SDC

SP-828

Mathematical Programming, Man-Computer Search and System Control

William Karush

16 May 1962

(SP Series)

SP-828

Mathematical Programming, Man-Computer Search and System Control

William Karush

16 May 1962

SYSTEM DEVELOPMENT CORPORATION, SANTA MONICA, CALIFORNIA

Mathematical Programming, Man-Computer Search and System Control*

William Karush

This analysis

~~This paper~~ falls into two main parts — the first~~, consisting of Sections I and II,~~ describes some current research work ~~of the author~~ in mathematical programming and related fields; the second~~, consisting of Sections III and IV,~~ presents some general aspects of computer methods and applications which may influence the future work of researchers in mathematical methods of system optimization throughout the country.

In Section I we summarize some results on the problem of "distribution of effort," in which it is required to distribute a given resource among several activities so as to maximize the total return from the several activities. This type of problem arises in various contexts; one is in the computer programming problem of allocating internal storage so as to minimize program running time; another is in the theory of search when the distribution of effort is sought which will maximize the probability of detection. In the common treatment of distribution, the assumption is made that the return, or pay-off, function for each activity is concave, that is, marginally decreasing.

*One of a series of talks given to SDC personnel by members of the Research Directorate describing the work of the Directorate.

The research considered here develops an algorithm for optimal distribution which does not impose this restriction on the pay-off functions, thereby permitting a wider application of the mathematical model.

This investigation highlights a basic operation on functions which we call the "maximum convolution" of two (pay-off) functions; this is the function whose value $f(x)$, for each value of the argument x, is the optimal return when x is distributed between the two given functions. This operation has important ramifications beyond the problem of distribution of effort, and in Section II of this paper we report on some research results dealing with this operation[1]. The maximum convolution is analogous to the well-known operation of (integral) convolution in classical applied mathematics and arises naturally in problems of optimization. In the classical case, the Laplace transform serves to convert ordinary convolution to an algebraic operation; what we introduce as the "maximum transform" achieves the same end for the new type of convolution.

Mathematical programming may be described as the study of mathematical models of systems for the purpose of realizing optimal performance of the systems. Research throughout the country on mathematical models in the past twenty years has produced a variety of optimization algorithms of importance in systems problems and operations research, and on-going investigations will

---

[1] The work is being carried out jointly with R. E. Bellman of the RAND Corporation.

extend these methods and their range of application. It is to be expected that advances in information-processing control systems and in the computerization of problem-solving will offer new possibilities of methods and applications which may have a significant effect on this mathematical research. Much preliminary study is required to formulate models and research problems in a manner that will make it likely that future research in optimization methods will be fruitful in its relation to these developments.

Section III of the paper deals with the possibility of conducting research on an unsolved problem by a closely coupled man-computer team. A cooperative search would be carried out with the researcher continually formulating conditional trials involving conjectures, special cases, sub-problems, partial algorithms, etc., and the computer assembling and processing the necessary information. The two would be tied together in short-response link by means of console and display apparatus through which the researcher would insert digital and graphical information by push-button and light pencil and on which the computer would display information in similar forms to the man. This joint activity might be regarded as the use of what has been called "man-computer symbiosis" in a form adapted to experimental mathematical research[2]. The

[2] See J. C. R. Licklider, "Man-Computer Symbiosis," IRE Trans. on Human Factors in Electronics, pp. 4-11, March 1960. This reference was drawn to the attention of the author by Lee S. Christie in his SDC document M-5642, March 13, 1962, to which a reprint of the article was attached.

eventual outcome of research of this sort on a given problem might be an algorithm which can be automated, i.e., carried out exclusively by the computer, or it might itself be a man-computer search procedure to be conducted with the computer by a trained operator.

The final Section IV of the paper treats the question of bringing mathematical models and algorithms into closer relationship with the realities of automatic and semi-automatic control of an operating system, where control is based on information about a system, and deals only indirectly with the system itself. Mathematical representation, roughly speaking, is a link connecting a controlled system with the information-based decision procedures of system control. Research in mathematical programming and models has not been centrally concerned with questions of informational requirements for the execution of theoretical models; these questions involve such factors as incomplete information, information lags and errors, cost of information, the capacity (human or mechanical) to handle information and execute decisions, and the effect of organization of information and decision-making on system optimization. The incorporation of such factors into research in mathematical optimization, where feasible, might conceivably call for evolving a research language which combines both mathematical and meta-mathematical features.

I. Optimal Distribution of Effort[3]

The problem to be considered is that of distributing a given amount w of a particular resource among several activities $i = 1,2,\ldots,m$, in such a way

[3] The research reported here is described more fully in the SDC document by the author entitled "A General Algorithm for the Optimal Distribution of Effort," TM-616, 49 pp. May 1961. This paper has been accepted for publication by Management Science.

as to maximize the total return, where the return from the i^{th} activity is a given function $f_i(x_i)$ of the amount x_i allocated to it. Formally, the problem is that of determining

$$(1) \qquad F(w) = \max\,[f_1(x_1) + f_2(x_2) + \ldots + f_m(x_m)], \qquad w \geq 0,$$

where the maximum is taken subject to the restraints

$$(2) \qquad \begin{aligned} & x_1 + x_2 + \ldots + x_m = w, \\ & 0 \leq x_i \leq A_i, \quad i = 1,2,\ldots,m. \end{aligned}$$

What is required is an algorithm for the effective computation of the maximum return function $F(w)$ and the corresponding optimal distributions $x_i(w)$ for arbitrary w.

By way of illustrating an application of this mathematical representation, consider the problem of searching a region R for a concealed object, as in searching a span of ocean by aircraft for a submarine[4]. Suppose R has been divided into m sub-regions $R_1, R_2, \ldots$ and an à priori probability p_i can be estimated for the presence of the object in the i^{th} sub-region. Suppose further that for each sub-region R_i there is given a function $g_i(x_i)$ which specifies the probability of detection in that region when the object is

[4] Cf. B. O. Koopmen "The Theory of Search: Part III, The Optimum Distribution of Searching Effort," Operations Research, vol. 5, pp. 613-626 (1957)

present there and x_i units of effort (e.g., flying hours) are expended in searching; that is, $g_i(x_i)$ is the conditional probability of detection in R_i, given the presence of the object in R_i. The total probability of detection for a given distribution $x_1, x_2, \ldots$ is given by

$$p_1 g(x_1) + p_2 g(x_2) + \ldots + p_m g(x_m).$$

If this quantity is taken as a measure of effectiveness, then what is to be determined is a distribution which maximizes this sum; writing $p_i g_i(x_i)$ as $f_i(x_i)$ identifies this problem with the previous one (A_i being the maximum effort that may be expended on R_i). Notice that the mathematical form allows the conditional probability $g_i(x_i)$ to depend on the nature of the sub-region as well as the effort expended.

Another illustration occurs in a model proposed for optimal storage allocation in computer programming[5]. Consider a computer program P which is to be assembled out of m sub-routines, $i = 1,2,\ldots,m$, each sub-routine realizing a given function G_i of the total program P. Suppose the available library of sub-routines permits a function G_i to be accomplished by alternative sub-routines, these alternatives allowing computer storage space to be traded off against running time. That is, associated with each G_i is a (mathematical) function $t_i(s_i)$ which expresses the time to execute G_i by a sub-routine requiring

[5]Cf. J. E. Kelly, Jr. "Techniques for Storage Allocation Algorithms," Comm. ACM, vol. 4, pp. 449-454, Oct. 1961.

s_i storage units, and $t_i(s_i)$ is a decreasing function of s_i. The time T to execute P by an assembly of sub-routines with individual storage requirements $s_1, s_2, \ldots$ is given by

$$T = t_1(s_1) + t_2(s_2) + \ldots + t_m(s_m).$$

With T as a measure of the comparative worth of alternative sets of sub-routines, this leads to the problem of minimizing T subject to the constraint that $s_1 + s_2 + \ldots + s_m$ does not exceed the total available storage in the computer. Apart from modifications, this problem is of the form (1) and (2) described above.

Let us return, then, to the general problem of distribution of effort as expressed in (1) and (2). The standard theory of optimal distribution imposes the requirement that the return functions f_i be marginally decreasing, or concave; this means that for each activity i, a small increase of effort added onto a given level of effort produces less incremental gain, the higher the level. When this assumption is granted, the algorithm for optimal distribution is comparatively simple; the solution is built up step-by-step by allocating the next increment of effort at each stage to that activity which is then marginally the most favorable, that is, which then has a largest derivative.

Besides being mathematically appealing, the assumption of marginal monotonicity would seem to be empirically sound in many cases[6]. Nevertheless,

[6]The proof of this property for general types of mathematical models raises an interesting research question; cf. W. Karush, "Marginal Analysis of Lost Sales," Management Sciences, Models and Techniques, Proc. Internat. Meeting of Man. Sci. Pergamon Press, N. Y., vol. I, pp. 523-538, 1960.

this assumption is by no means universally valid in applications, and it is not difficult to see how it might break down. To illustrate this, suppose an individual activity can be carried out by either of two means, the first yielding a return $r(x)$ and the second $s(x)$ for an allocation of effort x; in the above case of search, with x denoting, for instance, the man-hours spent in searching a given sub-region, then $r(x)$ might correspond to the use of one man in an aircraft and $s(x)$ to the use of two men in the aircraft for half the time. It is reasonable that $r(x)$ and $s(x)$ are individually concave, but the maximum of the two need not be, and it is this maximum which should be used as the return function $g(x)$ for the individual activity in question.

An algorithm has been developed by the author which allows the (increasing) return functions f_i in (1) to have arbitrary shapes. The method involves a sequential construction of maximized sums of functions in the spirit of dynamic programming and employs a technique of representing an arbitrary function as the maximum of a family of concave and convex "strings." The maximized sum of two functions is achieved by a simple table showing how to combine pairs of strings. The tabular construction is iterated until the final return function $F(w)$ is generated, and the optimal distribution functions $x_i(w)$ are found by tracing back through the tables to the original functions. The algorithm yields these functions for all values of w, and it is amenable to mechanized computation. It must be realized that in the general non-concave case the solution requires a "global" procedure and not a marginal (differential) approach; an increment which is locally disadvantageous may be eventually advantageous because of the non-decreasing behavior of the slope. One consequence of this is that the optimal allocation functions $x_i(w)$ may be discontinuous in w; for instance, at some critical value w^* the amount allocated to the first activity may drop from a positive value x_1^* to the value of 0 and the amount x_1^* assigned to some other activity.

II. The Maximum Transform[7]

A functional operator arising in the problem of distribution of effort is the "maximum sum" of two functions f and g, defined by

$$(3) \qquad h(x) = \max_{u+v=x} [f(u) + g(v)].$$

This operation is basic for a variety of problems in mathematical programming, and it would seem to be valuable to develop the theory and application of this operation. We shall outline some of the research results that have been obtained in this direction. The intended scope of the theory requires that f,g and h be taken as functions of n variables, so that x,u,v in

[7] This research is described in detail in the following SDC documents by the author and R. E. Bellman: (1) "On a New Functional Transformation in Analysis: The Maximum Transform," SP-370, 7 pp., June 1961; (2) "The Maximum Transform — I," TM-665, 28 pp., Nov. 1961; (3) "The Maximum Transform — II," TM-689, 26 pp., Feb. 1962; (4) "The Maximum Transform — III," TM-701, 15 pp., Feb. 1962; (5) "On the Maximum Transform and Semigroups of Transformations," SP-719, 4 pp., March 1962. Document (1) has appeared in the Bull. Amer. Math. Soc., vol. 67, pp. 501-503, 1961; document (5) has been submitted to the same journal. The remaining documents and related material will appear in two articles: "Mathematical Programming and the Maximum Transform," Journal of SIAM, and "On the Maximum Transform," J. of Math. Analysis and Applications.

(3) represent vectors of n (non-negative) components (e.g., $x = (x_1, x_2, \ldots, x_n)$, $x_i \geq 0$); however, the reader may continue to regard these as single real variables if he chooses. The function h is termed the <u>maximum convolution</u> of f and g and is written symbolically

$$h = f \oplus g.$$

This convolution has a multiplicative form which is obtained by a simple exponential change of variable and which we introduce temporarily for the purpose of pointing out an important analogy to classical applied mathematics. The multiplicative maximum convolution, symbolized by $F \otimes G$, is defined (for non-negative functions) by $\max\limits_{u+v=x} [F(u) \times G(v)]$, or

$$\max_{0 \leq y \leq x} [F(y) \times G(x - y)].$$

This expression is immediately suggestive of ordinary (integral) convolution $F * G$ given by

$$\int_0^x F(y)G(x - y)dy.$$

Ordinary convolution is of widespread occurrence in applied mathematics; to quote just two instances, it is used in combining probability distributions and in constructing solutions of partial differential equations out of "point source" solutions. Maximum convolution occupies a role in the optimization problems of mathematical programming which is comparable in some respects to ordinary convolution in classical analysis.

Fundamental for the application of ordinary convolution $F * G$ is the Laplace transform $\Phi = L(F)$ of a function F, given by

$$\Phi(\xi) = \int_0^\infty e^{-\xi x} F(x)dx.$$

This transformation has the property of converting convolution to algebraic multiplication, i.e.,

$$L(F * G) = L(F) \times L(G).$$

This is important in various applications; for example, it allows certain types of analytical problems involving functions to be transformed into (more manageable) algebraic problems involving transformed functions. The analogous process for mathematical programming would be a transform $T(F)$ with the property $T(F \otimes G) = T(F) \times T(G)$, and it is this kind of problem we shall consider.

We return now to the additive form $f \oplus g$ of the maximum convolution, which is more convenient for applications of mathematical programming, and raise the question of the existence of a functional transformation $\Phi = T(f)$ which converts this convolution to algebraic addition, i.e., which has the property

$$T(f \oplus g) = T(f) + T(g). \tag{4}$$

The answer is a positive one: Such a transformation is given by

$$\varphi(\xi) = - \max_{x \geq 0} [- \xi x + f(x)]$$

(when x is a vector of n components, then so is ξ and ξx is understood to mean the inner product $x_1\xi_1 + x_2\xi_2 + \ldots + x_n\xi_n$). This particular transformation is denoted by M and the function $\varphi = Mf$ is called the maximum transform of f. It should be noted that unlike the Laplace transformation M is not a linear transformation: that is, in general, $M(f + g) \neq M(f) + M(g)$ and $M(k \cdot f) \neq k \cdot M(f)$, for a constant k; in particular, if $\varphi(\xi)$ is the transform of f(x), than $k\varphi\left(\frac{\xi}{k}\right)$, not $k\varphi(\xi)$, is the transform of kf(x) for non-negative k. This non-linearity is a complicating factor in the theory of the maximum convolution.

An important question is that of determining the inverse M^{-1} of the transformation M, that is, determining the general solution of the equation $Mf = \varphi$ for the unknown function f (this is needed, for example, to convert an answer derived in the space of transforms back to the space of an original problem). The answer to this question is that M is its own inverse M^{-1}; the required solution of $Mf = \varphi$ is $f = M\varphi$, the maximum transform of φ. More precisely, the inversion is given by $\overline{f} = M\varphi$, where $\overline{f}$ is the increasing concave "cap" of f (identical to f when f is an increasing concave function). Other questions have been considered; for example, it has been shown that although there may be various transformations T having the property $T(f \oplus g) = T(f) + T(g)$, the particular transformation M is the "best" of these, in an appropriately formulated sense. Details of this and other results can be found in the references cited above.

As an application of the maximum transform M, consider the following problem in multi-stage decision-making over a sequence of time periods. At the beginning of the first period, an amount x of a given resource is at hand of which some portion y is to be committed that yields a return $g(y)$, discounted to this point in time; the remainder $x - y$ is to be withheld as available reserve for the beginning of the second period. When the latter point in time is reached, however, the available amount $x - y$ is assumed to have deteriorated (or been enhanced) by a factor k so that actually $k\cdot(x - y)$ is at hand for partition into an immediate commitment and a reserve. The problem is to determine the allocations $y_1, y_2, \ldots, y_N$ for N successive (equal) time periods which will maximize the total return, discounted to the initial point of the entire process.

This problem can be expressed by means of a functional equation of dynamic programming. Let $f_n(z)$ be the maximum return over n intervals beginning with an initial amount z and discounted to the initial point of the n intervals. Consider the $(n + 1)$-period problem; if y out of x units is committed in the first period, then $k\cdot(x - y)$ is the initial amount for the subsequent n-period problem and $f_n(k\cdot(x - y))$ is the most that can be realized in the n remaining periods. Hence $g(y) + rf_n(k\cdot(x - y))$ is the total return for $n + 1$ periods, where r is the discount factor over one period. This results in the desired functional equation

$$f_{n+1}(x) = \max_{0 \leq y \leq x} [g(y) + rf_n(k\cdot(x - y))],$$

with $f_1(x) = g(x)$. The right side of the displayed equation is a maximum convolution of the two functions in the sum, and we apply the maximum transformation M to this equation. By the fundamental relation (4), this yields

$$\varphi_{n+1}(\xi) = \psi(\xi) + r\varphi_n(\rho\xi), \qquad \text{with } \rho = \frac{1}{rk},$$

where φ_i and ψ are the transforms of f_i and g (we have used the property that $b\varphi(\xi/ab)$ is the transform of $bf(ax)$ for positive constants a,b). Unwinding this recursion gives the following explicit solution in the space of transforms:

$$\varphi_N(\xi) = \psi(\xi) + r\psi(\rho\xi) + \ldots + r^{N-1}\psi(\rho^{N-1}\xi).$$

The maximum return is obtained by the inversion $f_N = M(\varphi_N)$; more generally, we have $f_n = M(\varphi_n)$, $n = 1,2,\ldots,N$. The optimal allocations $y_1, y_2, \ldots$ for an initial amount x_N are obtained by recursive solution of

$$g(y_1) + rf_{N-1}(k\cdot(x_N - y_1)) = f_N(x_N)$$

$$g(y_2) + rf_{N-2}(k\cdot(x_{N-1} - y_2)) = f_{N-1}(x_{N-1}), \qquad \text{with } x_{N-1} = x_N - y_1,$$

and so on.

To facilitate the transition between functions and transforms in various applications, it would be useful to have available a table of transforms of common functions, just as with the Laplace transform. A brief list is given in the SDC document TM-665; an example is

$$f(x) = a + b \log x, \ \varphi(\xi) = a + b(\log \frac{b}{\xi} - 1).$$

In practical cases, a wide class of functions can be handled through approximation by special classes of simpler functions and tabulating transforms of the simpler functions.

III. Man-Computer Search

Research in mathematical programming has produced effective computational algorithms for many useful optimization models and has had an important effect in various computer applications. The more common mathematical models utilize as an underlying collection of entities among which preferred alternatives are to be selected, a continuous space described by a finite number of real variables $x = (x_1, x_2, \ldots, x_n)$, and the typical problem requires the maximization (minimization) of a real-valued function of x with the admissible, or feasible, x described by certain relations involving the components of x. Formulation in terms of continuous real variables, real-valued functions of these, and arithmetical operations and relations, permits recourse to a large body of existing mathematical knowledge, and this advantage has been exploited in developing mathematical methods. Applied mathematical research of this kind has great value and will continue to discover solution procedures for important classes of optimization problems.

In recent years, the need has grown for optimal procedures for classes of problems which do not fall within the framework of ordinary analysis. One general class involves such problems as sequencing of operations, network flow, line-balancing, etc., which are expressed in combinatorial terms such as permutations of elements, partitions of sets of elements, partial orderings, nesting, network relations, and the like. Mathematical structures easily become very complex that require the optimization of functions whose arguments are permutations, or networks, or similar entities which are constrained in various ways to lie in certain subsets of permutations, networks, etc. Such structures are often of especial interest in behavioral research and in problems of the automation of decision-making[8].

The following "machine sequencing" problem falls in this combinatorial class. There are m objects $i = 1,2,\ldots,m$ to process through n centers, or machines, $j = 1,2,\ldots,n$; object i requires a time t_{ij} in center j, and process $j + 1$ on object i cannot begin until process j is complete on i; each center can process one object at a time, but any number of centers can operate simultaneously on separate objects. The problem is to sequence the m objects through the n centers so that the processing of all objects is completed in a minimum time. A complete solution has been obtained for m objects and two machines[9]; the problem is unsolved for m objects and three machines, and a mathematical solution seems out of the question for the general problem of m objects and n machines at the present time. Another example of a model of the general kind under discussion concerns optimal

---

[8] For a discussion of this and related matters see W. Karush "On the Use of Mathematics in Behavioral Research," SDC document SP-807, May 1962

[9] S. M. Johnson "Optimal Two-and Three-Stage Production Schedule with Setup Times Included," Naval Res. Log. Qtrly, vol. 1, pp. 61-68, March 1954

procedures for "PERT-like" management control systems; here, expressing that certain activities must be completed before others can begin leads to a network, or partially ordered arrangement, of activities. If the cost c_i of a given activity i is a suitably limited function of the time t_i allowed for carrying out the activity, and the objective is to assign times t_i so as minimize total cost $\Sigma c_i(t_i)$, then a solution algorithm can be prescribed[10]. The problem quickly becomes a mathematically unsolved one by imposing certain features of greater realism, such as prescribing an upper bound or some regularity on the rate of consumption of a resource.

These difficult problems are often finite in character, requiring a best selection to be made among only a finite number N of alternatives. In principle, it is only necessary to compute the objective function for the N cases and pick the one with highest functional value; in practice N is so enormous that this procedure is out of the question even with the fastest computers. The finite nature of the problem indeed raises the question of what is to be meant strictly by a solution algorithm; what is generally sought for is a procedure which requires a search of a number of cases of a considerably lower order of magnitude than N and is computationally feasible.

A fundamental difficulty in the class of problems being discussed is that typically no systematic method is at hand for moving from one feasible

[10] D. R. Fulkerson "A Network Flow Computation for Project Cost Curves," Management Science, vol. 7, pp. 167-178, January 1961

alternative to another that is sure to be better, quite apart from the larger question of tracking down the optimal alternative itself. This is to be contrasted with optimizing a function of continuous real variables, where a "gradient" is available to point the direction to an improved alternative[11]. In less well understood problems, such as combinatorial problems, what needs to be developed is a "feel" for how various possible systematic schemes of changes incrementally affect the objective function and other dependent features — an analogy might be a complex system of linkages in which one wants to learn how moving certain end rods in certain ways affects the motion of other end rods.

There is important basic research to be done in new concepts and methods of mathematical programming if we are to make substantial progress in some of the outstanding questions in this field, as the preceding discussion has suggested. What is of interest here are means of exploiting the high-speed digital computer as an instrument in this research. The use of the computer as an experimental laboratory to simulate and explore mathematical structures in order to gain insights into these structures is an existing and developing technique; the term "experimental arithmetic" has been used to refer to it, or at least certain aspects of it[12]. The methods have almost exclusively been of a nature in which the computer experiments are pre-programmed, and the

---

[11] Cf. Edward L. Pugh "Gradient Techniques for Nonlinear Programming," SDC document TM-695, 21 pp. February 1962.

[12] A "Symposium on Experimental Arithmetic" was recently sponsored by the Institute for Defense Analyses at the University of Chicago in conjunction with the 589th meeting of the American Mathematical Society.

experiments are run off-line with respect to the activities of the researcher; experimental results, often presented in unwieldy form, need to be analyzed and studied in preparation for further experimental ideas. This method would seem to require that the structure of a mathematical system be well enough understood to make this intermittent interaction with the computer a feasible one.

The principal interest in this paper is the possibility of a more intimate interaction between researcher and computer in which exploratory search would be carried out as an on-line exchange between man and computer. The lack of knowledge of the structure under study would require a highly conditional interaction where the researcher would require quick answers to various probing questions and would be thinking through aspects of the problem and trying out ideas as the search went along. The power of the computer to search through a large number and a great variety of cases as specified by the man would be combined with the intuition and intelligence of the man in a cooperative attempt to learn more about the unsolved problem at hand. It might be argued that heuristic problem-solving carried out exclusively by the computer is a preferable approach. However, it would be difficult to give credence to claims of likely machine solution of mathematical research problems which defy solution by talented mathematicians; but there is no reason that advances in computer heuristics could not be incorporated in an extension of the computer's role in a man-computer research team.

Many questions arise in considering the possibility of matching man capabilities with those of the computer to form an effective man-computer team. These are discussed in the reference cited earlier, which treats the feasibility and requirements of this approach from a broader point of view and with comparative emphasis on its uses in empirical studies and operating systems. We wish to consider briefly some aspects more specific to the investigation of mathematical models, where the formulation of the problem is precise but the discovery of means of solution requires search and experimentation whose course cannot be anticipated.

It may be imagined that communication between researcher and computer is realized by means of a console and display unit which is under the observation and control of the researcher. Computer-generated information called for by the operator is displayed to him in graphical forms and digital arrays; in turn, he designates data and computer instructions by depressing push-buttons and by drawing graphical forms with a light pencil on a display scope. The scale of graphical forms is modifiable so that finer structure can be examined, and a long continuous graph or network can be unrolled before the man, as a film might be. The man can readily draw in new graphs or change existing forms to serve as new inputs in an iteration procedure. The display scope can accommodate a number of similar displays simultaneously so that present results can be compared with past ones as a basis for proceeding further.

In dividing functions between man and computer, the tasks assigned to the latter are to require lengths of time which are in balance with the periods of time required for guessing and decision-making by the former. Tasks are also to be partitioned with respect to characteristics other than the time they require. For example, a computer might well be assigned construction of one-dimensional elements such as a curves (representing functions f(x) of one variable, say) or a number of curves (representing functions f(x,y) of two variables as families of one-dimensional relations f(x,y) = const.). Also, the computer can be required to search such an element or a family of elements for certain well-defined features as, for example, the maximum of f(x). The researcher may take over searching for features that cannot be anticipated or recognizing more complex patterns of curves, families of curves, or arrays of numbers. A parametric space of several dimensions might be searched by a series of one-dimensional paths, a single path (or finite set of paths) being accomplished quickly by the computer but the selection of the next path being made by the researcher whose judgment is conditioned by the pattern of past results.

The computer is to be programmed to allow the researcher to re-organize and create procedures with speed and flexibility. Push-button or similar actions are to be available to the operator for directing an executive routine to assemble quickly the sub-routines for any one of a variety of possible computer procedures. That is, in addition to changing readily the entities such as graphical representations which serve as input data for computer processing, the researcher must be able to manipulate with appropriate freedom the operations to be applied to the data. Also, it should be possible to direct the computer to begin recording and organizing certain kinds of information which will be presented to the operator at certain times, either automatically or at his instruction. Post-problem analysis by the computer or by the man-computer team of successful courses of solution in particular cases can serve as a basis for abstraction of trial-and-error procedures into more systematic procedures in later stages.

There has been little experience in the use of this method of research, and it is not an easy matter to decide whether a given mathematical problem is appropriate for study by such a method. One early question is what display form should be used to represent a mathematical structure in order to give the greatest assistance to the intuition of the researcher at the console. Alternative forms are possible; for example, in the machine scheduling problem referred to earlier, information on a given arrangement and processing of the objects may be represented by a scheduling chart of line segments arranged in parallel rows, or a "maximum walk" through a matrix of numbers, or a set of values of certain algebraic formulas, etc.: what is the "natural" way to present the information? The form of representation in the computer is likewise important; mathematical systems such as network structures or partially ordered sets of entities should be coded so they may be manipulated easily. The form of coding may need to be devised keeping in mind a non-trivial question of how to determine when two different designations represent, in fact, "isomorphic" mathematical structures (a simple example is provided by the names "1/4" and "0.25" specifying the same number in ordinary arithmetic).

A common exploratory approach to a mathematical research problem is to devise and examine equivalent formulations of the problem in various mathematical or applied contexts with the hope that one of these will yield a key idea. A similar preliminary analysis would be important in deciding whether there are feasible forms of the problem to work with in a man-computer research enterprise, and what the best form might be; in a given set-up with the computer the researcher would not expect to have at his disposal all equivalent formulations, nor would he want this if it were possible to see in advance that one is better than others. If man-computer methods prove to be of major significance in mathematical research, they will demand as much hard creative thinking as research does now, except that new considerations will come to the fore.

A general kind of problem that suggests itself for man-computer research is one of the "fixed point" type, where it is required to find a solution C of an equation $C = F(C)$. In this general form, C is restrained to lie in a class of admissible elements which may be functions, permutations, partially ordered systems, or any well-defined structures, and the transformation F operates in a well-defined, though possibly involved, way on an input element X to produce as output element $Y = F(X)$. In a man-computer enterprise, a search would be made through the space of elements X in an attempt to home in on an input C which is the same as its output; the computer would have the task of computing $F(X)$ for each trial element X, and the man would perform the selection of a next trial element X_{n+1} conditioned by the trials $X_1,\ldots,X_n$ and outputs $Y_1,\ldots,Y_n$ to date. He would attempt to "tune in" the computer to the stationary state $C = F(C)$.

This method is being used in some research work being carried out[13] on the solution of non-linear integral equations in theoretical physics. The desired element C is a function of a single real variable, and the transformation F is a non-linear integral operator which maps one function X into another Y. The elements X and Y are presented on a display scope as graphs in rectangular coordinates. The shape of a last input X_n can be modified in a more or less arbitrary manner by marking various points of interpolation on the grid; these points are inserted by manual control of horizontal and vertical hairlines which are used to mark cross-hair points; a light pencil is used to erase such points. The next input X_{n+1} can be taken as such a modified graph or it can be taken as a linear combination of X_n and Y_n with coefficients designated by push-button. The display scope can be requested to present simultaneously any combination of the four graphs X_{n-1}, Y_{n-1}, X_n, Y_n. The mathematical problem is initiated by selecting three numerical parameters which fix the operator F; the operator F, in technical terminology, has norm greater than 1 for some parametric values, so that blind iteration with the output at one stage taken as the input at the stage need not converge, even if a good initial approximation X_1 is used. This first phase of the research is now being generalized to allow console programming of a variety

[13] By Dr. Glenn J. Culler and others at the Ramo-Wooldridge Division of Thompson Ramo Wooldridge; cf. G. J. Culler and R. W. Huff "Solution of Non-Linear Integral Equations Using On-Line Computer Control" reported at the AFIPS Spring Joint Computer Conference, San Francisco, May 1-3, 1962.

of functional operators built out of elementary operators under the control of the researcher. In this way, he can compose many functions and many ways of modifying functions for experimentation in the search for solutions or methods of solution. What are involved in the recent work are forms of composition of operators; the computer program which builds an operator C out of operators A and B in a certain way as directed by console action can thereafter make available any C' built in the same way out of other operators A' and B'; for example, A' can be taken as B and B' as C. These constructions have the advantage that they can be checked out as they are built by the on-line techniques of a man-computer team.

In this discussion we have viewed man-machine search as a technique of research. The purpose of such a research effort applied to a given mathematical problem need not be the eventual discovery of an algorithm which itself requires a man-computer team for actual execution. The algorithm might turn out to be of this sort, or it might turn out to be one that can be done efficiently by hand computation or by automatic machine computation. Cost considerations, for example, would be factors in deciding whether or how the outcome of the research might be put to practical use.

IV. System Control[14]

In this final section of the paper we wish to draw attention to certain

[14] Some of the ideas described in this part of the paper were developed jointly with Dr. Andrew Vazsonyi who served as a consultant with the Research Directorate.

possible implications of the development of information-processing methods of control upon research in optimal operation of systems.

Broadly speaking, the growth of man's knowledge and control of his environment has proceeded by a division of labor into scientific theory and engineering. As made here, the division is intended in a general sense which implies that the former deals with the discovery of first principles and the latter with putting these principles to work in a practical context. This separation seems to have been an effective way of employing human intelligence, and scientific theory has evolved in forms which are not centrally concerned with general problems of means of application, leaving to engineering skill the resolution of the special manifold difficulties of execution.

We can recognize a counterpart on a less grand scale in comparing research in mathematical programming and modeling with the design of operational control by information processing. In its role as scientific theory, the mathematical formulation employs idealizations and abstractions adapted to the search for general principles by mathematical methods; even with the motivation for selecting a given research question directed by practical needs, the problem is necessarily isolated out of its total complex in order to express a researchable question. The matter of re-imbedding results coming out of this research into a system of operational control subject to a host of technological and other restraints is a job of the "software" engineer and designer.

The point to be made is that the rise of modern computerized control methods with the mechanization of control through the communication and centralized processing of information is a force which is acting against this separation of theory and the means of using it. Control is exercised through information about a controlled system, not through direct contact with the system; the processes of information handling and decision-making are becoming regularized to a point which suggests the theoretical conceptualization and study of these phenomena. This formalization of operational control removes some of the obstacles to incorporation of informational control features into mathematical models of system behavior, where previously such features were too irregular and unsystematic for mathematical scientific consideration.

The matter of particular interest here is the connection between theoretical algorithms of optimization and the informational and decision-making requirements they impose for execution. In this connection it is important to consider the relationship between the state of a controlled system at any time and the state as known to the control mechanism through the information it has at hand. As an illustration, consider a mathematical formulation of a stochastic system in which entities belonging to different classes E_i arrive at random and are to be processed through a network of "action" points c_j; the passage of entities from one c_j to a next until exit from the network is to be effected by a set of conditional priority rules which depend upon the state of the system at any time, and we may take the state to be the number of entities in each class E_i present at each point c_j. We may imagine a mathematically optimal set of priority rules R has been determined relative to some acceptable measure of performance. However, to incorporate the mathematical procedure R into an operational context we must take into account the means by which R is to be executed; we cannot assume, as in the mathematical theory, that an omniscient decision-maker is present who has total knowledge of the state of the system at any moment and who can instantly and accurately act on the system. Is it possible to obtain complete, accurate and timely information on the state of the system? What feasible decision centers are to be set up to carry out the actions triggered by R? What information is to be made available to the decision centers and how shall they be interrelated? With the clarification of the nature and limitation of allowable control operations through the systemization and automation of these operations, it would appear that theoretical formulation can begin to take questions of this kind into consideration in defining and discovering optimal methods.

A preliminary requirement for the formulation of feasible research problems, which combine theoretical features of both a mathematical and informational character and which can have significance for operational control, is a precise and formalized description of types of underlying controlled system S, together with a similar description of types of information systems S^* which exercise control over it. The following is an outline of one general approach.

As noted, we distinguish between the underlying system S and the decision-making system S^* that envelopes and manipulates it. S is to attain certain objectives (possibly of a continuous nature) and S^* is to guide S toward them.

The system S is described principally in terms of certain physical components, operational procedures, and interactions which are to be taken as fixed for a given S (within possible parametric variation). At each moment in time, S exists in one of a class of many possible states; these states are well-defined and theoretically identifiable. In the course of time, S undergoes transitions between states, and the behavior of S is expressed as a history of transitions of states. We distinguish two types of transitions, internal transitions and external (control) transitions. The internal ones come about through dynamic relations inherent to the system which would describe the course of behavior of the system if it were left to itself, i.e., if it were not subjected to any control actions; these transition relations may be either deterministic or stochastic in nature. The external transitions are those super-imposed by the control system S^*, and these are subject to certain restraints of feasibility in terms of the structure of S and of a given S^*. As an illustration, we can imagine S as a toy mechanism on wheels which gropes toward a beam of light and left to itself would follow a path along the floor determined by the initial state (position and velocity) of the mechanism and the dynamic laws of transition, or motion, built into it; the imposed control S^* might be represented by a person who can pick up the toy at any time and set it down anywhere with a shove in a selected direction. This may be analogized to the general situation where we imagine the history of S as consisting of a sequence of time segments of behavior with the initial state of S at the beginning of a segment imposed by S^* and the course of behavior within a segment unfolding according to the internal laws of transition, until a next control action imposed by S^* puts S in a new initial state.

We shall not go into details here concerning the structure of S and S^*, but it may be appropriate to go a little further in the description of S^*. The control system contains decision-making units which receive information, come to decisions, and initiate action on these decisions; decision units may be represented by humans or by automated procedures. The design of S^* must take into account the limited capacity of a unit to take in information, to perform decision-making, and to manipulate control variables. The decision units are structured in an hierarchy. At the first level is the operator whose input information represents observation on the state variables of S and whose control actions are intended to impinge directly on S to produce state transitions of an immediate kind. The range of decision rules to which operators are restricted in transforming observations into the fixing of control values is determined and varied by the next-level decision-maker, who exercises control by fixing higher-order parameters that define classes of allowable decision-rules. The higher-level decision-maker himself exercises certain decision procedures for his task; there might be called "indirect" decision rules to distinguish them from the "direct" rules of the operators whose output produces a direct reaction on the system S. Input information for indirect rules would include aggregated or integrated effects over time, the occurrence of exceptional states or patterns of states, or other information appropriate to this level. Third and higher level decision units may be specified.

With the emergence of a formalized description of control systems it becomes possible to begin to delineate possible research questions. One question which has been considered[15] treats the trade-off between the cost of information and performance of the controlled system: How does performance improve with increased availability of information, even assuming optimal use of information? How do the constraints of a control organization S^* limit the attainable optimal performance of S, and how may alternative S^*'s be compared? The partition of decision-making functions raises the question of balancing conflicting sub-optimal procedures pursued by decision units at one level through the control actions of a higher decision unit so as to realize over-all maximum performance; this situation would be expected to occur in a realistic modification of the idealizations ordinarily implicit in mathematical models[16]. An operational control system does not work directly with the controlled system but with informational content that presents an image of it; this image may be "blurred" for many reasons. How shall the correspondence between true states and images of states be conceptualized? How does lack or inaccuracy of information affect performance of the controlled system S? The time lag of information is an ever-existing phenomenon: Can such lags produce increasing oscillations in S?

[15] Cf. J. Marschak, "Elements for a Theory of Teams," Management Science, vol. 1, pp. 127-137, January 1955.

[16] Cf. R. Radner, "The Application of Linear Programming to Team Decision Problems," Management Science, vol. 5, pp. 143-150, January 1959.

It is to be expected that some questions of these sorts can be formulated as precise research problems by appropriate modification of existing mathematical models and concepts. It is also possible to speculate on a greater divergence from present mathematical methods in which modeling which mixes mathematical and informational elements would lead to the development of novel concepts. Increased research in the area between mathematical formulation and informational control should generate symbolism and theoretical methods designed for the needs of such research. The study of new types of models may lean heavily on computer research methods involving experimental simulation by computer or men and computers (among which might appear experimental man-computer research methods of the kind described in the preceding section); the theoretical format that emerges may well be oriented toward the use of such methods. The simulation methods referred to here would be keyed to the study of general models and theoretical constructs; they are not to be identified with the type of simulation commonly used to duplicate a comparatively rigid, though complex, system. The research techniques would represent an extension of scientific method to be used to uncover general principles related to the optimal performance of operational systems.