MATHEMATICAL ASPECTS OF SCHEDULING THEORY

by

Richard Bellman

P-651

Revised 23 May 1955

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

74

SUMMARY

The purpose of this paper is to formulate a number of significant mathematical problems which have arisen in connection with the theory of scheduling, and to discuss the methods which have been devised to treat these problems. A brief bibliography is appended.

MATHEMATICAL ASPECTS OF SCHEDULING THEORY

by

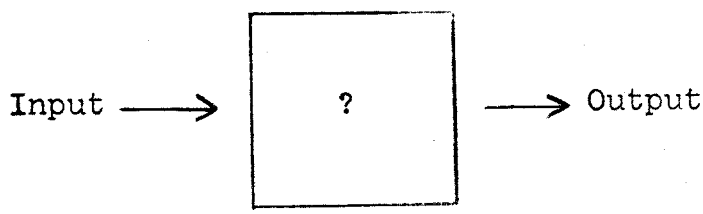
Richard Bellman

Introduction

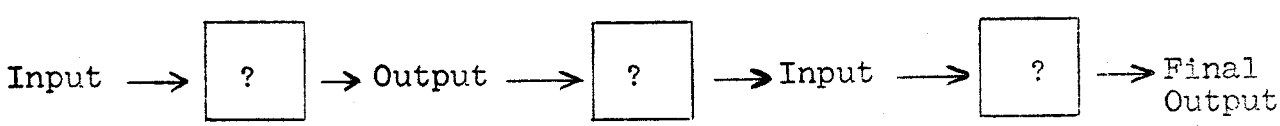
The purpose of this expository paper is to describe a number of representative problems in the field of scheduling which will furnish some idea of the types of mathematical questions which confront us in this domain. In doing so we shall touch upon some of the analytic and computational techniques which have been utilized to date to treat these problems, and thus unavoidably display the deplorable state of the art.

There is little point at this stage of development of the theory in attempting to give any precise definition of what we mean by a "scheduling" problem, since any definition which can include the spectrum of problems arising from all phases of economic and industrial activity, and from parts of mathematics itself ranging from algebra and topology to analysis and mathematical physics, will necessarily be sufficiently vague to be fairly useless. We can, however, make the following general remarks.

In characterizing a physical system, we may begin with a phenomenological approach: a certain cause produces a certain effect, or, alternatively, a certain input produces a certain output. Thus, as a first approach, we use the traditional "black box" concept: the physical system is regarded as inexplicable in its workings, and its mechanism is taken to be completely sealed from view. We agree to consider only the "response curves". Schematically

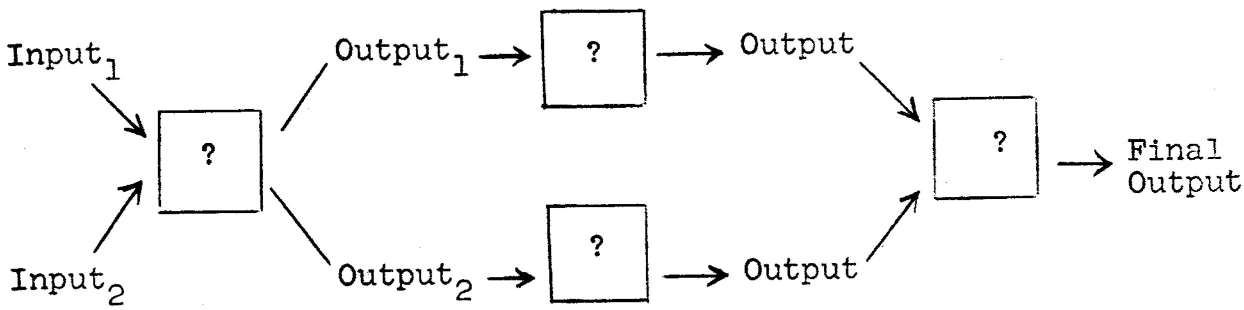


As we become more ambitious, we begin to peer into the murky confines of the black box, and our next step is to divide the original box into a set of small boxes in the following way

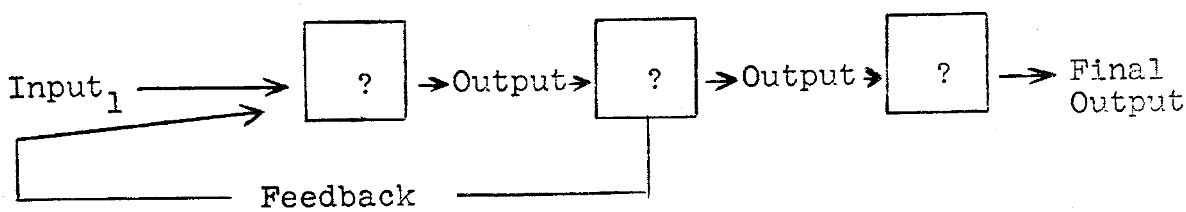


We now have a multi-stage process.

A further degree of sophistication yields more complex block diagrams, such as



Eventually we may be forced to consider even more complicated circuits involving "feedbacks":



Classical physics, as well as classical economics, is concerned with the general problems of determining the output of a system as a function of the input, and of determining the structure of the system. This is an inverse problem of most interesting type. We observe some inputs and some outputs and it is up to us to wield Occam's razor so as to deduce the structures that explain the observations in the most esthetic fashion.

Simultaneously with general problems of the type discussed above, we meet the problems of control. Starting with a fixed structure, we wish to regulate the inputs and outputs so as to achieve some desired state of the system. These problems combine all the difficulties of the problem of the previous stage with the additional complications presented by this new optimization problem. On the other hand, we may note, that in return the optimal system will possess certain simplifying properties.

Continuing in the direction of controlling the system, let us suppose that we have the power to alter the basic structure of the system in various ways. We can interchange the position of various boxes, bypass others, introduce new boxes, and so on. The problem is now to determine the structures which are optimal according to various criteria of efficiency.

Problems of the last two types we call "scheduling problems". The first type, difficult as it is, has been intensively studied in recent years, and a number of mathematical techniques has been devised to treat these questions with some degree of success.

The structural problems involving combinatorial considerations have only recently been studied in an intensive manner. They involve mathematical difficulties of the highest order even in what seem to be the simplest cases. These are the problems we shall focus upon in the following pages.

We have spoken above of input-output curves, tacitly assuming a deterministic situation where the output is a fixed function of the input. In many situations, a convenient mathematical device for by-passing ignorance is to assume that the output is stochastic. In some cases, e.g. statistical mechanics, this simplifies the problem; in other cases, e.g. assembly line production, it greatly complicates it.

Finally, in our hierarchy of problems, let us mention one further stage. We took it for granted in posing the problem of determining optimal arrangements of blocks that we knew the response curves, the outputs as functions of inputs. This is sometimes not true. In some situations we must simultaneously determine response curves and optimal structures. Since the optimal structure depends upon the response curves, and the information we obtain concerning the response curve depends upon the structure we assume, we see that this type of operation is quite involved.

Processes of this type we call "learning processes". They arise in connection with the design of experiments and in the theory of sequential analysis.

Since it is impossible in a survey article of even this size to give any adequate coverage of all or even of some of the most promising aspects of this fascinating and fundamental field, we have compromised by discussing a very few topics in small detail and by referring briefly to a multitude of others, with a bibliography for the interested reader.

Further references to these interesting processes may be found in R. Bellman, "A problem in the Sequential Design of Experiments", RAND Corporation, Paper, P-586, 1954, and R. Bellman, T. E. Harris, and H. N. Shapiro, "Studies in Functional Equations Occurring in Decision Processes", RAND Corporation Paper, P-382, 1952.

In Part I we shall consider a number of simple prototype problems associated with multi-stage production processes, and various mathematical models and methods which have been used to treat these problems. In passing, we shall note the connection between these problems and the problems of organization theory, problems encountered in the theory of switching, and the construction of computing machines, insofar as these are all problems of determining optimal structure.

Part II is devoted to survey of some of the work that has been done on transportation problems in connection with determining most efficient routings. The Hitchcock-Koopmans problem is typical of the questions that arise.

Many of these problems can be treated by techniques developed in the topological theory of graphs and networks due to Konig, Menger, and others, which in turn grew out of the work on electrical network theory by Kirchoff.

Iterative techniques for solving various classes of these problems have been given by M. Flood and A. Boldyreff. In particular, the new technique of Boldyreff, the "flooding" technique, seems to have great possibilities in connection with many similar types of problems. Linear programming techniques are also applicable in many cases.

The third part is devoted to a short discussion of "smoothing" problems of various types. Here we shall meet the "caterer" problem and the "optimal inventory" problem as representative problems in that area, and discuss some closely associated control problems occurring in engineering control.

In Part IV, we mention briefly some scheduling problems arising in computational analysis. Here we have been quite arbitrary in our choice, and any expert on the field can readily add a score more of vital questions.

Part V is devoted to a brief survey of some applications of the theory of linear programming to the field of scheduling theory.

Finally, in Part V we do some crystal-gazing and discuss various parts of mathematics which we feel will have to be intensively cultivated before essential progress can be made in the field of scheduling theory.

We have omitted any detailed discussion of the lower level scheduling processes, such as those which occur in multi-stage productive processes, allocation and investment processes, the calculus of variations, the theory of multi-stage games, and generally any topics which can be treated by mathematical tools of a fairly established type. In a sense then, we are being masochistic and deliberately restricting ourselves to those fields where we must plead "nul contendere".

Part I

Assembly Line and Related Problems

§1. Preliminaries

In this part of the paper we shall consider some problems arising from the study of multi-stage production processes.

The first problem we shall consider involves a two-stage production process in which we have a number of books in manuscript form which must be printed and bound. We wish to determine the order of processing the books which minimizes the total time required to complete all the books. Two solutions will be presented, the original one due to S. Johnson, based upon an explicit formula of interest in itself, and a solution based upon the functional equation approach of the theory of dynamic programming.

Continuing in this direction, we shall consider the analogous three-stage process. Formidable difficulties appear in this case, and there is at the present no solution for the general case. This sad statement will appear repeatedly in the lines below.

The corresponding explicit formula of Johnson for the three-stage process may be used to obtain a continuous version of some particular cases of importance. In the continuous case we can derive a very simple solution which throws some light on the discrete problem.

Following this we shall consider some natural extensions of the simply-formulated problems mentioned above. These problems are closely related to problems in organization theory of the type formulated by J. Marshak, [7], and these in turn are connected to the theory of switching, to the theory of automata, and all similar problems involving the construction of complicated objects from simpler components, [4], [15].

The last section is devoted to the "assignment" problem. This is a simpler problem involving the maximization of a function of permutations*, and may be treated by two methods, due to Egervary, [3], and von Neumann [14]. Here we meet the fundamental technique of imbedding permutation matrices in the semi-group of doubly-stochastic matrices.

§2. The Book-binding Problem

Let us begin by considering the following simple problem. We possess one printing press, one binding machine, and the manuscripts of a number of possibly different books. Assuming that we know the times required to perform the printing and binding operations for each book, we wish to determine the order in which the books should be processed in order to minimize the total time required to turn out all the books.

This is a problem in rearrangements, or permutations, which in any particular case can conceivably be determined by a direct

---

* In the sense that it has been effectively solved, while the others have not been so far.

enumeration of possibilities. Lest anyone too blithely refer to modern computing machines, let us quote some figures.

The number of possible arrangements of ten different items is

(1) $$10! = 3,628,800 \quad ,$$

while the number of possible arrangements of twenty different items is

(2) $$20! = 2,432,902,008,176,640,000 \quad .$$

To give some idea of the magnitude of this number, let us merely observe that at one sorting per microsecond an enumeration of the 20! possibilities would require well over a half a million years. Clearly we must use some native "inwit".

§3. Lemmas of S. Johnson

Let us introduce some notation. Let

(1) $$a_i = \text{time required to print the } i^{\underline{th}} \text{ book,}$$
$$b_i = \text{time required to bind the } i^{\underline{th}} \text{ book,}$$

for $i = 1, 2, \ldots, n$, where n is the number of different books.

We wish to determine the arrangements of the books which yield the minimum time required to print and bind all the books. It is assumed that the printing will precede the binding in each book.

The first result we shall require is

Lemma 1. To obtain a minimal ordering, it is sufficient to consider that the books are processed in the same order through both machines.

A possibility which cannot be discounted without investigation is that it may be more efficient to feed the books into the binding machine in some order different from that used on the printing press. As we shall see below, this result concerning identical ordering carries over to the three-stage process, but not to four-stage processes.

Let us begin by taking the items in chronological order, i.e. the $i^{\underline{th}}$ item in the $i^{\underline{th}}$ position, and derive an expression for the total time required to complete the process.

Define

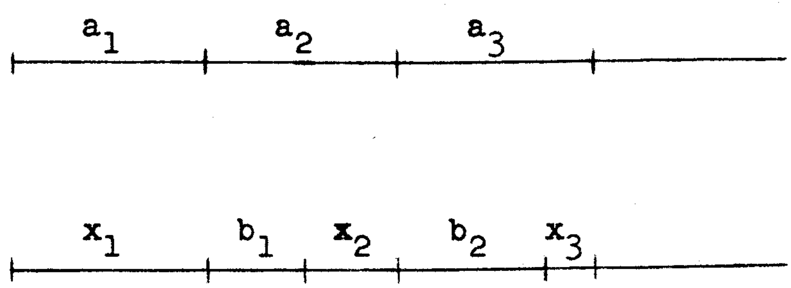
(2) x_i = inactive time on the second machine, the binding machine, immediately before processing the $i^{\underline{th}}$ item.

In many cases, of course, x_i will be zero. The total "idle" time on the second machine is the sum

(3) $$I_n = \sum_{i=1}^{n} x_i$$

and thus the total time required for the process is $I_n + \sum_{i=1}^{n} b_i$. It follows that we may take I_n as our measure of the efficiency of a scheduling of operations, and that it is this quantity which we wish to minimize.

Schematically,



An important result for our purposes is

<u>Lemma</u> 2.

$$(4) \qquad I_n = \sum_{i=1}^{n} x_i = \max_{1 \leq u \leq n} \left(\sum_{i=1}^{u} a_i - \sum_{i=1}^{u-1} b_i \right)$$

<u>Proof</u>: We have

$$(5) \qquad \begin{aligned} x_1 &= a_1 \\ x_2 &= \max(a_1 + a_2 - b_1 - x_1, 0) \end{aligned}$$

whence

$$(6) \qquad x_1 + x_2 = \max(a_1 + a_2 - b_1, a_1)$$

Similarly

$$(7) \qquad x_3 = \max\left(\sum_{i=1}^{3} a_i - \sum_{i=1}^{2} b_i - \sum_{i=1}^{2} x_i,\ 0\right)$$

$$\sum_{i=1}^{3} x_i = \max\left(\sum_{i=1}^{3} a_i - \sum_{i=1}^{2} b_i,\ \sum_{i=1}^{2} x_i\right)$$

$$= \max\left(\sum_{i=1}^{3} a_i - \sum_{i=1}^{2} b_i,\ \sum_{i=1}^{2} a_i - b_1,\ 0\right)$$

It is clear now how the proof proceeds by induction for $n > 3$.

§4. The Result of S. Johnson

The problem of determining the optimal order is equivalent to determining the arrangement of the n items which minimizes the expression for I_n given in (3.4). We thus have a numerical function defined over the permutation group of order n, and we wish to determine the minimum of this function.

The solution is given by

Theorem 1. *The optimal ordering is determined by the following rule: item i precedes item j if*

$$(1) \qquad \min(a_i, b_j) < \min(a_j, b_i)$$

If there is equality, either ordering is optimal, provided that it is consistent with the definite preferences.

For Johnson's derivation of his result and the remainder of the proof that this criterion yields the absolute minimum, we refer to his paper, [5]. We shall give another derivation below.

The importance of this result resides in the fact that it shows that in this problem the minimum time may be found by determining the effect of the interchange of any two neighboring items. This result is not true in general and accounts for the difficulty of the multi-stage problem for more than two stages.

§5. An Example

Let us now illustrate the simple way in which this criterion may be applied. We follow the steps given below:

1. List the a_i and b_i in two vertical columns

i	a_i	b_i
1	a_1	b_1
2	a_2	b_2
⋮		
n	a_n	b_n

2. Determine the minimum of all the a_i and b_i.

3. If it is an a_i, place the corresponding item first.

4. If it is a b_i, place the corresponding item last.

5. Cross off both times for that item.

6. Repeat the steps on the reduced set of $(n-1)$ items.

7. In case of ties, order the items with smallest subscript first, for the sake of definiteness. If a tie between a_i and b_i, order the item according to the a-rule.

To illustrate the method, consider the following example

i	a_i	b_i
1	4	5
2	4	1
3	30	4
4	6	30
5	2	3

The rule yields (5, 1, 4, 3, 2) as the minimal order with a total time of 47 units, and 4 units of idle time. For the reversed order, the total time is 78, the longest time.

§6. An Alternate Derivation of the Decision Function

Let us now give another derivation of the decision function given above in (4.1) using the functional equation approach of the theory of dynamic programming, [1]. It has the merit of yielding this function without the use of the explicit formula, given in (3.4) above.

Let us define

(1) $f(a_1,b_1,a_2,b_2,\ldots,a_N,b_N,t)$ = the time consumed in processing the n items, when the second machine is committed t hours ahead, and an optimal policy is employed.

If the first item is processed first, we have

$$(2)\qquad f(a_1,b_1,a_2,b_2,\ldots,a_N,b_N,\ t) =$$

$$a_1+f\left[0,0,a_2,b_2,\ldots,a_N,b_N,\ b_1+\max(t-a_1,0)\right] \ .$$

Choosing the second item to follow, we obtain

$$(3)\qquad f(a_1,b_1,a_2,b_2,\ldots,a_N,b_N,\ t) =$$

$$a_1+a_2+f\left[0,0,0,0,a_3,b_3,\ldots,a_N,b_N,g(a_1,b_1,a_2,b_2,\ t)\right] \ ,$$

where

$$(4)\qquad g(a_1,b_1,a_2,b_2,\ t) =$$

$$b_2+\max\left[b_1-a_2+\max(t-a_1,0),\ 0\right] \ .$$

On the other hand, interchanging the orders, we obtain

$$(5)\qquad f(a_1,b_1,a_2,b_2,\ldots,a_N,b_N,\ t) =$$

$$a_2+a_1+f\left[0,0,0,0,a_3,b_3,\ldots,a_N,b_N,g(a_2,b_2,a_1,b_1,\ t)\right] \ .$$

It follows from this that the order of operations which minimizes the new t-term is optimal. Hence we choose the order which yields the minimum of $g(a_1,\ b_1,\ a_2,\ b_2,\ t)$ and $g(a_2,b_2,\ a_1,\ b_1,\ t)$. It is not immediately obvious that this choice will be independent of t, and it is actually quite surprising that this is so. We have

$$(6)\qquad \begin{aligned} g(a_1,b_1,a_2,b_2,t) &= b_2+\max\left[b_1-a_2+\max(t-a_1,0),\ 0\right] \\ &= b_1+b_2-a_2+\max\left[\max(t-a_1,0),\ a_2-b_1\right] \\ &= b_1+b_2-a_2+\max\left[t-a_1,\ 0,\ a_2-b_1\right] \\ &= b_1+b_2-a_1-a_2+\max\left[t,\ a_1,\ a_1+a_2-b_2\right] \ . \end{aligned}$$

It follows from this that regardless of the value of t, the order of operations which yields the minimum of max $[a_1, a_1+a_2-b_2]$ and max $[a_2, a_1+a_2-b_1]$ cannot increase the total time, and may decrease it.

It is easily seen that this criterion is equivalent to Johnson's criterion above.

§7. The Three-stage Process

Let us now consider the three stage process obtained from the previous two stage process by the addition of a third required operation, which we can think of as the typing of the manuscript. We now have n books in manuscript form which require successively typing, printing and binding. Let a_i, b_i, c_i represent the respective times required by the i^{th} book for each of these processes.

The first result we require is

Lemma 3. <u>To obtain a minimal ordering, it is sufficient to consider the case where the books are processed in the same order on the three machines.</u>

A double application of Lemma 1 yields this result.

The example

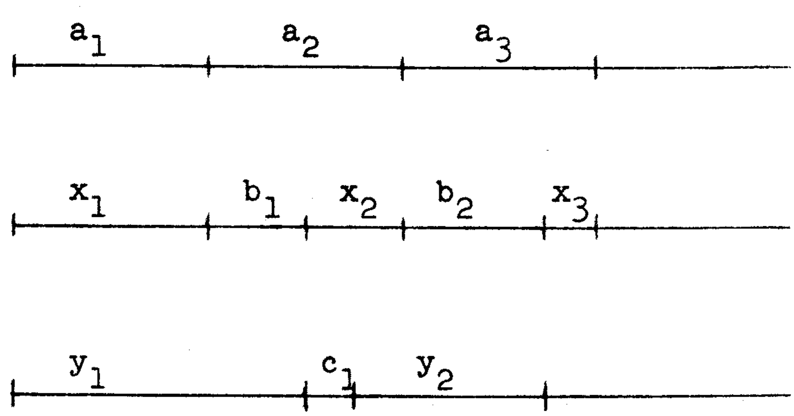
i	a_i	b_i	c_i	d_i
1	3	3	3	3
2	3	1	1	3

of two items going through a four-stage process shows that uniform ordering is <u>not</u> universally valid. It is easy to verify that in this case the optimal arrangement changes in going from the second to the third stages. However, the ordering on the first two machines and the last two machines may always be taken to be the same regardless of the number of stages.

Let us now present a formula for the total idle time on the third machine similar to that given above for the two-stage process. Let

(1) y_i = the idle time on the third machine immediately preceding the processing of the $i^{\underline{th}}$ item.

Schematically



The same type of argumentation as above yields the formula

<u>Lemma</u> 4.

$$(2) \qquad \sum_{i=1}^{n} y_i = \max_{1 \leq u \leq v \leq n} (H_v + K_u)$$

where

$$(3) \qquad H_v = \sum_{i=1}^{v} b_i - \sum_{i=1}^{u-1} c_i \ , \qquad v = 1, 2, \ldots, n \ ,$$

$$K_u = \sum_{i=1}^{u} a_i - \sum_{i=1}^{u-1} b_i \ , \qquad u = 1, 2, \ldots, n \ .$$

As yet, no one has been able to use this formula to derive the optimal ordering for the three stage problem. As we shall see from the discussion of the continuous version below, the general solution must have a quite complicated form.

§8. A Continuous Approximation

In view of the lack of success in treating the general problem, it seems worthwhile to consider various special cases in the hopes that the solution of these may throw light upon the general case.

Let us consider first, to illustrate the method we shall employ, the particular case where we have two stages and only two types of items, each occurring in large quantities. As an approximation to the expression

$$(1) \qquad S(u) = \sum_{i=1}^{u} a_i - \sum_{i=1}^{u-1} b_i$$

we shall consider the integral

(2) $$I(u) = \int_0^u (a(t) - b(t))dt$$

As an analogue of the arrangement of these n items of only two distinct types, we shall consider a characteristic function, $\phi(t)$, defined over the interval $[0,T]$, which is to say a function $\phi(t)$ with the property that

(3) $$\begin{aligned}\phi(t) &= 1 \text{ for } t \epsilon S, \text{ a subset of } [0,T]\\ &= 0 \text{ for } t \text{ in the complement of } S.\end{aligned}$$

Since there are only two types of items, $a(t)$ and $b(t)$ are determined by the relations

(4) $$\begin{aligned}a(t) &= a_1\phi + a_2(1-\phi)\\ b(t) &= b_1\phi + b_2(1-\phi)\end{aligned}$$

We now wish to determine $\phi(t)$ so as to minimize the functional

(5) $$I(T) = \max_{0 \leq u \leq T} \left[\int_0^u (a(t) - b(t))dt \right]$$

subject to the conditions

(6) a. $\phi(t) = 0 \text{ or } 1$,

b. $\int_0^T \phi(t)dt = k < T$.

The last condition is the continuous analogue of the condition that k of the items are of the first type and n - k of the second type.

Using (4), the problem is that of determining

(7) $$\min_{\phi} \max_{0 \leq u \leq T} \left[\alpha \int_0^u \phi dt + \beta u \right]$$

where

(8) a. $\alpha = (a_1 - a_2) + (b_2 - b_1)$

b. $\beta = (a_2 - b_2)$

It is easy to see that a solution is given by

(9) $$\alpha > 0: \quad \phi(t) = 0 \; , \quad 0 \leq t \leq T - k$$
$$\phi(t) = 1 \; , \quad T - k < t \leq T \; .$$

$$\alpha < 0: \quad \phi(t) = 1 \; , \quad 0 \leq t \leq k$$
$$\phi(t) = 0 \; , \quad k < t \leq T \; .$$

We see that the form of the solution depends upon only the ordering of the quantities $a_1 - b_1$ and $a_2 - b_2$, which is precisely what we might expect. A similar solution is obtained for any number of distinct types of items, see [2].

§9. Continuous Version—three Machines

Let us now discuss the continuous version of the three-machine or three stage-process, again for the case of two types of items. As the continuous analogue of the idle time on the third machine we have the functional

$$(1) \qquad I(T) = \max_{0 \le u \le v \le T} \left[\int_0^u (a(t)-b(t))dt + \int_0^v (b(t)-c(t))dt \right]$$

$$= \max_{0 \le u \le v \le T} \left[\alpha \int_0^u \phi dt + \beta u + \gamma \int_0^v \phi dt + \delta v \right] ,$$

where

$$(2) \qquad a(t) = a_1\phi + a_2(1-\phi), \; b(t) = b_1\phi + b_2(1-\phi) ,$$
$$c(t) = c_1\phi + c_2(1-\phi) ,$$

and

$$(3) \qquad \begin{aligned} \alpha &= a_1 - a_2 + b_2 - b_1 \\ \beta &= a_2 - b_2 \\ \gamma &= b_1 - b_2 + c_2 - c_1 \\ \delta &= b_2 - c_2 . \end{aligned}$$

We wish to determine the minimum of I_T for all ϕ subject to the conditions

(4) a. $\phi(t) = 0$ or 1 ,

b. $\int_0^T \phi dt = k$.

As we shall see below the minimum over this class of functions may not exist. Let us then consider the more extensive class of functions satisfying (4b) and the weaker condition

$$(5) \qquad 0 \le \phi(t) \le 1 .$$

The solution to this problem is contained in

<u>Theorem</u> 4. <u>The minimum value of</u> I <u>is</u>

(6) $$V(k,T) = \max\left[0,\ \gamma k + \delta T,\ (\alpha+\gamma)k + (\beta+\delta)T\right] .$$

<u>A minimizing</u> ϕ <u>is given by</u>

(7) $$\phi^*(t) = k/T \quad \underline{\text{for}} \quad 0 \leq t \leq T .$$

<u>In general, the solution is non-unique</u>.

The proof, which is quite simple, may be found in [2].

The particular solution given above, ϕ^*, involves mixing which in the discrete case is impossible, although we can always approximate to it. Since the solution is non-unique in general, it is possible that in many cases the solution will have the simpler form

(8) $$\begin{aligned} \phi(t) &= 1 , & 0 \leq t \leq k , \\ &= 0 , & k < t \leq T , \end{aligned}$$

which corresponds to a very simple solution in the general case.

As we shall see below we can find values of the parameters for which the solution cannot have this simple form, and, as a matter of fact, for which the solution given above is the unique solution.

The importance of this result is that it shows that the three stage process presents a genuinely difficult problem.

§10. Example of Unique Solution*

Let us take $k/T = 1/2$, so that we may take $k = 1/2$ and $T = 1$. Choose

$$(1)\qquad \begin{array}{lll} a_1 = 2 \ , & b_1 = 3 \ , & c_1 = 1 \ , \\ a_2 = 2 \ , & b_1 = 1 \ , & c_1 = 3 \ . \end{array}$$

Then

$$(2)\qquad \alpha = -2 \ , \quad \beta = 1 \ , \quad \gamma = 4 \ , \quad \delta = 2 \ .$$

Hence

$$(3)\qquad \begin{aligned} V &= \max\,(0,\ \gamma k + \delta T,\ (\alpha+\gamma)k + (\beta+\delta)T) \\ &= \max\,(0,\ 0,\ 0) = 0 \ . \end{aligned}$$

Thus, if $\phi(t)$ is a minimizing function we must have

$$(4)\qquad \max_{0\leq u\leq v\leq T} \left[\alpha \int_0^u \phi dt + \beta u + \gamma \int_0^v \phi dt + \delta v\right] = 0$$

which for the set of parameters chosen above reads

$$(5)\qquad \max_{0\leq u\leq v\leq 1} \left[-2 \int_0^u \phi dt + u + 4 \int_0^v \phi dt - 2v\right] = 0$$

For all u and v in the range $0 \leq u \leq v \leq 1$ we must have

$$(6)\qquad -2 \int_0^u \phi dt + u + 4 \int_0^v \phi dt - 2v = 0 \ .$$

Setting $v = u$ the result is

$$(7)\qquad 2 \int_0^u \phi dt \leq u \ ,$$

for $0 \leq u \leq 1$.

* Due to Oliver Gross

On the other hand, setting v = T = 1, and using the relation

(8) $$\int_0^1 \phi dt = k = 1/2 \quad ,$$

we obtain from (5) the condition

(9) $$-2 \int_0^u \phi dt + u \leq 0 \quad .$$

Comparing (7) and (9), we see that we must have equality

(10) $$\int_0^u \phi dt = u/2$$

From this we see that $\phi(t) = 1/2$ for almost all u, which is to say the solution is unique in the set of Lebesgue integrable functions satisfying the constraints.

§11. Stochastic Versions

Let us now consider the case in which the processing times are stochastic rather than fixed parameters. In other words, each item has associated with it a set of distributions for the times required on the various machines.

The problem is now that of determining arrangements which minimize the expected value of the total time required to process all the books, or some other mean of the total time. Whatever the difficulty of the deterministic versions we have discussed above, the stochastic version seems to transcend them. Nothing is known about the solution even in the two stage process.

Since the total time will be a nonlinear function of the individual processing times, it is clear that a knowledge of expected values of individual processing times will not be sufficient to determine the minimum total expected time.

This problem introduces some interesting questions concerning the "stability" or "stiffness" of a scheduling policy. An ordering which minimizes the expected total time may very well permit of a large variance. Consequently, for many purposes a "looser" solution with more flexibility and slack, which is comparatively unaffected by minor variations, may be more desirable.

Although these ideas are, of course, well-known in practice, a precise mathematical formulation seems difficult.

§12. Queuing Theory

If we fix the order and fasten our attention upon the distribution of idle times, the distribution of waiting times, and similar questions, we enter the domain of queuing theory, see [6].

In this connection, we would like to point out that the explicit formulas of Johnson may be of some utility in determining limiting distributions.

§13. Extensions

Although it may seem quite academic to discuss extensions when the simplest problems defy analysis at the moment, it is actually worthwhile to scan the horizon if only for the sake of inspiration.

In realistic problems we will be dealing with processes in which there are many machines of each type at each stage. A first question that arises is that of our policy of feeding items into the machines at any particular stage. For example, we may retain a rigid order, so that in the case of three machines, the first, fourth and seventh items, and so forth, go to the first machine, the second, fifth and eighth go to the second machine and so on. Or we may use a first-come first-serve principle and allocate each new item to the machine which is free, or committed for the least time ahead.

The determination of a feeding policy is part, of course, of the general scheduling problem, but it will be simpler to solve scheduling problems of this type if, arbitrarily, we restrict ourselves to certain types of sub-policies.

If many of the processes are interchangeable, although not all, we have the important problem of determining the arrangement of the stages.

We may also consider ourselves to have available a semi-fluid labor force which can be assigned in various combinations to man the machines at various stages, thereby decreasing the processing time at that stage. The question is now that of determining the optimal partitioning of the labor force among the various activities.

Further, more detailed discussions will be found in Salveson, [9], [10], and Vasonyi, [11], [12], [13].

Finally there is the overall problem of allocating money for the purchase of various types of machines and various catagories of labor. It is easy to see that we can construct hierarchies of problems of increasing orders of difficulty. We can, however, obtain approximate solutions to the larger problems by simplifying the behavior of the component parts, with the result that from a certain stage on, the solution of the overall problems may be simpler, if less exact. This is, after all, the usual technique employed in the treatment of the physical world.

§14. Organization Theory

It is clear that in considering the general problem of arranging combinations of men and machines to perform certain multi-component tasks, we are encroaching upon the domain of organization theory.

The mathematical problems encountered here are of precisely the same general nature as those discussed above, and abstractly, there is no difference. For those interested in some of the questions in this field which have been treated recently, we refer to the work of J. Marschak, [7], [8], on the theory of teams and similar problems.

A particularly interesting set of problems arise in connection with the design of automatic control circuits, computing machines and related mechanisms. The mathematical techniques employed here range over Boolean algebra, mathematical logic, topology and other regions of abstract mathematics which, at first glance, would seem far removed from applications. For a discussion of the "switching" problem, see Hohn [4]; for a discussion of the automaton problem, see von Neumann, [15].

§15. The Assignment Problem

In the sections above we have attempted to determine the minimum over all permutations of n objects of a certain function associated with these permutations. As we know, the group of n! permutations on n objects is equivalent to the group of permutation matrices, P, of the representative form,

$$(1) \qquad \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \end{pmatrix}$$

characterized by the property that there is precisely one non-zero element in each row and column.

Consequently, in place of maximizing over all permutations, we can think of maximizing a function f(P) over all matrices P. The simplest such function is a linear function

$$(2) \qquad f(P) = \operatorname{tr}(AP) \quad ,$$

where $\mathrm{tr}\,(B) = \sum_{i=1}^{n} b_{ii}$, the sum of the diagonal elements of a matrix B.

This function actually occurs in connection with an important problem, the "assignment problem", which reads as follows:*

"Given n men and n jobs with the utility of the assignment of the $i^{\underline{th}}$ man to the $j^{\underline{th}}$ job equal to a_{ij}, $i, j = 1, 2, \ldots, n$, determine the allocation of assignments which maximizes the total utility."

Crudely we see that we have n! possibilities to contemplate. There are two ways of overcoming this difficulty. The first is due to von Neumann, [14], who converts the problem into an $n^2 \times 2n$ game which is a generalized version of "hide-and-seek". The second is due to Egervary, [3], and is based upon some work of D. Konig in the theory of graphs.

The basic principle involved in each is the fact that we can imbed the permutation matrices in the continuous set of doubly-stochastic matrices of order n, i.e. matrices $P = (p_{ij})$ characterized by the properties

(3) a. $0 \le p_{ij} \le 1$,

b. $\sum_{j=1}^{n} p_{ij} = 1$,

c. $\sum_{i=1}^{n} p_{ij} = 1$,

* Some very interesting generalizations of these problems, dealing with teams, have been formulated by L. Shapley, (unpublished).

and that the permutation matrices are the extreme elements of this convex set, a result due to G. Birkhoff. Another proof is contained in von Neumann, [14].

Other problems involving maximization over the set of permutation matrices arise in various parts of transportation theory. We mention the "travelling salesman" problem, and a problem of Beckmann and Koopmans which requires the maximization of tr (PAPB).

Part II.

Transportation Problems

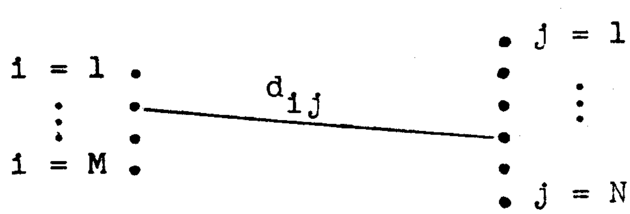
§1. Introduction

A large and important class of problems with strong topological overtones arise from the determination of efficient transportation schedules for buses, trains, airplanes, cargo ships, and so on.

Since the subject is a vast one, with most of what is known to date in readily available papers, we shall content ourselves with a brief reference to the work on the Hitchcock-Koopmans transportation problem, which is representative of the type of question that arises, and then discuss the question of determining the capacity of a rail network. This problem is particularly interesting since it may be approached by means of the theory of graphs, using the results of Menger, or by means of linear programming, or by means of various iterative techniques such as the "flooding" technique of Boldyreff.

§2. The Hitchcock-Koopmans Transportation Problem

Suppose that we have M ports containing cargoes and N ports as destinations of these cargoes.



Let d_{ij} be the distance between the $i^{\underline{th}}$ cargo port and $j^{\underline{th}}$ destination, and let x_{ij} be the cargo transmitted between these two ports. The cost of transmitting this cargo is taken to be $x_{ij}d_{ij}$. Assume that we have an amount, c_i, of cargo at the $i^{\underline{th}}$ port and a requirement r_j at the $j^{\underline{th}}$ port, with

$$(1) \qquad \sum_i c_i = \sum_j r_j \quad ,$$

and that it is desired to transmit all the cargo at a minimum total cost.

The mathematical problem is that of minimizing the linear form

$$(2) \qquad \sum_{i,j=1}^{M,N} d_{ij}x_{ij}$$

subject to the constraints

$$(3) \quad \text{a.} \quad \sum_{i=1}^{M} x_{ij} = r_j$$

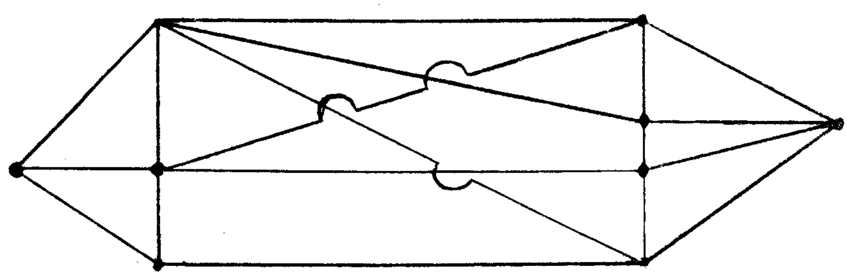
$$\text{b.} \quad \sum_{j=1}^{N} x_{ij} = c_i$$

$$\text{c.} \quad x_{ij} \geq 0$$

A thorough discussion of an iterative technique for solving this problem is given by Flood, [3], and a discussion of related classes of problems is contained in Koopmans, [7], and in Beckmann, McQuire, and Winsten, [1].

§3. On a Simplified Transportation Network

Suppose that we have an idealized network which can be represented schematically by diagrams of the following type



The vertices represent terminals and the connecting arcs represent rail lines between these terminals. The arc connecting the $i^{\underline{th}}$ terminal to the $j^{\underline{th}}$ terminal is assigned a numerical quantity, a "capacity", a_{ij}, which we consider to be the maximum number of cars it can transmit in unit time, and the $i^{\underline{th}}$ vertex is assigned a number, t_i, the maximum rate of flow through the terminal in unit time.

An important quantity associated with a network of this type is the maximum rate of flow from O to T in unit time, or, more generally, from any vertex to any other vertex. That the problem is non-trivial is due to the fact that bottlenecks may develop at various junctions if the outgoing lines have smaller capacity than the incoming lines or if the terminals cannot handle maximum incoming and outgoing traffic. The problem is that of determining a systematic procedure for ascertaining the maximum rate of flow from O to T and the allocation of cars which yields this maximum rate.

§4. Topological Approach

As has happened so frequently in the past, a mathematical theory which treats problems of this genre is already in existence.

It is the theory of graphs, an outgrowth of the electrical network theory of Kirchoff. The particular results required are due to Menger [8], with the application to railway networks suggested by A. Hoffman, and worked out in detail by Boldyreff, Robacker and others.*

§5. Linear Programming

Since the problem of determining the maximum rate of flow may be formulated as a maximization of a linear function, namely the rate of flow to T from the immediately adjacent terminals, subject to the input-output relations and the capacity restraints, it is clear that the theory of linear inequalities or linear programming is applicable. It has been shown by Ford and Fulkerson, [4], that the results derived from Menger's theory may be obtained in this way from the duality theorems of linear programming.

§6. The Flooding Technique

Since the maximization problem is derived from a dynamic process with certain characteristic features, it is to be expected that we can find an iterative scheme of solution which will make use of these features and therefore be simpler to use than any of the standard algorithms of linear programming.

A particularly interesting technique which seems to have many wider applications is the "flooding" technique recently developed by Boldyreff, which we shall describe briefly.

* A very interesting paper by W. Prager, "On the Role of Congestion in Transportation Problems", Brown University (1955), has just appeared showing that a very natural nonlinearization leads to a unique solution of the Hitchcock-Koopmans problem.

We start out from O by routing trains at maximum capacity along the lines emanating from O. At each new terminal we may or may not encounter a bottleneck. If we encounter no bottleneck, we continue in this fashion using the maximum number of trains available and routing in a fixed fashion if we have a choice, e. g. maximum capacity to uppermost line and continuing clockwise in this fashion. If we encounter a bottleneck, we use the maximum allowable number of trains and continue the process.

Having gone through the whole network in this way, we now remove the bottlenecks one at a time by working backwards either from T or from the stages closest to O.

We see that the method possesses the essential features of the "relaxation" method which is widely used in applied mathematics. Although no proof has been given of the convergence of this technique as yet, there seems to be little doubt that it is valid. A complete discussion of the technique together with examples will be found in a forthcoming paper by Boldyreff, [2].

It is worthwhile noting that there is a strong analogy between a railway network and a production network and thus that this technique may be equally useful in estimating the potentialities of industrial networks.*

* R. Fulkerson has shown that some structural theorems for partially ordered sets, due to R. P. Dilworth, Ann. of Math (1950), may be derived from Menger's theorems. A. Hoffman and G. Dantzig have given another proof based upon the duality theorem of the theory of linear inequalities.

There are a number of interesting and important mathematical problems concerned with the transmission of power from one source to another, and with the interconnection of transmission systems. Let us refer the interested reader to

G. Krön, Tensorial Analysis of Integrated Transmission Systems, I, II, III, IV, Vol. 70, 71, AIEE Proceedings.

A. F. Glimm, R. Haberman, L. K. Kirchmeyer, R. W. Thomas, Automatic Digital Computer Applied to Generation Scheduling, Paper 54-276, AIEE Meeting, June 21-25, 1954.

————————————, Loss Formulas Made Easy, Paper 53-209, AIEE Meeting, June 15-19, 1953.

Part III.

Smoothing Problems

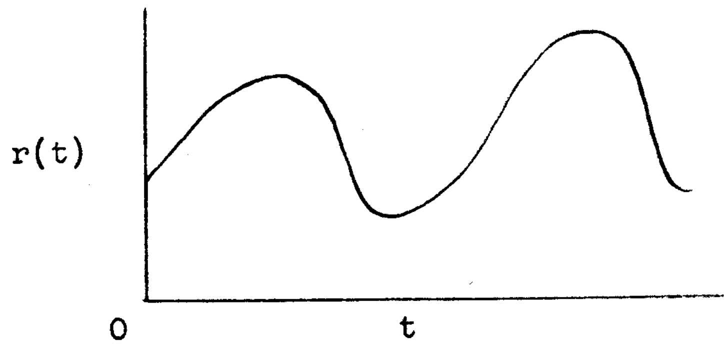
§1. Preliminaries

In this part of the paper we shall discuss a class of problems which are occasionally called "smoothing" problems. The general problem has the following character: we wish to maintain a system in a given state, with penalties for deviations from this state. In some cases, the penalty is the same regardless of the direction of deviation; in other cases, we have one type of penalty for overshooting the requirements, and another type of penalty for not being able to meet the requirements.

We shall discuss some simple problems of this type, arising from economic and industrial situations, and then discuss some "control" problems arising in engineering practice.

§2. An Industrial Smoothing Problem

Let us suppose that at any time t we have a have a staff of employees capable of turning out a certain quantity of work $x(t)$, and committments requiring a quantity $r(t)$. In general, the function $r(t)$ will be oscillatory as a function of time,



We shall assume that at all times we are required to have $x(t) \geq r(t)$, which is to say we are required to meet our committments come what may.

If $x(t)$ exceeds $r(t)$, we consider ourselves to be losing an amount of money in excess wages proportional to $(x(t)-r(t))dt$ over the interval $(t,t+dt)$. However, if we decrease $x(t)$ to the level of $r(t)$, we are faced with the prospect of having possibly to increase $x(t)$ if $r(t)$ increases. This cannot be done without cost, and we shall take the cost of doing this proportional to dx/dt. Assuming that it costs us nothing to decrease $x(t)$, the cost of changing the level of labor supply over the interval $(t,t+dt)$ will be taken to be proportional to

$$\max\,(dx/dt,\ 0)dt \quad . \tag{1}$$

The total cost over an interval $[0,T]$ can then be taken to be

(2) $$J(x) = \int_0^T [(x(t)-r(t)) + \alpha \text{ Max } (dx/dt, 0)]dt$$

where α is some positive constant.

The problem is now to choose an absolutely continuous $x(t)$ which minimizes $J(x)$ while satisfying the requirement $x(t) \geq r(t)$.

The solution turns out to have a quite simple form, see [4].

§3. Discrete Version

If we now take time to be measured in discrete units, the analogue of the expression in (2.2) is

(1) $$J\{x\} = \sum_{k=1}^{N} [x(k)-r(k) + \alpha \text{ Max } (x(k+1)-x(k), 0)] \quad .$$

The constraint is now

(2) $$x(k) \geq r(k) \quad .$$

The solution to this problem is similar to that given for the continuous version, see Karlin [8].

§4. Expansion Limitation

An interesting version of the above problem is one in which we do not allow arbitrarily rapid increase in $x(k)$ or $x(t)$. Thus for the discrete case we may impose a condition of the type

(1) $$x(k+1) \leq \lambda x(k) \quad ,$$

or

(2) $$x(k+1) - x(k) \leq b \quad .$$

A problem of the first kind has been treated by Baldwin and Shephard, in an unpublished work.

§5. The Caterer Problem

Let us now state another smoothing problem in the form given by W. Jacobs.

"A caterer knows that in connection with the meals he has arranged to serve during the next n days, he will require $r_j \geq 0$ fresh napkins on the $j^{\underline{th}}$ day, $j = 1, 2, \ldots, n$. Laundering normally takes p days; that is a soiled napkin sent for laundering immediately after use on the $j^{\underline{th}}$ day is returned in time to be used again on the $(j+p)^{\underline{th}}$ day. However, the laundry also has a higher cost service which returns the napkins in $q < p$ days (p and q integers). Having no usable napkins on hand or in the laundry, the caterer will meet his early needs by purchasing napkins at a cents each. Laundering costs b and c cents a napkin for the normal and high cost service respectively. How does the caterer arrange matters so as to meet his needs and minimize his total outlay for the n days?"

The solution for the case where $q = p-1$ is given by Jacobs in [7].

§6. The Optimal Inventory Problem

The preceding problems have been of deterministic type. Let us now consider a stochastic version.

The situation is as follows. At various specified times, we have an opportunity to order supplies of a certain set of items, where the cost of ordering depends upon the number ordered of each item, and where there may or may not be some fixed administrative costs which are independent of the number ordered. At various other times, demands are made upon the stocks of these items. The interesting case is where the demands are not known in advance, but where we do know the joint distribution of demands. The incentive for ordering lies in a penalty which is assessed whenever the demand of an item exceeds the supply. Different penalties are levied in different fields of activity.

We wish to determine the ordering policy which minimizes the expected cost of the total process.

Here is a case where we have one type of penalty for being unable to supply the demand and different penalty for being overstocked. This last penalty may be expressed in terms of frozen assets, in storage cost, etc.

The problem was first formulated by Arrow, Harris and Marschak [1], and has been subsequently treated by Dvoretzky, Kiefer and Wolfowitz, [6], Bellman, [2], and Bellman, Glicksberg and Gross, [4].

§7. Learning Processes

In formulating the problem above, we have assumed that we knew the distribution of demand. In many cases this is not true, and we have the additional problem of determining the probability distribution and making decisions at the same time. A process of this type we call a "learning process", cf. our discussion in the Introduction.

Problems of this type arise most frequently in statistical investigation where they have given rise to the theory of sequential analysis of Wald.

An interesting survey of this general area is contained in the paper of Robbins, [9].

§8. Control Processes

Let us close this part with a brief description of the types of mathematical problems arising in the theory of control processes. Let us consider a physical, economic, or engineering system, whose state at any time is determined by the vector $x(t)$. If left to itself, the system will be determined by the linear differential equation

(1) $$dx/dt = Ax \quad , \qquad x(0) = c \quad ,$$

where A is a constant matrix and c is some initial value.

Let us suppose, however, that we want the system to behave in a different way, one specified by the vector $y(t)$. To force the system into this desired state, we must introduce some external influence, which we call "control".

For our purposes we assume that this external influence manifests itself by way of an inhomogeneous term, a "forcing" term, so that (1) above becomes

$$(2) \qquad dx/dt = Ax + f(t) \quad , \qquad x(0) = c \quad .$$

As usual, it costs us something to exert this control. In determining the amount of control we will exert, we must balance the cost of control against the cost of deviation of the system from its desired state. Depending upon the way we measure these various costs, we obtain various classes of mathematical problems.*

Problems of precisely similar mathematical type arise in mathematical economics in connection with reinvestment policy. Here it is a question of determining the rate at which profits should be put back into a business so as to maximize the total profit we obtain over a given period.

For other aspects of control processes we refer to the book by N. Wiener, [10].

* See, R. Bellman, I. Glicksberg and O. Gross, "Some Variational Problems in the Theory of Dynamic Programming", Rendiconte del Palermo, (to appear), Proc. Nat. Acad. Sci., Vol. 39 (1953).

Part IV
Computational Processes

§1. Preliminaries

In discussing scheduling processes associated with the use of computing machines we are entering a field of great importance where little has been done in comparison to the large number of important and difficult problems remaining.

We shall make no effort to cover the vast area of problems encountered in determining efficient coding procedures, but rather mention briefly some problems of particular interest from the standpoint of scheduling theory. First, we shall consider a problem concerned with the evaluation of polynomials posed by Ostrowski. This is a particular case of the general problem of determining uniform procedures for computing square roots, solving polynomial equations, and so on.

Then we shall discuss the sorting problem. Here we are given a set of items in some jumbled array and we wish to arrange them in some assigned order, say alphabetically, or chronologically, or with respect to other properties. A particular case of this is the problem of determining the maximum of n quantities, which arises in computing the solutions of the functional equations occurring in the theory of dynamic programming.

§2. Horner's Rule

If we wish to calculate the value of the polynomial

$$(1) \qquad f(x) = a_0x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n$$

for a particular value of x, we can do it in the following uninspired fashion. We first calculate the powers of x, namely x^2, x^3, ..., x^n, requiring n-1 multiplications, and then compute the products, a_0x^n, a_1x^{n-1}, ..., $a_{n-1}x$, n additional multiplications. Having performed these 2n-1 multiplications, we now require n additions to complete the evaluation. Thus we require a total of 2n-1 multiplications and n additions, following this procedure.

There are many shorter methods. One method, which is usually called Horner's rule, or synthetic division, employs the sequence of polynomials

$$(2) \qquad f_k(x) = a_0x^k + a_1x^{k-1} + \ldots + a_k$$

connected by the recurrence relation

$$(3) \qquad f_k(x) = x\, f_{k-1}(x) + a_k \quad .$$

It is easy to see that $f_n(x)$ computed according to this algorithm requires only n multiplications and n additions.

It seems intuitively clear that no method can improve upon this, but the proof seems difficult. See Ostrowski, [2].

§3. Universal Algorithms

The problem we have posed above is a simple example of a class of problems that we meet in coding for high speed computing machines.

Various criteria are employed to measure the efficiency of a coding procedure. We may wish to minimize the possibility of error, minimize the computing time, minimize the memory requirements, and so on.

There is a vast literature on these topics in connection with solving systems of linear equations, systems of differential equations, partial differential equations and so on.

§4. Determining the Maximum of n Quantities

An interesting function of n quantities is the maximum quantity. The use of computing machines in determining the maximum of a given number of quantities is extremely important in connection with functional equations of the type

$$(1) \qquad f(x) = \max_{0 \leq y \leq x} \left[g(y) + h(x-y) + f(ay+b(x-y)) \right] ,$$

which arise in the theory of dynamic programming, see [1].

§5. Sorting

Determining the maximum or minimum of n quantities is a special case of the problem of sorting n quantities according to some preassigned ordering relation.

There are really many special classes of problems of this type. For example, one problem is that of determining a uniform procedure which will work for any given set of quantities. Another problem is that of determining a sorting procedure which will minimize expected sorting time when we are given the items one at a time or in small batches and the information that there is a given distribution governing their ordering. A further problem is that of simultaneously ordering and determining the distribution as we go along, which is to say a learning process.

A discussion of related problems is contained in Seward, [3], where a number of further references may be found.

Part V

Applications of the Theory of Linear Programming

§1. Introduction.

In the previous chapters, we have discussed a number of different classes of scheduling processes together with various analytic and computational techniques which may be utilized to treat these problems to a greater or lesser degree. In this chapter our theme is the theory of linear programming. We shall illustrate by means of a number of examples chosen from different fields how wide is the range of application of this important mathematical tool.

The theory of linear programming has as its central purpose the problem of obtaining the maximum or minimum of the linear form

$$(1) \qquad L(x) = \sum_{i=1}^{n} c_i x_i$$

subject to the series of constraints

$$(2) \qquad \sum_{j=1}^{n} a_{ij} x_j \geq b_i, \quad i=1,2,\ldots,m.$$

The theory of nonlinear programming is concerned with the corresponding problem for nonlinear functions. However, this study has not achieved the same stage of advancement as the linear theory, for obvious reasons, and hence we shall not discuss it here. The interested reader may consult Kuhn and Tucker, [13],

The classical theory of inequalities, as developed thirty to fifty years ago by Dines, Farkas, Motzkin, Stiemke, and others, furnishes a number of elegant theoretical results, together with computational methods. The essential aim of the newer theory of linear programming is to discover and apply rapid computational algorithms which will yield the numerical solution of problems of the above type of large dimension either by hand or machine computation.

The most useful and flexible, in the sense of being widely applicable, algorithm devised to date is the "simplex" method of G. Dantzig, together with its modifications and extensions by Beale, Charnes, and Cooper, Dorfman, Lemke, and Dantzig himself.

To illustrate the range of application of these techniques, we shall consider a number of problems in some small detail. These involve the routing of aircraft, the determination of the number and arrangement of toll collectors at bridges, the scheduling of military tanker fleets, network problems, production smoothing problems, the "fixed charge" problem, and finally some examples involving uncertainty. These last two, however, can only be treated in very special cases, and at the moment are a challenge to the ingenuity of the mathematician.

§2. Production Smoothing.

We have discussed a few smoothing problems above. Let us now consider one that may be formulated as a linear programming problem. A single item is to be produced over a given number of time periods to satisfy certain requirements at each of these periods. We wish to produce this item so as to minimize the total cost which is composed of costs of production, costs of storage, and costs for change in production rate.

To formulate the problem mathematically, let T be the total number of time periods and define

(1) r_t = requirement at time t, assumed known.

x_t = quantity produced over the time period

$y_t = x_{t+1} - x_t \geq 0$, the increase in production rate at time t.

Let

(2) $X_i = \Sigma x_t$ = the total production from t = 0 to t = i

$R_i = \Sigma r_t$ = the total requirement over the period t = 0 to t = i, with $R_o = 0$. Then the excess of accumulated production over accumulated requirements up to time i is given by

(3) $$u_i = u_o + X_i - R_i \geq 0,$$

where u_o is a given constant, the excess production at the start of the process.

To express the costs, let

(4) c_i = the cost of producing each unit in the period i-1 to i,

d_i = the cost of storing each unit of excess u_i for one period,

e_i = the cost of increasing production rate one unit per unit time at time i.

The problem is then to minimize the total cost of the process,

(5) $$\sum_{i=0}^{T} (c_i x_i + d_i u_i + e_i y_i)$$

subject to the constraints

(6) (a) $$\sum_{t=1}^{i} x_t \geq \sum_{t=1}^{i} r_t$$

(b) $x_i, y_i, u_i \geq 0$.

In their paper, [5], of Part III, Dantzig and Johnson exhibit a rapid graphical method, involving only intersections and rotations of straight lines, which seems to require only a few iterations. See also A. J. Hoffman and W. Jacobs, [12].

§3. Traffic Delay at Toll Booths.

An interesting problem in the general theory of queuing is treated by L. C. Edie in [10]. It concerns the collection of tolls at Port Authority tunnels and bridges in New York City. The Port Authority desires to handle traffic with the minimum number of toll collectors that is consistent with both adequate service to the public and a sufficient number of relief periods for the toll collector. These relief periods are required since the work is continuous and exacting. We then have the usual conflict between economy and service.

Having determined the traffic characteristics by observation over a period of time, Edie reduced the problem to one of scheduling. The solution obtained theoretically was found in actual practice to be very satisfactory.

At the suggestion of E. W. Paxson, the problem was tackled by G. Dantzig, [5]. His formulation of the problem leads to the problem of determining the minimum of L(x) subject to the restrictions

$$(1) \qquad \sum_{i=1}^{m} a_{ij}x_i \geq b_j, \quad i=1,2,\ldots, m+14$$
$$x_i \geq 0,$$

where the essential feature of the problem is that the a_{ij} are either 0 or 1. As a consequence of this, problems of moderate size can often be solved by hand computation in a few hours, using the dual simplex technique. The observation that the problem is a variant of a transportation-type problem, see [6], enables large systems to be solved rapidly with the aid of computing machines.

§4. Scheduling a Military Tanker Fleet.

In a previous chapter we mentioned the Hitchcock-Koopmans transportation problem, and noted the fact that independent formulations and computational schemes had been given by a number of different people. A degenerate form of the transportation problem is the assignment problem, which has been treated by the simplex method in a paper by Votaw and Orden, see [6].

An interesting application of the Koopmans-Dantzig approach to the general transportation problem, using the simplex method of computation, is contained in a paper by M. M. Flood, [11], discussing the scheduling of a military tanker fleet.

Another discussion of the problem is contained in a paper by G. B. Dantzig and D. R. Fulkerson, [9], where they show that fairly large problems of this nature can be solved by hand computation requiring only a few iterations. An alternate computational scheme had previously been given by Robinson and Walsh, [15].

A similar problem occurs in connection with the routing of aircraft where there are a number of alternative routes, a number of different types of aircraft, and different payloads and traffic on different routes. For a discussion of this problem see, [8].

§5. Fixed Charge Problems.

In the previous sections we have discussed a number of problems which can be resolved readily using the computational methods of linear programming. Let us now present some problems which are either intractable or can only be handled in special cases.

Consider the following situation. We have a factory containing m different machines each capable of performing any of n distinct operations. Let h_{ij} denote the time consumed by the $j^{\underline{th}}$ machine performing the $i^{\underline{th}}$ operation on a unit quantity of goods. Let x_{ij} denote the quantity of goods on which the $i^{\underline{th}}$ operation is performed by the $j^{\underline{th}}$ mathine, and b_i the total quantity of goods which require the $i^{\underline{th}}$ operation. The following relations are satisfied

(1) a. $\sum_{j=1}^{n} x_{ij} = b_i,\ i=1,2,\dots,m,$

b. $x_{ij} \geq 0.$

Let us assume furthermore that there is a constraint on the total man hours available for the $j^{\underline{th}}$ machine; i.e.

(2) $$\sum_{i=1}^{m} h_{ij}\, x_{ij} \leq c_j,\ j=1,2,\dots,n.$$

The most efficient operation is taken to be the choice of the x_{ij} which minimizes the total cost of operation,

(3) $$L(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}x_{ij}.$$

In this formulation, the problem falls within the domain of linear programming.

Let us, however, now assume that there is a fixed charge, k_{ij}, for preparing the $j^{\underline{th}}$ machine for the $i^{\underline{th}}$ operation, regardless of the length of time that the machine is used. The total cost is now

(4) $$L(x) = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[c_{ij}x_{ij}+k_{ij}\,\delta(x_{ij})\right]$$

where $\delta(x)$ is the discontinuous function defined by the conditions

(5) $$\delta(x) = 1, \ x > 0$$
$$= 0, \ x = 0$$

The problem of minimizing the total cost, as given by (4) is one that at the moment escapes any of the known efficient computational algorithms. It is interesting to note, however, that the special case where $k_{ij} = k$ may be treated by an extension of the conventional techniques, see Dantzig and Hirsch,* .

A similar difficulty arises in the "optimal inventory" problem, where the k_{ij} correspond to paperwork or "red tape" costs in some cases and to physical set-up costs in others, see [2], [3].

§6. Convex Functions and Uncertainty.

Most problems involving uncertainty or stochastic processes lie outside the range of linear programming techniques. In some cases they can be treated by the methods of dynamic programming, see [1], where further references may be found, and in most cases they remain unsolved.

There are, however, a class of problems involving uncertain demand which lead to the problem of minimizing a sum of the form

(1) $$M(x) = \sum_{i=1}^{n} \phi_i(x_i)$$

subject to linear constraints, where the $\phi_i(x)$ are convex functions of the x_i, or alternatively, of maximizing the sum when the $\phi_i(x)$ are concave functions. These can be treated by linear programming techniques if we approximate to $\phi_i'(x)$ by a step function. Details may be found in [4] or [7].

Let us consider a simple example involving uncertain demand,

*The Fixed Charge Problem, RM-1383, RAND Corporation, 1954.

occurring in the walnut-growing industry, which leads to a problem of the above type. Each year, the walnut crop consists of walnuts of different grades, say $G_1, G_2, \ldots, G_k$ in quantities $q_1, q_2, \ldots, q_k$. Using various quantities of each grade assortments of walnuts are put together for commercial sale. Let us assume that there are N different types of packets selling for prices $p_1, p_2, \ldots, p_N$ per packet respectively. If we assume that there are fixed demands for these assortments, d_i for the $i^{\underline{th}}$ assortment, then the problem may be formulated vary easily as a generalized transportation problem.

Let

(2) $\quad x_{ij}$ = the amount of the $i^{\underline{th}}$ grade of walnuts that is used to make up the $j^{\underline{th}}$ assortment.

Then we have

(3) $$\sum_{j=1}^{N} x_{ij} \leq g_i, \quad x_{ij} \geq 0$$

The number of packets of the $i^{\underline{th}}$ kind that can be put together is

(4) $$u_j = \operatorname*{Min}_{i} \; x_{ij}$$

where u_j satisfies the restriction

(5) $$u_j \leq d_j.$$

The total profit is

(6) $$P = \sum_{j=1}^{N} u_j p_j.$$

The problem of maximizing the profit can be resolved numerically using the standard techniques.

Consider, however, the case where the demand is uncertain. In this case we may assume that, on the basis of experience, we can predict the distribution of demand for each type of assortment.*

Let $dG_i(z)$ be the distribution function for the demand for the $i^{\underline{th}}$ packet. If u_i packets are made, the expected profit will be

$$(7) \qquad p_i \int_0^{u_i} z\,dG_i(z) + p_i u_i \int_{u_i}^{\infty} dG_i(z)$$

Hence the total expected profit is

$$(8) \qquad \sum_{i=1}^{N} p_i \;\phi_i(u_i),$$

where

$$(9) \qquad \phi_i(u) = \int_0^{u} z\,dG_i(z) + u \int_u^{\infty} dG(z)$$
$$= u + \int_0^{u} (z-u)\,dG(z)$$

It is easy to see that each $\phi_i(u)$ is a concave function. Hence the approximation method used above may be employed.

A problem of similar type, arising from the problem of allocation of a carrier fleet to airline routes to meet an uncertain demand is treated by G. Dantzig in [7].

Let us mention in passing that maximization of expected profit may be undesirable if there is a considerable risk involved. We may prefer a smaller expected profit and a smaller risk, or a smaller variance. An interesting discussion of this general problem is contained in Markowitz, [14].

*This must be taken cum grano salis.

Part VI.

Crystal Gazing

§1. Introduction

In the previous parts, we have examined a number of problems of various types and discussed a number of mathematical techniques that have been employed to handle these problems. In this part we shall turn the spotlight on a number of regions of mathematics which we feel must be intensively explored before we can hope to master the field of scheduling theory.

Taking subjects in no particular order as far as priority or importance or difficulty are concerned, we shall discuss the extremum properties of functions defined over discontinuous groups, non-classical aspects of the calculus of variations, non-commutative stochastic processes, non-markovian processes, and iterative techniques.

Each of these topics is well worth studying in its own right apart from any possible applications. However, it is probably as true here as in other parts of mathematics that the best entry to a new field is by way of an important and natural physical problem. Scheduling theory contains an abundance of these, almost all difficult and challenging.

§2. Extremum Properties of Functions on Discontinuous Groups

The classical problems in variational analysis may be handled uniformly by the principle of continuous variation of the independent variables in the neighborhood of the extremum. In many of the problems we have discussed above, as for example the assembly line problems of Part I, we are confronted by the problem of maximizing over the set of all permutations on n objects. Here the concept of continuous variation is missing, and this accounts for a great deal of the difficulty of the problem.

It would seem therefore that a study of the extremum properties of functions defined over discontinuous groups would have interesting and important consequences for scheduling theory, and that, conversely, a good point to begin this study would be in connection with the permutation group on n objects, the fundamental group of scheduling theory.

§3. Non-classical Calculus of Variations

As we have seen in Part I, the continuous versions of some problems involving permutations lead to variational problems over the space of characteristic functions.

Problems of this type are not amenable to the classical variational techniques and require new techniques. A particular problem which has not been previously treated, and which arises in the study of production processes involving mutually exclusive operations, is that of determining the extrema of

$$(1) \qquad J(\phi) = \int_0^T f(\phi_1, \phi_2, \ldots, \phi_k, t)dt$$

where the $\phi_i(t)$ are subject to the constraints

$$(2) \quad a. \quad 0 \leq \phi_i(t) \leq 1 \ ,$$

$$b. \quad \int_0^T \phi_i(t)dt = k_i \ , \qquad \sum_i k_i = T$$

$$c. \quad \phi_i(t)\phi_j(t) = 0 \qquad \text{for } i \neq j \ .$$

Problems of this type arise also in the theory of multi-stage production processes involving mutually exclusive activities, [2].

§4. Non-commutative Stochastic Processes

The classical theory of probability is occupied almost exclusively with the study of commutative processes, in the following sense. Let x_i be a sequence of vectors having a common distribution, and let

$$(1) \qquad z_N = \sum_{i=1}^{N} x_i$$

The classical limit theorems, such as the central limit theorem, are concerned with the question of determining the asymptotic distribution of various functions of z_N.

If we consider z_N to represent the state of a physical system at time N, the x_k represent certain random disturbances of the system. In writing z_N as a sum of these disturbances, we are tacitly assuming that these disturbances commute as far as their effects are concerned, i.e. their order of occurrence of no importance.

In many situations, this is not true. Consider, for example, a system whose state at any time is specified by the vector x. Let an event correspond to a transformation of this system into another vector x' and let the transformation be a linear one. Then $x' = X_1 x$, and if the events are stochastic the matrix X_1 is a stochastic matrix. The result of n successive events will be a stochastic vector x_n given by the relation

$$x_n = X_n X_{n-1} \cdots X_2 X_1 x \tag{2}$$

where the X_i are stochastic matrices.

The problem of determining the limiting distribution of the vector x_n given the distribution of the random matrix X_i seems to be a very difficult one which has been discussed up to the present in very brief detail, cf [1].

The difficulties of queuing theory are in a large part due to the non-commutative aspects of the problem and the resultant non-linear functions which occur.

§5. Non-Markovian Processes

Consider a sequence of random variables $\{x_i\}$, where $i = 1, 2, \ldots$ The three most important types of sequences which have been studied to date are sequences of independent variables, Markoff processes, and stationary processes. In each of these cases, the sequence possesses a special structure which enables us to bypass any necessity for the complete past history of the sequence, in predicting the behavior of future elements in the sequence.

In scheduling theory, however, and in many other stochastic processes as well, the structure is more complicated. We must study various classes of stochastic sequences where the distribution of x_i depends upon the distribution of all the preceding x_j. Posed in this way the problem is too broad, and it would seem that once again the best approach to an extension of the present mathematical theory lies in a study of the natural problems which arise from queuing theory, scheduling theory, and related fields.

§6. Further Applications of Game Theory

As we noted in §15 of Part I, the solution to the assignment problem, which crudely requires $n!$ calculations, can be obtained as the solution to an $n^2 \times 2n$ game, which is an enormous decrease in dimensionality for large n.

An important problem which immediately springs to mind is that of cataloguing the classes of problems involving permutations, or more generally variation over the elements of a discrete group, which can be transformed into two-person games of lower dimensions.

An interesting problem which includes the scheduling problems discussed in Part I is that of determining the minimum over all permutations of the x_i of the function

$$f(x) = \max_i \sum_{j=1}^{n} a_{ij} x_j \tag{1}$$

§7. Iterative Techniques

It is intuitively clear that each problem that arises will have certain characteristic features which can be utilized to speed the computation of a solution. This is one of the advantages of considering problems arising from the physical world. Of all the universes within pencil reach of the mathematician, the real one is the most likely to furnish important, interesting and tractable problems.

Essentially, the argument goes that whatever exists cannot be too complicated, if viewed properly. This, of course, is an article of faith and not to be interpreted too literally.

The idea of exploiting the intrinsic structure of a process is the guiding concept of the theory of dynamic programming.

Similarly, in employing the algorithm of the "simplex method", G. Danzig has introduced many devices to take account of dynamic processes. The "flooding technique" of Boldyreff is another example of an iterative technique particularly adapted to the problem it treats.

A certain amount of compromise must always be made in formulating computing procedures. If we are interested in coding for modern high speed machines, we want fairly uniform procedures so that recoding will not be necessary for every new problem that comes along. On the other hand if an appropriate modification, or an entirely new method, can save an appreciable amount of time on problems of a special but important type then it is worth employing.

A great deal of work in this direction has been done in connection with solving systems of linear equations and in solving polynomial equations. However, very little has been done from the purely abstract point of view of classifying structural properties of processes and correlating them with appropriate computational techniques.

BIBLIOGRAPHY*

Part I

1. R. Bellman, "Dynamic Programming of Continuous Processes", The RAND Corporation, Report No. R-271 (1954).

2. R. Bellman and O. Gross, "Some Combinatorial Problems Arising in the Theory of Multi-stage Processes", J. Soc. Indust. Appl. Math., Vol. 2 (1954), pp. 175-183.

3. E. Egervary, "On Combinatorial Properties of Matrices", Matematikai és Fizikai Lapok 38 (1931), pp. 16-28.

4. F. Hohn, "Some Mathematical Aspects of Switching", Amer. Math. Monthly, Vol. 62 (1955), pp. 75-91.

5. S. Johnson, "Optimal Two- and Three-stage Production Schedules with Setup Times Included", Naval Logistics Research Quarterly, March 1954.

6. D. G. Kendall, "Some Problems in the Theory of Queues", Journal of the Royal Statistical Society, Series B, 13 (1951), pp. 151-185.

7. J. Marschak, "Towards an Economic Theory of Organization and Information, Decision Processes", ed. by Thrall, Coombs and Davis, J. Wiley and Sons, 1954, pp. 187-220.

8. ————, "Elements for a Theory of Teams", Management Science, Vol. 1, No. 2, January 1955, pp. 127-137.

9. M. Salveson, "On a Qualitative Method in Production Planning and Scheduling", Econometrica, Vol. 20 (1952), pp. 571-4.

10. ————, "The Assembly Line Balancing Problem", presented at A.S.M.E. Annual Meeting, 1954.

11. A. Vasonyi, "The Use of Mathematics in Production and Inventory Control", Management Science, Vol. 1 (1954), pp. 70-85.

12. ————, Part II, presented to Second Annual Meeting of Operations Research Society, Chicago, 1954.

* This is a skeleton bibliography with no pretensions to any sort of completeness. A more extensive list is being prepared.

13. A. Vasonyi, "Industrial Operations Research and Mathematical Models", Seventh Annual International Systems Meeting, Systems and Procedures Association of America, October 1954, New York City.

14. J. von Neumann, "A Certain Zero-sum Two-person Game Equivalent to the Optimal Assignment Problem, Contributions to the Theory of Games", Princeton University Press, 1953.

15. ———————, "Probabilistic Logics", Calif. Inst. of Tech., 1952, notes by R. S. Pierce.

Part II

1. M. Beckmann, C. B. McGuire and C. B. Winsten, "Studies in the Economics of Transportation, (to appear).

2. A. Boldyreff, "An Iterative Technique for Determining Rail Capacity", (to appear).

3. M. Flood, "On the Hitchcock Distribution Problem", Pacific Journal of Mathematics 3, 1953, pp. 369-396.

4. L. R. Ford and R. Fulkerson, "Maximal Flow Through a Network", The RAND Corporation, Paper No. P-605, 1954.

5. F. L. Hitchcock, "The Distribution of a Product from Several Sources to Numerous Localities," Journal of Mathematics and Physics, 20 (1941), pp. 224-230.

6. L. Kantorovitch, "On the Translocation of Masses", Doklady Akad Nauk SSSR, 37 (1942), pp. 199-201.

7. T. C. Koopmans and S. Recter, "A Model of Transportation, Chapter XIV, Activity Analysis of Production and Allocation", J. Wiley and Sons, 1951.

8. A. Menger, see D. Konig, "Theorie der Endlichen und Unendlichen Graphen", Chelsea, 1950, p.244.

Part III

1. K. Arrow, T. E. Harris and J. Marschak, "Optimal Inventory Policy", Cowles Commission Paper No. 44, 1951.

2. R. Bellman, "An Introduction to the Theory of Dynamic Programming", The RAND Corporation, Report No. R-245, 1953.

3. R. Bellman, I. Glicksberg and O. Gross, "The Theory of Dynamic Programming as Applied to a Smoothing Problem", J. Soc. Indust. Appl. Math., Vol. 2 (1954), pp. 82-88.

4. ————————————————, "On Some Mathematical Problems Arising in the Theory of Optimal Inventory and Stock Control", (to appear).

5. G. Dantzig and S. Johnson, "Notes on Linear Programming: Part XXIII, A Production Smoothing Problem", The RAND Corporation, Report No. RM-1432, 1955.

6. A. Dvoretzky, J. Kiefer and J. Wolfowitz, "The Inventory Problem", Econometrica, Vol. 20 (1955), pp. 187-222.

7. W. Jacobs, "The Caterer Problem", Naval Research Logistics Quarterly, Vol. 1 (1954), pp. 154-165.

8. S. Karlin, "A Discrete Smoothing Problem", (to appear).

9. H. Robbins, "Some Aspects of the Sequential Design of Experiments", Bull. Amer. Math. Soc., Vol. 58 (1952), pp. 527-536.

10. N. Wiener, "Cybernetics", J. Wiley and Sons, 1948.

Part IV

1. R. Bellman, "The Theory of Dynamic Programming", Bull. Amer. Math. Soc., Vol. 60 (1954), pp. 503-516.

2. A. Ostrowski, "On Two Problems in Abstract Algebra Connected with Horner's Rule", Studies in Mathematics and Mechanics, presented to R. Von Mises, Academic Press, New York, 1954.

3. H. H. Seward, "Information Sorting in the Application of Electronic Digital Computers to Business Operations", R-232, Project Whirlwind, Digital Computer Laboratory, Mass. Inst. Tech.

4. T. S. Motzkin, "Evaluation of Rational Functions", Bull. Amer Math. Soc., abstract 314t, Vol. 61 (1955), p. 163

5. ————————————————, "Evaluation of Polynomials, ibid.

A number of papers devoted to the logical design of computers and the application of Boolean algebra may be found in the two new journals:

Computers and Automation, 1953 et. seq. — E. C. Berkeley and Associates, 36 West 11th Street, New York 11, New York

The Journal of Computing Systems, 1953 et. seq. — The Institute of Applied Logic, 45 West Water Street, St. Paul 1, Minnesota

An important bibliography is contained in the recent paper by J. Todd, "Motivation for Working in Numerical Analysis", Communications on Pure and Applied Mathematics, Vol. XIII (1955), pp. 97-116.

Part V

1. R. Bellman, "The Theory of Dynamic Programming", Bull. Amer. Math. Soc., Vol. 60 (1954), pp. 503-16.

2. R. Bellman, I. Glicksberg, and O. Gross, "On Some Mathematical Problems Arising in the Theory of Optimal Inventory and Stock Control", RAND Corporation, P-580, 1954.

3. ______________________, "Some Non-linear Functional Equations in the Theory of Dynamic Programming", Proc. Nat. Acad. Sci., to appear.

4. A. Charnes and C. E. Lemke, "Computational Theory of Linear Programming, II—Minimization of Non-linear Separable Convex Functions", O. N. R. Research Memorandum, No. 16, May, 1954.

5. G. B. Dantzig, "A Computational Procedure for a Scheduling Problem of Edie, RM-1290, RAND Corporation.

6. ______________________, "Application of the Simplex Technique to a Transportation Problem, Activity Analysis of Production and Allocation", Wiley, 1953.

7. ______________________, "Linear Programming Under Uncertainty, RM-1374, RAND Corporation, 1955.

8. G. B. Dantzig and A. R. Ferguson, "The Problem of Routing Aircraft—A Mathematical Solution", RM-1369, RAND Corporation, 1954.

9. G. B. Dantzig and D. R. Fulkerson, "Minimizing the Number of Carriers to Meet a Fixed Schedule", RM-1328, RAND Corporation, 1954.

10. L. C. Edie, "Traffic Delay at Toll Booths", Jour. Operations Research Society, Vol. 2 (1954), pp. 107-138.

11. M. M. Flood, "Application of Transportation Theory to Scheduling a Military Tanker Fleet", Jour. Operations Research Soc., Vol. 2 (1954), pp. 150-162.

12. A. J. Hoffman and W. Jacobs, "Smooth Patterns of Production", Management Science, Vol. 1 (1954), pp. 86-91.

13. H. W. Kuhn and A. W. Tucker, "Nonlinear Programming", Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, U. of California Press, 1951.

14. H. Markowitz, "Portfolio Selection", Thesis, U. of Chicago, 1954.

15. J. Robinson and J. Walsh, "Routing of Empties for Fixed-Schedule Transportation", RM-406, RAND Corporation, 1950.

Part VI

1. E. W. Barankin, "The Scheduling Problem as an Algebraic Generalization of Ordinary Linear Programming", Industrial Logistics Research Project Discussion Paper, UCLA, Nov., 1952.

2. R. Bellman, "Limit Theorems for Non-Commutative Operations—I", Duke Math. Jour., Vol. 21 (1954), pp. 491-500.

3. R. Bellman, I. Glicksberg, and O. Gross, "On some Non—Classical Problems in the Calculus of Variations (to appear).

A recent paper devoted to the enumerative aspects of graph theory, with its many applications is

F. Harary, "The Number of Linear, Directed, Rooted and Connected Graphs", Trans. Amer. Math. Soc., Vol 78 (1955), p. 445-463.

Further references will be found there. The basic paper in the field is

G. Polya, "Kombinatorische Anzahlbestimmungen fur Gruppen, Graphen, and chemische Verbunderungen", Acta Math, Vol. 68 (1937), pp. 145-254.