Los Angeles, USA

A. O. ESOGBUE
Georgia Institute of Technology
Atlanta, USA

and

I. NABESHIMA
University of Electro-Communications
Tokyo, Japan

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

UK	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
USA	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon Press Canada Ltd., Suite 104, 150 Consumers Road, Willowdale, Ontario M2J 1P9, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., P.O. Box 544, Potts Point, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, 6242 Kronberg-Taunus, Hammerweg 6, Federal Republic of Germany

First edition 1982

Library of Congress Cataloging in Publication Data
Bellman, Richard Ernest, 1920–
Mathematical aspects of scheduling and applications.
(International series in modern applied mathematics
and computer science; v. 4)
Includes bibliographies.
1. Scheduling (Management) I. Esogbue, Augustine O.
II. Nabeshima, Ichiro, 1921– III. Title. IV. Series.
TS157.5.B44 1982 658.5′3 81–15809

British Library Cataloguing in Publication Data
Bellman, R.
Mathematical aspects of scheduling and applications
(International series in modern applied
mathematics and computer science; v. 4)
1. Scheduling (Management)—Mathematics
I. Title II. Esogbue, A. O. III. Nabeshima, I.
IV. Series
658.5′3 TS157.5

ISBN 0–08–026477–8 Hardcover
ISBN 0–08–026476–X Flexicover

Printed in Great Britain by A. Wheaton & Co. Ltd., Exeter

Preface

THROUGHOUT the two major epochs in human history, in both the mechanical age and now the systems age, mankind has been beset with the arduous problem of optimal utilization of the usually limited resources in accomplishing the variegated tasks or objectives. The problem is thus not new although contemporary manifestations of this problem in forms such as the energy crisis or inflation may proffer it that coloration. Further, the problem appears recurrent and eternal. An attempt to cope with such problems demands that one develop some plan or schedule. The analysis and study of such plans particularly, with respect to their optimality and other properties under various scenarios of objectives and constraints, constitutes an exciting field known as Scheduling Theory. This theory thus deals with a concretization of continuing human experience.

The purpose of the book is, however, not a prosaic rendition of history. Rather, it is an attempt to syphon out of both the mechanical age (when man was building machines to analyze and do work) and the systems age (when the synthesis of the total problem and interrelationships are the focus of attention) certain pervasive problems that occur and are important enough to warrant the focus of our attention. We then present the kernel of these problems in mathematical terms and devise mathematical solutions to them. As always, we study their intrinsic properties and, whenever we can, generalize to other situations. Similar objectives appear in numerous works in the field. What we have done that is unique is to integrate our experiences of over two decades of research work in the field, from different vistas of experience with problems and techniques into one compendium.

The book is particularly timely. First, as we said, the problem of scheduling is a perennial one and nearly as old as history. Second, mankind is currently going through a special period in which critical shortages of important resources needed to continue civilization have been flung to the limelight. The plea to conserve is not only a plea to schedule one's consumption. It is more than that. It is also a plea for efficient utilization of available resources including optimal combinations of resources. These are issues of central interest in scheduling theory. Third, the existing books in the field are either considerably outdated or are severely limited in their scope of treatment. Fourth, the problem of optimal systems design and control bears striking interdependence with that of optimal scheduling.

We have attempted to capture the essential problems in scheduling as well as treat them via the most effective techniques. We have introduced material in this book that are considerably current, novel, and well researched. Sometimes, these techniques and problems have not even appeared in the open literature. Other times, they may have only appeared in the form of technical reports of some research institute or in scientific journals. Further, we have exploited our familiarity with the literature of such fields as mathematics, economics, operations research, management science, computer science, engineering and medicine, to motivate our problems and methods of attack.

This book does not and is not intended to discuss everything there is to know about scheduling. For example, dynamic problems involving stochastic arrival times of jobs are not considered. Inclusion of such topics here would have led to a much lengthier book than this one or to the sacrifice of the depth of coverage of each of the topics in this book. We decided to err on the side of depth than breadth.

This book is intended as a graduate text or reference work in a course usually entitled Scheduling Theory or Control Theory in most universities' departments of mathematics, operations research, management science, computer science or engineering. It will also be useful to economists and planners. It gives an almost didactic treatment of the subject. The central techniques employed can be picked up along the way although previous familiarity with dynamic programming and integer programming is useful particularly with the advanced algorithms. The reader will find the exercises at the end of each chapter instructive and challenging. The comments and bibliography section will be useful in pursuing an in-depth study of the material covered in each chapter.

The book consists of twelve chapters. It begins with network problems—a very special subgroup of mathematical programming problems. The shortest-path problem, one of the fundamental problems in mathematics, is the problem of tracing a path of shortest length through a finite network. This problem occurs in many fields and we devote the first two chapters to it. In Chapter 1, we show how a simple dynamic programming argument yields a fundamental nonlinear equation. Thus, dynamic programming converts a combinatorial problem into an analytic one. Simple approximate techniques yield upper and lower bounds for the solution of this equation. These bounds can be easily interpreted in terms of approximate policies. We also exploit such bounds and other fathoming criteria suggested by branch and bound strategies to solve large-scale traveling-salesman problems. Applications to control theory and other parts of the calculus of variations are also discussed. In Chapter 2 we show that this problem can be used to treat many problems in artificial intelligence and that many popular mathematical games can be interpreted in these terms. The role and use of computers is stressed.

In Chapter 3 we classify scheduling problems and briefly discuss their solution approaches. The machine scheduling problem is the problem of processing n items on m machines in an efficient manner. The problem is remarkably difficult. We discuss the many methods that can be applied, and give many examples. Among these methods are the branch and bound method, backtrack programming, dynamic programming, and various combinatorial methods.

The only analytic result known is due to Selmer Johnson for the case of two machines. If there are more than two machines, it is quite likely that no simple analytic result exists. However, for the permutation flow-shop problem with makespan objective, there have been many efficient analytic results under some restricting conditions, as described in Chapter 7.

Since these are finite problems, it might be thought they could be handled by enumeration, particularly with the speed of the digital computer. This is not the case, mathematical analysis is definitely needed, to identify and/or extend the solvable cases. Also we require an approximation method with guaranteed accuracy, or with simple efficient heuristics.

To see this, two numbers are convenient to keep in mind. First of all, $10! = 3{,}628{,}800$; secondly, a year has approximately 3×10^7 seconds.

Consequently, if each case takes one second to examine, we see that 10! cases takes about a month. Since 20! is more than $10^{10} \times 10!$, we see that 20! cases cannot be examined by enumeration. Even if each case requires only a microsecond, it is not feasible to examine 20! cases. In combinatorial problems of this type, it is not uncommon to meet numbers such as 100! or 1000!. Thus mathematical analysis, whenever possible, helps us reduce the computational drain tremendously.

Chapters 4 through 6 discuss different problems and techniques of machine sequencing. Chapter 5 discusses capacity expansion problems and introduces the new technique of imbedded state space dynamic programming for reducing dimensionality so that larger problems can be solved. Chapter 6 considers an important class of network problems with nonserial phase structures and exploits dimensionality reduction techniques such as the pseudo-stage concept, branch compression and optimal order elimination methods to solve large-scale nonlinear network scheduling problems.

In Chapters 7 through 11 we consider the flow-shop scheduling problem under different objectives and constraints. We present several novel analytic results and employ various ingenious techniques of branch and bound including backtrack programming, lexicographical search method and unified multi-stage combinatorial algorithms. Applications to the increasingly important area of parallel processing are discussed. Approximate solutions for these interesting cases are also presented. In Chapter 11 we address machine scheduling problems involving sequence-dependent set-up times and present novel efficient dynamic programming formulations especially for the three-machine problem.

The book is concluded with a chapter on the job-shop-scheduling problem. Using the disjunctive graph viewpoint and the techniques of BAB, EXTBAB, dynamic programming, important analytical and computational results are derived.

What we have attempted in this work is to present a unified treatment of the many problems and techniques of solution. We want to point out how many problems exist in these domains and what opportunities there are. These are new parts of applied mathematics. What is particularly interesting about these problems is that they require a blend of analysis, algebra, topology, computer science, and a knowledge of how the problems arise.

We are grateful to a long list of friends, colleagues, and graduate students who have collaborated with us in various phases of our research efforts. In particular, we acknowledge the contributions of Burton Corwin and Thomas Morin.

RICHARD BELLMAN
Santa Monica, California

AUGUSTINE ESOGBUE
Atlanta, Georgia

ICHIRO NABESHIMA
Chofu-shi, Tokyo, Japan

Contents

CHAPTER 1

Network Flow, Shortest Path and Control Problems

1.1. Introduction

In this chapter we wish to discuss briefly an interesting class of problems with wide ramifications. On the one hand, they can be viewed as abstractions and natural generalizations of a control process. Furthermore, they provide us with an entry into the rugged terrain of scheduling processes and into the study of some other types of decision processes of combinatorial type, such as pattern recognition and numerous other problems arising in the field of artificial intelligence. The application of these ideas to artificial intelligence will be given in Chapter 2. Finally, in our continuing quest for feasible computational procedures, they furnish considerable motivation for the creation of sophisticated decomposition techniques based upon topological considerations. We will do this in Section 1.16 when we make an application of these techniques to the calculus of variations. Let us also mention that questions of this nature occur more and more frequently in connection with the execution of complex computer problems.

In abstract terms, we are interested in a discrete control process of the following type: Let p be a generic element of a finite set S and $T(p, q)$ be a family of transformations with the property that $T(p, q) \in S$ whenever $p \in S$ and $q \in D$, the decision space, again taken to be discrete. We wish to determine a sequence of decisions, $q_1, q_2, \ldots,$ which transform p, the initial state, into p_N, a specified state, in an optimal fashion.

We are really interested in questions of feasibility. However, in order to handle this imprecise concept we narrow our sights and consider optimization.

A problem of great contemporary interest, the "routing problem", is a particular case of the foregoing. We will use it as our leitmotif.

1.2. The Routing Problem

Consider a set of N points, numbered $1, 2, \ldots, N$, with N the terminal point, as shown in Fig. 1.1.

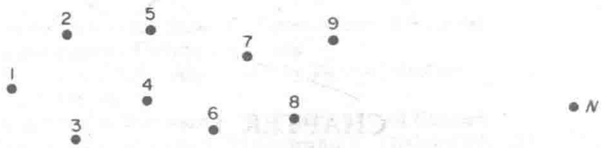

FIG. 1.1

We suppose that there exists a direct link between any two points i and j which requires a time t_{ij} to traverse, with $t_{ii} = 0$, $i = 1, 2, \ldots, N$. In all that follows we take $t_{ij} \geq 0$. What path do we pursue, starting at 1, passing through some subsets of the points $2, 3, \ldots, N-1$, and ending at N, which requires a minimum time?

Two possible paths are shown in Figs. 1.2 and 1.3. The first goes directly to N; the second goes through every point always moving from a lower order number to a higher order one before reaching N.

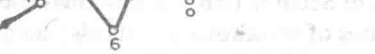

FIG. 1.2

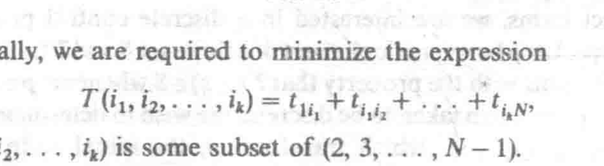

FIG. 1.3

Analytically, we are required to minimize the expression

$$T(i_1, i_2, \ldots, i_k) = t_{1i_1} + t_{i_1 i_2} + \ldots + t_{i_k N}, \tag{1.1}$$

where $(i_1, i_2, \ldots, i_k)$ is some subset of $(2, 3, \ldots, N-1)$.

Exercises

1. Does a minimizing path ever contain a loop, i.e. pass through the same point twice?

2. How many different admissible paths are there?

3. Can the problem presently be solved by direct enumeration with a digital computer for $N = 100$? Assume that we can enumerate and compare paths at the rate of one per microsecond.

1.3. Dynamic Programming Approach

To treat this problem by means of dynamic programming, we imbed the original problem within the family of problems consisting of determining the optimal path from i to N; $i = 1, 2, \ldots, N-1$. Let

$$f_i = \text{minimum time to go from } i \text{ to } N,\ i = 1, 2, \ldots, N-1, \tag{1.2}$$

and set $f_N \equiv 0$. Then the principle of optimality yields the relation

$$f_i = \min_{j \neq i} [t_{ij} + f_j], \quad i = 1, 2, \ldots, N-1, \tag{1.3}$$

with the "boundary condition" $f_N = 0$.

This functional equation is different in form from the usual dynamic programming equation, being implicit rather than explicit. Further, we note its dependence on one (the current node) variable rather than two (the stage and state) variables. The unknown function occurs on both sides of the equation which means that the solution cannot be obtained by any direct iteration. Consequently, some interesting analytic questions arise:

(a) Does (1.3) possess a solution?

(b) Does it possess a unique solution?

(c) Provided that it does possess a unique solution, how can the equation be used to provide the optimal path?

(d) How can the solution be obtained computationally?

The question of computational feasibility is, as usual, the thorniest one and one that forces us to consider new and alternate methods. Some aspects of this question will be discussed now; others will be described in the references.

Exercises

1. In some cases we are interested not only in the path of shortest time, but also in the next best path and generally the kth best path. Let $f_i(2)$ be defined as time required to go from i to N following a second best path. Obtain a functional equation similar in form connecting $f_i(2)$ and f_i.

2. Similarly, obtain a functional equation for $f_i(k)$, the time associated with the kth best path.

3. Obtain a functional equation corresponding to (1.3) for the case where the time to go from i to j depends upon the "direction" in which i is entered, which is to say upon the point from which i is reached.

4. What is the form of the equation if we cannot go from every point to every other point?

1.4. Upper and Lower Bounds

Prior to a demonstration of the existence and uniqueness of the solution of (1.3), let us show how easy it is to obtain upper and lower bounds for the solution (or solutions) of (1.3). An "experimental proof" of uniqueness is thus available in any particular case if we can show computationally that the upper and lower bounds coincide.

Let the sequence $\{\phi_i^{(r)}\}, r = 0, 1, 2, \ldots$, be defined for each i in the following fashion:

$$\phi_i^{(0)} = \min_{j \neq i} t_{ij}, \quad i = 1, 2, \ldots, N-1, \quad \phi_N^{(0)} = 0,$$

$$\phi_i^{(r+1)} = \min_{j \neq i} [t_{ij} + \phi_j^{(r)}], \quad i = 1, 2, \ldots, N-1, \quad \phi_N^{(r+1)} = 0. \tag{1.4}$$

Let us recall that we are assuming that $t_{ij} \geq 0$.

It is clear that $\phi_i^{(r)} \geq 0$ and thus that

$$\phi_i^{(1)} = \min_{j \neq i} [t_{ij} + \phi_j^{(0)}] \geq \min_{j \neq i} t_{ij} = \phi_i^{(0)}. \tag{1.5}$$

Hence, inductively,

$$\phi_i^{(0)} \leq \phi_i^{(1)} \leq \ldots \leq \phi_i^{(r)} \leq \phi_i^{(r+1)} \leq \ldots. \tag{1.6}$$

Let us now show inductively that

$$\phi_i^{(r)} \leq f_i, \tag{1.7}$$

where f_i is any nonnegative solution of (1.3). We have

$$\phi_i^{(0)} = \min_{j \neq i} t_{ij} \leq \min_{j \neq i} [t_{ij} + f_j] = f_i, \tag{1.8}$$

whence (1.7) follows from (1.4) via an induction.

Since the sequence $\{\phi_i^{(r)}\}, r = 1, 2, \ldots$, is uniformly bounded and monotone increasing, we have convergence for each i. Let

$$\phi_i = \lim_{r \to \infty} \phi_i^{(r)}. \tag{1.9}$$

Then (1.7) yields

$$\phi_i \leq f_i, \tag{1.10}$$

and (1.4) shows that ϕ_i is itself a solution. Hence, ϕ_i is a lower bound for any solution.

To obtain an upper bound, let us introduce the sequence $\{\psi_i^{(r)}\}$ where

$$\psi_i^{(0)} = t_{iN}, \quad i = 1, 2, \ldots, N,$$

$$\psi_i^{(r+1)} = \min_{j \neq i} [t_{ij} + \psi_j^{(r)}], \quad i = 1, 2, \ldots, N. \tag{1.11}$$

Then

$$\psi_i^{(0)} = t_{1N} \geq \min_{j \neq i} [t_{ij} + f_j] = f_i, \tag{1.12}$$

and again an induction establishes

$$\psi_i^{(0)} \geq \psi_i^{(1)} \geq \ldots \geq \psi_i^{(r)} \leq f_i. \tag{1.13}$$