Nonlinear Programming: A Historical Note

Harold W. Khun

1 Introduction

The paper [1] that first used the name 'nonlinear programming' was written 41 years ago. In the intervening period, I have learned a number of things about the influences, both mathematical and social, that have shaped the modern development of the subject. Some of these 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 41 years ago when this differentiation took place.

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. These statements will allow the comparison of the results of various mathematicians who made early contributions to nonlinear programming. In reconstructing this story, I had the benefit of personal Communications from A. W. Tucker, W. Karush, and F. John, who shared their memories of the relevant events.

In Section 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 Section 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 Section 2 is subjected to a natural linearization which yields a set of likely necessary conditions for a local optimum in Section 4. Of course, these conditions do not hold in full generality without a regularity condition (conventionally called the *constraint qualification).* When it is invoked, the result is a theorem which has often been attributed to Kuhn and Tucker. This section is completed by a description of the background of the 1939 work of W. Karush [2] in which the theorem first appeared.

As will be seen in Section 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 Section 5, an attempt is made to reconstruct the influences on Kuhn and Tucker that led them to their formulation. 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 Section 6. Some conclusions are drawn in Section 7.

2 What is a Nonlinear Program?

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

maximize $f(x_1, \ldots, x_n)$ for 'feasible' solutions to

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

$$\ldots$$

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

for given functions $f, g_1, \ldots, g_m$ and real constants $b_1, \ldots, b_m$. 'Feasible' means that each x_j and y_i is required to be nonnegative, zero, or free. Naturally, it would be unreasonable to require an independent variable x_j to be zero or a dependent variable y_j to be 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:

   maximize $f(x_1, \ldots, x_n)$ subject to

   $$g_1(x_1, \ldots, x_n) = -0,$$

   $$\ldots$$

   $$g_m(x_1, \ldots, x_n) = 0.$$

   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):

   maximize $c \cdot x$ subject to $Ax \leq b, x \geq 0.$

   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:

   maximize $c \cdot x$ subject to $Ax = b, x \geq 0$.

   This is a linear program in *standard* form.

4. Let S be any set in $\mathbb{R}^N$ 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:

   maximize $f(x)$ 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 represent 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 *feasible solutions* 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 $\left(0, \frac{1}{2}\right), (1, 0)$, 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

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

   $$x_1 \geq 0, \ x_2 \geq 0, \ y_1 \geq 0, \ y_2 \geq 0.$$

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

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

   $$x_1 \geq 0, \ x_2 \geq 0, \ y_1 \geq 0.$$

   Note that, if f is linear, then 5a 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 optimally to be given in the next section, let us place ourselves in the position of mathematical programmers in the late 1940's. After G. B. Dantzig visited John von Neumann in Princeton in May, 1948, von Neumann circulated privately a short typewritten note that was first published fifteen years later [4]. This note formulated the dual for a linear program and gave a flawed proof of the equality of optimal objective values based on an invalid inhomogeneous form of Farkas' Lemma. (This error is corrected in the published version [4].) Motivated by this note, Gale, Kuhn, and Tucker provided rigorous duality theorems and generalizations [5]. (Recently, Dantzig has revealed that, at the same time, he constructed his own proof of von Neumann's result; his note on this subject has not been published.) The results of Gale, Kuhn, and Tucker can be stated in a compact form using the terminology of Section 2.

Let us start with a *linear program*, 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 $h\,(x) = v \cdot b$ 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. The feasibility requirements are that paired variables (at the ends of the same row or column) are *either* both nonnegative *or* one is zero and the other is free.

$$\begin{array}{c|c|c|l} & x & -1 & \\ \cline{2-3} v & A & b & = -y \\ \cline{2-3} -1 & c & 0 & = f\,(max) \\ \cline{2-3} & = u & \begin{array}{c} = \\ h\,(min) \end{array} & \end{array}$$

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

$$h - f \geq 0$$

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 italicization 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 occurrence 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 $h \geq f$;

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 seventeenth 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, 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 Lycee 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 in *canonical form* (compare Example 2 of Section 2):

maximize $f(x)$ for 'feasible' solutions to

$$g(x) - b = -y,$$

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:

maximize $df = f'(\bar{x})\,dx$ for 'feasible' solutions to

$$g'(\bar{x})\,dx = -dy.$$

(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):

$$
\begin{array}{c|c|c|l}
 & dx & -1 & \\
\hline
v & g'(\bar{x}) & 0 & = -dy \\
\hline
-1 & f'(\bar{x}) & 0 & = df\,(max) \\
\hline
 & = u & = 0\,(min) &
\end{array}
$$

The specification of feasible (dx, dy) given above induces the following specification of feasible (u, v): the variable u_j (v_i) 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 $df \leq 0$ for all feasible (dx, dy) for the linearized nonlinear program in canonical form at a feasible $(\bar{x}, \bar{y})$. Then there exists $(\bar{u}, \bar{v}) \geq 0$ 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 *constraint qualifications.*

Definition 4.1. *A nonlinear program satisfies the constraint qualification (CQ) at a feasible solution $(\bar{x}, \bar{y})$ if for every feasible (dx, dy) for the linearized problem there exists a sequence $\left(x^k, y^k\right)$ of feasible solutions and a sequence λ^k of nonnegative numbers sucha that*

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

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

Proof. By the differentiability of f,

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

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

$$0 \geq f'(\bar{x})\, \lambda_k \left(x^k - \bar{x}\right) + \epsilon_k \left|x^k - \bar{x}\right|$$

for k large enough. Taking limits,

$$0 \geq f'(\bar{x})\, dx + \left(\lim_{k\to\infty} \epsilon_k\right) |dx| = df.$$

□

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 $(\bar{u}, \bar{v}) \geq 0$ such that*

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

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

It may be appropriate to note that the proof that we have given was constructed to honor A. W. Tucker in the year of his retirement (1973-74) when the author and Tucker taught a course in mathematical programming to undergraduates at Princeton University. Furthermore, although Theorem 4.3 has been stated for programs in canonical form for the sake of simplicity, the same proof holds for general nonlinear programs and asserts the existence of dual feasible $(\bar{u}, \bar{v})$ satisfying the necessary conditions. This version covers all of the possible cases of feasibility specifications for the primal nonlinear program.

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 prior to 1974 brought forth citations in [12, 13, 14, 15] to add to Takayama's book referenced above. Of course, one reason is that Karush's work has not been published.

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 flourished 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 PhD, the thought of publication never occurred to Karush and he was not encouraged to publish by Graves. Also, at that time, no one anticipated the future interest in these problems and their potential practical application.

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 $(\bar{x}, \bar{y})$ 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 Section 2 illustrates this phenomenon in a striking way. If $f(x_1, x_2) = x_1$, then the problem as formulated in 5a 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 5b, the 'same' problem does not satisfy the constraint qualification at this optimal solution and the conditions of Theorem 4.3 cannot be satisfied. (The reader should note that this example, contrary to all textbook examples known to the author, does not utilize a feasible set with a cusp.)

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 1930'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 May, 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 short 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 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 Neumann [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 $x \geq 0$, $v \geq 0$. 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 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. Neyman 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 saddlepoint of any differentiable function $\phi(x,y)$ with nonnegative arguments, that is, for a pair $(\bar{x},\bar{v}) \geq 0$ such that

$$\phi(x,\bar{v}) \leq \phi(\bar{x},\bar{v}) \leq \phi(\bar{x},v) \text{ for all } x \geq 0, v \geq 0.$$

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 Section 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 $\hat{x} \geq 0$ 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 condition. 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 Caratheodory [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 $\mathbb{R}^n$ lies between two homothetic ellipsoids of ratio $\leq n$, 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 Mathematical 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 sixtieth 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 motivated Fritz John, but rather the direct desire to find a method that would help to prove inequalities as they occur in geometry.

7 Conclusion

Apart from its differentiation as a separate subject, the first results in nonlinear programming were the extensions of Lagrange's multiplier rule to the case where the constraints are inequalities instead of equations. Since Lagrange's multiplier rule was well known by the early nineteenth century (see, for example [24]), it is somewhat surprising that the extension to inequality constraints seems not to have been studied until the twentieth century. Once the conditions were needed for the distinct applications to the calculus of variations, geometrical inequalities, and economic problems, it is not surprising that similar conditions were discovered by several mathematicians. With hindsight, the idea behind these conditions is very simple. At a point $\bar{x}$ achieving a minimum of a function f on a set S, the one-sided tangential derivatives must be nonnegative. This implies that the negative of the gradient of f is an outward normal to S at $\bar{x}$. These facts are then given expression as a Lagrange multiplier rule.

Although all three works considered here are concerned exclusively with the finite-dimensional case, it can be argued that this case is a special instance of results for general variational problems known prior to any of this research (see, for example, [25]). Questions of priority aside, these results which have proved to be useful, at least in the sense of suggesting computational algorithms, were sought and found first with no thought given to its application to practical situations. They were 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 half century. A scattering of isolated results on linear inequalities has been replaced by a respectable area of pure mathematics. 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.

References

[1] H. W. Kuhn, A. W. Tucker (1951). Nonlinear programming. J. Ney-man (ed.). *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 481-492.

[2] W. Karush (1939). *Minima of Functions of Several Variables with Inequalities as Side Conditions*, MSc Thesis, Department of Mathematics, University of Chicago, Chicago. [A summary of this paper was published as an appendix to: H. W. Kuhn (1976). Nonlinear programming: a historical view. R. W. Cottle, C. E. Lemke (eds.). *Nonlinear Programming*, SIAM-AMS Proceedings 9,1-26.]

[3] F.John (1948). Extremum problems with inequalities as subsidiary conditions. *Studies and Essays, Presented to R. Courant on his 60th Birthday January 8, 1948*, Interscience, New York, 187-204.

[4] J. von Neumann (1963). Discussion of a maximum problem. A. H. Taub (ed.). *John von Neumann, Collected Works, Vol. VI, Theory of Games, Astrophysics, Hydrodynamics and Meteorology*, Pergamon, New York, 89-95.

[5] D. Gale, H. W. Kuhn, A. W. Tucker (1951). Linear programming and the theory of games. T. C. Koopmans (ed.). *Activity Analysis of Production and Allocation*, Wiley, New York, 317-329.

[6] R. Courant, D. Hilbert (1953). *Methods of Mathematical Physics*, I, Inter-science, New York, 231.

[7] H. W. Kuhn (1967). On a pair of dual nonlinear programs. J. Abadie (ed.). *Nonlinear Programming*, North-Holland, Amsterdam, 37-54.

[8] H. W. Kuhn (1973). A note on Fermat's problem. *Math. Programming 4*, 98-107.

[9] H. W. Kuhn (1974). "Steiner's" problem revisited. G. B. Dantzig, B. C. Eaves (eds.). *Studies in Optimization*, Studies in Mathematics 10, Mathematical Association of America, 52-70.

[10] E. Fasbender (1846). Über die gleichseitigen Dreiecke, welche um ein gegebenes Dreieck gelegt werden köonnen. *J. Reine Angew. Math. 30*, 230-231.

[11] A. Takayama (1974). *Mathematical Economics*, Dryden Press, Hinsdale.

[12] L. L. Pennisi (1953). An indirect sufficiency proof for the problem of Lagrange with differential inequalities as added side conditions. *Trans. Amer. Math.Soc. 74*,177-198.

[13] M. A. El-Hodiri (1971). *Constrained Extrema: Introduction to the Differentiable Case with Economic Applications*, Springer, Berlin. [Originally: M. A. El-Hodiri (1966). *Constrained Extrema of Functions of a Finite Number of Variables: Review and Generalizations*, Krannert Institute Paper 141, Purdue University, Purdue.]

[14] M. A. El-Hodiri (1967). *The Karush Characterization of Constrained Extrema of Functions of a Finite Number of Variables*, Research Memorandum A.3, UAR Ministry of Treasury.

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

[16] F. A. Valentine (1937). The problem of Lagrange with differential inequalities as added side conditions. *Contributions to the Calculus of Variations (1933-1937)*, University of Chicago Press, Chicago.

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

[18] M. Slater (1950). *Lagrange Multipliers Revisited*, Cowles Commission Discussion Paper 403.

[19] A. W. Tucker (1957). Linear and nonlinear programming. *Oper. Res. 5*, 244-257.

[20] G. A. Bliss (1938). Normality and abnormality in the calculus of variations. *Trans. Amer. Math. Soc. 43*, 365-376.

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

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

[23] O. B. Ader (1938). An affine invariant of convex regions. *Duke Math. J. 4*, 291-299.

[24] J.-L. Lagrange (1813). *Théorie des fonctions analytiques* (nouvelle édition), Ve Courcier, Paris [reprinted as: J.-A. Serret (ed.) (1881). *Oeuvres de Lagrange, Tome IX*, Gauthier-Villars, Paris], Seconde partie, Chapitre XI, Section 58, p. 292.

[25] O. Bolza (1914). Über Variationsprobleme mit Ungleichungen als Nebenbedingungen. H. A. Schwarz (ed.). *Mathematische Abhandlungen*, 1-18.