

A Contextualized Historical Analysis of the Kuhn–Tucker Theorem in Nonlinear Programming: The Impact of World War II

Tinne Hoff Kjeldsen

Section for Mathematics, University of Roskilde, 02, P.O. Box 260, DK-4000 Roskilde, Denmark

E-mail: thk@ruc.dk

When Kuhn and Tucker proved the Kuhn–Tucker theorem in 1950 they launched the theory of nonlinear programming. However, in a sense this theorem had been proven already: In 1939 by W. Karush in a master's thesis, which was unpublished; in 1948 by F. John in a paper that was at first rejected by the *Duke Mathematical Journal*; and possibly earlier by Ostrogradsky and Farkas. The questions of whether the Kuhn–Tucker theorem can be seen as a multiple discovery and why the different occurrences of the theorem were so differently received by the mathematical communities are discussed on the basis of a contextualized historical analysis of these works. The significance of the contexts both mathematically and socially for these questions is discussed, including the role played by the military in the shape of Office of Naval Research (ONR) and operations research (OR). © 2000 Academic Press

En démontrant, en 1950, le théorème qui porte aujourd'hui leur nom, Kuhn et Tucker ont donné naissance à la théorie de la programmation non-linéaire. Cependant, en un sens, ce théorème avait été démontré auparavant, d'abord par W. Karush en 1939 dans un mémoire de maîtrise inédit, par la suite par F. John en 1948 dans un article qui avait d'abord été rejeté par le *Duke Mathematical Journal*, et peut-être même plus tôt par Ostrogradsky et aussi par Farkas. Le présent article cherche à élucider deux questions: Peut-on considérer le théorème Kuhn–Tucker comme un exemple de découverte multiple? Et pourquoi le théorème a-t-il été reçu si différemment dans les diverses communautés mathématiques? Notre discussion se base sur une analyse historique contextuelle des différents ouvrages. Nous examinons ici l'importance du contexte, tant du point de vue des mathématiques que du point de vue social, y compris le rôle joué par le secteur militaire dans le cadre de l'Office of Naval Research et de la recherche opérationnelle. © 2000 Academic Press

MSC 1991 subject classification: 01A60; 49-03; 52-03; 90-03; 90C30.

Key Words: nonlinear programming; the Kuhn–Tucker theorem; Lagrange multiplier method; the calculus of variations; the theory of convexity; multiple discovery; operations research; ONR.

1. INTRODUCTION

In the summer of 1950 at the Second Berkeley Symposium on Mathematical Statistics and Probability, held in Berkeley, California, a mathematician from Princeton, Albert W. Tucker, who was generally known as a topologist, gave a talk with the title 'Nonlinear Programming.' It was based on a joint work of Tucker and a young mathematician, Harold W. Kuhn, who had just finished his Ph.D. study at Princeton University. The talks were published in a conference proceedings, and for the first time the name "nonlinear programming"—the title Kuhn and Tucker chose for their paper—appeared in the mathematical literature [Kuhn and Tucker, 1950]. In the paper Kuhn and Tucker introduced a nonlinear programming problem (to be explained below) and proved the main theorem of the theory—the so-called

“Kuhn–Tucker theorem.” This theorem, which gives necessary conditions for the existence of an optimal solution to a nonlinear programming problem, launched the mathematical theory of nonlinear programming.

The result is famous, and not long after its publication people began to talk about it as the Kuhn–Tucker theorem, but apparently Kuhn and Tucker were not the first mathematicians to prove it. In modern textbooks on nonlinear programming there will often be a footnote telling that William Karush proved the theorem in 1939 in his master’s thesis from the University of Chicago, and that Fritz John derived (almost) the same result in a paper published in 1948 in an essay collection for Richard Courant’s 60th birthday. Today one often sees the theorem referred to as the ‘Karush–Kuhn–Tucker theorem’ to acknowledge the work of Karush. But when he handed in his master’s thesis in December 1939 nothing happened: the work was not published, nobody encouraged him to publish his result, and apparently it was not very interesting. Fritz John’s paper came out only two years before Kuhn and Tucker’s paper; again nobody noticed it. In fact John tried to get it published earlier in the *Duke Mathematics Journal* but they rejected the paper! It is striking that only two years later when Kuhn and Tucker derived the result, it became famous almost instantaneously and caused the launching of a new mathematical research area.

These historical facts leads to the following questions. Was it really the same result they had derived? Is it fair here to talk about a multiple discovery, and in what sense is it or is it not a multiple discovery? Why were the reactions of the mathematical community so different in the three cases? Why did nothing happen the first two times? Or, maybe more interesting, why did Kuhn and Tucker’s work have such an enormous impact?

This paper is centered on these questions. They will be addressed and discussed on the basis of a contextualized historical analysis of the work of John, Karush, Kuhn, and Tucker. Both mathematical and social contexts will be considered, and the paper will end with a discussion of the role played by the military through the Office of Naval Research (ONR) and operations research (OR).

1.1. Mathematical Prerequisites

Let me very briefly explain what is to be understood by the concept of a nonlinear programming problem and state more precisely the Kuhn–Tucker theorem. A nonlinear programming problem is an optimization problem of the following type:

$$\begin{aligned} & \text{Minimize} && f(x) \\ & \text{subject to the constraints} && g_i(x) \leq 0 \quad \text{for } i = 1, \dots, m \\ & && x \in X. \end{aligned}$$

Here X is a subset of \mathbf{R}^n , the functions f, g_1, \dots, g_m are defined on X , and x is an n -dimensional vector (x_1, \dots, x_n) .¹

Thus a nonlinear programming problem is a finite-dimensional optimization problem where the variables have to fulfil some inequality constraints. A variable, $x \in \mathbf{R}^n$, which satisfies all the constraints is said to be “feasible.”

¹ For an exposition on the mathematical theory of nonlinear programming see, for example, [Bazaraa *et al.*, 1979, 1993; Luenberger, 1973; Peressini *et al.*, 1988].

THE KUHN–TUCKER THEOREM. *Suppose X is a nonempty open set in \mathbf{R}^n . Let \bar{x} be feasible and the functions f, g_1, \dots, g_m differentiable at \bar{x} . Suppose the gradient vectors $\nabla g_i(\bar{x})$ for the binding—or active—constraints, i.e., the constraints g_i for which $g_i(\bar{x}) = 0$, are linearly independent. Then the following will be true:*

Necessary conditions for $f(\bar{x})$ to be a minimum for the nonlinear programming problem above are that there exist scalars (multipliers) u_1, \dots, u_m such that

$$\nabla f(\bar{x}) + \sum_{i=1}^m u_i \nabla g_i(\bar{x}) = 0, \tag{1}$$

$$u_i g_i(\bar{x}) = 0 \quad i = 1, \dots, m, \tag{2}$$

$$u_i \geq 0 \quad i = 1, \dots, m.$$

These necessary conditions are called “the Kuhn–Tucker conditions.”

The first of these conditions, (1), is recognizable as saying that the corresponding Lagrangian function, $\phi(x, u) = f(x) + \sum_i^m u_i g_i(x)$, has a critical point in (\bar{x}, u) . The second condition, (2), ensures that if $g_i(\bar{x}) \neq 0$, that is, if g_i is not active in \bar{x} , then the corresponding multiplier u_i is equal to 0.

2. THE THEOREM OF KARUSH: A RESULT IN THE CALCULUS OF VARIATIONS

In December 1939 William Karush received a master’s degree in mathematics from the University of Chicago. His master’s thesis had the title “Minima of Functions of Several Variables with Inequalities as Side Conditions” [Karush, 1939].² Today we would say that such an optimization problem subject to inequality constraints belongs to the domain of nonlinear programming. But since the latter did not exist at that time, we need to take a closer look at Karush’s thesis in order to determine the field of mathematics to which it was considered a contribution. This student project was proposed by Karush’s supervisor Lawrence M. Graves [Karush, 1975]; so how did it fit in with the activities in the Department of Mathematics at Chicago at the time? Why was this problem interesting and what was Karush trying to do?

In the introduction to his thesis Karush stated the purpose of his work, and he also gave a hint where to look for the motivation behind the proposal of the problem. He wrote:

The problem of determining necessary conditions and sufficient conditions for a relative minimum of a function $f(x_1, \dots, x_n)$ in the class of points $x = (x_1, \dots, x_n)$ satisfying the equations $g_\alpha(x) = 0$ ($\alpha = 1, \dots, m$), where the functions f and g_α have continuous derivatives of at least the second order, has been satisfactorily treated [1]. This paper [Karush’s thesis] proposes to take up the corresponding problem in the class of points x satisfying the inequalities

$$g_\alpha(x) \geq 0 \quad (\alpha = 1, 2, \dots, m),$$

where m may be less than, equal to, or greater than n . [Karush, 1939, p. 1]

The reference ‘[1]’ in the above quotation is to a paper titled ‘Normality and Abnormality in the Calculus of Variations’ [Bliss, 1938]. It had been published just the year before by

² I am very grateful to the late Professor W. Karush for providing me with a copy of his thesis.

Gilbert Ames Bliss, who was the head of the department at Chicago. The problem that Karush's supervisor proposed for the thesis originated from this paper by Bliss. So the roots of the problem Karush set out to work on was buried in the calculus of variations, a field in mathematics that had a special connection to the department.

2.1. *The Chicago School in the Calculus of Variations*

The mathematical department at the University of Chicago was founded with the opening of the university in 1892. The first leader of the department was Eliakim H. Moore (1862–1932), who in cooperation with the two Germans Oskar Bolza (1857–1936) and Heinrich Maschke (1853–1903) created a mathematical environment that soon became the leading department of mathematics in the USA [Parshall and Rowe, 1994].

It was Bolza who introduced the calculus of variations as a major research field at the department. His own interest in the topic stemmed from Weierstrass's famous lectures in 1879, and Bolza taught the subject to graduate students at Chicago. From 1901 Bolza also turned his own research toward the calculus of variations. This indicated a shift in research direction, caused by a series of talks Bolza gave at the third American Mathematical Society (AMS) symposium. The purpose of these AMS meetings was to give an overview of selected mathematical topics for a broader audience of mathematicians and thereby suggest directions for new research. Chosen as one of the main speakers for the 1901 meeting, Bolza was asked to talk about hyperelliptic functions; but instead he chose to give talks on the calculus of variations. Interesting unsolved problems became visible, and from then on Bolza was deeply involved in research in that field [Parshall and Rowe, 1994, p. 394].

Bolza was very popular as a thesis advisor, often guiding his students to work in the field in which he was currently doing research himself. The result was that he created a solid foundations for research in the calculus of variations at Chicago—the so-called Chicago School of the calculus of variations [Parshall and Rowe, 1994, p. 393].

In 1908 Maschke died, and two years later Bolza returned to Germany. Chicago thus lost two of its leading mathematicians, and from 1910 on there seems to have been a decline in the reputation of the mathematics department. According to some Chicago mathematicians, this decline was caused by a too narrow focus on the calculus of variations.³ The “new team” at Chicago consisted of Bliss, Dickson, and Wliczynski. It was Bliss who, as a student of Bolza, continued the calculus of variations tradition.

Bliss was head of the department from 1927 to 1941 and this period in the life of the institute was characterized by intensive research in the calculus of variations. In the 10-year period from 1927 to 1937 the department produced 117 Ph.D. theses. Bliss supervised 35 of these, and 34 fell within the calculus of variations [MacLane, 1989, p. 138]. Several mathematicians connected with Chicago later held a very critical view of Bliss's program in the calculus of variations. They seem to share the following view put forward by A. L. Duren, who himself was a student of Bliss and wrote a Ph.D. in this field:

The subject itself had come to be too narrowly defined as the study of local, interior minimum points for certain prescribed functionals given by integrals of a special form. Generalization came only at the cost of excessive notational and analytical complications. It was like defining the ordinary calculus to consist exclusively of the chapter on maxima and minima [Duren, 1976, p. 245].

³ See for example [MacLane, 1989; Browder, 1989; Stone, 1989; Duren, 1976].

This is of course a characterization of the Chicago School under Bliss with hindsight, but it tells something about how extensive the research in the field was in the department at the time, and that it was quite narrowly defined there. As a student in Chicago Karush was a product of this tradition, and his master's thesis must be analyzed and discussed within that context.⁴

2.2. Karush's Master's Thesis

The purpose of Karush's work was to determine necessary and sufficient conditions for a relative minimum of a function $f(x_1, \dots, x_n)$ in the class of points $x = (x_1, \dots, x_n)$ satisfying the inequalities $g_\alpha(x) \geq 0$ for $(\alpha = 1, 2, \dots, m)$, where the functions f and g_α are subject to various continuity and differentiability conditions. He carried out this work in 1939 at a time when the research in Chicago was centered on variational calculus problems with *inequalities* as side conditions. Viewed in that context, Karush's problem can be interpreted as a finite-dimensional version of such a problem.

At first sight it can seem a little strange to have asked Karush—who was a promising student—to work on a finite-dimensional version of the real focus of attention, which lay in infinite dimensions. Karush did not explain the importance of his work in a broader perspective, but from his introduction it is clear that he viewed it as an extension of the work of Bliss, mentioned above, from the year before. From the mid-1930s Bliss had been interested in some properties called “normality” and “abnormality” for the minimizing arc of an *equality*-constrained problem in the calculus of variations. The purpose of the paper by Bliss which Karush took as point of departure was to

[...] analyze, more explicitly than has been done before, the meaning of normality and abnormality for the calculus of variations. To do this I have emphasized in §1 below the meaning of normality for the problem of a relative minimum of a function of a finite number of variables. [Bliss, 1938, p. 365]

Because as Bliss wrote,

The significance of the notion of abnormality in the calculus of variations can be indicated by a study of the theory of the simpler [finite-dimensional] problem. [Bliss, 1938, p. 367]

Hence, Bliss's idea was that valuable insight into the general more complicated cases could be obtained through a thorough study of the finite-dimensional case. In the light of this it is reasonable to presume that the same would hold true for the *inequality*-constrained case. This shows that even though the problem proposed for Karush's thesis did not fall directly in the main research area in the calculus of variations at the department, it would still have made sense to examine it.

The theorem which relates to the Kuhn–Tucker theorem appears in the third section of the thesis. Here Karush examined the minimum problem under the condition that the functions f and g_α , that is, the objective function and the constrained functions, are C^1 -functions near a point x^0 .

Before he proved the theorem which is now recognized as the Kuhn–Tucker theorem he showed its less restricted version:

⁴ For more information on the mathematical institute at Chicago under the leadership of Moore see [Parshall and Rowe, 1994; Duren, 1989]. For the history of the calculus of variations see [Fraser, 1992, 1994].

THEOREM 3.1. *If $f(x^0)$ is a minimum then there exist multipliers l_0, l_α not all zero such that the derivatives F_{x_i} of the function*

$$F(x) = l_0 f(x) + l_\alpha g_\alpha(x)$$

all vanish at x^0 . [Karush, 1939, pp. 12–13]⁵

Note that there is no sign restriction on the multipliers in these first necessary conditions. Also, the multiplier l_0 associated with the objective function f can take the value zero, in which case x^0 is called an “abnormal” point. In order to avoid the abnormal case some kind of regularity conditions or “constraint qualification,” as Kuhn and Tucker later called it, is needed.

The concepts which Karush introduced to construct such a regularity condition were “admissible direction,” “admissible curve,” and “normal point.” By an admissible direction Karush understood a nonzero vector $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ that solved the inequality system

$$\sum_{i=1}^n \frac{\partial g_\alpha}{\partial x_i}(x^0) \lambda_i \geq 0$$

[Karush, 1939, p. 11]. In other words, he considered a direction admissible if the directional derivatives of the constrained functions g_α in the direction of λ are nonnegative, which means that “you stay” in the feasible area if “you walk” from x^0 in the direction of λ . He called a regular arc $x_i(t)$ ($i = 1, 2, \dots, n$; $0 \leq t \leq t_0$) an admissible arc if

$$g_\alpha(x(t)) \geq 0 \quad \text{for all } \alpha \text{ and } t$$

[Karush, 1939, p. 11]. This means that a regular arc is admissible if “you stay” feasible when “you move” along the arc. Finally, he called a point x^0 normal if the Jacobian matrix for g has rank m at x^0 , that is, if the gradients

$$\nabla g_1(x^0), \quad \nabla g_2(x^0), \dots, \quad \nabla g_m(x^0)$$

are linearly independent.

Karush then formulated the “Kuhn–Tucker theorem” in the following way:

THEOREM 3.2. *Suppose that for each admissible direction λ there is an admissible arc issuing from x^0 in the direction λ . Then a first necessary condition for $f(x^0)$ to be a minimum is that there exist multipliers $l_\alpha \leq 0$ such that the derivatives F_{x_i} of the function*

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

all vanish at x^0 . [Karush, 1939, p. 13]

By a curve $x_i(t)$ ($0 \leq t \leq t_0$), “issuing from x^0 in the direction λ ” he meant that $x_i(0) = x_i^0$ and $x_i'(0) = \lambda_i$ [Karush, 1939, p. 13].

⁵ Karush used the Einstein summation symbolism: i.e., $F(x) = l_0 f(x) + l_\alpha g_\alpha(x)$ means $F(x) = l_0 f(x) + \sum_{\alpha=1}^m l_\alpha g_\alpha(x)$.

His idea was to use Farkas's lemma⁶ to guarantee the existence of nonpositive multipliers l_α , and the assumptions in the theorem—the regularity condition—ensure precisely that this lemma can be brought into action.

2.3. *The Acknowledgment of Karush's Thesis in Nonlinear Programming*

Karush's theorem looks indeed very much like the version of the Kuhn–Tucker theorem shown in the Introduction. That is, there should exist multipliers (l_α) such that the Lagrangian function F has a critical point at $x^0, (l_\alpha)$. The condition $l_\alpha g_\alpha(x^0) = 0$ is missing because Karush only considered the active constraints, i.e., constraints for which $g_\alpha(x^0) = 0$.

In 1975 Harold Kuhn wrote a letter to Karush saying:

First let me say that you have clear priority on the results known as the Kuhn–Tucker conditions (including the constraint qualification). I intend to set the record as straight as I can in my talk. [Kuhn, 1975a]

Kuhn was referring to a talk that he had been asked to give on the history of nonlinear programming at an AMS symposium. He became aware of the work of Karush through Takayama's book "Mathematical Economics" [Takayama, 1974]; [Kuhn, 1976, p. 10]. During the research for the AMS talk Kuhn made contact with Karush and offered a partial publication of the master's thesis as an appendix to Kuhn's historical paper in the AMS proceedings that was to be published after the meeting. In this paper Kuhn announced Karush's thesis as an unpublished classic in the field of nonlinear programming [Kuhn, 1976].

Just looking at Karush's result independent of the context of discovery, one can only agree with Kuhn and say that Karush actually had the later Kuhn–Tucker theorem. In the light of its later importance one is then naturally led to the questions: Why was Karush's result not valued at the time? Why was it not published?

As we have seen, the main interest in Chicago at the time was variational calculus with inequality constraints, and if Karush's work is evaluated in this context then it was only a minor, finite-dimensional result, some "cleaning up" in a research direction where variational calculus with inequality constraints was the main field. Neither the posed problem nor the result was special. The interesting questions in this field were different from those that were to be important and later guided the research in nonlinear programming.

The letter from Kuhn to Karush quoted above also suggests how important the theorem was considered to be in the community of mathematicians working in nonlinear programming. Kuhn tells in the letter that Richard Cottle, who was among the organizers of the AMS symposium, made the following remark about Karush when he heard about Kuhn's intentions of "setting the record straight":

"you must be a saint" not to complain about the absence of recognition. [Kuhn, 1975a]

Kuhn also writes about Tucker's reaction when he learned of the result in Karush's thesis. Tucker was truly amazed that Karush had never told him about his work when they met at the RAND Coperation [Kuhn, 1975a].⁷ Richard Bellman wrote the following to Kuhn when he learned about Kuhn's forthcoming talk:

⁶ For a short historical account of Farkas's lemma see [Brentjes, 1976a]. To consult Farkas's own work see [Farkas, 1901].

⁷ Project RAND emerged just after the Second World War with the purpose of continuing the cooperation between researchers in academia and in industry and the military which took place during the war. For further information see [Smith, 1969; Hourshell, 1997].

I understand from Will Karush that you will try and set the record straight on the famous Kuhn–Tucker condition. I applaud your effort. Fortunately, there is enough credit for everybody. It would certainly be wonderful if you wrote it as the Kuhn–Tucker–Karush condition. Like many important results, it is not difficult to establish, once observed. That does not distract from the importance of the condition. [Bellman, 1975]

Also the mathematician Phil Wolfe informed Kuhn how pleased he was that Karush's work would now be recognized [Kuhn, 1975b].

From the letters it is clear that the mathematicians working in the field were truly amazed that Karush had not come forward to claim if not priority then at least recognition. To this Karush himself gave the following explanation:

That does not answer the question of why I did not point to my work in later years when nonlinear programming took hold and flourished. The thought of doing this did occur to me from time to time, but I felt rather diffident about that early work and I don't think I have a strong necessity to be "recognized." In any case, the master's thesis lay buried until a few years ago when Hestenes urged me to look at it again to see if it shouldn't receive its proper place in history. . . . So I did look at the thesis again, and I looked again at your work with Tucker. I concluded that you two had exploited and developed the subject so much further than I, that there was no justification for my announcing to the world, "Look what I did, first." [Karush, 1975]

From the point of view of the history of mathematics I think Karush is right here. He did derive a result that was comparable to the Kuhn–Tucker theorem, but he did not explore the subject further, and his work was *not* nonlinear programming but occurred in a completely different context. The department at Chicago had become under Bliss a place with focus on a very narrowly defined calculus of variations research programme, and within this research direction nobody was interested in exploring the possibilities for applications of Karush's result.

3. THE THEOREM OF FRITZ JOHN: A CONTRIBUTION TO THE THEORY OF CONVEXITY

Fritz John's version of the Kuhn–Tucker theorem appeared in his essay "Extremum Problems with Inequalities as Subsidiary Conditions," which was published in 1948 in the Courant anniversary volume [John, 1948].

John was a student of Richard Courant in Göttingen, where he received a Ph.D. in 1933. He had Jewish ancestors, and Courant worked hard to find him a position outside of Germany. In 1934 he succeeded in getting John a research scholarship at Cambridge, England. John moved to the United States a year later, where he received an offer from the University of Kentucky. He worked there until 1943, and after some years of war-related work at the Ballistic Research Laboratory at Aberdeen Proving Ground, he returned "home" to Courant at his institute at New York University [Reid, 1976, pp. 131–132, 154–155].

Fritz John was a world-class mathematician. His list of publications counts 101 mathematical texts, papers as well as monographs, and he has received many prizes and fellowships. Today he is probably most recognized for his work on partial differential equations, but he has also made important contributions in the fields of geometry, analysis, and nonlinear elasticity. At the time when the Courant anniversary volume was published John had mostly been working within the theory of convexity—more than half of his mathematical publications until this one of 1948 were in that field, and quite a few are now considered "classics" in the theory of convexity [Gårding, 1985; Moser, 1985].

3.1. John’s Paper

What was John’s intention in this paper? In the Introduction he wrote:

This paper deals with an extension of Lagrange’s multiplier rule to the case, where the subsidiary conditions are inequalities instead of equations. Only extrema of differentiable functions of a finite number of variables will be considered. [John, 1948, p. 187]

Like Karush, John only looked at the finite-dimensional case; so, judging from the title and the Introduction, it sounds very much as if John was interested in the same kind of questions as Karush. This impression is reinforced later in the introduction where John pointed to further directions of research on the problem:

from the point of view of applications it would seem desirable to extend the method used here to cases, where the functions involved . . . do not depend on a finite number of independent variables. [John, 1948, p. 187]

This extension of the problem clearly belonged to the calculus of variations; but if John considered his work as a contribution to this field, it would seem unlikely that he did not know the work of the Chicago School in the calculus of variations—well known at the time—who had already carried out this work for the general case.

However, apparently John did not know the Chicago work; there is no reference to the calculus of variations in his paper. What was his real interest, then? In the following I will scrutinize his paper to see *what* he actually did and *how* he did it.

The paper is divided into two parts; the first is concerned with the question of necessary and sufficient conditions for the existence of a minimum and the second is devoted to two geometrical applications of the theoretical result in part one.

John formulated the result that later was acknowledged as a version of the Kuhn–Tucker theorem in the following way:

Let R be a set of points x in \mathbf{R}^n ,⁸ and $F(x)$ a real-valued function defined in R . We consider a subset R' of R , which is described by a system of inequalities with parameter y :

$$G(x, y) \geq 0,$$

where G is a function defined for all x in R and all “values” of the parameter y . . . we assume that the “values” of the parameter y vary over a set of points S in a space H . . . We are interested in conditions a point x^0 of R' has to satisfy in order that

$$M = F(x^0) = \min_{x \in R'} F(x).$$

[John, 1948, p. 187–188]

Under some further continuity and differentiability conditions John was able to prove the following theorem:

THEOREM I. *Let x^0 be an interior point of R , and belong to the set R' of all points x of R , which satisfy the constraints $G(x, y) \geq 0$ for all $y \in S$. Let*

$$F(x^0) = \min_{x \in R'} F(x).$$

⁸ Instead of \mathbf{R}^n John wrote . . . in a space E , but in the following he restricted himself to the case where the space E containing the set R is the n -dimensional Euclidean space, which I have called \mathbf{R}^n [John, 1948, p. 188].

Then there exists a finite set of points y^1, \dots, y^s , in S and numbers $\lambda_0, \lambda_1, \dots, \lambda_s$, which do not all vanish, such that

$$G(x^0, y^r) = 0 \quad \text{for } r = 1, \dots, s$$

$$\lambda_0 \geq 0, \quad \lambda_1 > 0, \dots, \lambda_s > 0,$$

$$0 \leq s \leq n,$$

the function

$$\phi(x) = \lambda_0 F(x) - \sum_{r=1}^s \lambda_r G(x, y^r)$$

has a critical point at x^0 i.e., the partial derivatives are zero at x^0 :

$$\phi_i(x^0) = 0 \quad \text{for } i = 1, \dots, n.$$

(See [John, 1948, pp. 188–189])

John's way of attacking the problem was the same one Karush used, but where Karush invoked Farkas's lemma as his main tool John used other similar results from the theory of convexity, with which he was familiar through various recent works (for example [Dines, 1936; Stokes, 1931]).

John's formulation of the theorem looks a little different from Karush's, but the stated conditions are the Kuhn–Tucker conditions. The differences are the appearance of the parameter y in the parameter set S , and that the multiplier λ_0 associated with the objective function F can become zero as in Karush's first theorem. The latter difference is caused by the fact that John did not have the constraint qualification, as Kuhn and Tucker called it, or the normality condition, as Karush would have said.

3.2. The Two Geometrical Applications

From reading the second part of the paper, which is concerned with the two geometrical applications, it becomes clear why John chose this construction with the parameter y and a parameter set S . It also explains why John did not touch upon the problem of abnormality and thereby did not consider the problem of constraint qualification.

More than half of the paper is devoted to these geometrical applications. The first is "Application to Minimum Sphere Containing a Set" and the second concerns the ellipsoid of least volume containing a set S in \mathbf{R}^m [John, 1948, p. 193]. In the first one John considered the following problem:

Let S be a bounded set in \mathbf{R}^m . Find the sphere of least positive radius enclosing S . [John, 1948, p. 193–194]

John was not interested in the *existence* of such a sphere. If the assumption is made that the bounded set S contains at least two distinct points, it is quite clear that such a sphere exists [John, 1948, p. 194].

To be able to use his theorem derived in the first part of the paper John characterized spheres in \mathbf{R}^m as points in \mathbf{R}^{m+1} ,

$$x = (x_1, \dots, x_{m+1}),$$

where (x_1, \dots, x_m) are the coordinates of its center and x_{m+1} the square of its radius. He could then rewrite the problem as an optimization problem subject to inequality constraints:

Minimize the function $F(x) = x_{m+1}$ subject to the constraints

$$G(x, y) = x_{m+1} - \sum_{i=1}^m (x_i - y_i)^2 \geq 0 \quad \text{for all } y \in S.$$

The constraints ensure that the minimum is only sought among spheres containing S .

John used a similar procedure in the second application about the ellipsoid. In both cases he knew that a minimum, x^0 , existed, so the necessary conditions of the theorem were fulfilled. He then used these conditions to derive significant properties of the minimum sphere and the minimum ellipsoid. From this he also derived several general properties of closed convex sets [John, 1948, pp. 201–202].

3.3. *The Link to the Theory of Convexity*

In the application part of John's paper and especially in the last one it becomes clear that his main interest was in the results about closed convex sets that he developed through the applications of his theoretical result: the extension of Lagrange's multiplier method to problems with inequality constraints. In connection with Kuhn's talk on the history of nonlinear programming Kuhn also had a brief correspondence with John.⁹ According to Kuhn, John should have revealed that he was led to the theorem when he was

trying to prove the theorem ... that asserts that the boundary of a compact convex set S in \mathbf{R}^n lies between two homothetic ellipsoids of ratio $\leq n$, and that the outer ellipsoid can be the ellipsoid of least volume containing S . [Kuhn, 1976, p. 15]

Even though in his title and introduction John gives the impression that he is concerned with problems in the calculus of variations, it is my opinion that his paper rather should be viewed as a contribution to the theory of convexity, to which he had made fine contributions. All the references in the paper are either to the theory of convexity or to less general works—by John and others—on the two applications.¹⁰ In considering the applications it becomes quite clear that they have a justification in themselves, for they serve a deeper purpose than just as illustrations of the theoretical result. The conclusion must be that the guiding questions—the important issues for John—were the applications and the results he could derive from these.

3.4. *The Status of the Theorem*

In Karush's work the theorem was important in itself. The whole purpose of his work was to derive these necessary conditions for the existence of a minimum or maximum. In John's work, on the other hand, the theorem was only derived as a tool for deriving general results about convex sets. The applications guided the formulation of the theorem, which explains John's construction of the "parameter set" which clearly is dictated by the applications.

⁹ This correspondence is apparently lost.

¹⁰ [John, 1936, 1942; Behrend, 1937, 1938; Ader, 1938].

Another difference between Karush's and John's work is the "normality" condition, as Karush called it, or the "constraint qualification," as Kuhn and Tucker will call it; John does not touch upon that feature. This can also be explained from the fact that both applications are actually examples of the normal case.

In his paper on the history of nonlinear programming Kuhn wrote about John's work that it "very nearly joined the ranks of unpublished classics in our subject" [Kuhn, 1976, p. 15]. But John himself apparently did not view this work in this way, and he never came forward with priority claims.

4. THE THEOREM OF KUHN AND TUCKER: AN EXTENSION OF LINEAR PROGRAMMING

Albert W. Tucker was born in Canada in 1905 and died in Princeton, New Jersey, in 1995. He received a bachelor's degree in mathematics from the University of Toronto in 1928, and a year later began Ph.D. study at Princeton University. This turned out to be the beginning of a lifelong connection to the Mathematics Department at Princeton. In 1932 he received the Ph.D. on a thesis in the field of topology, and two years later he was appointed assistant professor. In 1938 he became associate professor, and then full professor in 1946. An important figure in the maintenance of Princeton in the 1930s and 1940s as a prestigious place for mathematical research, he served as head of the department from 1953 to 1963. He had a tremendous influence on the students who came in contact with him, and he is often characterized as a very good teacher and leader [Tucker, 1980; Kuhn, 1995].

Harold W. Kuhn—20 years younger than Tucker—was born in California. He received a bachelor's degree in science from the California Institute of Technology in 1947, and then moved on to Princeton where he wrote a Ph.D. thesis on "Subgroup Theorems for Groups Presented by Generators and Relations" in 1950 [Kuhn, 1952]. After some travelling and a seven-year appointment at Bryn Mawr College, Kuhn returned to Princeton as associate professor. He was connected to both the mathematics and the economics departments [Kuhn, 1986].

4.1. *The Nonlinear Programming Paper*

The main point in Kuhn and Tucker's paper was to find necessary and sufficient conditions for the existence of a solution to the following "maximum problem," as they called it:

To find an x^0 that maximizes $g(x)$ constrained by $Fx \geq 0$, $x \geq 0$ [Kuhn and Tucker, 1950, p. 483].

Here $x^0 \in \mathbf{R}^n$ and $x \rightarrow u = Fx$ is a differentiable mapping of nonnegative n -vectors x into m -vectors u . That is, Fx is an m -vector whose components $f_1(x), \dots, f_m(x)$ are differentiable functions of x defined for $x \geq 0$, and $g(x)$ is a differentiable real function of $x \in \mathbf{R}^n$ defined for $x \geq 0$ [Kuhn and Tucker, 1950, p. 483].

Kuhn and Tucker handled this problem by taking the so-called "saddle value problem" as their point of departure. They defined it as the problem of finding nonnegative vectors $x^0 \in \mathbf{R}^n$ and $u^0 \in \mathbf{R}^m$, such that

$$\phi(x, u^0) \leq \phi(x^0, u^0) \leq \phi(x^0, u) \quad \text{for all } x \geq 0, u \geq 0,$$

where $\phi(x, u)$ is a differentiable function of an n -vector x with components $x_i \geq 0$ and an m -vector u with components $u_h \geq 0$.

They let ϕ_x^0, ϕ_u^0 denote the partial derivatives, evaluated at a particular point x^0, u^0 . That is, ϕ_x^0 is an n -vector,

$$\phi_x^0 = \left(\frac{\partial \phi}{\partial x_1}(x^0), \dots, \frac{\partial \phi}{\partial x_n}(x^0) \right),$$

and ϕ_u^0 is an m -vector,

$$\phi_u^0 = \left(\frac{\partial \phi}{\partial u_1}(u^0), \dots, \frac{\partial \phi}{\partial u_m}(u^0) \right).$$

They used the $'$ notation to denote the transposed vector.

The first theorem Kuhn and Tucker proved in the paper concerned the question of necessary and sufficient conditions for the existence of a solution to the saddle value problem. They proved that the conditions

$$\phi_x^0 \leq 0, \quad \phi_x^{0'} x^0 = 0, \quad x^0 \geq 0 \tag{1}$$

$$\phi_u^0 \geq 0, \quad \phi_u^{0'} u^0 = 0, \quad u^0 \geq 0 \tag{2}$$

are necessary for x^0, u^0 to provide a solution [Kuhn and Tucker, 1950, pp. 482–483]. For the second part of the question they proved that the conditions (1), (2) together with the two conditions

$$\phi(x, u^0) \leq \phi(x^0, u^0) + \phi_x^{0'}(x - x^0) \tag{3}$$

$$\phi(x^0, u) \geq \phi(x^0, u^0) + \phi_u^{0'}(u - u^0) \tag{4}$$

for all $x \geq 0, u \geq 0$, are *sufficient* [Kuhn and Tucker, 1950, p. 483].

Equipped with these conditions, Kuhn and Tucker phrased their theorem in the following way:

THEOREM 1. *In order that x^0 be a solution of the maximum problem, it is necessary that x^0 and some u^0 satisfy conditions (1) and (2) for $\phi(x, u) = g(x) + u'Fx$. [Kuhn and Tucker, 1950, p. 484]*

If the condition $x^0 \geq 0$ is incorporated into the constraint function, F , the first and last conditions in (1) together mean that the Lagrangian function $\phi(x, u)$ has a critical point at x^0, u . The second condition in (1) ensures that the multipliers associated with the nonbinding components of x^0 are equal to zero. The first condition in (2) ensures that x^0 is feasible, the second ensures that the multipliers associated with nonbinding constraints are equal to zero, and the last is the sign-restriction on the multipliers. These conditions later became known as “the Kuhn–Tucker conditions,” and they constitute one of the main results in the mathematical theory of nonlinear programming.

Actually the first time Kuhn and Tucker announced this theorem was not at the Berkeley Symposium but a few months earlier at a seminar held at the RAND Corporation in May 1950. Among the audience was C. B. Tompkins, who came up with something as unpleasant as a counterexample [Kuhn, 1976, p. 14]. The result—as it stood—could not rule out the

“abnormal” case, as Karush would have called it. Kuhn and Tucker got back to work and realized the need for some regularity conditions on the constraint functions. This led them to introduce the term “constraint qualification.” The constraint qualification they used in their paper was the same as Karush’s: that for each x^0 of the boundary of the set determined by the constraints and for any vector differential dx for which the directional derivatives of the binding constraints in the direction of dx are nonnegative, there corresponds a differentiable arc $x = a(\theta)$, $0 \leq \theta \leq 1$, contained in the constrained set, with $x^0 = a(0)$, and some positive scalar λ such that $a'(x^0) = \lambda dx$ [Kuhn and Tucker, 1950, p. 483].

As Kuhn and Tucker pointed out in their paper, it can seem artificial to introduce the conditions (3) and (4) that occurred in the sufficiency part of the saddle value problem; but these conditions are satisfied if $\phi(x, u^0)$ is a concave function of x and $\phi(x^0, u)$ is a convex function of u [Kuhn and Tucker, 1950, p. 483]. In order to gain full equivalence between solutions of the maximum problem and the saddle value problem Kuhn and Tucker then required that the functions involved, g, f_1, \dots, f_m , be concave as well as differentiable for $x \geq 0$. With these extra requirements they showed that

x^0 is a solution of the maximum problem if, and only if, x^0 and some u^0 give a solution of the saddle value problem for $\phi(x, u) = g(x) + u'F(x)$. [Kuhn and Tucker, 1950, p. 486]

4.2. *The Saddle Value Problem: A Detour?*

Kuhn and Tucker’s formulation of the theorem is different from that of Karush and John, neither of whom considered the concept of saddle points. Why did Kuhn and Tucker choose the saddle point formulation, and why were they looking for equivalence between the maximum problem and the saddle value problem? The mathematical context of their work can provide an answer to these questions.

Their cooperation had begun two years earlier, in 1948, where they had examined the relation between game theory and the linear programming model that had just been developed by George B. Dantzig for the U.S. Air Force. Kuhn was still a student at the time and together with another student, David Gale, they worked out the mathematical foundations for linear programming [Gale *et al.*, 1951]. They formulated the corresponding dual problem, proved the duality theorem, and showed the relation between linear programming and game theory.¹¹

When he was introduced to the linear programming problem, Tucker was at first reminded of Kirchoff’s laws for electrical networks [Albers and Anderson, 1985, pp. 342–343]. In the autumn of 1949 just after Kuhn, Gale, and Tucker had presented their work on linear programming and game theory at the first conference on linear programming, held in Chicago in June 1949, Tucker went on leave to Stanford. Here he dug deeper into his first association, and discovered the underlying optimization problem of minimizing heat loss. According to Kuhn, this knowledge led Tucker to the recognition that the Lagrangian multiplier method which is normally used to solve *equality*-constrained optimization problems

¹¹ A linear programming problem is a nonlinear programming problem where all the involved functions are linear functions. For a linear programming problem one can formulate another linear programming problem on the same data called the *dual* program. The duality theorem says that the original, *primal*, problem has a finite optimal solution if and only if the dual problem has a finite optimal solution, and the optimum values will be the same.

could be adapted to optimization problems subject to *inequality* constraints [Kuhn, 1976, pp. 12–13]. Tucker then wrote to Kuhn and Gale and invited them to continue the work and extend their duality result for linear programming to quadratic programming, i.e., to problems where the involved functions would be quadratic in form [Kuhn, 1976, pp. 12–13]. David Gale declined the offer but Kuhn accepted, and he and Tucker developed the theory in correspondence between Stanford and Princeton.¹²

Thus, the original purpose of Kuhn and Tucker’s work was to extend the duality result from linear programming to quadratic programming, and the idea was to adapt the classical Lagrangian multiplier method. In the introduction to their paper Kuhn and Tucker explained how this would work for linear programming. From a linear programming problem

$$\text{maximize } g(x) = \sum c_i x_i, \quad c_i \in \mathbf{R},$$

where x_1, \dots, x_n are n real variables constrained by $m + n$ linear inequalities,

$$f_h(x) = b_h - \sum a_{hi} x_i \geq 0, \quad x_i \geq 0,$$

with $h = 1, \dots, m, i = 1, \dots, n, a_{hi}, b_h \in \mathbf{R}$, they formed the corresponding Lagrangian function,

$$\phi(x, u) = g(x) + \sum u_h f_h(x), \quad u_h \in \mathbf{R}.$$

They realized that $x^0 = (x_1^0, \dots, x_n^0)$ will maximize $g(x)$ subject to the given constraints if and only if there exists a vector $u^0 = (u_1^0, \dots, u_m^0) \in \mathbf{R}^m$ with components $u_i^0 \geq 0$ for all i , such that (x^0, u^0) is a saddle point for the Lagrangian function $\phi(x, u)$ [Kuhn and Tucker, 1950, p. 481]. The really interesting feature of this saddle point result for linear programming was, as Kuhn and Tucker phrased it,

The bilinear symmetry of $\phi(x, u)$ in x and u yields the characteristic duality of linear programming. [Kuhn and Tucker, 1950, p. 481]

Thus a linear programming problem has a solution if and only if the corresponding Lagrangian function has a saddle point; this saddle point then constitutes a solution not only to the linear programming problem but also to the dual program. Considering now that Kuhn and Tucker actually were searching for a way to extend the duality theorem for linear programming to more general cases,¹³ it seems perfectly natural to take the saddle point for the Lagrangian function as the starting point.¹⁴

Until now I have only explained and interpreted the content, the structure, and the underlying mathematical ideas of the results in Kuhn and Tucker’s paper. The important question raised in the Introduction concerning why their work had such an enormous impact that it could launch a new research field in applied mathematics can only be understood in a

¹² This correspondence is lost; I know about it from an interview with Kuhn, who also mentioned it in [Kuhn, 1976, p. 13].

¹³ Somewhere during the process they shifted the focus from the quadratic case to the general nonlinear case.

¹⁴ It is striking then that Kuhn and Tucker did not mention duality for nonlinear programming in the paper. The first duality result for nonlinear programming was derived by Werner Fenchel in 1951, published in 1953 [Fenchel, 1953].

broader perspective that takes into account the relation between the military and science during World War II as well as the postwar organization of science support in the United States. These questions will be dealt with in Section 6.

5. THE ASPECT OF MULTIPLE DISCOVERY

The reason a question of multiple discovery arises in connection with a historical study of the Kuhn–Tucker theorem is that the result today in textbooks and in papers on the history of mathematics is ascribed to all of them—Karush, John, and Kuhn and Tucker.¹⁵

One can also see the result ascribed to the Russian mathematician Mikhail Ostrogradsky (1801–1862) and the Hungarian mathematician Julius Farkas (1847–1930). In three papers Franksen discusses Fourier’s extension of the principle of virtual work in mechanics and how it sheds new light on the development of the second law of thermodynamics and mathematical programming [Franksen, 1985a, 1985b, 1985c]. He concludes that the Kuhn–Tucker theorem is an independent rediscovery, by Kuhn and Tucker, of a theorem derived by Ostrogradsky in a paper which was read for the French Academy in 1834 and published four years later, in 1838 [Franksen, 1985c, pp. 337–338, 353, 355]. Prékopa gives an account of the development of optimization theory in a paper of 1980. He had searched for the first appearance of the Kuhn–Tucker conditions in the literature and he found it in Ostrogradsky and Farkas [Prékopa, 1980, p. 528].

Before I return to the question whether Karush’s, John’s, and Kuhn and Tucker’s work can be said to count as a multiple discovery I will briefly deal with these older sources which discuss questions belonging to the field of analytical mechanics—questions that came out of Fourier’s extension of the principle of virtual work.

5.1. *The Kuhn–Tucker Theorem in Analytical Mechanics*

John as well as Kuhn and Tucker mentioned explicitly that their work in one way or another was connected with the Lagrangian multiplier method. John wrote directly in his introduction that the purpose of his work was to extend this method to problems with inequality constraints. Tucker associated the network nature of linear programming with Kirchoff’s laws for electrical networks and got the idea that maybe the Lagrangian multiplier method could be adapted to inequality constraint cases.

Lagrange developed his multiplier method in “*Méchanique analytique*” (1788) as a method for finding an equilibrium for a mechanical system [Lagrange, 1788]. He founded his theory of equilibrium on what is now called the principle of virtual work, which he took as an axiom. In modern terms the principle states that in order for an equilibrium to take place the virtual work of the applied forces acting on the system must be equal to zero. This principle was stated in terms of reversible displacements which means that if a virtual displacement δr is allowed then the opposite displacement $-\delta r$ is also possible without breaking the constraints on the system. This means that the mechanical system is subject only to constraints that can be formulated as equations [Franksen, 1985a, p. 137].

Inequalities entered the picture in 1798 where Fourier extended the principle of virtual work to irreversible displacements, that is, to mechanical systems subject to inequality

¹⁵[Bazaraa *et al.*, 1993, p. 149; Peressini *et al.*, 1988, p. 169]. For an account of the prehistory of linear and nonlinear programming see [Grattan-Guinness, 1970, 1994]. For an account on the history of nonlinear programming see [Kuhn, 1976; Kjeldsen, 1999].

constraints [Fourier, 1798]. Based on arguments concerning “le moment de la force” [Fourier, 1798, p. 479], he formulated the conditions for equilibrium for such systems as inequality conditions, as he realized that such systems are in equilibrium if, and only if, the virtual work of the applied forces is nonpositive [Fourier, 1798, p. 494]. This inequality is often called “the Fourier inequality.”

Ostrogradsky derived the conditions for equilibrium for such a system in [Ostrogradsky, 1838]. He denoted the applied forces acting on a system by P, Q, R, \dots . He then wrote the equilibrium condition, that is, the Fourier inequality, in the following way: the total work

$$Pdp + Qdq + Rdr + \dots$$

has to be nonpositive for every feasible displacement. The constraints were named L, M, \dots and because these constraints were given by inequalities Ostrogradsky argued that dL, dM, \dots “... can only change sign in cases where one moves from feasible to infeasible displacements” [Ostrogradsky, 1838, p. 131].

Ostrogradsky’s maneuver was to change the coordinates by introducing so-called “generalized” coordinates; instead of considering $dp, dq, dr \dots$ he introduced some other variations $d\xi, d\eta, d\psi, \dots$ which are functions of dp, dq, dr, \dots and in number equal the number of the original variables. Since dL, dM, \dots are also functions of dp, dq, dr, \dots , he took these to be the first of the new generalized coordinates (this means that Ostrogradsky’s method can be used only when the number of constraints does not exceed the number of variables). He then reformulated the whole thing with these new coordinates and obtained the equilibrium condition

$$\lambda dL + \mu dM + \dots + Ad\xi + Bd\eta + Cd\zeta + \dots \leq 0$$

for every feasible displacement [Ostrogradsky, 1838, p. 131]. Using arguments about the impossibility of changing signs for dL, dM, \dots and the possibility of sign changing for $d\xi, d\eta, d\psi, \dots$ Ostrogradsky concluded that $A = B = C = \dots = 0$. This meant that the total work, $Pdp + Qdq + Rdr + \dots$, equals $\lambda dL + \mu dM + \dots$; i.e.

$$Pdp + Qdq + Rdr + \dots = \lambda dL + \mu dM + \dots$$

for all feasible displacements. Since dL, dM, \dots cannot change sign, the equilibrium condition can only take place, he concluded, if the multipliers λ, μ, \dots have signs opposite to those of the corresponding constraints, dL, dM, \dots [Ostrogradsky, 1838, p. 132].

Ostrogradsky then ended up by concluding that:

[...] les conditions de l’équilibre d’un système quelconque seront exprimées

1^{mo} par l’équation

$$0 = Pdp + Qdq + Rdr + \dots + \lambda dL + \mu dM + \dots$$

qui doit avoir lieu pour tous les déplacements imaginables,

2^{do} par la condition que les quantités λ, μ, \dots aient respectivement les mêmes signes que les différentielles dL, dM, \dots pour les déplacements possibles. [Ostrogradsky, 1838, p. 132–133]

Today, in terms of potential theory, if $P = -\frac{\partial V}{\partial p}$, $Q = -\frac{\partial V}{\partial q}$, $R = -\frac{\partial V}{\partial r}$, . . . , one can “translate” the question of finding an equilibrium into a problem about minimizing the potential energy. So the conclusion of Franksen and Prékopa that Ostrogradsky here formulated as well as argued for what we call the Kuhn–Tucker theorem in nonlinear programming can only be understood with this interpretation and “translation” of Ostrogradsky’s work. My opinion is that in ascribing the Kuhn–Tucker theorem to Ostrogradsky too much has been read into the sources. In the next section I will provide further reasons for this conclusion.

The mathematical foundations for the extension of Lagrange’s multiplier method to equilibrium for mechanical systems subject to irreversible displacements was treated by Farkas. The main mathematical result that came out of this is Farkas’s lemma about linear inequality systems [Farkas, 1901]. Farkas developed it in some earlier papers [Farkas, 1895, 1897, 1899] whose main focus was

[. . .] zu erweisen, dass mit einer passenden Modifikation die Methode der Multiplikatoren von Lagrange auch auf das Fourier’sche Princip betragen werden kann. [Farkas, 1895, p. 266]

There is a remarkable resemblance to the goal stated in John’s introduction, but here in a context of analytical mechanics.

Farkas knew the work of Ostrogradsky, and he made a remark about the limitation of the method used by Ostrogradsky to situations where the number of constraints does not exceed the number of variables [Farkas, 1895]. Farkas wanted to find a method that could be used in any problem no matter what relationship between the numbers of constraints and of variables [Farkas, 1895, p. 266]. He was very much concerned with the mathematical foundations of the method and had a clear insight that homogeneous linear inequalities could provide a satisfactory form; so he began his 1895 paper with such a theory:

I. enthält eine algebraische Einleitung über die homogenen linearen Ungleichheiten als mathematische Grundlage der weiteren Betrachtungen. [Farkas, 1895, p. 266]

This “algebraische Einleitung” consists of a proof of what we now call “Farkas’s lemma.” With the help of it Farkas was able to reach the same conclusion as Ostrogradsky but this time for the general problem where there was no restriction on the relation between the numbers of variables and of constraints on the system. Again, if a potential V exists, Farkas’s results can be translated and interpreted as the Kuhn–Tucker conditions, but the conclusion drawn for Ostrogradsky also holds here.

The work of Ostrogradsky and Farkas had no direct influence on the development of nonlinear programming. It is true that Farkas’s lemma functions as an important tool in both the work of Karush and that of Kuhn and Tucker, but they used a version of Farkas’s lemma that was completely removed from analytical mechanics and equilibrium conditions. Indeed the title of Farkas’s 1901 paper, “Theorie der einfachen Ungleichungen,” shows that here he was concerned solely with the pure theory of inequalities [Farkas, 1901].

5.2. Theories of Multiple Discoveries

The mathematical community does not ascribe the Kuhn–Tucker theorem to Ostrogradsky and Farkas, but it does consider the work of Karush and John as papers belonging to nonlinear programming, and both names now appear in textbooks. The Kuhn–Tucker theorem is now often renamed the Karush–Kuhn–Tucker theorem and there is also a Fritz John theorem [Bazaraa *et al.*, 1993]. Also, my analysis of Karush’s, John’s, and Kuhn and Tucker’s work

seems to indicate that we may actually have a multiple discovery. What I find particularly interesting is the fact that three occurrences of a result which the scientific community later viewed as the same, developed within a time span of only 11 years, were received so differently. In order to examine and understand this phenomenon I turned to theories of multiple discoveries.

A central figure in the literature on multiple discoveries in science is Robert K. Merton. His main criterion for talking about a multiple discovery is independent discovery of the same scientific result, and his theory is that they are not something special in science; on the contrary, it is the discoveries that on the surface appear to be single that deserve special attention. Merton's hypothesis states that a thorough investigation will show that these singletons will turn out to be if not multiple then at least potentially so. According to him, "all scientific discoveries are in principle multiples" [Merton, 1973, p. 356]. He has 10 different arguments for this hypothesis. First of all he points to the huge class of singletons which later turn out to be rediscoveries of results found in earlier work—unpublished or published in "obscure" places. Then he has six arguments that all are concerned with the problem of "being anticipated." He describes situations where the scientist for some reason suddenly realizes that someone else already has developed the result he or she is working on. If the scientist then lets go of the result, the discovery is an example of a singleton which in reality was a potential multiple discovery. If the scientist goes ahead and publishes anyway there will typically be a footnote saying that this or that person arrived at this conclusion in this or that source. The last three of Merton's arguments deal with the way scientists behave. In Merton's view the behavior reveals that they themselves believe that all scientific discoveries are potentially multiple. Here he refers to all the different things that scientists do in order to secure that they will not be anticipated by another scientist: they carefully date their notes, they "leak" information about their ideas and circulate incomplete versions of their work [Merton, 1973, p. 358–361]. This behavior, Merton points out, is based on a wish to ensure priority, which is very important in the scientific world:

the culture of science puts a premium not only on originality but on chronological firsts in discovery, this awareness of multiples understandably activates a rush to ensure priority. [Merton, 1973, p. 361]

Evaluated according to Merton's theory, the Kuhn–Tucker theorem is a triple discovery. Some of the circumstances Merton points out can be found in the work of Kuhn and Tucker. Tucker presented their work at a meeting before they had the theory thoroughly worked out; Kuhn told me that he felt that the Berkeley Symposium on Mathematical Statistics and Probability was an odd place to present their work but explained it by arguing that it provided an opportunity to get the result published fast [Kuhn, 1998]. Another of Merton's points also holds for Kuhn and Tucker. They do not say that Fritz John had worked on the same problem, but they give a reference to his paper; in an interview Kuhn told me that the reference to John was made in the proofreading stage when someone told them about his work [Kuhn, 1998].

Merton's hypothesis has not survived undisputed. It has been criticized by Don Patinkin, who points out especially two issues which he finds have not received proper attention: first, what is it actually that has been discovered; second, to what degree does the discovery form part of the central message of the scientist? [Patinkin, 1983, p. 306]. Patinkin claims that a lot of so-called multiple discoveries will turn out to be singletons if they are subject to

an analysis that takes these two issues seriously. Patinkin's own "central message" is that a scientist cannot be considered as having made a discovery unless this discovery forms part of the central message of the scientist. The question now is of course how to identify that message. Patinkin sets up the following criteria:

[...] the central message of a scientific work is announced by its presentation early in the work (and frequently in its title) and by repetition, either verbatim or modified in accordance with the circumstances. [Patinkin, 1983, p. 314]

Patinkin's reason for the importance of the central message is first the scientific reward system. In order for this system to be "fair" Patinkin finds that it is important that

[...] its rewards must go to the true discoverers: to those who brought about a cognitive change. [Patinkin, 1983, p. 316]

Second, in Patinkin's view the function of a scientific discoverer is to

stimulate a new research program on the part of colleagues in his field of inquiry, for only in that way can the full scientific potential of the discovery be efficiently exploited. [Patinkin, 1983, p. 316]

Using Patinkin's criteria for multiplicity the picture becomes a little more subtle. Using his method for uncovering the central message of the scientist and taking John's introduction at face value, it must be said that the Kuhn–Tucker theorem is indeed part of the central message in all three papers. The titles of both Karush's and John's paper indicate that the subject is optimization constrained by inequality conditions. The title of Kuhn and Tucker's paper is simply "Nonlinear Programming"; but at that time linear programming was well known in the circles to which Kuhn and Tucker belonged, so in 1950 this word could not refer to anything but finite-dimensional optimization subject to inequality constraints. So using only this criterion we must once again conclude that the Kuhn–Tucker theorem is a triple discovery.

This, however, is not very satisfactory, and if one is also considering Patinkin's reasons for putting such a high emphasis on the central message, namely that the purpose of scientific discoveries is to stimulate further research in the field, it becomes clear that only Kuhn and Tucker can be said to be the true discoverer of the Kuhn–Tucker theorem in nonlinear programming. Neither Karush's nor John's work stimulated any further research. Their work had no influence on the development of any discipline.

This however does not shed light on why the three different versions of the result were so differently received in the scientific community. I think that Patinkin's second essential point—what is it exactly that has been discovered—analyzed with respect to the different contexts the three papers originated in is a more fruitful approach to understanding this phenomenon.

6. THE SIGNIFICANCE OF THE CONTEXT

In the following I shall distinguish between a mathematical and a sociological context. I shall make a further division of the mathematical context into what I call the context of "pure mathematical content," which refers to analysis of mathematical results without taking into account the context of discovery or the mathematical environment in which they are presented, and the context of mathematical subdisciplines such as the theory of convexity or the calculus of variations.

Today mathematicians conceive of Karush's and Kuhn and Tucker's result as the same one—as the Kuhn–Tucker theorem—and of John's result as the Kuhn–Tucker theorem without the constraint qualification. The reason for this is an analysis of the results in relation to “pure mathematical content” based on the theoretical knowledge of today. In such an analysis mathematicians disregard the differences and focus solely on the similarities between the three results. They look at the theorems independent of the context within which they were developed.

An analysis which instead focuses on the *differences* in the three formulations of the theorem and takes the context of the subdisciplines into account can provide an explanation for the different influences on the mathematical development and the different reception in the mathematical community at the time of the three occurrences of the result. As was argued in Section 2 and 3, the reason the works of Karush and John were “overlooked” was not that their result did not form part of the central message of their work but rather because they were not central in relation to the internal mathematical—and maybe also sociological—context in which they appeared.

In order to understand the fame and recognition that almost immediately followed the work of Kuhn and Tucker one must also understand its origin in applied mathematics and the importance of the postwar organization of science support in the United States, both of which were consequences of World War II.

6.1. *The Social Context of Kuhn and Tucker's Work*

Introduction. Before World War II applied mathematics had a very bad reputation among professional mathematicians in the United States. From the beginning of the 20th century this country had witnessed a growing community of professional mathematicians. The kind of research that was pursued and the mathematical interests were mainly in what traditionally is called pure mathematics. Only a very few of the mathematicians working in academia were interested in applied mathematics. In the academic environment there was a hierarchy among mathematicians and generally, mathematicians working in applied areas were not ranked very high on the scale. The state of affair before World War II can be summarized by the words of Professor Prager, who gave the following description in 1972:

[...] their number [professional mathematicians interested in the applications] was extremely small. Moreover, with a few notable exceptions, they were not held in high esteem by their colleagues in pure mathematics, because of a widespread belief that you turned to applied mathematics if you found the going too hard in pure mathematics. [Prager, 1972, p. 1]

Some of this changed as a consequence of the Second World War.¹⁶ During this period a huge number of American scientists took part in the war effort. Some of them were hired directly by the armed forces, but most of them were organized through Office of Scientific Research and Development (OSRD) which was established in May 1941 under the leadership of Vannevar Bush and financed by Congress.¹⁷ But it was not until 1943 when the Applied Mathematics Panel (AMP) was founded as a subsection under OSRD that the mathematicians got involved in great numbers. The mathematicians organized through AMP worked on war related issues bounded by contracts. Thus, the AMP provided the mediating

¹⁶ See, e.g., [Dalmedico, 1996].

¹⁷ See, e.g., [Zachary, 1997].

link between the military and the mathematicians, who stayed in the universities and the industries. This activity during the war served to stimulate the involvement of professional mathematicians in solving applied problems, some of which were subsequently made the subject of theoretical mathematical research and development, and in some cases—as with nonlinear programming—became a new discipline in mathematics.

The Air Force programming problem. The work on what became linear programming began during the war. The main person was the mathematician George B. Dantzig who was hired in 1941 by the Air Force to work on the so-called “programming planning methods”—a tool in the Air Force for handling huge logistic planning. In a paper presented in December 1948 and published in 1951, Dantzig and Marshall K. Wood, who also worked for the Planning Research Division at the Air Force, gave the following definition:

Programming, or program planning, may be defined as the construction of a schedule of actions by means of which an economy, organization, or other complex of activities may move from one defined state to another, or from a defined state toward some specifically defined objective. [Dantzig and Wood, 1951, p. 15]

In this definition there is a possibility of moving towards a defined objective. This was not the case during the war. Here the focus was to make sure that the plan for activities was consistent:

The levels of various activities such as training, maintenance, supply, and combat had to be adjusted in such a manner as not to exceed the availability of various equipment items. Indeed, activities should be so carefully phased that the necessary amounts of these various equipment items were available when they were supposed to be available, so that the activity could take place. [Dantzig, 1951, p. 18]

Wood and Murray A. Geisler described the procedure behind the Air Force wartime program scheduling in [Geisler and Wood, 1951, p. 189]. They emphasized that “the major difficulty with this procedure was that it took too long. Even with the most careful scheduling, it took about seven months to complete the process.” [Geisler and Wood, 1951, p. 191].¹⁸

Postwar organization of science support. The end of the war also meant the end of OSRD. Bush’s organization was an emergency organization and it had been clear right from the beginning that OSRD would disappear with the war. There was a common concern that the scientists would just go back to their university duties after the war. There also was a strong belief that America had to be strong scientifically in order to be strong militarily. A lot of people were concerned about the further financing of science after the war, military related science as well as basic science.¹⁹

In his annual report to the President in 1945 the Secretary of the Navy, James V. Forrestal expressed the concern of the Navy on the further relationship between science and the military in peacetime. He stressed the need for an independent agency established by law and devoted to long-term, basic military research, securing its own funds from the Congress and responsive to, but not dominated by, the Army and the Navy. On the request of President Roosevelt, Vannevar Bush prepared a plan for the organization of postwar research and

¹⁸ For further readings on the origin of linear programming see the memoirs by Dantzig [Dantzig, 1963, 1968, 1982, 1988, 1991] and [Dorfman, 1984]. For historical accounts on the development of linear programming in the USSR see, e.g., [Brentjes; Brentjes, 1976b; Charnes and Cooper, 1961; Koopmans, 1961; Isbell and Marlow, 1961; Leifman; 1990; Kantorovich, 1939].

¹⁹ See [Rees, 1977a; Schweber, 1988; Dupree, 1986].

education. In his report “The Endless Frontier” Bush—like Forrestal—called for a governmental supply of money for independent research in the universities and industries. In contrast to Forrestal, who lobbied for basic *military* research, Bush wanted the government to supply basic research without necessary regards to the military. Bush’s main point was that

basic research leads to new knowledge. It provides scientific capital. It creates the fund from which the practical applications of knowledge must be drawn. . . . today it is truer than ever that basic research is the pacemaker of technological progress. . . . A nation which depends upon others for its new basic scientific knowledge will be slow in its industrial progress and weak in its competition in world trade, regardless of its mechanical skill. (Citation from [Schweber, 1988, p. 14])

This gives an impression of the spirit just after the war. There was a willingness to offer money on basic science and a philosophy that basic science was a necessity that automatically would lead to something that eventually could be applied for practical and therefore military purposes.

Bush wanted a National Science Foundation to support research in the universities and the industries but it took some time to establish such a foundation. In the mean time the Navy established the Office of Naval Research (ONR) the purpose of which was to continue the research practice established by the OSRD.²⁰

Towards linear programming. The different military sections also hired scientists on their own. George B. Dantzig was hired—again—by the Air Force where he—from 1946 until 1952—functioned as mathematical advisor for the U.S.A.F. headquarters. The assignment he was hired to work on was to

[. . .] develop some kind of analog devise which would accept, as input, equations of all types, basic data, and ground rules, and use these to generate as output a consistent Air Force plan. [Dantzig, 1988, p. 12]

Still, no objective was formulated: the programs were built on personal experience and a lot of ad hoc ground rules were issued by those in authority [Dantzig, 1968, p. 4]. This changed with the emergence of the computer, which had a profound influence on the work of Dantzig and his group. The idea of an “analog device” was rejected. Instead the work took a turn towards the development of what is now called linear programming. In the spring of 1947 the Air Force established project SCOOP (Scientific Computation of Optimum Programs) where Dantzig, Wood, and Geisler were the main figures. The purpose of this project was twofold: to build a mathematical model for the programming problem and the development and construction of computers.²¹

Wood and Geisler described the problems and the prospects in [1951, p. 194]:

These complexities [of the Air Force programming problem] have been spelled out to indicate a whole range of planning problems which, because of the present difficulties of computing alternative programs, receive little or no consideration. *So much time and effort is now devoted to working out the operational program that no attention can be given to the question whether there may not be some better program that is equally compatible with the given conditions.* It is perhaps too much to suppose that this difference between programs is as much as the difference between victory and defeat, but it is certainly a significant difference with respect to the tax dollar and the division of the total national product between military and civilian uses.

²⁰ For historical accounts on ONR see, e.g., [Old, 1961; Sapolsky, 1979; Schweber, 1988].

²¹ See [Brentjes, p. 177].

Consideration of the practical advantages to be gained by comparative programming, and particularly by the selection of “best” programs, leads to a requirement for a technique for handling all program elements simultaneously and for introducing the maximization process directly into the computation of programs. Such a technique is now in prospect. [Geisler and Wood, 1951, p. 194]

The possibilities of high-speed computers made it realistic to implement the notion of an objective in the programming problem, because there now seemed to be a possibility of computing alternative programs in order to choose the “best” one. Thus, there was a prospect of an effective decision tool useable not only for war-time activities but also “in planning for organizations or economic systems, where relationships are largely technological and decision making is centralized.” [Geisler and Wood, 1951, p. 189].

The model for the programming planning problem that the group ended up with was reflected in the following mathematical problem:

[...] the minimization of a linear form subject to linear equations and inequalities. [Dantzig, 1982, p. 44]

This is now known as a linear programming problem. Originally Dantzig called it “Programming in a Linear Structure.”

The involvement of John von Neumann. Dantzig was advised to make contact with the economist T. C. Koopmans, who had been working with a transportation model during the war, and with John von Neumann. Koopmans did a lot to introduce linear programming, especially to economists but it was the involvement of John von Neumann that was crucial for the further development.

John von Neumann was involved with almost everything related to mathematics that went on during the war. He was a member of many military scientific advisory boards and he also held a lot of military consulting jobs.²² In October 1947 Dantzig and von Neumann met in Princeton. This was the first time von Neumann heard about linear programming and—not surprisingly—he recognized the relationship to two-persons zero-sum games. In 1944 he had published the famous book “Theory of Games and Economic Behavior” together with the Austrian-American economist Oskar Morgenstern [von Neumann and Morgenstern, 1944]. Both models—the linear programming model and the game model—can be formulated as questions about linear inequalities. According to Dantzig, at this meeting von Neumann showed that a zero-sum two-person game can be reduced to a linear programming problem and conjectured the reverse relationship [Dantzig, 1982, 1988]. That the interest of von Neumann was caught can be seen from a note “Discussion of a maximum problem” that he wrote in November 1947 [von Neumann, 1947]. In this note he worked on a linear maximum problem subject to linear inequality constraints:

$$\begin{aligned} \max_x \quad & a \cdot x \\ \text{s. t.} \quad & x \geq 0 \\ & xA \leq \alpha \end{aligned}$$

where a , x are n -dimensional vectors, A is an n by m matrix, and α is an m -dimensional vector. He almost—almost because he used an incorrect version of Farkas’s lemma—proved

²² See [Ulam, 1958, p. 42].

that if there exists a finite maximum with a maximum point x_0 , satisfying the constraints $x_0 \geq 0, x_0 A \leq \alpha$, then there exists an m -dimensional vector ξ with $A\xi \geq a$ and $a \cdot x \geq \xi \cdot \alpha$ [von Neumann, 1947, p. 91]. This ξ will actually minimize the linear form $\xi \cdot \alpha$ and then von Neumann's result in this note can be—and has been—interpreted as the duality theorem for linear programming. But von Neumann did not state that conclusion, and he did not formulate the dual linear programming problem or the duality theorem in this note. He did introduce what are now known as dual variables even though he did not call them that. Whether von Neumann was fully aware of the relationship between the primal— x —and the dual— ξ —variables cannot be decided from the note.

Anyway, this note is the first sign of developing linear programming into a theory. Von Neumann had an enormous influence in speeding up this process. First, he was a member of the committee²³ set up by the National Academy of Sciences to act as advisers to the Mathematics Branch of ONR on questions connected with projects in pure mathematics; second, he had considerable influence in promoting game theory as a major research area at the RAND in the immediate postwar period.²⁴

The mathematics division of ONR. Kuhn and Tucker's work, which was a direct consequence of these circumstances, took place under contract with the mathematics division of ONR. Mina Rees who had served as technical assistant to Warren Weaver—the leader of AMP—during the war, was asked by ONR just after the war to set up a mathematics program. Even though from the outset she had expressed her doubt about the success of such a program (she did not think that mathematicians would let the military finance their peacetime research), she took the position as head of the mathematics branch because she found it extremely important for the further development of mathematics in the United States to be actively involved in the ONR program [Rees, 1977b; Albers and Alexanderson, 1985]. As such she was a very influential person in the mathematics community in the post-war period. The program she prepared for the ONR was one she had discussed with most of the leading mathematicians and mathematics departments in the country. She was very much concerned that the ONR mathematics program should reflect what mathematicians thought would help mathematics. The question was of course whether the Navy would support basic research and especially research in pure mathematics without any relevance for the Navy. Rees was very concerned with this, for she wanted the program to strengthen the mathematical research in the USA and not to fragment the field [Rees, 1977a].

In 1948 the mathematics department of the ONR had been functioning for a little over a year and Rees had a note in the Bulletin of the American Mathematical Society where she announced “the philosophy which has determined the mathematical research projects which ONR is sponsoring.” She stated that:

The Office of Naval Research is committed primarily to the support of fundamental research in the sciences, as contrasted with development, or with applications of known scientific results It is natural, however, that the most obvious types of mathematical research which would seem to warrant Navy support would be research in applied directions. [Rees, 1948, p. 1]

The state of affairs when it came to money was that 4/5 of the annual expenditure went to research in applied mathematics, mathematical statistics, numerical analysis, and computing

²³ Consisting of John von Neumann, G. C. Evans, H. M. Morse, H. M. Stone, H. Whiney, and O. Zariski.

²⁴ See [Mirowski, 1991; Leonard, 1992, 1995].

devices. But the number of contracts with theoretical objectives stood for more than 1/3 of the entire group [Rees, 1948]. She emphasized that basic research in mathematics proper was deemed important and was receiving funding from ONR.

The support of ONR. The prospects of the programming planning methods that the Air Force group was developing were—as explained above—considerable. Rees remembered it like this in 1977:

when, in the late 1940's the staff of our office became aware that some mathematical results obtained by George Dantzig, . . . could be used by the Navy to reduce the burdensome costs of their logistics operations, the possibilities were pointed out to the Deputy Chief of Naval Operations for Logistics. His enthusiasm for the possibilities presented by these results was so great that he called together all those senior officers who had anything to do with logistics, as well as their civilian counterparts, to hear what we always referred to as a "presentation." The outcome of this meeting was the establishment in the Office of Naval Research of a separate Logistics Branch with a separate research program. This has proved to be a most successful activity of the Mathematics Division of ONR, both in its usefulness to the Navy, and in its impact on industry and the universities. [Rees, 1977a, p. 111]

In the spring of 1948 Dantzig went to Princeton on behalf of ONR to meet with John von Neumann in order to discuss the possibilities for a university-based project on linear programming and its relationship to game theory financed by ONR [Albers and Alexanderson, 1985, pp. 342–343]. During this visit Dantzig was introduced to Tucker, who gave him a ride to the train station. During this short car trip Dantzig gave Tucker a brief introduction to the linear programming problem. Tucker made a remark about a possible connection to Kirchoff–Maxwell's law of electric networks; because of it Tucker was contacted by the ONR a few days later and asked if he would set up such a mathematics project [Albers and Alexanderson, 1985, pp. 342–343].

Until this moment Tucker had been absorbed in research in topology. He agreed to become principal investigator, and this completely changed his research direction. The same happened for Kuhn, who at the time was finishing a Ph.D. project on group theory. In the summer of 1948 Kuhn went to Tucker to ask for summer employment because he needed the additional income. Tucker hired him, together with David Gale, who was also a graduate student, to work with him on the ONR project [Kuhn, 1998]. The three of them presented the results of their work on the project at the first conference on linear programming, which took place in Chicago in June 1949 [Gale *et al.*, 1951]. The most prominent among their results was the duality theorem for linear programming. After that Kuhn and Tucker became committed to the project; the duality theorem caught their attention as interesting from a mathematical point of view. From then on, proceeding according to the "inner" rules for research in pure mathematics, they tried to extend this result to more general cases, which resulted in the "nonlinear programming" paper and the Kuhn–Tucker theorem. This work was also sponsored by the ONR, which continued to support Tucker's project until 1972, when the National Science Foundation took over.

Another social factor also related to the military was the development of operations research (OR) during the war and the establishment of OR as a scientific discipline at the universities after the war.²⁵ ONR also played a major role in this process. Fred Rigby, the head of the logistics program of ONR, later described its significance:

²⁵ For historical accounts on OR in the United States see, e.g., [Rau; Fortun and Schweber, 1993].

We did indeed influence the introduction of operations research into business schools. The subdiscipline called management science is our invention, in quite a real sense. That is, we and our contract researchers recognized its potentials, planned its early growth, and, as it turned out, set the dominant pattern in which it has developed. (Quoted in [Rees, 1977a, p. 111]).

After the war several people—especially Philip Morse from M.I.T.—who had been involved in the operations research groups during the war, with the help of ONR and the National Research Council, introduced operations research into the universities. Already from 1948 Morse had two courses running at M.I.T. [Morse, 1956]. From 1952 on Johns Hopkins had a program in operations research and from 1954 on it was possible to take a Ph.D. in the field [Roy, 1956].

Morse was a key figure in the shaping of operations research as an academic discipline in the United States. From the beginning he emphasized the importance of the newly developed linear programming for operations research. His main point was that basic research in mathematical programming was vital for operations research [Morse, 1955, p. 383].

From the journals and from the proceedings of the international conferences in operations research which began in 1957 it can be seen that mathematical programming was quite well represented. But all the time there was a continuously running debate about what OR actually was and not everybody held the opinions of Morse. In 1956 W. N. Jessop warned against “the placing of emphasis on mathematical methods and on highly abstract treatments of general situations” in the journal of the American Operations Research Society [Jessop, 1956]. When it came to linear programming Jessop also held the opinion that there was too much focus on developing “a subject so delightful to the pure mathematician that many papers appear to have had their origin in sheer exuberance unsullied by any thought of a factual situation” [Jessop, 1956, p. 51].

Linear programming was immediately incorporated into the toolbox of OR, which meant that OR also stood ready to provide a “home” for nonlinear programming as soon as it was developed. In this way it can be seen that the Office of Naval Research had an enormous influence in creating a scientific community of people doing linear programming, and in this community it was almost inevitable that the nonlinear programming paper of Kuhn and Tucker would give rise to the new research field of nonlinear programming.

During the first two decades of its existence mathematical programming established itself as a discipline with conferences, monographs, and textbooks. The question of a journal and a society for mathematical programming was discussed on and off and in 1971 the first journal, “Mathematical Programming,” was founded, and two years later came the “Mathematical Programming Society.”

7. CLOSING REMARKS

The Kuhn–Tucker theorem shows that a mathematical theorem in itself, its “pure mathematical content,” does not always decide whether it will stimulate further research or not. Social contexts can also play an essential role. Even though the three results today are viewed as the same theorem, they were in practice very different. The significance of a result and its potential for stimulating further research in its area are determined by the mathematical—and sometimes also the social—context within which it was developed. The Kuhn–Tucker theorem was an important result in the mathematical discipline in which Kuhn and Tucker were working, a discipline which also received huge financial

support. This was not the case in the subdisciplines where the papers of Karush and John appeared.

The fact that Karush, John and Kuhn and Tucker all receive credit for the theorem in the scientific community of nonlinear programming is due to the influence of “third parties”—a notion introduced by Susan Cozzens. In her book “Social Control and Multiple Discoveries in Science: The Opiate Receptor Case,” she focuses on how discoveries later become established as multiples [Cozzens, 1989]. She points out that it is often due to an “after-the-fact process” where the case is settled by influence from third parties, that is, members of the scientific community who are not directly involved in the discovery. Through later references and acknowledgement the third parties establish the discoveries as multiple. The quotations I showed in the section on Karush show that this also was the case for the establishment of the Kuhn–Tucker theorem as a multiple discovery, even though Kuhn himself, that is, one of the involved scientists, here played a major role in the recognition of predecessors.

REFERENCES

- Ader, O. B. 1938. An affine invariant of convex regions. *Duke Mathematics Journal* **4**, 291–299.
- Albers, J. D., and Alexanderson, G. L. 1985. *Mathematical People, Profiles and Interviews*. Boston: Birkhäuser.
- Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. 1979. *Nonlinear Programming, Theory and Algorithms*. New York: Wiley.
- Bazaraa, M. S., Sherali, H. D., and Shetty, C. M. 1993. *Nonlinear Programming, Theory and Algorithms*. 2nd ed. New York: Wiley.
- Behrend, F. 1937. Über einige Affinvarianten konvexer Bereiche. *Mathematische Annalen* **113**, 713–747.
- Behrend, F. 1938. Über die kleinste umbeschriebene und die grösste einbeschriebene Ellipse eines konvexen Bereichs. *Mathematische Annalen* **115**, 379–411.
- Bellman, R. 1975. Letter to Harold W. Kuhn, dated 11 February, 1975. [Unpublished, copy in possession of the author]
- Bliss, G. A. 1938. Normality and abnormality in the calculus of variations. *Transactions of the American Mathematical Society* **43**, 365–376.
- Brentjes, S. 1975. *Untersuchungen zur Geschichte der linearen Optimierung (LO) von ihren Anfängen bis zur Konstituierung als selbständige mathematische Theorie—Eine Studie zum Problem der Entstehung mathematischer Disziplinen im 20. Jahrhundert*. Ph.D. dissertation, Leipzig. [Unpublished]
- Brentjes, S. 1976a. Bemerkungen zum Beitrag von Julius Farkas zur Theorie der linearen Optimierung. *NTM-Schriftenreihe zur Geschichte der Naturwissenschaften, Technik und Medizin* **13**, 21–23.
- Brentjes, S. 1976b. Der Beitrag der sowjetischen Wissenschaftler zur Entwicklungen der Theorie der linearen Optimierung. *NTM-Schriftenreihe zur Geschichte der Naturwissenschaften, Technik und Medizin* **13**, 105–110.
- Browder, F. E. 1989. The stone age of mathematics on the midway. In *A Century of Mathematics in America, Part II*, Peter Duren, Ed., History of Mathematics, Vol. 2, pp. 191–193. Amer. Math. Soc., Providence, RI.
- Charnes, A., and Cooper, W. W. 1961. On some works of Kantorovich, Koopmans and Others. *Management Science* **8**, 246–263.
- Cozzens, S. E. 1989. *Social Control and Multiple Discovery in Science: The Opiate Receptor Case*. Albany, NY: State Univ. of New York Press.
- Dalmedico, A. D. 1996. L'essor des mathématiques appliquées aux États-Unis: L'impact de la seconde guerre mondiale. *Revue d'histoire des mathématiques* **2**, 149–213.
- Dantzig, G. B. 1951. Linear programming. In *Problems for the Numerical Analysis of the Future*, Institute for Numerical Analysis, Eds., Applied Mathematics Series, Vol. 15, pp. 18–21. Washington, D.C.: National Bureau of Standards.
- Dantzig, G. B. 1963. *Linear Programming and Extensions*. Princeton, NJ: Princeton Univ. Press.

- Dantzig, G. B. 1968. Linear programming and its progeny. In *Applications of Mathematical Programming Techniques*, E. M. L. Beale, Ed., pp. 3–15. London: English Universities Press.
- Dantzig, G. B. 1982. Reminiscences about the origins of linear programming. *Operations Research Letters* **1**, 43–48.
- Dantzig, G. B. 1988. Impact of linear programming on computer development. *OR/MS Today*, 12–17.
- Dantzig, G. B. 1991. Linear programming. In *History of Mathematical Programming, A Collection of Personal Reminiscences*, Jan Karel Lenstra, Alexander H. G. Rinnooy Kan, and Alexander Schrijver, Eds., pp. 19–31. Amsterdam: North-Holland.
- Dantzig, G. B., and Wood, M. K. 1951. The programming of interdependent activities: General discussion. In *Activity Analysis of Production and Allocation*, Tjalling C. Koopmans, Ed., Cowles Commission Monographs, Vol. 13, pp. 15–18. New York: Wiley.
- Dines, L. L. 1936. Convex extension and linear inequalities. *Bulletin of the American Mathematical Society*, **42**, 353–365.
- Dorfman, R. 1984. The discovery of linear programming. *Annals of the History of Computing* **6**, 283–295.
- Dupree, A. H. 1986. National security and the post-war science establishment in the United States. *Nature* **323**, 213–216.
- Duren, A. L. 1976. Graduate student at Chicago in the twenties. *The American Mathematical Monthly* **83**, 243–248.
- Duren, P. (Ed.). 1989. *A century of Mathematics in America, Part II*, History of Mathematics, Vol. 2. Providence, RI: Amer. Math. Soc.
- Farkas, G. (J.) 1895. Über die Anwendungen des mechanischen Princips von Fourier. *Mathematische und Naturwissenschaftliche Berichte aus Ungarn* **12**, 263–281.
- Farkas, G. (J.) 1897. Die algebraischen Grundlagen der Anwendung des Fourier'sschen Princips in der Mechanik. *Mathematische und Naturwissenschaftliche Berichte aus Ungarn* **15**, 25–40.
- Farkas, G. (J.) 1899. Die algebraische Grundlage der Anwendungen des mechanischen Princips von Fourier. *Mathematische und Naturwissenschaftliche Berichte aus Ungarn* **16**, 154–157.
- Farkas, G. (J.) 1901. Theorie der einfachen Ungleichungen. *Journal für die reine und angewandte Mathematik* **124**, 1–27.
- Fenchel, W. 1953. *Convex Cones, Sets, and Functions*. Lecture Notes, Department of Mathematics, Princeton University, 1953.
- Fortun, M., and Schweber, S. S. 1993. Scientists and the legacy of World War II: The case of operations research (OR). *Social Studies of Science* **23**, 595–642.
- Fourier, J. B. J. 1798. Mémoire sur la statique, contenant la démonstration du principe de vitesses virtuelles, et la théorie des momens. In *Oeuvres*, Gaston Darboux, Ed., Vol. 2, pp. 475–521. Paris: Gauthier–Villars.
- Franksen, O. I. 1985a. Irreversibility by inequality constraints. I. On Fourier's inequality. *Systems Analysis, Modelling, Simulation* **2**, 137–149.
- Franksen, O. I. 1985b. Irreversibility by inequality constraints. II. The second law of thermodynamics. *Systems Analysis, Modelling, Simulation* **3**, 251–273.
- Franksen, O. I. 1985c. Irreversibility by inequality constraints. III. Towards mathematical programming. *Systems Analysis, Modelling, Simulation* **4**, 337–359.
- Fraser, C. G. 1992. Isoperimetric problems in the variational calculus of Euler and Lagrange. *Historia Mathematica* **19**, 4–23.
- Fraser, C. G. 1994. The origin of Euler's variational calculus. *Archive for History of Exact Sciences* **47**, 103–141.
- Gale, D., Kuhn, H. W., and Tucker, A. W. 1951. Linear programming and the theory of games. In *Activity Analysis of Production and Allocation*, Tjalling C. Koopmans, Ed., Cowles Commission Monographs, Vol. 13, pp. 317–329. New York: Wiley.
- Gårding, L. 1985. Foreword. In *Fritz John, Collected Papers*, J. Moser, Ed., Vol. 1. Boston: Birkhäuser.
- Geisler, M. A., and Wood, M. K. 1951. Development of dynamic models for program planning. In *Activity Analysis of Production and Allocation*, Tjalling C. Koopmans, Ed., Cowles Commission Monographs, Vol. 13, pp. 189–215. New York: Wiley.

- Grattan-Guinness, I. 1970. Joseph Fourier's anticipation of linear programming, *Operational Research Quarterly* **21**, 361–364.
- Grattan-Guinness, I. 1994. "A new type of question": On the prehistory of linear and non-linear programming, 1770–1940. In *The History of Modern Mathematics. III*, Eberhard Knobloch and David E. Rowe, Eds., pp. 43–89. Boston: Academic Press.
- Hourshell, D. 1997. The Cold War, RAND, and the generation of knowledge, 1946–1962. *Historical Studies in the Physical and Biological Sciences* **27**, 237–267.
- Isbell, J. R., and Marlow, W. H. 1961. On an industrial programming problem of Kantorovich. *Management Science* **8**, 13–17.
- Jessop, W. N. 1956. Operational research methods: What are they? *Operational Research Quarterly* **7**(2), 49–58.
- John, F. 1936. Moments of inertia of convex regions. *Duke Mathematics Journal* **2**, 447–452.
- John, F. 1942. An inequality for convex bodies. *University of Kentucky Research Club Bulletin* **8**, 8–11.
- John, F. 1948. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays, Presented to R. Courant on his 60th Birthday January 8, 1948*, pp. 187–204. New York: Interscience.
- Kantorovich, L. V. 1939. *Mathematical Methods of Organizing and Planning Production*. Leningrad: Leningrad University. [In Russian; an English translation can be found in *Management Science* **6**, 366–422, 1960.]
- Karush, W. 1939. *Minima of Functions of Several Variables with Inequalities as Side Conditions*. Dissertation, Department of Mathematics, University of Chicago.
- Karush, W. 1975. Letter to Harold W. Kuhn, dated 10 February, 1975. [Unpublished; copy in possession of the author]
- Kjeldsen, T. H. 1999. *A Contextualised Mathematico-Historical Analysis of Nonlinear Programming: Development and Multiple Discovery*. IMFUFA, Text 372, Roskilde University. [In Danish]
- Koopmans, T. C. 1961. On the evaluation of Kantorovich's work of 1939. *Management Science* **8**, 264–265.
- Kuhn, H. W. 1952. Subgroup theorems for groups presented by generators and relations. *Annals of Mathematics* **56**, 22–46.
- Kuhn, H. W. 1975a. Letter to William Karush, dated 4 February, 1975. [Unpublished; copy in possession of the author]
- Kuhn, H. W. 1975b. Letter to William Karush, dated 21 February, 1975. [Unpublished; copy in possession of the author]
- Kuhn, H. W. 1976. Nonlinear programming: A historical view. *SIAM-AMS Proceedings* **9**, 1–26.
- Kuhn, H. W. 1986. *Curriculum Vitae*. [Unpublished; copy in possession of the author]
- Kuhn, H. W., Tucker, A. W. 1995. Some reminiscences, prepared with the assistance of Alan and Tom Tucker. Based on an oration given by Harold W. Kuhn. *Notices of the AMS*, October 1995.
- Kuhn, H. W. 1998. Personal interview. Princeton University, Princeton, NJ, 23 April, 1998.
- Kuhn, H. W., and Tucker, A. W. 1950. Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman, Ed., pp. 481–492. Berkeley.
- Lagrange, J. L. 1788. *Mécanique analytique*, 1788. Reprint ed., 1965, Paris: Albert Blanchard.
- Leifman, L. J. (Ed.) 1990. *Functional Analysis, Optimization, and Mathematical Economics: A Collection of Papers Dedicated to the Memory of Leonid Vital'evich Kantorovich*. Oxford: Oxford Univ. Press.
- Leonard, R. J. 1992. Creating a context for game theory. In *Towards a History of Game Theory*, E. Roy Weintraub, Ed., pp. 29–76. Durham/London: Duke University Press.
- Leonard, R. J. 1995. From parlor games to social science: Von Neumann, Morgenstern, and the creation of game theory 1928–1944. *Journal of Economic Literature* **33**, 730–761.
- Luenberger, D. G. 1973. *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley.
- MacLane, S. 1989. Mathematics at the University of Chicago, A brief history. In *A Century of Mathematics in America, part II*, Peter Duren, Ed., American Mathematical Society, History of Mathematics, Vol. 2, pp. 127–154. Providence, RI: Amer. Math. Soc.
- Merton, R. K. 1973. Singletons and multiples in science. In *The Sociology of Science, Theoretical and Empirical Investigations*, Robert K. Merton, Ed., pp. 343–382. Chicago: Univ. of Chicago Press.

- Mirowski, P. 1991. When games grow deadly serious: The military influence on the evolution of game theory. In *Economics and National Security*, D. G. Goodwin, Ed., Annual Supplement to History of Political Economy, Vol. 23, pp. 227–255. Durham/London: Duke Univ. Press.
- Morse, P. M. 1955. Where is the new blood? *Journal of the Operations Research Society of America* **3**, 383–387.
- Morse, P. M. 1956. Training in operations research at the Massachusetts Institute of Technology. *Operations Research* **4**, 733–735.
- Moser, J. (Ed.) 1985. *Fritz John, Collected Papers*, Vol. 1, Boston: Birkhäuser.
- Old, B. S. 1961. The evolution of the Office of Naval Research. *Physics Today* **14**, 30–35.
- Ostrogradsky, M. V. 1838. Considérations générales sur les moments des forces, *Mémoires de l'Académie impériale des sciences de St.-Pétersbourg, Sixième Série* **1**, 129–150. [Read 1834]
- Parshall, K. H., and Rowe, D. E. 1994. *The Emergence of the American Mathematical Research Community, 1876–1900: J. J. Sylvester, Felix Klein, and E. H. Moore*, History of Mathematics, Vol. 8. Providence, RI: Amer. Math. Soc.
- Patinkin, D. 1983. Multiple discoveries and the central message. *American Journal of Sociology* **89**, 306–323.
- Peressini, A. L., Sullivan, F. E., and Uhl, J. J. 1988. *The Mathematics of Nonlinear Programming*. New York: Springer-Verlag.
- Prager, W. 1972. Introductory Remarks. *Quarterly of Applied Mathematics* **30**, 1–9.
- Prékopa, A. 1980. On the development of optimization theory, *American Mathematical Monthly* **87**, 527–542.
- Rau, E. The adoption of operations research in the United States during World War II. In *Systems, Experts, and Computers*, A. C. Hughes and Tom P. Hughes, Eds., Dibner Series, Jed Z. Buchwald, Ed., Cambridge, MA: MIT Press (to appear).
- Rees, M. S. 1948. The mathematics program of the Office of Naval Research. *Bulletin of the American Mathematical Society*, 1–5.
- Rees, M. S. 1977a. Mathematics and the government: The post-war years as augury of the future. In *The Bicentennial Tribute to American Mathematics, 1776–1976*, D. Tarwater, Ed., pp. 101–116. Buffalo, NY: The Mathematical Association of America.
- Rees, M. S. 1977b. Early years of the mathematics program at ONR. *Naval Research Reviews* **30**, 22–29.
- Reid, C. 1976. *Courant in Göttingen and New York. The Story of an Improbable Mathematician*. New York: Springer-Verlag.
- Roy, R. H. 1956. Operations research at the Johns Hopkins University. *Operations Research* **4**, 735–738.
- Sapolsky, H. M. 1979. Academic science and the military: The years since the Second World War. In *The Sciences in the American Context: New Perspectives*, Reingold, Ed., pp. 379–399. Washington, D.C.: Smithsonian Institution Press.
- Schweber, S. S. 1988. The mutual embrace of science and the military: ONR and the growth of physics in the United States after World War II. In *Science, Technology and the Military*, E. Mendelsohn, M. R. Smith, and P. Weingart, Eds., pp. 3–45. Dordrecht: Kluwer Academic.
- Smith, B. 1969. *The RAND Corporation, Case Study of a Nonprofit Advisory Corporation*. Cambridge, MA: Harvard Univ. Press.
- Stokes, R. W. 1931. A geometric theory of linear inequalities. *Transactions of the American Mathematical Society* **33**, 782–805.
- Stone, M. H. 1989. Reminiscences of mathematics at Chicago. In *A Century of Mathematics in America, Part II*, Peter Duren, Ed., History of Mathematics, Vol. 2, pp. 183–190. Providence, RI: Amer. Math. Soc.
- Takyama, A. 1974. *Mathematical Economics*. Hinsdale, IL: Dryden.
- Tucker, A. W. 1980. *Curriculum Vitae*. [Unpublished; copy in possession of the author]
- Ulam, S. 1958. John von Neumann, 1903–1957. *Bulletin of the American Mathematical Society* **64**, 1–49.
- von Neumann, J. 1947. Discussion of a maximum problem. In *John von Neumann Collected Works*, A. H., Taub, Ed., Vol. 6, pp. 89–95. Taub, Oxford: Pergamon.
- von Neumann, J., and Morgenstern, O. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press.
- Zachary, P. G. 1997. *Endless Frontier: Vannevar Bush, Engineer of the American Century*. New York: The Free Press.