SIAM-AMS Proceedings
Volume 9
1976

NONLINEAR PROGRAMMING: A HISTORICAL VIEW

Harold W. Kuhn[1]

ABSTRACT

A historical survey of the origins of nonlinear programming is presented with emphasis placed on necessary conditions for optimality. The mathematical sources for the work of Karush, John, Kuhn, and Tucker are traced and compared. Their results are illustrated by duality theorems for nonlinear programs that antedate the modern development of the subject.

1. INTRODUCTION AND SUMMARY

The paper [1] that gave the name to the subject of this symposium was written almost exactly twenty five years ago. Thus, it may be appropriate to take stock of where we are and how we got there. This historical survey has two major objectives.

First, it will trace some of the influences, both mathematical and social, that shaped the modern development of the subject. Some of the sources are quite old and long predate the differentiation of nonlinear programming as a separate area for research. Others are comparatively modern and culminate in the period a quarter of a century ago when this differentiation took place.

AMS (MOS) subject classifications (1970). 90-C30, 01-A60.

[1]Research supported by the National Science Foundation under grant NSF-MPS72-04983 A02.

Secondly, in order to discuss these influences in a precise context, a few key results will be stated and "proved". This will be done in an almost self-contained manner in the spirit of the call for this symposium which announced that the lectures would be pedagogical. The definitions and statements should help to set the stage for some of the papers to follow by providing a formal framework. In addition, these statements will allow the comparison of the results of various mathematicians who made early contributions to nonlinear programming. This will also give the pleasant opportunity to rewrite some history and give W. Karush his proper place in the development.

In §2, a definition of a nonlinear program is given. It will be seen to be a straightforward generalization of a linear program and those experienced in this field will recognize that the definition is far too broad to admit very much in the way of results. However, the immediate objective is the derivation of necessary conditions for a local optimum in the differentiable case. For this purpose, it will be seen that the definition includes situations in which these conditions are well-known. On the other hand, it will be seen that the definition of a nonlinear program hides several implicit traps which have an important effect on the form of the correct necessary conditions.

In §3, an account is given of the duality of linear programming as motivation for the generalization to follow. This duality, although it was discovered and explored with surprise and delight in the early days of linear programming, has ancient and honorable ancestors in pure and applied mathematics. Some of these are explored to round out this section.

With the example of linear programming before us, the nonlinear program of §2 is subjected to a natural linearization which yields a set of likely necessary conditions for a local optimum in §4. Of course, these conditions do not hold in full generality without a regularity condition (conventionally called the <u>constraint qualification</u>). When it is invoked, the result is a theorem which has been incorrectly attributed to Kuhn and Tucker. This section is completed by a description of the background of the 1939 work of W. Karush [2] (which is further amplified by an Appendix to this paper).

As will be seen in §4, the motivation for Karush's work was different from the spirit of mathematical programming that prevailed at the end of the 1940's. In §5, an attempt is made to reconstruct the influences on Kuhn and Tucker that led them to Karush's result. These include such diverse sources as electrical networks, game theory, and the classical theory of Lagrange multipliers.

Independent of Karush, and prior to Kuhn and Tucker, John had published a result [3] giving necessary conditions for the local optimum of a function subject to inequalities. His motivation was different from either of the other works and is described in §6. A crucial example that is typical of the type of geometric optimization problem that influenced John is Sylvester's Problem. This is given a modern and concise treatment in §7.

The conclusion of the paper, contained in §8, is a sermon on the nature of applied mathematics. It may be appropriate in that it was delivered at the first session of the symposium on a Sunday morning.

2. WHAT IS A NONLINEAR PROGRAM?

With malice aforethought and considerable historical hindsight, a <u>nonlinear program</u> will be defined as a problem of the following form:

$$\text{Maximize } f(x_1,\dots,x_n) \text{ for "feasible" solutions to}$$

$$g_1(x_1,\dots,x_n)-b_1 = -y_1$$

$$\dots \quad \dots \quad \dots$$

$$g_m(x_1,\dots,x_n)-b_m = -y_m$$

for given functions $f, g_1, \dots, g_m$ and real constants $b_1, \dots, b_m$. "Feasible" means that each x_j and y_i is required to be nonnegative, zero, or free.

The following examples show that this definition encompasses in a natural way a host of important special cases.

(1) If we specify that all x_j are free, all y_i are zero, and all b_i are zero, then the problem reads:

$$\text{Maximize } f(x_1,\ldots,x_n) \text{ subject to}$$
$$g_1(x_1,\ldots,x_n) = 0$$
$$\cdots \quad \cdots$$
$$g_m(x_1,\ldots,x_n) = 0.$$

This is the classical case of equality constrained (nonlinear) optimization treated first by Lagrange.

(2) If $f(x_1,\ldots,x_n) = c_1x_1+\cdots+c_nx_n$ is linear, each $g_i(x_1,\ldots,x_n) = a_{i1}x_1+\cdots+a_{in}x_n$ is linear, and all x_j and y_i are required to be nonnegative, then the problem reads (in customary vector-matrix notation):

$$\text{Maximize } c\cdot x \text{ subject to } Ax \leqq b, \quad x \geqq 0.$$

This is the familiar case of a linear program in canonical form.

(3) If f and all of the g_i are linear functions as in (2) and we require all x_j to be nonnegative and all y_i to be zero, then the problem reads:

$$\text{Maximize } c\cdot x \text{ subject to } Ax = b, \quad x \geqq 0.$$

This is a linear program in standard form.

(4) Let S be any set in R^n and let $g_1(x)$ be the characteristic function of S (that is, $g_1(x) = 1$ for $x\in S$ and $g_1(x) = 0$ otherwise). Then, if $m = 1$, $b_1 = 1$, all x_j are free, and $y_1 = 0$, the problem reads:

$$\text{Maximize } f(x) \text{ subject to } x\in S.$$

Of course, the generality of this statement reveals in rather stark form that the definition of a nonlinear program is too broad for any but the most superficial results.

A final example will illustrate an important distinction which must be kept in mind when a nonlinear program is studied. Example (4) shows that, for any set S, we can present the problem: "Maximize $f(x)$ subject to $x\in S$," as a nonlinear program in at least one way. The set S is called the set of <u>feasible solutions</u> for the problem and will be the same however the problem is presented. However the same problem may have several presentations and some may be better behaved than others.

(5) Let S be the triangle in the (x_1,x_2) plane with vertices $(0,1/2)$, $(1,0)$, and $(0,1)$. Consider the problem: Maximize $f(x_1,x_2)$ subject to $x \in S$. This has two simple algebraic presentations that follow:

(a) Maximize $f(x_1,x_2)$ subject to

$$x_1+x_2-1 = -y_1, \quad -x_1-2x_2+1 = -y_2$$

$$x_1 \geqq 0, \quad x_2 \geqq 0, \quad y_1 \geqq 0, \quad y_2 \geqq 0.$$

(b) Maximize $f(x_1,x_2)$ subject to

$$(x_1+x_2-1)(x_1+2x_2-1) = -y_1$$

$$x_1 \geqq 0, \quad x_2 \geqq 0, \quad y_1 \geqq 0.$$

Note that if f is linear then (a) is a linear program in canonical form, and so is as well-behaved as one could desire.

3. DUALITY IN LINEAR PROGRAMMING AND BEFORE

To motivate the derivation of the necessary conditions for optimality to be given in the next section, let us place ourselves in the position of mathematical programmers in the late 40's. Von Neumann had given a formulation of the dual for a linear program [4] and Gale, Kuhn, and Tucker had provided rigorous duality theorems and generalizations [5]. These are easily stated in a compact form using the terminology of the preceding section.

Let us start with a <u>linear program</u>, that is, with f and all g_i linear. As before, this may be written:

Maximize $f(x) = c \cdot x$ for "feasible" solutions to

$$Ax-b = -y.$$

Here, as before, "feasible" is a requirement that each x_j and y_i be nonnegative, zero, or free. This specification induces a notion of "dual feasible" for a related dual minimum problem on the same data. This problem reads:

Minimize $h(v) = v \cdot b$ for "dual feasible" solutions to

$$vA-c = u.$$

In this dual linear program, each u_j and v_i is required to be nonnegative, zero, or free if the corresponding variable x_j or y_i has been required to be nonnegative, free, or zero, respectively, in the original (or primal) linear program.

The pair of programs can be displayed conveniently by a diagram due to A. W. Tucker.

	x	-1	
v	A	b	=-y
-1	c	0	= f(max).
	=u	=h(min)	

The feasibility requirements are that paired variables (at the ends of the same row or column) are <u>either</u> both nonnegative <u>or</u> one is zero and the other is free.

With this diagram available, it is obvious that for all solutions, feasible or not,

$$h-f = u \cdot x + v \cdot y$$

while the definition of feasibility for the dual pair implies that

$$h-f \geqq 0$$

for all feasible solutions. Hence, trivially, $h-f = 0$ is a sufficient condition for the optimality of a pair of feasible solutions. Necessary conditions are contained in the following theorem:

Theorem 3.1: If $(\bar{x},\bar{y})$ is an optimal feasible solution for the primal program then there exists a feasible solution $(\bar{u},\bar{v})$ for the dual program with $\bar{u} \cdot \bar{x} + \bar{v} \cdot \bar{y} = 0$ (and hence an optimal feasible solution for the dual program).

As was said in the introduction, this duality theorem "was discovered and explored with surprise and delight in the early days" of our subject. In retrospect, it should have been obvious to all of us. Similar situations had been recognized much earlier, even in nonlinear programs. The phenomenon had even been raised to the level of a method (that is, a trick that has worked more than once) by Courant and Hilbert [6] in the following passage (slightly amended and with underlining added):

"The Lagrange multiplier method leads to several transformations which are important both theoretically and practically.

By means of these transformations new problems equivalent to a given problem can be so formulated that stationary conditions occur simultaneously in equivalent problems. In this way we are led to transformations of the problems which are important because of their symmetric character. Moreover, for a given maximum problem with maximum M, we shall often be able to find an equivalent minimum problem with the same value M as minimum; this is a useful tool for bounding M from above and below."

It is a scholarly challenge to discover the first occurence of the elements of such duality in the mathematical literature. These elements are:

(a) A pair of optimization problems, one a maximum problem with objective function f and the other a minimum problem with objective function h, based on the same data;

(b) For feasible solutions to the pair of problems, always $h \geqq f$;

(c) Necessary and sufficient conditions for optimality are $h = f$.

Surely one of the first situations in which this pattern was recognized originated in the problem posed by Fermat early in the 17th century: Given three points in the plane, find a fourth point such that the sum of its distances to the three given points is a minimum. Previously, on several occasions ([7], [8], and [9]), I have incorrectly attributed the dual problem to E. Fasbender [10], writing in 1846. Further search has led to earlier sources. In a remarkable journal, not much read today, The Ladies Diary or Woman's Almanack (1755), the following problem is posed by a Mr. Tho. Moss (p. 47): "In the three Sides of an equiangular Field stand three Trees, at the Distances of 10, 12, and 16 Chains from one another: To find the Content of the Field, it being the greatest the Data will admit of?" While there seems to have been no explicit recognition of the connection with Fermat's Problem in the Ladies Diary, the observation was not long in coming. In the Annales de Mathématiques Pures et Appliquées, edited by J. D. Gergonne, vol. I (1810-11), we find the following problem posed on p. 384: "Given any triangle, circumscribe the largest possible equilateral triangle about it." In the solutions proposed by Rochat, Vecten, Fauguier, and Pilatte in vol. II (1811-12), pp. 88-93, the observation is made: "Thus the largest equilateral triangle circumscribing a given triangle has sides perpendicular to the lines joining the vertices of the given triangle to the point such that the sum of the distances to these vertices is a minimum. (p. 91). One can conclude that the altitude of the largest equilateral triangle that can be circumscribed about a given triangle is equal to the sum of distances from the vertices of the given triangle to the point at which the sum of distances is a minimum. (p. 92)". The credit for recognizing this duality, which has all of the elements listed above, appears to be due to Vecten, professor of mathématiques speciales at the Lycée de Nismes. Until further evidence is discovered, this must stand as the first instance of duality in nonlinear programming!

4. THE KARUSH CONDITIONS

The generalization of Theorem 3.1 will be derived for a nonlinear program in canonical form (compare Example 2 of §2):

$$\text{Maximize } f(x) \text{ for feasible solutions of}$$
$$g(x)-b = -y$$

where feasible means all x_j and y_i are nonnegative. (Here we have used $g(x)$ as a natural notation for the column vector of values $(g_1(x),\ldots,g_m(x))$.) We seek necessary conditions that must be satisfied by a feasible solution $(\bar{x},\bar{y})$ to be locally optimal. Therefore, it is natural to linearize by differentiating to yield a linear program:

$$\text{Maximize } df = f'(\bar{x})dx \text{ for feasible solutions of}$$
$$g'(\bar{x})dx = -dy.$$

(Here, we have further restricted the nonlinear program to have differentiable f and g_i. Furthermore, we have used $f'(\bar{x})$ and $g'(\bar{x})$ as the customary notations for the gradient of f and the Jacobian of g, respectively, evaluated at $\bar{x}$.)

Some care must be taken with the specification of feasibility in this linear program. Intuitively, we are testing directions of change (dx,dy) from a feasible solution $(\bar{x},\bar{y})$ and we want the resulting position $(\bar{x}+dx,\bar{y}+dy)$ to be feasible (or feasible in some limiting sense). This leads naturally to the following specification of feasibility for the linearized problem:

The variable dx_j (dy_i) is nonnegative if $\bar{x}_j = 0$ $(\bar{y}_i = 0)$; otherwise dx_j and dy_i are free.

The fact that the linearized problem is a linear program can be presented as the following diagram (which includes the variables for the dual linear program):

	dx	-1	
v	$g'(\bar{x})$	0	$=-dy$
-1	$f'(\bar{x})$	0	$= df(\max)$
	$=u$	$=0(\min)$	

The specification of feasible (dx,dy) given above induces the following specification of feasible (u,v):

The variable u_j (v_i) is nonnegative if $\bar{x}_j = 0$ $(\bar{y}_i = 0)$; otherwise u_j and v_i are zero.

Noting the fact that $(\bar{x},\bar{y})$ is feasible and hence nonnegative, the specification of feasible (u,v) can be rephrased as nonnegativity and orthogonality to $(\bar{x},\bar{y})$:

The variables (u,v) are feasible if and only if they are nonnegative and $u\cdot\bar{x}+v\cdot\bar{y} = 0$.

Theorem 4.1: Suppose $df \leqq 0$ for all feasible (dx,dy) for the linearized nonlinear program in canonical form at a feasible $(\bar{x},\bar{y})$. Then there exist $(\bar{u},\bar{v}) \geqq 0$ such that

$$\bar{v}g'(\bar{x})-f'(\bar{x}) = \bar{u}$$

$$\bar{u}\cdot\bar{x}+\bar{v}\cdot\bar{y} = 0.$$

Proof: With the hypothesis of the theorem, the primal linear program has the optimal solution $(dx,dy) = (0,0)$. Hence, by Theorem 3.1, there exists a feasible solution $(\bar{u},\bar{v})$ for the dual program. The conditions of the theorem combine the linear equations from the diagram and the characterization of feasibility given above. ∎

To complete the derivation of the necessary conditions, we need to introduce assumptions that insure that the linearized problem correctly represents the possibilities for variation near $(\bar{x},\bar{y})$. Since the work of Kuhn and Tucker, these assumptions have been called <u>constraint</u> <u>qualifications</u>.

Definition 4.1: A nonlinear program satisfies the <u>constraint qualification</u> (CQ) at a feasible solution $(\bar{x},\bar{y})$ if for every feasible (dx,dy) for the linearized problem there exists a sequence (x^k,y^k) of feasible solutions and a sequence λ_k of nonnegative numbers such that

$$\lim_{k\to\infty} x^k = \bar{x} \quad \text{and} \quad \lim_{k\to\infty} \lambda_k(x^k-\bar{x}) = dx.$$

Theorem 4.2: Suppose a nonlinear program satisfies the CQ at a feasible solution $(\bar{x},\bar{y})$ at which f achieves a local maximum. Then $df \leqq 0$ for all feasible solutions (dx,dy) for the linearized problem.

Proof: By the differentiability of f,

$$f(x^k)-f(\bar{x}) = f'(\bar{x})(x^k-\bar{x})+\varepsilon_k|x^k-\bar{x}|$$

where $\lim_{k\to\infty} \varepsilon_k = 0$. Since $(\bar{x},\bar{y})$ is a local maximum,

$$0 \geqq f'(\bar{x})\lambda_k(x^k-\bar{x})+\varepsilon_k\lambda_k|x^k-\bar{x}|$$

for k large enough. Taking limits

$$0 \geqq f'(\bar{x})dx+(\lim_{k\to\infty} \varepsilon_k)|dx| = df. \qquad \square$$

These two theorems are combined to yield the necessary conditions that are sought.

Theorem 4.3: Suppose a nonlinear program in canonical form satisfies the CQ at a feasible solution $(\bar{x},\bar{y})$ at which f achieves a local maximum. Then there exist $(\bar{u},\bar{v}) \geqq 0$ such that

$$\bar{v}g'(\bar{x})-f'(\bar{x}) = \bar{u}$$

$$\bar{u}\cdot\bar{x}+\bar{v}\cdot\bar{y} = 0.$$

The result just stated is customarily called the Kuhn-Tucker conditions. The following quotation from Takayama [11] gives a more accurate account of the history of these conditions:

"Linear programming aroused interest in constraints in the form of inequalities and in the theory of linear inequalities and convex sets. The Kuhn-Tucker study appeared in the middle of this interest with a full recognition of such developments. However, the theory of nonlinear programming when the constraints are all in the form of equalities has been known for a long time -- in fact, since Euler and Lagrange. The inequality constraints were treated in a fairly satisfactory manner already in 1939 by Karush. Karush's work is apparently under the influence of a similar work in the calculus of variations by Valentine. Unfortunately, Karush's work has been largely ignored."

Although known to a number of people, especially mathematicians with a connection with the Chicago school of the calculus of variations, it is certainly true that Karush's work has been ignored. A diligent search of the literature has brought forth citations in [12], [13], [14], and [15] to add to Takayama's book referenced above. Of course, one reason is that Karush's work has not been published; to allow the reader to see for himself that Karush was indeed the first to prove Theorem 4.3, the Appendix to this paper provides excerpts from the original work. Precisely, THEOREM 3:2 is equivalent to Theorem 4.3.

Karush's work was done as a master's thesis at the University of Chicago under L. M. Graves, who also proposed the problem. It was written in the final years of the very influential school of classical calculus of variations that had florished at Chicago. One may suppose that the problem was set as a finite-dimensional version of research then proceeding on the calculus of variations with inequality side conditions [16]. G. A. Bliss was chairman of the department and M. R. Hestenes was a young member of the faculty; both of these men influenced Karush. (It is amusing to note that this group also anticipated the work in optimal control theory, popularized under the name of the "Pontryagin" maximum principle. For details, see [17].) As a struggling graduate student meeting requirements for going on to his Ph.D., the thought of publication never occurred to Karush. Also, at that time, no one anticipated the future interest in these problems and their potential practical application. We shall return to this question in the last section of this paper.

The constraint qualification employed by Karush is identical to that used by Kuhn and Tucker and hence is slightly less general than Definition 4.1. Precisely, he required that there exist arcs of feasible solutions issuing from $(\bar{x},\bar{y})$ tangent to every (dx,dy). The need for some such regularity condition was familiar from the equality constrained case. As the proof of Theorem 4.3 given above shows, the inequality constrained case requires the equality of a cone generated by directions that are feasible from $(\bar{x},\bar{y})$ and the cone of feasible directions (dx,dy) from $(\bar{x},\bar{y})$. Since the latter cone depends on the nature of $g(x)$, two problems with the same objective function and the same feasible set but specified in two different ways may behave differently. Example 5 at the end of §2 illustrates this phenomenon in a striking way. If $f(x_1,x_2) = x_1$ then the problem as formulated in (a) is a linear program with the unique optimal solution $\bar{x}_1 = 1$, $\bar{x}_2 = 0$, $\bar{y}_1 = 0$. However, it is easily verified that, as formulated in (b), the "same" problem does not satisfy the constraint qualification at this optimal solution and the conditions of Theorem 4.3 cannot be satisfied.

A full discussion of constraint qualifications and their historical antecedents would take us too far afield. However it is appropriate to cite at this point another early and important but unpublished contribution to this area. This is the work of Morton Slater [18], issued as a Cowles Commission Discussion Paper in November, 1950, and often referenced since then. Slater's main result is an elegant regularity condition that implies saddlepoint necessary conditions for nonlinear programs without differentiability of f and g. We shall return to this in the next section.

5. THE KUHN-TUCKER PAPER

The background of the work of Karush was so different from that of Kuhn and Tucker that one must marvel that the same theorem resulted. From the mid 30's, Tucker had sustained an interest in the duality between covariant and contravariant that arises in the tensor calculus and in the duality between homology and cohomology that arises in combinatorial topology. He was also aware of the pre-topology appearance of such phenomena in the development of the theory of electrical networks. However, this intellectual awareness might have lain fallow except for a happy historical accident. In the May of 1948, G. B. Dantzig visited John von Neumann in Princeton to discuss potential connections between the then very new subject of linear programming and the theory of games. Tucker happened to give Dantzig a lift to the train station for his return trip to Washington. On the way, Dantzig gave a five minute exposition of what linear programming was, using the Transportation Problem as a simple illustrative example.

This sounded like Kirkhoff's Laws to Tucker and he made this observation during the short ride, but thought little about it until later. Dantzig's visit to Princeton resulted in the initiation of a research project which had as its original object the study of the relations between linear programs and matrix games. (Staffed in the summer of 1948 by David Gale and Kuhn, graduate students at Princeton, with Tucker as principal investigator, this project continued in various forms under the generous sponsorship of the Office of Naval Research until 1972.) Stimulated by a note circulated privately by von Neuman [4], the duality theorem for linear programming (Theorem 3.1 above) was proved [5] and various connections were established between the solutions of matrix games and linear programs. As an example, in the summer of 1949, Kuhn produced a one-page working note expressing the duality of linear programming as a saddlepoint property of the Lagrangian expression:

$$L(x,v) = c \cdot x + v(b - Ax)$$

defined for $x \geqq 0$, $v \geqq 0$. Thus formulated, the optimization problems involved (maximize in x and minimize in v) yielded familiar necessary conditions with only minor modifications to take account of the boundaries at 0. Of course, this expression generalizes naturally to

$$L(x,v) = f(x) - v \cdot g(x)$$

in the nonlinear case and this saddlepoint problem was later chosen as the starting point for the exposition of the Kuhn-Tucker analysis.

On leave at Stanford in the fall of 1949, Tucker had a chance to return to question: What was the relation between linear programming and the Kirkhoff-Maxwell treatment of electrical networks? It was at this point that he recognized the parallel between Maxwell's potentials and Lagrange multipliers and identified the underlying optimization problem of minimizing heat loss (see [19]). Tucker then wrote Gale and Kuhn, inviting them to do a sequel to [5] generalizing the duality of linear programs to quadratic programs. Gale declined, Kuhn accepted and the paper developed by correspondence between Stanford and Princeton. As it was written, the emphasis shifted from the quadratic case to the general nonlinear case and to properties of convexity that imply that the necessary conditions for an optimum are also sufficient. In the final version, the quadratic programming case that figured so prominently in Tucker's research appears beside the duality of linear programming as an instance of the application of the general theory. A preliminary version (without the constraint qualification) was presented by Tucker at a seminar at the RAND Corporation in May 1950. A counterexample provided by C. B. Tompkins led to a hasty revision to correct this oversight. Finally, this work might have appeared in the published literature at a much later date were it not for a fortuitous invitation from J. Neymann to present an invited paper at the Second Berkeley Symposium on Probability and Statistics in the summer of 1950.

The paper [1] formulates necessary and sufficient conditions for a saddle-point of any differentiable function $\varphi(x,v)$ with nonnegative arguments, that is, for a pair $(\bar{x},\bar{v}) \geqq 0$ such that

$$\varphi(x,\bar{v}) \leqq \varphi(\bar{x},\bar{v}) \leqq \varphi(\bar{x},v) \quad \text{for all} \quad x \geqq 0,\ v \geqq 0.$$

It then applies them, through the Lagrangian $L(x,v) = f(x)-v\cdot g(x)$ introduced above, to the canonical nonlinear program treated in §4 of this paper. The equivalence between the problems, subject to the constraint qualification, is shown to hold when f and all g_i are concave functions. It is noted, but not proved in the paper, that the equivalence still holds when the assumption of differentiability is dropped. Of course, for this to be true, the constraint qualification must be changed since both Karush's qualification and Definition 4.1 use derivatives. As noted above, Slater's regularity condition [18] is an elegant way of doing this. It merely requires the existence of an $\hat{x} \geqq 0$ such that $g(\hat{x}) < 0$, and makes possible a complete statement without differentiability. Of course, for most applications, the conditions of the differentiable case (Theorem 4.3) are used.

6. THE JOHN CONDITIONS

To establish the relation of the paper of F. John [3] to the work discussed earlier, we shall paraphrase Takayama again [11]:

"Next to Karush, but still prior to Kuhn and Tucker, Fritz John considered the nonlinear programming problem with inequality constraints. He assumed no qualification except that all functions are continuously differentiable. Here the Lagrangian expression looks like $v_0 f(x) - v \cdot g(x)$ instead of $f(x) - v \cdot g(x)$ and v_0 can be zero in the first order conditions. The Karush-Kuhn-Tucker constraint qualification amounts to providing a condition which guarantees $v_0 > 0$ (that is, a normality condition)."

This expresses the situation quite accurately for our purposes, except to record that Karush also considered nonlinear programs without a constraint qualification and proved the same first-order conditions. Karush's proof is a direct application of a result of Bliss [20] for the equality constrained case, combined with a trick used earlier by Valentine [16] to convert inequalities into equations by introducing squared slack variables. For the equality constrained case, the result also appears in Carathéodory [21] as Theorem 2, p. 177.

Questions of precedence aside, what led Fritz John to consider this problem? Marvelously, his motives were quite different from those we have met previously. The main impulse came from trying to prove the theorem (which forms the main application in [3]) that asserts that the boundary of a compact convex set S in R^n lies between two homothetic ellipsoids of ration $\leqq n$, and that the outer ellipsoid can be taken to be the ellipsoid of least volume containing S. The case $n = 2$ had been settled by F. Behrend [22] with whom John had become acquainted in 1934 in Cambridge, England. A student of John's, O. B. Ader, dealt with the case $n = 3$ in 1938 [23]. By that time, John had become deeply interested in convex sets and in the inequalities connected with them. Stimulation came also from the work of Dines and Stokes, in which the duality that pervades systems of linear equations and inequalities appears prominently. Ader's proof strongly suggested that duality was the proper tool for this geometrical problem in the n-dimensional case, and John was able to use these ideas to write up the problem for general n. The resulting paper was rejected by the Duke Mathematics Journal and so very nearly joined the ranks of unpublished classics in our subject. However, this rejection only gave more time to explore the implications of the technique used to derive necessary conditions for the minimum of a quantity (here the volume of an ellipsoid) subject to inequalities as side conditions.

It is poetic justice that Fritz John was aided in solving this problem by a heuristic principle often stressed by Richard Courant that in a variational problem where an inequality is a constraint, a solution always behaves as if the inequality were absent, or satisfies strict equality. It was the occasion of Courant's 60th birthday in 1948 that gave John the opportunity to complete and publish the paper [3].

In summary, it was not the calculus of variations, programming, optimization, or control theory that motiviated Fritz John but rather the direct desire to find a method that would help to prove inequalities as they occur in geometry. In the next section, we shall treat such a problem, used by John as an illustrative example, from our present point of view.

7. SYLVESTER'S PROBLEM

In 1857, J. J. Sylvester published a one sentence note [24]: "It is required to find the least circle which shall contain a given set of points in the plane." The generalization to an arbitrary bounded set in R^m was used by John in [3] as an illustration of the application of his necessary conditions. Our purposes in this section are similar to his; we have the advantage of the cumulative research in nonlinear programming over the last quarter century. Although the problem has an extensive literature (see [25] for some of its history), it is only recently that it has been recognized as a quadratic program by Elzinga and Hearn [26]. More precisely, Sylvester's problem can be formulated as a hybrid program (that is, a linear program with a sum of squares added to the objective function [27]). As such, it has a natural dual which is also a hybrid program. This fact can be discovered very naturally by constructing a dual using the theory of conjugate functions [28], then recognizing that the dual is a hybrid program. Therefore, Sylvester's Problem must be a hybrid program in disguise. The treatment given below reverses this process in traditional mathematical style.

Let $a_1,\ldots,a_m$ be n given points in R^m. Then Sylvester's Problem in m-space asks for $x \in R^m$ minimizing $\max_j |x-a_j|$, where $|x-a_j|$ denotes the Euclidean distance from x to a_j. This is clearly equivalent to:

(1) $$\text{Minimize } \max_j (x-a_j)^2/2 \text{ for } x \in R^m,$$

where $(x-a_j)^2 = |x-a_j|^2$ and the factor of $1/2$ has been inserted for simplification later. Problem (1) is equivalent, in turn, to:

(2) $$\text{Minimize } v+x^2/2 \text{ subject to}$$
$$v+x^2/2 \geqq (x-a_j)^2/2$$
$$\text{for } v \in R,\ x \in R^m, \text{ and } j = 1,\ldots,n.$$

We may rewrite (2), introducing slack variables y_j and explicit coordinates for a_j and x, as:

(3) $$\text{Minimize } v+\Sigma_i x_i^2/2 \text{ subject to}$$
$$y_j = v+\Sigma_i x_i a_{ij} - \Sigma_i a_{ij}^2/2 \geqq 0$$
$$\text{for } v \in R,\ (x_i) \in R^m, \text{ and } j = 1,\ldots,n.$$

Thus Sylvester's problem is equivalent to a hybrid program in the sense of Parsons and Tucker [27]. This program is displayed with its dual in the following schema:

	λ_1	$\cdots$	λ_n	-1	
v	1	$\cdots$	1	1	$= 0$
x_1	a_{11}	$\cdots$	a_{1n}	0	$= z_1$
$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot$		$\cdot$	$\cdot$	$\cdot$
x_m	a_{m1}	$\cdots$	a_{mn}	0	$= z_m$
-1	$a_1^2/2$	$\cdots$	$a_n^2/2$	0	$= f$
	$=y_1$	$\cdots$	$=y_n$	$=h$	

This schema displays two programs (the first of which is exactly (3)):

(4) $$\text{Minimize } H = h+x^2/2 \text{ for}$$
$$h = v,\ \ y_j = v+\Sigma_i x_i a_{ij} - a_j^2/2 \geqq 0, \text{ all } j.$$

(5) $$\text{Maximize } F = f-z^2/2 \text{ for}$$
$$f = \Sigma_j \lambda_j a_j^2/2,\ \ \Sigma_j \lambda_j = 1,\ \ z_i = \Sigma_j a_{ij}\lambda_j,\ \ \lambda_j \geqq 0, \text{ all } i \text{ and } j.$$

Following the results of Parsons and Tucker, these programs are coupled by a duality equation, an identity that is valid for all v, x, λ:

$$H-F = \Sigma_j y_j \lambda_j + (x-z)^2/2. \tag{6}$$

Therefore, $H \geqq F$ for feasible solutions of (4) and (5) and $H = F$ for feasible solutions if, and only if:

$$y_j \lambda_j = 0 \text{ for all } j \text{ and } x = z. \tag{7}$$

Conditions (7) are necessary and sufficient for feasible solutions to be optimal. The sufficiency is obvious; the necessity is a direct application of Theorem 4.3. Since the constraints are linear inequalities the constraint qualification is trivially satisfied. We have proved (dropping the factor of 1/2):

Theorem 7.1:

$$\max_{\lambda}[\Sigma_j \lambda_j a_j^2 - (\Sigma_j a_{ij}\lambda_j)^2] = \min_{x}(\max_j (x-a_j)^2)$$

for $\lambda \geqq 0$, $\Sigma_j \lambda_j = 1$ and $x \in R^m$.

Of course, by expressing optimality for Sylvester's Problem as the solution of conditions (7), we have cast it as a linear complementarity problem. See Eaves' paper in these proceedings. Explicitly, conditions (7) ask for the solution of

$$\left[\begin{array}{cc|ccc} 0 & 0 & 1 & \cdots & 1 \\ 0 & 0 & -1 & \cdots & -1 \\ \hline 1 & -1 & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & A^TA & \\ \cdot & \cdot & & & \\ 1 & -1 & & & \end{array}\right] \begin{bmatrix} v_1 \\ v_2 \\ \lambda_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_n \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \\ -a_1^2/2 \\ \cdot \\ \cdot \\ \cdot \\ -a_n^2/2 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$$

such that

$$v_1 \geqq 0, v_2 \geqq 0, \lambda_1 \geqq 0, \ldots, \lambda_n \geqq 0$$
$$w_1 \geqq 0, w_2 \geqq 0, y_1 \geqq 0, \ldots, y_n \geqq 0$$

and

$$v_1 w_1 + v_2 w_2 + \lambda_1 y_1 + \cdots + \lambda_n y_n = 0.$$

This formulation opens a number of possibilites for computation.

Finally, by algebraic manipulation, the maximum program can be given a slightly different form with the result:

$$\max_{\lambda} \Sigma_j \lambda_j (\Sigma_k \lambda_k a_k - a_j)^2 = \min_{x} \max_{j} (x-a_j)^2 \tag{8}$$

where $\lambda \geqq 0$, $\Sigma_j \lambda_j = 1$, and $x \in R^m$. In this form, both programs conceal their nature as hybrid programs but exhibit a saddlepoint property that could have been discovered by studying the following 0-sum 2-person game: Player 1 chooses $\lambda \geqq 0$, $\Sigma_j \lambda_j = 1$. Player 2 chooses x in the convex hull of $\{a_1, \ldots, a_n\}$. Player 2 pays Player 1 the amount $\Sigma_j \lambda_j (x-a_j)^2$. If this payoff function is denoted by $\psi(\lambda,x)$, then $\max_{\lambda} \min_{x} \psi(\lambda,x) = \min_{x} \max_{\lambda} \psi(\lambda,x)$ since the strategy sets are compact and convex, and the payoff function ψ is concave in λ and convex in x. Since, for given λ, the minimum over x is achieved at $x = \Sigma_k \lambda_k a_k$ and, for given x, the maximum over λ is achieved at a pure strategy chosen by $\max_j (x-a_j)^2$, this saddlepoint statement is exactly (8) again.

Finally, the expression on the left side of (8) admits a physical interpretation. We wish to distribute weights on the points $\{a_1, \ldots, a_n\}$ so that the second moment about the center of gravity of those weights is a maximum. This moment can be interpreted as the moment of inertia about an axis perpendicular to the space in which the points lie. The duality relation then says that the radius of the minimal circle enclosing the points is the maximum radius of gyration of the system, the maximum being taken over all possible distributions of the unit mass among the points $a_1, \ldots, a_n$. Here the radius of gyration is as discussed in Goldstein [29]. It would be interesting to know if this duality has been studied in the literature of mechanics or of geometrical optimization.

There are a number of other observations that could be made about this ancient problem. However, it should be clear by now that we can probe the mysteries, both theoretical and computational, of such classical optimization problems more efficiently today than we could 25 years ago.

8. A SERMON

This sermon will be short. We have seen that the same result, which is central to the subject of nonlinear programming, was found independently by mathematicians who found their inspiration in the calculus of variations, geometrical inequalities, the theory of games, duality in topology, network theory, and linear programming. This result which has proved to be useful, at least in the sense of suggesting computational algorithms, was sought and found first with no thought given to its application to practical situations. It was rediscovered and recognized as important only in the midst of the development of the applied field of mathematical programming. This, in turn, had a beneficial effect. With the impetus of evident applicability, the mathematical structure of the subjects neighboring mathematical programming has deepened in the last quarter century. A scattering of isolated results on linear inequalities has been replaced by a respectable area of pure mathematics to which this symposium bears witness. Notable achievements have been recorded in the subjects of convex analysis, the analysis of nonlinear systems, and algorithms to solve optimization problems. This has been possible only because communication has been opened between mathematicians and the potential areas of application, to the benefit of both. The historical record is clear and I believe that the moral is equally clear: the lines of communication between applied fields such as mathematical programming and the practitioners of classical branches of mathematics should be broadened and not narrowed by specialization. This symposium is a constructive step in this direction.

APPENDIX

The purpose of this appendix is to place in print the precise results obtained by Karush in his pioneering work [2]. With the exception of some preliminary results on linear inequalities, complete statements are given for all of the theorems and corollaries. No proofs are included since these are now readily available in the literature or are easy to reconstruct. The notation of the original has been conserved; in particular, the convention that a repeated subscript indicates summation is followed. The titles of the sections, the statements of results, and the references are unchanged; some connecting text has been freely rendered.

1. Introduction

This paper treats the problem of determining necessary and sufficient conditions for a relative minimum of a function $f(x)$ subject to $g_\alpha(x) \geqq 0$ for $x = (x_1,\ldots,x_n)$ and $\alpha = 1,\ldots,m$, where the functions f and g_α are required to have continuous derivatives of order one or two. By a well-known argument, we may restrict our attention to minimizing points x^o where $g_\alpha(x^o) = 0$ for all α. Differentiability assumptions on f and g_α near x^o are as follows: Class C' for Theorem 1.1, Sections 3 and 4; Class C" for the other theorems of Section 1, Sections 5 and 6.

The results of Bliss [20] for minimizing $f(x)$ subject to equations $h_\alpha(x) = 0$ for $\alpha = 1,\ldots,m$ are used in the proofs. They are listed here for comparison with the results of this paper.

THEOREM 1:1. A first necessary condition for $f(x^o)$ to be a minimum is that there exist constants ℓ_o, ℓ_α not all zero such that the derivatives H_{x_i} of the function

$$H = \ell_o f + \ell_\alpha h_\alpha$$

all vanish at x^o.

LEMMA 1:1. If $\|h_{\alpha x_i}(x^o)\|$ has rank m, then for every set of constants η_i $(i = 1,2,\ldots,n)$ satisfying the equations

$$h_{\alpha x_i}(x^o)\eta_i = 0$$

there exists a curve $x_i(t)$ having continuous second derivatives near $t = 0$, satisfying the equations $h_\alpha[x(t)] = 0$, and such that

$$x_i(0) = x_i^{\,o}, \qquad x_i'(0) = \eta_i.$$

THEOREM 1:2. If $\|h_{\alpha x_i}(x^o)\|$ has rank m and $f(x^o)$ is a minimum then the condition

$$H_{x_i x_k}(x^o)\eta_i\eta_k \geqq 0$$

must hold for every set η_i satisfying $h_{\alpha x_i}(x^o)\eta_i = 0$, where $H = f + \ell_\alpha h_\alpha$ is the function formed with the unique set of multipliers $\ell_o = 1$, ℓ_α belonging to x^o.

Our final excerpt from Bliss's paper is a sufficiency theorem.

THEOREM 1:3. If a point x^o has a set of multipliers $\ell_o = 1$, ℓ_α for which the function $\Pi = f + \ell_\alpha h_\alpha$ satisfies the conditions

$$H_{x_i}(x^o) = 0, \qquad H_{x_i x_k}(x^o)\eta_i\eta_k > 0$$

for all sets η_i satisfying the equations

$$h_{\alpha x_i}(x^o)\eta_i = 0,$$

then $f(x^o)$ is a minimum.

2. Preliminary theorems on linear inequalities

This section contains properties of systems of linear inequalities that are used for the proofs in the later sections. These include Farkas' Lemma [30], various results from Dines and McCoy [31], and Dines [32].

3. Necessary conditions involving only first derivatives

We make some preliminary definitions. A solution $\lambda = (\lambda_1, \lambda_o, \ldots, \lambda_n)$ of

$$g_{\alpha x_i}(x^o)\lambda_i \geqq 0 \qquad (\alpha = 1, 2, \ldots m),$$

will be called an admissible direction if λ is not the zero vector. A regular arc $x_i(t)$ $(i = 1, 2, \ldots, n;\ 0 \leqq t \leqq t_o)$, will be called admissible in case $g_\alpha[x(t)] \geqq 0$ for every α and t. A point x^o is a normal point in case the matrix

$$\| g_{\alpha x_i}(x^o) \|$$

has rank m.

THEOREM 3:1. If $f(x^o)$ is a minimum then there exist multipliers ℓ_o, ℓ_α not all zero such that the derivatives F_{x_i} of the function

$$F(x) = \ell_o f(x) + \ell_\alpha g_\alpha(x)$$

all vanish at x^o.

THEOREM 3:2. Suppose that for each admissible direction λ there is an admissible arc issuing from x^o in the direction λ. Then a first necessary condition for $f(x^o)$ to be a minimum is that there exist multipliers $\ell_\alpha \leqq 0$ such that the derivatives F_{x_i} of the function

$$F = f + \ell_\alpha g_\alpha$$

all vanish at x^o.

The condition that there exist multipliers $\ell_\alpha \leqq 0$ satisfying the conclusion of Theorem 3:2 will be referred to as "the first necessary condition". For brevity, the property that for each admissible direction λ there is an admissible arc issuing from x^o in the direction λ will be called property Q.

COROLLARY. Suppose that for every admissible direction λ it is true that $g_{\alpha x_i}(x^o)\lambda_i = 0$ implies that $g_{\alpha x_i x_k}(x^o)\lambda_i\lambda_k > 0$. Then if $f(x^o) = \text{minimum}$ the first necessary condition is satisfied.

THEOREM 3:3. Suppose there exists an admissible direction $\bar{\lambda}$ for which $g_{\alpha x_i}(x^o)\bar{\lambda}_i > 0$ for every α. Then if $f(x^o) = \text{minimum}$ the first necessary condition is satisfied.

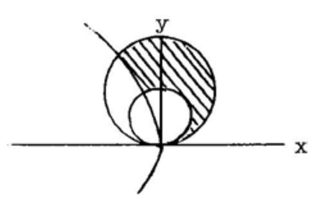
COROLLARY. Suppose $m = n$ and determinant $\|g_{\alpha x_i}(x^o)\| \neq 0$. Then a necessary condition for $f(x^o)$ to be a minimum is that

$$f_{x_i}(x^o)G_{i_\alpha} \geqq 0 \qquad (\alpha = 1,2,\ldots,n),$$

where $\|G_{i\alpha}\|$ is the inverse matrix of $\|g_{\alpha x_i}(x^o)\|$.

It is easy to give an example in which the functions g_α satisfy neither the hypothesis of the corollary to Theorem 3:2 nor the hypothesis of Theorem 3:3, but in which the hypothesis of Theorem 3:2 is satisfied. Let

$$g_1(x,y) = x^2 + (y-1)^2 - 1 \geqq 0$$
$$g_2(x,y) = 4 - [x^2+(y-2)^2] \geqq 0$$
$$g_3(x,y) = y^2 + x \geqq 0$$



determine the class of points (x,y) under consideration. At $(0,0)$ we have

$$\begin{Vmatrix} g_{1x} & g_{1y} \\ g_{2x} & g_{2y} \\ g_{3x} & g_{3y} \end{Vmatrix} = \begin{Vmatrix} 0 & -2 \\ 0 & 4 \\ 1 & 0 \end{Vmatrix}.$$

The only admissible direction is $(a,0)$ with $a > 0$. There is no solution of $g_{\alpha x}(0,0)\bar{\lambda}_1 + g_{\alpha y}(0,0)\,\bar{\lambda}_2 > 0$ for all α. Also $g_{2xx}(0,0)a^2 < 0$ so that the hypothesis of the corollary to Theorem 3:2 is not satisfied. However, it is obvious that there is an admissible arc issuing from $(0,0)$ in the direction $(a,0)$.

4. Sufficient conditions involving only first derivatives

THEOREM 4:1. Suppose $m \geqq n$ and $\|g_{\alpha x_i}(x^o)\|$ has maximum rank n. If x^o is a point satisfying $g_\alpha(x^o) = 0$ for which there exist multipliers $\ell_\alpha < 0$ such that $F = f + \ell_\alpha g_\alpha$ has $F_{x_i}(x^o) = 0$, then $f(x^o)$ is a minimum.

COROLLARY. Suppose $m = n$ and determinant $\|g_{\alpha x_i}(x^o)\| \neq 0$. We let $\|G_{i\alpha}\|$ be the inverse matrix of $\|g_{\alpha x_i}\|$. If x^o is a point satisfying $g_\alpha(x^o) = 0$ such that

$$f_{x_i}(x^o)G_{i\alpha} > 0 \qquad (\alpha = 1,2,\ldots,n),$$

then $f(x^o)$ is a minimum.

THEOREM 4:2. Suppose $m \geqq n$ and $\|g_{\alpha x_i}(x^o)\|$ has rank n. If x^o is a point satisfying $g_\alpha(x^o) = 0$ such that $f_{x_i}(x^o)\lambda_i > 0$ for every admissible direction λ, then $f(x^o)$ is a minimum.

5. A necessary condition involving second derivatives

THEOREM 5:1. Suppose $f(x^o)$ is a minimum and there exist multipliers ℓ_α such that $F = f + \ell_\alpha g_\alpha$ has $F_{x_i}(x^o) = 0$. Suppose, further, that $\|g_{\alpha x_i}(x^o)\|$ has rank $r < n$ with the first r rows linearly independent. Then for every admissible direction η satisfying $g_{\alpha x_i}(x^o)\eta_i = 0$ $(\alpha = 1,2,\ldots,m)$, such that there is an admissible arc $x(t)$ of class C'' issuing from x^o in the direction η and satisfying $g_\alpha[x(t)] = 0$ for $\alpha = 1,2,\ldots,r$, it is true that

$$F_{x_i x_k}(x^o)\eta_i\eta_k \geqq 0,$$

where F is formed with the unique set of multipliers ℓ_α belonging to the first r rows of $\|g_{\alpha x_i}(x^o)\|$.

COROLLARY. Suppose x^o is a normal point. Then necessary conditions for $f(x^o)$ to be a minimum arc that the first necessary condition be satisfied and that

$$F_{x_i x_k}(x^o)\eta_i\eta_k \geqq 0$$

be satisfied for every admissible direction η satisfying

$$g_{\alpha x_i}(x^o)\eta_i = 0 \qquad (\alpha = 1,2,\ldots,m).$$

6. A sufficiency theorem involving second derivatives

THEOREM 6:1. If a point x^o satisfying $g_\alpha(x^o) = 0$ has a set of multipliers $\ell_\alpha < 0$ for which the function $F = f + \ell_\alpha g_\alpha$ satisfies

$$F_{x_i}(x^0) = 0, \quad F_{x_i x_k}(x^0)\eta_i\eta_k > 0$$

for all admissible directions η satisfying

$$g_{\alpha x_i}(x^0)\eta_i = 0,$$

then $f(x^0)$ is a minimum.

REFERENCES

1. H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in Jerzy Neyman (ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, U. of Calif. Press, 1950), 481-492.

2. William Karush, "Minima of Functions of Several Variables with Inequalities as Side Conditions," Master's Thesis, Department of Mathematics, University of Chicago, December, 1939, 26 pp.

3. Fritz John, "Extremum Problems with Inequalities as Subsidiary Conditions," Studies and Essays, Courant Anniversary Volume (New York, Interscience, 1948), 187-204.

4. John von Neumann, "Discussion of a Maximum Problem," in A. H. Taub (ed.), John von Neumann, Collected Works, Vol. VI (New York, Pergamon, 1963), 89-95.

5. David Gale, H. W. Kuhn, and A. W. Tucker, "Linear Programming and the Theory of Games," in T. C. Koopmans (ed.), Activity Analysis of Production and Allocation (New York, Wiley, 1951).

6. R. Courant and D. Hilbert, Methods of Mathematical Physics, I (New York, Interscience, 1953), p. 231.

7. H. W. Kuhn, "On a Pair of Dual Nonlinear Programs," in J. Abadie (ed.), Methods of Nonlinear Programming (Amsterdam, North-Holland, 1967), 38-54.

8. H. W. Kuhn, "A Note on Fermat's Problem," Mathematical Programming 4 (1973), 98-107.

9. H. W. Kuhn, "'Steiner's' Problem Revisited," in G. B. Dantzig and B. C. Eaves (eds.), Studies in Optimization (Math. Assoc. Amer., 1974), 52-70.

10. E. Fasbender, "Über die gleichseitige Dreiecke, welche um ein gegebenes Dreiecke gelegt werden können," J. f. Math. 30 (1846), 230-1.

11. A. Takayama, Mathematical Economics (Hinsdale, Illinois, Dryden Press, 1974).

12. L. L. Pennisi, "A Indirect Sufficiency Proof for the Problem of Lagrange with Differential Inequalities as Added Side Conditions," Trans. Amer. Math. Soc. 74 (1953), 177-198.

13. M. A. El-Hodiri, Constrained Extrema: Introduction to the Differentiable Case with Economic Applications (Berlin, Springer-Verlag, 1971), originally "Constrained Extrema of Functions of a Finite Number of Variables: Review and Generalizations," Krannert Institute Paper No. 141, Purdue University, 1966.

14. M. A. El-Hodiri, "The Karush Characterization of Constrained Extrema of Functions of a Finite Number of Variables," UAR Ministry of Treasury Research Memoranda, Series A, No. 3, July 1967.

15. A. V. Fiacco and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques (New York, Wiley, 1968).

16. F. A. Valentine, "The Problem of Lagrange in the Calculus of Variations with Inequalities as Added Side Conditions," in Contributions to the Calculus of Variations (1933-1937) (Chicago, U. of Chicago Press, 1937).

17. M. R. Hestenes, Calculus of Variations and Optimal Control Theory (New York, Wiley, 1966).

18. M. Slater, "Lagrange Multipliers Revisited," Cowles Commission Discussion Paper No. 403, November, 1950.

19. A. W. Tucker, "Linear and Nonlinear Programming," Operations Research 5 (1957), 244-257.

20. G. A. Bliss, "Normality and Abnormality in the Calculus of Variations," Trans. Amer. Math. Soc. 43 (1938), 365-376.

21. C. Carathéodory, Calculus of Variations and Partial Differential Equations of the First Order, Part II, Calculus of Variations (San Francisco, Holden Day, 1967), translated by R. Dean from the German original, 1935.

22. F. Behrend, "Über die kleinste umbeschriebene und die grösste einbeschriebene Ellipse eines konvexen Bereiches," Math. Ann. 115 (1938), 379-411.

23. O. B. Ader, "An Affine Invariant of Convex Regions," Duke Math. J. 4 (1938), 291-299.

24. J. J. Sylvester, "A Question in the Geometry of Situation," Quarterly J. Pure and Appl. Math. 1 (1857), 79.

25. L. M. Blumenthal and G. E. Wahlin, "On the Spherical Surface of Smallest Radius Enclosing a Bounded Subset of n-Dimensional Euclidian Space," Bull. Amer. Math. Soc. 47 (1941), 771-777.

26. D. J. Elzinga and D. W. Hearn, "The Minimum Covering Sphere Problem," Management Science 19 (1972), 96-104.

27. T. D. Parsons and A. W. Tucker, "Hybrid Programs: Linear and Least Distance," Math. Programming 1 (1971), 153-167.

28. R. T. Rockafellar, Convex Analysis (Princeton, N.J., Princeton U. Press, 1970).

29. H. Goldstein, Classical Mechanics (Cambridge, Mass., Addison-Wesley, 1950).

30. J. I. Farkas, "Theorie der einfachen Ungleichungen," Crelle 124 (1902), 1-27.

31. L. L. Dines and N. H. McCoy, "On Linear Inequalities," Trans. Royal Soc. Canada 27 (1933), 37-70.

32. L. L. Dines, "Systems of Linear Inequalities," Ann. Math. 20 (1920), 191-199.