Giorgio Giorgi
Tinne Hoff Kjeldsen
Editors

Traces and Emergence of Nonlinear Programming

Birkhäuser

Birkhäuser

Giorgio Giorgi • Tinne Hoff Kjeldsen
Editors

Traces and Emergence of Nonlinear Programming

Birkhäuser

Editors
Giorgio Giorgi
Dipto. Ricerche Aziendali
Sez. Matematica Generale
Universita Pavia
Pavia, Italy

Tinne Hoff Kjeldsen
IMFUFA, NSM
Roskilde University
Roskilde, Denmark

ISBN 978-3-0348-0438-7 ISBN 978-3-0348-0439-4 (eBook)
DOI 10.1007/978-3-0348-0439-4
Springer Basel Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013950039

Printed on acid-free paper

Springer Basel is part of Springer Science+Business Media (www.birkhauser-science.com)

To the beloved memory of my mother Olga
Giorgio Giorgi

Preface

The history of mathematics has undergone, in the last decades, a very fast evolution process. At first cultivated by a narrow group of researchers, from the second half of the past century it has attracted the interest of a growing number of scholars, several of them being also researchers in various areas of mathematics. A subsequent period has been marked by a growing attention towards the specific methods of the historical research, with the formation of a professional community of historians of mathematics whose scientific production has often reached an important level, both from a quantitative and from a qualitative point of view. In the present period the best products of the historians of mathematics are tending to emphasize the *cultural* dimension of their subjects of study, that is the cultural dimension of mathematics itself, through the historical approach.

Among all the historical contributions, the studies on the evolution of the so-called applied mathematics are very numerous, even if perhaps more recent than the studies on the history of the so-called pure mathematics. Nevertheless, the history of mathematical optimization can cover a long period, as optimization problems date back to antiquity: the human race, for his daily requirements of surviving has always met implicit problems of maximum or minimum. This can be summed up in the words of Leonhard Euler, who wrote: “For since the fabric of the universe is the most perfect and the work of a most wise Creator, nothing takes place in the universe in which some rule of maximum or minimum does not appear.” There are also some references to optimization problems in literature and poetry; a famous example is the so-called *Dido’s Problem*, contained in Virgil’s “Aeneid,” and which may be considered as the first literary appearance of the *isoperimetric problem* of the Calculus of Variations.

The words “minimum” and “maximum” are typical of several problems, not only of mathematics but also of physics, chemistry, engineering, economics, etc. They are central in some very general principles, such as the “principle of least action” of Maupertuis. By the mid-seventeenth century, the invention of the “Calculus” has shifted the barycenter of optimization problems from a geometrical field to a predominantly analytical field. All the analytical results subsequently obtained (from Fermat onwards) have given birth to a mathematical theory, more and more refined, both from the point of view of its theoretical formulations and from the point of view of the solution techniques.

The history of the so-called *mathematical programming* (or also *nonlinear programming,* where the word "nonlinear" means "not necessarily linear") is on the contrary more recent. If, by mathematical programming or nonlinear programming, we intend those static optimization problems (usually defined in finite-dimensional spaces), where a function of n variables ("objective function") is to be minimized or maximized, subject to a certain number of constraints, *not necessarily given as equality constraints,* we can say that its history began, roughly speaking, in the twentieth century and in particular, from the years of the Second World War. Obviously, for mathematical programming problems there is also a sort of "prehistory," with individual contributions of high scientific level.

At present mathematical programming problems (and in general optimization problems) are pervasive in several sciences, such as economics, engineering, operations research, chemistry, physics, biology, social sciences, and management sciences. Moreover, they have many important applications in these areas and promise to have even wider usage in the future.

The present book collects some papers that are, in our opinion, the first basic stepping stones in nonlinear programming (see the list in the index of the book and at the end of the introductory chapter). Here we have excluded those papers exclusively concerned with linear programming, such as the contributions of G. B. Dantzig and L. V. Kantorovich. Obviously, our choice is subjective and does not claim to be complete; however, we believe that some contributions could not have been neglected: this is the case for the Master Thesis of W. Karush (1939), here published in its full length, the papers of F. John (1948) and of H. W. Kuhn and A. W. Tucker (1951). Some basic papers of K. J. Arrow, L. Hurwicz, and H. Uzawa have been included in our list. The paper of Arrow and Hurwicz of 1956 treated, with a more general approach, the equivalence between the usual nonlinear programming problem and the saddle-point problem of the Lagrangian function, a question previously analyzed by Kuhn and Tucker in their 1951 paper. Arrow and Hurwicz (1951) devised a gradient technique for approximating saddle-points and constrained optima. This is one of the earlier gradient techniques offered for solving the constrained nonlinear optimization problem. The paper of Arrow, Hurwicz, and Uzawa (1961) on constraint qualifications is perhaps the first contribution concerned with the important problem of locating "regular" constraints and of establishing the relationships between the various constraint qualifications proposed, a problem which has not lost its importance after 50 years. Equally important is the paper of Hurwicz of 1958 (but written before), one of the first treatments of nonlinear programming problems, both for the scalar and for the vector case, in topological spaces. The papers of L. L. Pennisi (1953) and G. P. McCormick (1967) are important for their treatment of second-order optimality conditions for a general nonlinear programming problem. The papers of W. Fenchel (1949), M. L. Slater (1950), and Uzawa (1958) are important for the case of convex nonlinear programming problems; in fact, convex analysis is in itself a basic tool for the development of mathematical programming, as the works of W. Fenchel,

J. J. Moreau, and R. T. Rockafellar have shown. The paper of 1963 by R. W. Cottle is perhaps the first "translation" of the original Fritz John optimality conditions into the setting of the usual nonlinear programming problems. The papers of Bliss (1938) and Valentine (1937) have been included, due to their historical relevance in connection with the birth of optimality conditions for a nonlinear programming problem.

In the introductory chapter, entitled *A Historical View of Nonlinear Programming: Traces and Emergence,* we have attempted to recapitulate the basic facts in the history of mathematical programming: "classical" mathematical programming (i.e., with only equality constraints), linear programming, and nonlinear programming. We have also added an Appendix, containing three further contributions, useful to illuminate further the growth process of optimization theory.

The main purpose of this book is to offer researchers direct access to the original sources and classics in nonlinear programming, together with an examination of the historical context regarding the emergence and development of this field of research. We hope that this collection will be useful and stimulating for all those who are interested in deepening their knowledge of the emergence and first developments of nonlinear programming and in general of mathematical optimization.

Pavia, Italy — Giorgio Giorgi
Roskilde, Denmark — Tinne Hoff Kjeldsen

Acknowledgements

The editors thank the following organizations for their permissions to reproduce the various articles and extracts in this collection.

– The RAND Corporation—Santa Monica, CA—USA, for

K. J. Arrow and L. Hurwicz, A gradient method for approximating saddle points and constrained maxima, RAND Corporation, paper P-223, June 1951.

R. W. Cottle, A theorem of Fritz John in mathematical programming, RAND Corporation Memorandum RM-3858-PR, 1963.

– The University of California Press—Berkeley—Berkeley, CA—USA, for

K. J. Arrow and L. Hurwicz, Reduction of constrained maxima to saddle-point problems; in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Vol. 5, 1955, 1–20.

H. W. Kuhn and A. W. Tucker, Nonlinear programming; in J. Neyman (Ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1951, 481–492.

– The American Mathematical Society—Providence, RI—USA, for

G. Bliss, Normality and abnormality in the calculus of variations, Trans. Amer. Math. Soc., 43, 1938, 365–376.

L. L. Pennisi, An indirect sufficiency proof for the problem of Lagrange with differential inequalities as added side conditions, Trans. Amer. Math. Soc., 74, 1953, 177–198.

H. W. Kuhn, Nonlinear programming: a historical view; in R. W. Cottle and C. E. Lemke (Eds.), Nonlinear Programming, SIAM-AMS Proceedings, Vol. XI, American Mathematical Society, Providence, 1976, 1–26.

– The Stanford University Press—Palo Alto, CA—USA, for

L. Hurwicz, Programming in linear spaces; in K. J. Arrow, L. Hurwicz and H. Uzawa (Eds.), Studies in Linear and Nonlinear Programming, Stanford University Press, Stanford, 1958, 38–102.

H. Uzawa, The Kuhn-Tucker theorem in concave programming; in K. J. Arrow, L. Hurwicz and H. Uzawa (Eds.), Studies in Linear and Nonlinear programming, Stanford University Press, Stanford, 1958, 32–37.

– The Society for Industrial and Applied Mathematics—Philadelphia, PA—USA, for

G. P. McCormick, Second-order conditions for constrained minima, SIAM J. on Applied Mathematics, 15, 1967, 641–652.

– The Cowles Foundation—New Haven, CT—USA, for

M. Slater, Lagrange multipliers revisited, Cowles Commission Discussion Paper 403, 1950.

– Birkhäuser (Division of Springer Basel AG)—Basel, Switzerland, for

F. John, Extremum problems with inequalities as subsidiary conditions; in Collected Papers, Vol. 2 (J. Moser Ed.), Birkhäuser, 1985, 543–560. Originally published in K. O. Friedrichs, O. E. Neugebauer and J. J. Stoker (Eds.), Studies and Essays presented to R. Courant on his 60th Birthday, Wiley Interscience, New York, 1948, 187–204.

– The Canadian Mathematical Society—Ottawa, Canada, for

W. Fenchel, On Conjugate Convex Functions, Canadian Journal of Mathematics, 1, 1949, 73–77. Reprinted with the permission of the Canadian Mathematical Society, this article was originally published in the Canadian Journal of Mathematics by Fenchel, W., *On conjugate convex functions*, vol. 1, no. 1, 1949, p. 73–77.

– University of Chicago Press—Chicago, USA, for

F. A. Valentine, The Problem of Lagrange with Differential Inequalities as added Side Conditions. In Contributions to the Calculus of Variations (1933–1937), University of Chicago Press, Chicago, 1937, 407–448.

– The Royal Society of Canada—Ottawa, Canada, for

L. L. Dines and N. H. McCoy, On Linear Inequalities. Trans. Roy. Soc. Canada, 27, 1933, 37–70.

The editors thank Barbara Karush and Larry Karush for their permissions to reproduce the Master Thesis of their father:

W. Karush: Minima of Functions of Several Variables with Inequalities as Side Conditions. MSc Thesis, Department of Mathematics, University of Chicago, 1939.

Finally, we thank Dr. Anna Mätzener of Birkhäuser (Basel) for her support in the production of this anthology. Any error or inaccuracy that may appear in the final version of the book is of our sole responsibility.

Pavia, Italy Giorgio Giorgi
Roskilde, Denmark Tinne Hoff Kjeldsen

Contents

A Historical View of Nonlinear Programming: Traces and Emergence

Giorgio Giorgi and Tinne Hoff Kjeldsen

The historical view we propose in this introductory chapter will point out some of the technical difficulties of mathematical problems related to nonlinear programming and some features of the economic and social context (also military) that favored its rootedness in the years of the Second World War and the years immediately following the war. We recall some of the main definitions and basic results of mathematical programming and shortly address the "prehistory" of nonlinear programming. The main part of the chapter deals with the first ideas and developments of linear programming, first in the USSR and then in the USA and with the fundamental researches of W. Karush, Fritz John, H.W. Kuhn and A.W. Tucker which are analyzed and discussed with respect to their mathematical and historical features.

Introduction

It is well known that the central problem of *nonlinear programming* is that of minimizing or maximizing a given function of several variables subject to a finite set of inequality and/or equality constraints. The problems, both theoretical and practical, which can be translated into a nonlinear programming problem, are in fact countless and arise in different contexts, such as economics, game theory, operations research, statistics, physics, etc. Nonlinear programming can be viewed as that field of *optimization theory* which treats *static* and *finite-dimensional* optimization problems with emphasis on computational aspects. Also the term "mathematical

G. Giorgi (✉)
Dipto. Ricerche Aziendali, Sez. Matematica Generale, Universita Pavia, Via S. Felice 5, Pavia 27100, Italy
e-mail: ggiorgi@eco.unipv.it

T.H. Kjeldsen
IMFUFA, NSM, Roskilde University, Universitetsvej 1, Roskilde 4000, Denmark
e-mail: thk@ruc.dk

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*,
DOI 10.1007/978-3-0348-0439-4_1,

programming" is used. This term was first introduced by Robert Dorfman in 1949, according to the reminiscences of G.B. Dantzig on the birth of Linear Programming (see [91]). The term "nonlinear programming" appears for the first time in the title of the famous paper of Kuhn and Tucker [85].

As an autonomous field of research, mathematical programming, which draws on mathematical analysis, numerical analysis, linear algebra, operations research, etc., has encouraged or actually generated strong connections and exchanges with other theories and fields of research, such as nonsmooth analysis, decision theory, games theory, microeconomic theory, etc. Its main features make obvious its importance towards the whole economic theory and finance theory. Similarly, it must be stressed its importance from a didactic point of view: optimization theory stems, historically and logically, from the research of maximal and/or minimal points for a real function of one real variable. This is one of the classical subjects of any first-year course of a scientific university faculty. Then, this problem is generalized in second-year courses, to the research of free and constrained extremal points for a real function of several real variables.

This theory which is so "central" and so rich in its linkages to other important theories and applications, is relatively young: we recall again that the first time that the word "nonlinear programming" appeared in a printed paper is 1950 (date of printing, 1951). As is well known, though, aspects of some of the mathematical questions that are central to mathematical programming were dealt with in some of these other "connecting" fields of research before mathematical programming emerged as an autonomous mathematical discipline.

The historical view we propose in this introductory chapter will point out some of the technical difficulties of these mathematical problems and some features of the economic and social context (also military) that favored its rootedness in the years of the Second World War and the years immediately following the war. We begin by recalling the main definitions and basic results of mathematical programming theory, treating only finite-dimensional problems. Then, in section "The Prehistory of Linear and Nonlinear Programming", we shall be concerned with the "prehistory" of nonlinear programming, whereas section "Soviet Union and USA the First Years of Linear Programming" is concerned with the first ideas and developments of *linear programming,* first in the USSR and then in the USA. Section "The Birth of Nonlinear Programming" is the "central section" in which the fundamental researches of W. Karush, Fritz John, H.W. Kuhn and A.W. Tucker will be analyzed and discussed with respect to their mathematical and historical features. The final section presents some further considerations and conclusions.

Basic Results

The simplest mathematical programming problem is that of maximizing or minimizing a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ on a certain set $A \subseteq \mathbb{R}^n$, where A is a proper or improper subset of the domain of f:

$$\max_{x \in A} f(x) \quad \text{or} \quad \min_{x \in A} f(x) \tag{P1}$$

The set A is frequently called either the *feasible set*, the *feasible region* or the *opportunity set;* f is called the *objective function* or the *criterion function* and the vector $x = (x_1, x_2, \ldots, x_n)$ is the vector of the *decision variables* or *choice variables*; x is also called the *vector of instruments.* We can restrict our analysis only to one of the two problems (P_1), for example to the maximum problem, as we have that the problem $Max f(x)$ is equivalent to the problem $Min(-f(x))$ and obviously we have

$$f(x^\circ) = Max f(x) = -Min(-f(x))$$

We recall some general and well known results:

1. *Weierstrass theorem.* Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be continuous on the compact (i.e. closed and bounded) set $X \subset \mathbb{R}^n$. Then f admits maximum and minimum over X.

   This result, which however provides only a sufficient condition for the existence of maximum and minimum of f, can be generalized in several directions. We recall only the following generalization:
   Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be upper (lower) semicontinuous on the compact set $X \subset \mathbb{R}^n$. Then f admits maximum (minimum) over X.
2. Let $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined on $X \subseteq \mathbb{R}^n$. If X is a *convex set* (i.e. $\lambda x^1 + (1 - \lambda)x^2 \in X,\ \forall x^1, x^2 \in X,\ \forall \lambda \in [0,1]$) and f is *concave* on X, i.e.

   $$f(\lambda x^1 + (1-\lambda)x^2) \geq \lambda f(x^1) + (1-\lambda) f(x^2),\ \forall x^1, x^2 \in X,\ \forall \lambda \in [0,1]$$

   then:

   (a) The local maximum points of f are also global maximum points.
   (b) The set of (global) maximum points of f is a convex subset of X.
   (c) If f is *strictly concave* on X (i.e. the previous inequality holds with the order relation ">" for any $x^1 \neq x^2$ and any $\lambda \in (0,1)$), and admits a maximum point x°, this is *unique*.

   If in (P_1) we have to maximize or minimize the objective function f on an *open* feasible set A or if the extremal points x° are *interior points* of A, we shall call (P_1) an *unconstrained* or *free* mathematical programming problem. Otherwise we shall call (P_1) a *constrained* mathematical programming problem (or mathematical programming problem tout-court).

   For an unconstrained mathematical programming problem, the following results are well known:
3. If $x^\circ \in int(A)$ (the interior of A) and f is differentiable at x°, then we have (Fermat conditions) $\nabla f(x^\circ) = 0$, i.e. x° is a *stationary* or *critical point for* f.
4. Let $x^\circ \in A$ be a stationary point for f and let f be twice continuously differentiable on the open set A. Let us denote by $Hf(x^\circ)$ the Hessian matrix of f, evaluated at x°. Then:

(a) If $y^T Hf(x^\circ)y < 0,\ \forall y \in \mathbb{R}^n \setminus \{0\}$, i.e. if the quadratic form with matrix $Hf(x^\circ)$ is negative definite, then x° is a strict local maximum point for f. If $y^T Hf(x^\circ)y > 0,\ \forall y \in \mathbb{R}^n \setminus \{0\}$, then x° is a strict local minimum point for f.
(b) If the quadratic form $y^T Hf(x^\circ)y$ is indefinite, then x° is neither a maximum point, nor a minimum point. Sometimes it is called a "saddle point" for f.
(c) If $y^T Hf(x^\circ)y$ is a semidefinite quadratic form (positive or negative) we need further investigations to decide the nature of the stationary point x°.

5. If in the unconstrained problem (P_1) f is differentiable and concave on the open convex set $A \subseteq \mathbb{R}^n$, then every stationary point of f is a global maximum point for f.

Results (2) and (5) point out the importance of the concavity (and convexity) in mathematical programming problems. So, optimization theory has been also one of the main motivations for the introduction and study of various types of *generalized concavity (convexity)*, where by means of suitable extensions of the classical definition of concave (convex) functions, various classes of generalized concave (convex) functions that preserve several properties of concave (convex) functions have been established. In particular, the result (5) can be obtained also under the assumption that the (differentiable) function f is *pseudoconcave* on the open and convex set A, i.e. f must satisfy the following relation

$$\nabla f(x)(y - x) \le 0 \Longrightarrow f(y) \le f(x),\ \forall x, y \in A.$$

Another useful and important generalization of concavity (convexity), particularly useful in mathematical programming problems and meaningful in several questions of economic theory, is the class of *quasiconcave functions*, introduced by de Finetti [31]:

$$f(\lambda x + (1 - \lambda)y) \ge Min\{f(x), f(y)\},\ \forall x, y \in A,\ \forall \lambda \in [0, 1]$$

where $A \subseteq \mathbb{R}^n$ is a convex set.

If f is differentiable on the open and convex set A, f is quasiconcave on A if and only if

$$x, y \in A, f(y) \ge f(x) \Longrightarrow \nabla f(y)(y - x) \ge 0.$$

This result is due to Arrow and Enthoven [6]. It can be proved that if f is concave (and differentiable) then f is also pseudoconcave and if f is pseudoconcave, then it is also quasiconcave. The literature on generalized concavity is quite relevant; we quote only the books by Avriel et al. [10], by Cambini and Martein [17] and by Mishra and Giorgi [96]. For a history of quasiconcavity and its applications the reader may consult Guerraggio and Molho [64], Giorgi and Guerraggio [60] and Hadjisavvas et al. [66].

A first type of an unconstrained optimization problem is one where the set A is not open or the optimal point x° is not an interior point of A. In other words A is an *abstract constraint* or *set constraint*. In this case the Fermat necessary condition $\nabla f(x^\circ) = 0$ does not hold. A generalization of this condition can be obtained in the following way. Consider a subset A of $\mathbb{R}^n$ and a vector $x \in A$. A vector $y \in \mathbb{R}^n$ is a *feasible direction* of A at x if there exists an $\bar{\alpha} > 0$ such that $x + \alpha y \in A$ for all $\alpha \in [0, \bar{\alpha}]$. The set of all feasible directions of A at x is a cone, denoted $F(A, x)$, containing the origin but not necessarily closed nor open. If A is convex, $F(A, x)$ consists of the vectors of the form $\alpha(\bar{x} - x)$, with $\alpha > 0$ and $\bar{x} \in A$. Then we have the following generalized Fermat theorem:

If x° is a local minimum (maximum) point of the differentiable function f over A, then

$$\nabla f(x^\circ)y \geq 0 \ (\nabla f(x^\circ)y \leq 0), \ \forall y \in F(A, x^\circ).$$

If A is convex, then the above condition becomes

$$\nabla f(x^\circ)(x - x^\circ) \geq 0 \ (\leq 0), \ \forall x \in A$$

which is a type of *variational inequality.* Moreover, if f is convex (concave) on the convex set A, the said variational inequality is both a necessary and sufficient condition for x° to be a global minimum (maximum) point of f on A.

Sharper conditions can be obtained by means of the so-called *Bouligand tangent cone* (or *contingent cone*): see, e.g. Avriel [9], Bazaraa and Shetty [11], Giorgi et al. [59].

When the feasible set A is given by the solutions of a finite number of equations, we have a "classical" mathematical programming problem:

$$\max_{x \in S} f(x) \quad \text{or} \quad \min_{x \in S} f(x) \tag{P2}$$

where $S = \{x \in A, \ h_j(x) = 0, j = 1, \ldots, r < n\}$, $A \subseteq \mathbb{R}^n$ is open and $f, h_j \ (j = 1, \ldots, r) : \mathbb{R}^n \longrightarrow \mathbb{R}$.

The word "classical" makes reference to the works and discoveries of J.L. Lagrange, in the eighteenth century (see the next section). If for (P_2) we introduce the *Lagrangian function*

$$L(x, \lambda) = f(x) - \lambda h(x) = f(x) - \sum_{j=1}^{r} \lambda_j h_j(x),$$

where the numbers $\lambda_j \in \mathbb{R}$, $j = 1, \ldots, r$, are the well-known *Lagrange multipliers,* we have the following fundamental results.

6. Let $x^\circ \in S$ be a local solution of (P_2) and let the following assumptions be satisfied:

   (i) f is differentiable at x°;

   (ii) Every $h_j,\ j = 1, \dots r$, is continuously differentiable in a neighborhood of x° and the Jacobian matrix $Jh(x^\circ)$ has full rank, i.e. the vectors $\nabla_j h_j(x^\circ)$ are linearly independent (the constraints, in this evenience, are said to be "regular").

   Then there exists a unique vector $\lambda^\circ = (\lambda_1{}^\circ, \dots, \lambda_r{}^\circ)$ such that (x°, λ°) is a stationary point for the Lagrangian function, i.e.

$$\nabla f(x^\circ) - \lambda^\circ Jh(x^\circ) = 0. \tag{1}$$

   Under some additional assumptions (not too restrictive), the vector of multipliers λ° assumes an interesting meaning, in the sense that its components $\lambda_j{}^\circ,\ j = 1, \dots, r$, measure the effect that a marginal variation of the jth constraint gives rise to the optimal value of the objective function (see, e.g., Fiacco [48]). In economic problems the marginal variation of a constraint usually represents the variation of the available quantity of a given commodity, whereas the objective function has the meaning of profit or cost of production. So, economists call the Lagrange multipliers "shadow prices": through the values of $\lambda_j{}^\circ,\ j = 1, \dots, r$, it is then possible to get an economic evaluation of the "weight" that every constraint assumes in obtaining the optimal value of (P_2).
7. Let $x^\circ \in S$ and let (x°, λ°) verify relation (1). If $L(x, \lambda)$ is pseudoconcave with respect to x, then x° is a point of global maximum of f on S.

   Problem (P_2) has been the first constrained optimization problem to be considered by mathematicians (by J.L. Lagrange). When, long after Lagrange, optimization problems with a feasible region determined by inequality constraints (or by equality and inequality constraints), i.e., the modern mathematical programming problems, have been taken into consideration, the researches have been done towards two directions. On one hand we had the birth of *Linear Programming* problems, i.e. those mathematical programming problems where the objective function and the constraints are given by linear or affine functions. On the other hand we had the modern version of the problem (P_2), i.e. a *Nonlinear Programming* problem, of the form, e.g.,

$$\max_{x \in K} f(x) \tag{P3}$$

   $K = \{x \in A,\ g_i(x) \leq 0,\ i = 1, \dots, m\}$,where $A \subseteq \mathbb{R}^n$ is open and f, g_i $(i = 1, \dots, m) : \mathbb{R}^n \longrightarrow \mathbb{R}$.

   For this problem the *set of active (or effective) constraints* at $x^\circ \in K$ is given by

$$I(x^\circ) = \{i : g_i(x^\circ) = 0\}.$$

The two fundamental results for (P_3) are usually the Fritz John theorem and the Karush–Kuhn–Tucker theorem.

8. *Fritz John Theorem (1948)*. Let f and every g_i $(i = 1, \ldots, m)$ be differentiable at $x^\circ \in K$ and let x° be a local solution of (P_3). Then there exists a vector $(y_0, y_1, \ldots, y_m) \in \mathbb{R}^{m+1}$, *nonnegative and nonzero* (i.e. a *semipositive vector*), such that

   (i) $y_0 \nabla f(x^\circ) - \sum_{i=1}^{m} y_i \nabla g_i(x^\circ) = 0$
   (ii) $y_i g_i(x^\circ) = 0, \; i = 1, \ldots, m.$

   The proof of this theorem relies basically on the fact that an optimal point implies the empty intersection of certain sets, more precisely the inconsistency of a certain system of linear inequalities. Therefore, a classical theorem of the alternative applies and the thesis is obtained. This can be given also a geometrical interpretation: there exists a separating hyperplane (between two convex sets) whose equation contains, as coefficients, the John multipliers $y_i, \; i = 1, \ldots, m.$

If we compare the results (3) and (6), concerning respectively an unconstrained optimization problem and the problem (P_2), with the Fritz John theorem, we note that in this last one the sign of the multipliers is no longer "free": the multipliers must be all nonnegative but not all zero. Moreover, the relations (i) and (ii) above can be summarized in the following condition

$$y_0 \nabla f(x^\circ) - \sum_{i \in I(x^\circ)} y_i \nabla g_i(x^\circ) = 0,$$

where only the active constraints are considered, and where, as before, also the gradient of the objective function is associated to a multiplier. We have to say that the conditions of Fritz John for the problem (P_2) were anticipated by Caratheodory [18] and for a Pareto optimization problem with equality constraints (i.e. the objective function is a vector-valued function), by de Finetti [29, 30]. The contributions of this last author have been completely ignored by the mathematical literature (also Italian). See Giorgi and Guerraggio [60].

The multiplier y_0 can be zero. In order to avoid this "degenerate" case, where the objective function would play no role, some regularity condition must be imposed on the constraints of the problem (P_3). This is the problem of the "constraint qualifications". There are several constraint qualifications, varying in generality and complexity; see, e.g., Bazaraa and Shetty [11], Peterson [105], and Giorgi et al. [59]. The constraint qualification more similar to the regularity condition we have seen for (P_2) is: the active gradients $\nabla g_i(x^\circ), \; i \in I(x^\circ)$, are linearly independent.

9. *Karush–Kuhn–Tucker theorem (1939 and 1951)*. Let $x^\circ \in K$ be a local solution of (P_3), under the assumption of differentiability at x° of f and every $g_i, \; i = 1, \ldots, m$. If a constraint qualification is satisfied, then there exists a vector λ such that

   (i) $\nabla f(x^\circ) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x^\circ) = 0.$

(ii) $\lambda_i g_i(x^\circ) = 0,\ i = 1, \dots, m.$
(iii) $\lambda_i \geq 0,\ i = 1, \dots m.$

The next theorem provides a sufficient condition for $x^\circ \in K$ to solve problem (P_3).

10. Let $x^\circ \in K$ satisfy conditions (i)–(iii) of the theorem of Karush–Kuhn–Tucker. Let f be pseudoconcave on the open convex set $A \subseteq \mathbb{R}^n$ and let $g_i(x^\circ),\ i \in I(x^\circ)$ be quasiconvex (i.e. $-g_i$ is quasiconcave) on the same set. Then x° solves (P_3).

By this quite schematic introduction one cannot obviously be informed of all the results, even important and "classical", that have grown within mathematical programming theory. We shall make two exceptions: the first is concerned with duality theory, especially for linear programming problems, the second one with vector (or Pareto) optimization problems.

Duality theory allows to deduce some interesting features of a mathematical programming problem (P), by means of the analysis of another problem, in a certain sense "specular" to (P), called *dual problem*, built from (P) following certain rules.

Duality theory was born together with linear programming theory and game theory (matrix games) and subsequently has been extended also to nonlinear programming theory. If (P) represents the following linear programming problem

$$\begin{aligned} &Min\ cx \\ &Ax \leq b \\ &x \geq 0 \end{aligned}$$

where A is a (m, n) matrix, $c, x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, its dual problem (P') is

$$\begin{aligned} &Max\ by \\ &yA \geq c \\ &y \geq 0 \end{aligned}$$

If (P) is the following linear programming problem

$$\begin{aligned} &Min\ cx \\ &Ax = b \\ &x \geq 0, \end{aligned}$$

its dual problem (P') is

$$\begin{aligned} &Max\ by \\ &yA \geq c. \end{aligned}$$

Note that in this dual problem there are no sign constraints on y.

In general, given a linear programming problem (P), called also the "primal" problem, the following properties on (P) and its dual (P') hold:

(a) The dual (P') of (P) is unique.
(b) The dual of the dual is again the problem (P).
(c) If (P) is a minimum problem, then (P') is a maximum problem and conversely.
(d) Coefficients $b_1, b_2, \ldots, b_m$ in the objective function of the dual problem (P') are the constants of the right-hand side of (P) and the coefficients $c_{1,} c_{2,\ldots,} c_n$ in the objective function of (P) are the constants of the right-hand side of the dual (P').

Duality theory for linear programming is very important, not only from theoretical points of view, but also from computational points of view. A first basic result states that if (P) and (P') have both a nonempty feasible set, then both problems admit solution and the two optimal values are equal.

We give now a slight idea of what is a vector (or Pareto) optimization problem. If the objective function is given by a vector-valued function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, we have the class of *vector optimization problems* or *Pareto optimization problems* (from Vilfredo Pareto (1848–1923), the Italian economist and sociologist who first considered these kinds of problems). In this case, before considering the necessary and/or sufficient optimality conditions, one has to define and specify the notion of optimal solution, as the image space of a vector optimization problem is not a totally ordered set (as the space of real numbers), but only a partially ordered set. Therefore, it is not possible to "transfer" in an immediate way to the vector case the basic inequality $f(x) \leq f(x^\circ)$ which characterizes the definition of a maximum point for the scalar case.

The Prehistory of Linear and Nonlinear Programming

From what we have said in section "Introduction", it would be clear that we are treating a theory whose "official" birth lays at the hearth of the twentieth century, and precisely in the years immediately following the Second World War. However, all readers who have some recollections of a second course of Mathematical Analysis, are familiar with the term "Lagrange multipliers", named after Lagrange who treated *equality* constrained optimization problems, that is problems of type P_2, in the second half of the eighteenth century.

Joseph Louis Lagrange (Turin, 25/01/1736, Paris, 10/04/1813) is, in a sense, both Italian and French: his mother, Teresa Gros, was a daughter of a medical doctor of Cambiano (a spot near Turin) and his father was of a family coming from France (Touraine) but settled in Piedmont, in Turin, since the reign of Charles Emmanuel II, duke of Savoy. Gino Loria, the Italian historian of Mathematics, in his "Storia delle Matematiche" [92] writes that Lagrange was baptized as "Lagrangia Giuseppe Ludovico"; subsequently the famous mathematician used different forms of his family name: De la Grangia Tournier, Tournier de la Grangia, De la Grange, etc.

After his moving to Berlin and finally to Paris (in a time where a noble origin was not looked favourably upon) he signed his works with the name "Lagrange". This perhaps explains the reason why in several textbooks of mathematical analysis Lagrange is presented as a "French Mathematician".

Lagrange introduced "his" multipliers in 1788, in the fourth section of the first part of his famous book *Mécanique Analytique* [86]. These multipliers were, in the intention of Lagrange, the basic tool to state the configuration of a stable equilibrium for a Mechanical system. The problem is to deduce, from the general principle of Statics Theory [86, p. 77]:

> "des formules analytiques qui renferment la solution de tous les problèmes sur l'équilibre des corps [...], en reduisant en quelque manière tous les cas à celui d'un système entièrement libre".

In particular, one has to minimize the so-called *potential function*, taking into account that the system, individuated in its positions by $n+r$ coordinates, is subject to constraints of the type $h_1(x) = 0,\ h_2(x) = 0, \ldots,\ h_r(x) = 0$. By differentiating these equations we have $dh_1(x) = 0,\ dh_2(x) = 0, \ldots,\ dh_r(x) = 0$, and

> "comme ces équations ne doivent servir qu'à éliminer un pareil nombre de différentielles dans la formule générale de l'équilibre, après quoi les coefficients des différentielles restantes doivent être égalés chacun à zéro, il n'est pas difficile de prouver, par la théorie de l'élimination des équations linéaires, qu'on aura les mêmes résultats si l'on ajoute simplement à la formule dont s'agit les différentes équations [...] multipliées chacune par un cofficient indéterminé λ_j" [86, p. 78]

From this "algebraic" remark, it follows the "extrêmement simple" rule that allows to state the equilibrium configurations. Lagrange considers the equalities $df - \lambda_1 dh_1 - \ldots - \lambda_r dh_r = 0$, where df represents, following the words of Lagrange [86, p. 78], "la somme des *moments* de toutes les puissances qui doivent être en équilibre". Now, if we choose the multipliers λ_j in such a way that the coefficients of dx_{n+j} are zero, we have an equation in $dx_1, \ldots, dx_n$ whose coefficients must be all zero. So, we have obtained the $n+r$ conditions (to be added to the r equalities $h_1 = 0, \ldots, h_r = 0$) in the $n+2r$ variables $x_1, \ldots, x_{n+r}, \lambda_1, \ldots, \lambda_r$. The multipliers λ_j, introduced by Lagrange as an algebraic tool to obtain that "le nombre de ces équations sera égal à celui de toutes les coordonnées des corps" [86, p. 79] have also a physical meaning. Indeed Lagrange did not discuss the system of the $n+2r$ equations in the $n+2r$ unknowns, obtained by means of the "equations particuliéres de l'équilibre"; he only remarked that the value of the multipliers [86, p. 79]

> "pourra toujours exécuter par les moyens connus, mais il conviendra, dans chaque cas, de choisir ceux qui pourront conduire aux résultats les plus simples"

Moreover, he pointed out that the various $\lambda_j dh_j$ represent "les moments de différentes forces appliquées au même système" [86, p. 80]. It is by means of these forces, expressed by the constraints, that one can transform the constrained problem into an unconstrained one:

> "et de là on voit la raison métaphysique pourquoi l'introduction des termes $\lambda_1 dh_1 + \lambda_2 dh_2 + \ldots$ peut ensuite traiter cette équation comme si tous les corps du système étaient entièrement libres." [86, p. 82]

In the *Théorie des Fonctions Analytiques* [87] Lagrange presented his method of multipliers with a greater generality. Here he introduced the problem without making reference to some specific question of Mechanics, but directly with regard to a general constrained optimization problem with equality constraints. In Sect. 58 of Chap. XI of the second part, we find again the proof of the necessary optimality conditions, a proof performed with the standard of precision that is typical of Lagrange's exposition. Lagrange finished the section with the basic general principle:

> "il suffira d'ajouter à la fonction proposée les fonctions qui doivent être nulles, multipliées chacune par une quantité indéterminée, et de chercher ensuite le maximum ou minimum comme si les variables étaient indépendantes."

In the same years of Lagrange, we also find a few works on linear equality constrained optimization problems. We refer to Grattan-Guinness [61, 62] who records the names of Laplace [88, 89], of Boscovich [16] and of de Prony [33].

Lagrange developed his "multiplier rule" in dealing with optimization problems subjects to equality constraints. These problems came out of his considerations regarding stable equilibrium for mechanical systems that were subject to some equality constraints which where, again, subject to the principle of virtual work. Lagrange utilized this principle as an axiom for reversible displacements of the system. The French mathematician J. Fourier (1768–1830) extended the principle to irreversible displacements in his *Mémoire sur la Statique contenant la démonstration du principe des vitesses virtuelles et la théorie des moments* published in 1798. He thereby considered mechanical systems subject to inequality constraints. It is unclear what inspired Fourier to consider this extension to irreversible displacements, maybe it was just a very careful reading of Lagrange's *Mécanique analytique* (1788) that made Fourier aware that Lagrange disregarded displacements that, even though they did not fulfil the equality constraints, did not violate the constraints of the mechanical system—at least this was what the Russian mathematician and physicist M.V. Ostrogradsky (1801–1862) discovered when he read Lagrange's work: "…ce grand géometre [Lagrange] a incompletement énuméré les déplacements possibles dans la plupart des questions de la premiere partie de la Mecanique analytique, et il est facile de reconnaitre que les déplacements qu'il a négligé de considérer, ne sont empeché par aucune condition, …" [102, p. 130]. The displacements left out by Lagrange are precisely the irreversible ones. Instead of the virtual work, Fourier considered "le moment de la force" [49, p. 479], which results in a shift in sign (see Prekopa [107] and Franksen [54–56] for further details). But it was not only in mechanics that Fourier saw a need for a theory of linear inequalities. In his book *Analyse des Équations Déterminées* (1831) he explained that "The principle of the theory of inequalities will be expounded in the seventh and last of the books. This part of our work is concerned with a new kind of questions, which offers varied applications to geometry, to algebraic analysis, to mechanics, and to the theory of probability" [53, p. 71]. Fourier died before he could finish the project and he only published a short paper [52] and two summaries [50, 51] on linear inequalities. Even though Fourier did not get very far in his development of a theory for systems of linear inequalities these few publications show that he did indeed have a geometrical understanding of the set of solutions to such a system in three variables as a polyhedron. Despite Fourier's claim of the applicability of a theory of inequalities to various branches of mathematics the subject seems not to have raised any real interest at the time. A fact that Darboux clearly pointed out when he wrote the "Avertissement" of Fourier's collected work: "Nous avons aussi, par quelques emprunts à l'*Historie de l'Académie* pour années 1823 et 1824, pu faire connaitre d'une maniere assex precise certaines idées sur la theorie des inégalites auxquelles l'illustre géometre attachait une importance qu'il est permis, aujourd'hui, de trounver un peu exagérée" [28, pp. v–vi].

In the printed works of Fourier, and also in his unpublished notes, there are some anticipations of some basic subjects that we now a days think of as belonging to the modern theory of mathematical programming, but there is no theoretical deepening nor an organic treatment of the material: there are only some first acquisitions, as the convexity of the feasible region in optimization problems subject to linear inequality constraints and, above all, the study of several "practical situations", where the "inequality analysis" reveals its utility (for example, the "least squares method"). These suggestions, notwithstanding the scientific authority of Fourier, did not as Darboux's evaluation also made clear cause many reactions and interest at the time. Apparently only two of Fourier's students took his suggestions into consideration: the famous mathematician Navier [101] and the equally famous mathematical economist Cournot [25, 26], who, in the paper of 1827, without making reference to the work of Fourier (see [107]), rediscovered the principle of Fourier, giving the necessary conditions for equilibrium with ad hoc arguments which make specific reference to the mechanical interpretation of the question.

The same treatment appeared in more general terms in a paper of the Russian mathematician Ostrogradsky [102, 103]. Ostrogradsky (1801–1862) was a student in Paris and, before he returned to S. Petersburg, he had attended courses of Fourier, Poisson, Cauchy and other famous French mathematicians. Ostrogradsky, strangely without quoting Fourier nor Cournot, asserted that at the minimum point the gradient of the objective function can be represented as a linear combination, with nonnegative multipliers, of the gradients of the constraints. The multipliers are sign free when there are only equality constraints.

From then on there seems to have been practically no other interest, among mathematicians, for optimization problems with inequality constraints until the end of the nineteenth century. We can only quote a paper of Gauss of 1829 (see the complete reference in Grattan-Guinness [62]) where the inequality principle is "enunciated ... without mentioning Fourier" (see [107]). Other uses of linear inequalities were made by G. Boole (1815–1864) on questions regarding Statistics and the measure of Probability [15]. Also Gibbs [57] considered problems of dynamics with inequality constraints.

A work by Paul Gordan (1837–1912) from 1873 is included in Theodore Motzkin's (1908–1970) list of all known previous literature on linear inequality systems which he published in his master thesis *Beiträge zur Theorie der linearen Ungleichungen* in 1936. But historically, there seems to be nothing that indicates that Gordan was interested in the theory of systems of linear inequalities (see [78]). In the paper listed by Motzkin, Gordan proved a theorem about existence of positive solutions to a system of linear equations. His main concern seems to have been to develop a tool for determining when a system of linear Diophantine equations has a positive solution, a question that came up in his work on the so-called finite basis problem in invariant theory, which was his main area of interest at the time. Motzkin interpreted Gordan's paper in the following way:

> "He stated the elegant transposition theorem in disguised form and proved it in a roundabout way, but then confined himself to Diophantine problems" [100, p. 5].

Gordan did not, however, state a relationship between two systems of equations of the form of a transposition theorem. Such a theorem can be derived fairly easily, though, from Gordan's result, which is probably why later writers have credited a transposition theorem to Gordan.

The interest in inequality analysis began again towards the end of the nineteenth century, above all, with the publication of several papers of the Hungarian mathematician Julius Farkas, whose role became further emphasized when Kuhn and Tucker used what is now called the *Farkas lemma* or *Farkas–Minkowski lemma* to prove their famous theorem. Julius Farkas (1847–1930) was professor of theoretical physics at the University of Kolozvàr and a member of the Hungarian Academy of Sciences. He was well known in the scientific environments for his contributions to Mechanics and Termodynamics. His work on systems of linear inequalities was motivated by problems in physics, as was the case with Fourier. Farkas formulated the result, later known as Farkas' lemma, in the paper *Über die anwendung des mechanischen Princips von Fourier* [43]. Farkas developed a general method that could be used to treat all types of problems involving the inequality principle of Fourier: "Der Hauptzweck vorliegender Arbeit ist zu erweisen, dass mit einer passenden Modifikation die Methode der Multiplikatoren von Lagrange auch das Fourier'sche Princip übertragen werden kann" [43, p. 266]. Farkas focused on the mathematical foundation and developed a theory of homogenous linear inequalities in a series of papers, culminating with *Theorie der einfachen Ungleichungen* [44] which was published in 1901. Here we find the most complete and correct proof of his lemma, where he proves that every solution of the system of linear inequalities $Ax \geq 0$ (A is a (m, n) matrix, whose rows will be denoted by $A_i,\ i = 1, \ldots, m$) is also solution of the inequality $bx \geq 0$ if and only if there exist m nonnegative numbers λ_i such that $b = \sum_{i=1}^{m} \lambda_i A_i$.

The application of the Farkas "lemma" to the correct solution of the equilibrium problem (with inequalities) proposed by Fourier, appears quite immediate and "natural". Let x° be the minimum point of the potential function V. If $x^\circ + dx$ represents a different position in the feasible region (i.e. such that $g_i(x^\circ + dx) \geq 0$), under differentiability assumptions we have $dg_i(x^\circ) = \nabla g_i(x^\circ)dx \geq 0$. The converse does not hold, but if we impose a suitable constraint qualification (which obviously is missing in Farkas), we shall have that every dx such that $\nabla g_i(x^\circ)dx \geq 0$ belongs also to the feasible region and, moreover, $dV(x^\circ) = \nabla V(x^\circ)dx \geq 0$. When the problem has been "linearized", the Farkas lemma assures the existence of nonnegative multipliers λ_i such that $\nabla V(x^\circ) = \sum \lambda_i \nabla g_i(x^\circ)$.

Hermann Minkowski studied systems of linear inequalities at the same time but independent of Farkas and in a different context. Minkowski included an appendix about linear inequalities to the first chapter of his book *Geometrie der Zahlen* (first edition: 1896). He treated a system of homogeneous linear inequalities $\xi_1 \geq 0, \ldots, \xi_n \geq 0$ and proved that every non-trivial solution can be written as a positive linear combination of what he called "äusserste" (extreme or fundamental) solutions [95, p. 43]. By such solutions he understood non-trivial solutions that could not be written as the sum of two solutions that are, what he called, essentially different, meaning that they are not multiples of each other. Minkowski developed his theory of linear inequalities so it could be applied directly in his investigations of reduced forms of positive definite quadratic forms, which was his main motivation for developing a geometry of numbers (see [78, pp. 480–489]).

The Hungarian mathematician Alfred Haar (1885–1933) presented a paper on linear inequalities to the Hungarian Academy of Science in 1917 where he gave a new foundation for the work on inequalities by Farkas and Minkowski. Haar's work was published in 1918 and translated into German in 1924. He generalised the result of Farkas to nonhomogeneous systems of linear inequalities based on the theory of convexity [65].

Inspired by a paper on "preferential voting" published in *The American Mathematical Monthly* in 1916 [93] American mathematicians began to develop a theory for systems of linear inequalities beginning with work of Dines (1917) which was followed by further works of Dines [34–37], of Carver [19], of Stokes [112] and a paper of Dines and Mc Coy [38] that was quoted by Karush [75] in his Master Thesis (see Sect. 5). Another work which has been called a contribution to the prehistory of linear programming and is mentioned by Dantzig in his recollections on the origins of the simplex method [27] is a paper of the Belgian mathematician de la Vallée Poussin [32] who gave a method for finding minimum deviation solutions of systems of equations.

However, the first effective acknowledgement of the importance of the work of Farkas is given in the Master Thesis of T. Motzkin, accepted in 1933 at the University of Basel and published in 1936 [100]. See also the paper of Kjeldsen [78] for analyses and discussions of the different motivations and goals of the various contributions to the study of linear inequality systems. The work that led to mathematical programming was not motivated by these different developments in systems of linear inequalities, they were just important tools for proving fundamental results. The actual development of a theory, or theories, for solving inequality constrained optimization problems, was neither motivated by equilibrium problems in mechanics, nor by questions related to positive quadratic forms or to inequality systems as such. The unset of mathematical programming was spurred by a different set of problems that were "solved" by mathematicians who worked independently of each other under very different circumstances, first in the Soviet Union in the late 1930s and second in the USA after the Second World War.

Soviet Union and USA: The First Years of Linear Programming

Why does a branch of learning and a certain theory emerge and develop during a certain period, in a certain country rather than in another period and in another (equally developed) country? This is indeed a good question, almost always too general to receive satisfactory answers. Anyhow, the question may be useful and appropriate to understand the origin of some scientific theories which, usually, do not spurt out, like mushrooms, from evening to morning. For the case of mathematical programming, why did Farkas' lemma remain almost unused for nearly half a century? Why were there no substantial interests towards optimization theory until the 1950s of the last century even though the maximization or minimization of a certain function can be viewed as a "natural" curiosity of the "homo oeconomicus"? We recall that in those years (first half of twentieth century) Economics had already reached its status of science since longtime. Why and how did linear programming and nonlinear programming emerge and develop in the context of the Second World War?

Perhaps it is possible to give only partial answers and not one comprehensive answer. The birth of a new scientific field of research, such as mathematical programming, appears to be correlated with the development of other mathematical theories and tools and also to the ripening of certain political, sociological and economic situations. As for what concerns the mathematical tools, it is quite obvious that the differential calculus for functions of several variables (and also for functionals defined on abstract topological spaces) was well known since several years. The same cannot be said for what regards the tools of Convex Analysis, which are of primary importance in the study of inequalities (especially linear inequalities) and in establishing some fundamental results in mathematical programming and in mathematical economics. The war experience, with the natural presence of well determined tasks to reach, the necessity to consider at the same time a large number of variables, together with the necessity to get the solution of the problems in the least time possible, are surely some factors useful to explain the need of general methods for general but practical problems. Economic and military reasons are the natural motivations which led to the establishment of Linear Programming in the Soviet Union and in the USA.

For a good introduction to the history of mathematical programming in the USSR the reader is referred to Polyak [106]. In the USSR the father of linear programming methods is Leonid Vitalievich Kantorovich (1912–1996). He graduated from Leningrad University at the age of 18 and taught at the same university from 1934 to 1960. Within the international mathematical community he is remembered not only for his achievements in linear programming and mathematical economics, but also for his significant contributions to functional analysis. He was awarded the Stalin Prize (1949), the Lenin Prize (1965) and the Nobel Prize (1975), together with T.C. Koopmans. In the Spring of 1939, when Kantorovich was a young professor at Leningrad University, he was contacted for a scientific advice, by a state firm that produced ply-wood and wished to make more efficient the use of its machine tools. The aim was to increase the production level of five different types of ply-wood, carried out by eight factories, each with a different production capacity. Kantorovich realized that the proposed problem had a mathematical structure, common to other situations and problems one can find in, as he said, the organization and planning in the field of industry, construction, transportation and agriculture.

In the same year (1939) Kantorovich published a small book (Kantorovich [72] is the English translation of this essay) where he discussed and numerically solved variations of optimization problems under inequality constraints. Much later Kantorovich [73] (original Russian edition: 1959) developed further his ideas on the mathematical treatment of certain economic problems. This second book was translated into English and into French, so the Western researchers became acquainted with the early discoveries of Soviet mathematics on linear programming problems: in the Soviet Union linear programming was born in relation to requirements of industry productions, within the third "five-year plan", with the hope that it "will play a very useful role in the development of our socialist industry" [73]. In his booklet of 1939 Kantorovich presented several microeconomic problems (all coming from the production planning of certain industries), framed into three mathematical schemes. The first of these schemes (the other two are more general variations of the first one) considers the allocation of n machines which can produce items consisting of m different parts. If machine i is used for output k, it can produce a_{ik} units of part k per time unit. Let h_{ik} denote the number of time units of machine i allocated to produce part k and let z_k denote the total number of units produced of part k. Then $z_k = \sum_{i=1}^{n} a_{ik} h_{ik}$, and since one wants to produce complete items the requirement $z_1 = z_2 = \ldots = z_m$ need to be fulfilled.

Kantorovich then formulated the following problem, which he called problem A: Find h_{ik} $(i = 1, \ldots, n; k = 1, \ldots, m)$ such that

$$h_{ik} \geq 0$$

$$\sum_{i=1}^{n} h_{ik} = 1; \text{ for } i = 1, \ldots, n$$

$$z_1 = z_2 = \ldots = z_m, \text{ where } z_k = \sum_{i=1}^{n} a_{ik} h_{ik}, \text{ for } k = 1, \ldots, m$$

These are the constraints of the problem, and the objective is to maximize the common value of $z_1 = z_2 = \ldots = z_m$.

Kantorovich proved, both in analytical terms and in geometrical terms, the existence of what he called "resolving multipliers". His main attention was on numerical methods, based on the "resolving multipliers". The proof of the existence of these multipliers was postponed to the appendix, with the specification that "the ignorance of the proof [...] in no way interferes with mastering the method of its practical applications" [72, p. 419]. The method of "resolving multipliers" must be above all "sufficiently simple and effective", as the problems treated, not only have a theoretical interest but also a practical importance and require "the solution of ten of thousands or even millions of systems of equations for completion."

It has been discussed in the literature whether Kantorovich with his "resolving multipliers" actually introduced the idea of dual variables. On these matters we tend to agree with Charnes and Cooper [20, p. 249] when they write: "We cannot find any place in his piece where he appears to think of his "resolving multipliers" as anything more than devices for assisting in the solution of what is now called the "direct" (or primal) problem." Polyak [106] says that the 1939 revolutionary booklet of Kantorovich had "little response from economists or mathematicians" and Tikhomirov [116] adds that " according to some ideological doctrine, an abstract subject like mathematics was of no conceivable use to so life-related a subject as economics".

We have to wait for the second book of Kantorovich on linear programming [73] to get a more general formulation of his method in resolving a linear programming problem. This method, formally different from the simplex algorithm of G.B. Dantzig, is in fact equivalent to the one of Dantzig, as later proved by van de Panne and Rahnama [119].

We across Leonid Kantorovich, in the 1930s of the last century, as a young professor at Leningrad University. It is well known that, in the long run, his scientific reputation gained more and more recognition, within the Soviet mathematical community. But in the political environment the diffusion of his ideas on economic science was not so easy. Leifman in the introduction to the volume he edited in the memory of Kantorovich (Leifman [90], quoted also by Balinski in Lenstra et al. [91]) reports the following part of a speech Kantorovich delivered at the 1960 Conference on the Application of Mathematical Methods in Economics and Planning held in Moscow:

> "Here comrade Mstislavskii [the previous speaker] was talking about the necessity of applying mathematical methods in economics. But he did not always say so; not so long ago he was saying otherwise. And his friend and coauthor Yastremskii at one meeting said addressing me: "You are talking here about optimum. But do you know who is talking about optimum? The fascist Pareto is talking about optimum!" You know how that sounded in 1943. Nevertheless, I did not say that, not to be like the fascist Pareto, let us strive for maximum of costs and minimum of production". [90, pp. x–xi]

Through the history of the oppositions to the "new" mathematical economic methods of Kantorovich we can reconstruct a large part of the controversy (see, e.g., [42]) of the second post-war period, between the "optimal planners" and the political and government power, this last much more in line with Marxist dogma. It is well known that Stalin's conception of mathematical economics was that this is nothing but a "game with the numbers". By this expression, beyond its vulgarity, he wanted to specify that the planning and organization of economic resources are not problems of economic analysis or economic theory, but problems of political economy, and therefore these problems are exclusively pertaining to the political power. The role of economists (of theoretical economists) was another: as apologists of the USSR political system, their had the task to elaborate theoretical models that a posteriori would justify the choices made by politicians. This situation slowly began to change in the 1960s, and it is just in this period that the ideas of Kantorovich on economic science began to be considered also by the rulers of USSR. In 1958 the Central Economics-Mathematics Institute of the Academy of Sciences was founded, and in the same year Kantorovich was elected corresponding member of the Academy of Sciences.

The life of the Russian mathematical optimization community in the 1960s was full of events (see [106]): we mention only the conference in Moscow in 1960 on the use of mathematical methods in economic science and planning and the conference in Moscow in 1971 on optimal planning procedures. The work of Kantorovich was made available in the West in 1960, when Tjalling Charles Koopmans (1910–1985) managed to publish an English translation of Kantorovich's work of 1939 in "Management Science".

In the meantime, as is well known, a similar line of research in inequality constrained optimization took place in the USA independent of the work of the Russians. Koopmans is perhaps the economist who has contributed the most to the diffusion, especially among economists, of the results of the Western line of thoughts and results about linear programming, a theory that in the USA responded to military demands. Koopmans was born in the Netherlands where he studied at the University of Utrecht and the University of Leiden, and he moved to USA in 1940. He taught at Chicago University and from 1955 at Yale. From 1961 to 1967 he was director of the famous "Cowles Foundation", and in 1975 he was awarded the Nobel Prize in economics together with Kantorovich.

During the Second World War, from 1942 to 1944, Koopmans worked, as a statistician, at the "Allied Shipping Adjustment Board", being concerned, in particular, with some transportation models. In the same period George B. Dantzig (1914–2005), who is generally recognized as the "Western father" of linear programming, collaborated with the Pentagon, as an expert of programming methods, developed with the use of desk calculators. Dantzig finished his studies and became a Ph.D. in mathematics shortly after the war ended. Employment opportunities came from the University of California at Berkeley and, above all, from the Pentagon, where D. Hitchcock and M. Wood proposed him to find a way to mechanize the planning processes he had previously formalized, by using linear inequality systems as formal tools to describe the interindustry relations that in economic science had been studied by Wassily Leontief (1905–1999).

We can say, in this connection, that the development of linear programming was also influenced by the so-called "economic activity analysis" (see the volumes edited by Koopmans [81] and by Morgenstern [98]) and by certain aspects of the theory of games (matrix games). The basic book on the theory of games had been published by von Neumann and Morgenstern in 1944 [120]. We note also that another Hitchcock, F.L. Hitchcock, had published in 1941 an important paper on the so-called "transportation problem", which anticipated somewhat the subsequent work of Dantzig and was almost contemporaneous of the work of Kantorovich. At that time the paper of Hitchcock [69] had little impact on the scientific environments.

Another modern linear programming formulation was the so-called "diet problem", first considered by Stigler [110].

The first Leontief models on interindustry analysis (or input–output analysis) were, however, static models and "what the Air Force wanted was a highly *dynamic model*, one that could change over time." [27, p. 21]. Moreover, the military authorities insisted also on the possibility to have multiple choices between alternatives and to benefit from the possibility of numerical computations:

> "Once the model was formulated, there had to be a practical way to compute what quantities of these activities to engage in that was consistent with their respective input-output characteristics and with given resources. This would be no mean task since the military application had to be *large scale*, with hundreds and hundreds of items and activities."[27, p. 21]

The *simplex method* discovered by Dantzig, in order to solve a linear programming problem was presented for the first time in the Summer of 1947. In June of the same year Dantzig met with Koopmans, who at once understood the importance of this new algorithm and took it on his shoulders to bring its potentialities to the attention of the group of economists with whom he had several collaborations. These names include K.J. Arrow, P.A. Samuelson, H. Simon, R. Dorfman, L. Hurwicz, and H. Scarf. In the Fall of the same year Dantzig consulted John von Neumann at Princeton who, according to Dantzig, introduced him to Farkas' lemma, to duality, and to the problems connected with game theory ("matrix games").

Dantzig became acquainted with Albert Tucker, who later became head of the Mathematics Department at Princeton, at one of his visits, and when the Office of Naval Research (ONR) decided to set up a research project to study the connection between linear programming and game theory, as well as the underlying mathematical theory, Tucker was asked to undertake the project.

> "Soon Tucker and his students Harold Kuhn and David Gale and others like Lloyd Shapley began their historic work on game theory, nonlinear programming and duality theory. The Princeton group became the focal point among mathematicians doing research in this field" [27, p. 25]

The first official presentation of the simplex method took place at the conference Activity Analysis of Production and Allocation which was held at the University of Chicago in 1949 and organized by T.C. Koopmans. Later, this conference came to be called the Zero-th Symposium on Mathematical Programming [23]. In his recollections Dantzig underlined the importance of this meeting. A simple look at the list of the participants (T.C. Koopmans, L. Hurwicz, R. Dorfman, N. Georgescu-Roegen, A.W. Tucker, H.W. Kuhn, D. Gale, etc.) and their sponsors is sufficient to understand the significance of the Second World War for the emergence of linear programming and subsequently nonlinear programming in the USA. The conference and the research done by a majority of the participants were supported by the military. Dantzig summed it up as follows:

> "the advent or rather the *promise* that the electronic computer would exist soon, the exposure of theoretical mathematicians and economists to real problems during the war, the interest in mechanizing the planning process, and last but not least the availability of money for such applied research all converged during the period 1947-1949. The time was ripe. The

> research accomplished in exactly two years is, in my opinion, one of the remarkable events of history" [27, p. 26].

All the factors mentioned by Dantzig were directly or indirectly connected with the U.S. military and its activities during and after the war. (For an exposition of Dantzig's research we refer to Cottle [22]).

In the introduction to the proceedings of the conference, Koopmans pointed towards four distinct lines of research that played a significant role in the emergence and establishment of linear programming in the USA [81]: the debate on the construction of models of general economic equilibrium together with their mathematical formalization and the search for solutions, the new welfare economics, the analysis of the various input–output models of Leontief and, last but not least, the specific work of Dantzig and Wood, motivated by "the organization of defense, the conduct of the war, and other specifically war-related allocation problems". We can add also the work of von Neumann and Morgenstern on the theory of matrix games.

In his recollections [27, pp. 29–30], Dantzig says that the name "simplex method" arose out of a discussion with T. S. Motzkin, who gave a geometric interpretation of Dantzig's algorithm: this one is "best described as a movement from one simplex to a neighboring one". Also the term "primal problem" was new. It was proposed by Dantzig's father, the mathematician Tobias Dantzig, around 1954.

The Birth of Nonlinear Programming

With the ONR-financed project on game theory, linear programming and the underlying mathematical structure, which began in the summer of 1948 with Albert Tucker as principal investigator, the linear programming problem moved into academia and became exposed to academic mathematical research. In this environment linear programming acted as a starting point for further generalizations, "naturally" leading to the extension to the nonlinear case which was in fact pursued almost at once.

According to Tucker himself [2, pp. 342–343], he became involved with the project by coincidence:

> "I just happened to be introduced to him [Dantzig] and offered him a ride [to the train station after a meeting at Princeton between Dantzig and von Neumann], during which he gave me a five-minute introduction to linear programming, using as an example the transportation problem. What caught my attention was the network nature of the example, and to be encouraging, I remarked that there might be some connections with Kirchhoff's Laws for electrical networks, which I had been interested in from the point of view of combinatorial topology. Because of this five-minute conversation, several days later I was asked if I would undertake a trial project that summer, and I agreed. The two graduate students I got to work with me were Harold Kuhn and David Gale."

Towards the end of 1949 Tucker invited Gale and Kuhn to study a possible generalization of the results on duality theory, already obtained by the three of them for the linear case (results that were published in 1951 in the proceedings edited by

Koopmans) to the quadratic case. Gale declined the offer, but Kuhn accepted, and the first results of his and Tucker's joint work were presented in a

> "preliminary version (without the constraint qualification) [...] by Tucker at a seminar at the Rand Corporation in May 1950. A counterexample provided by C.B. Tompkins led to a hasty revision to correct this oversight. Finally, this work might have appeared in the published literature at a much later date were it not for a fortuitous invitation from J. Neyman to present an invited paper at the Second Berkeley Symposium on Probability and Statistics in the summer of 1950" [82, p. 14].

The paper was published the following year in a proceedings from the symposium and is the first paper that introduces the term "nonlinear programming".

It was later discovered that two other mathematicians earlier had proved results similar to Kuhn's and Tucker's, namely William Karush (1917–1997) in his master thesis from 1939 and Fritz John (1910–1994) who published his results in 1948. None of these works had any influence on the origin of nonlinear programming, but today they are considered classics in the field of modern nonlinear programming, and as such we will discuss their contributions in the following; see also Kjeldsen [76, 77, 80].

The Contribution of William Karush

In his paper of 1976 on the history of nonlinear programming, Kuhn promptly declares that the first organic contribution to the theory of nonlinear programming goes back to 1939, to William Karush's Master Thesis at the Department of Mathematics of Chicago University, a thesis that was never published. Kuhn [83, p. 83] speaks of "a theorem which has often been attributed to Kuhn and Tucker" and says that he learned for the first time about the existence of this Master Thesis, through the pages of the book of Takayama [115] on mathematical economics. In another paper Kuhn [84, p. 133] asserts that: "Takayama's book was the first citation of Karush's work that I read". In this book, at pages 61, 73 and 100 of the 1974 edition, the author makes several references to the thesis of Karush, writing in particular:

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

Takayama learned of the existence of Karush's work from El Hodiri [39, 40]. At page 100 (1974 edition) Takayama again says:

> "Unfortunately, this work [of Karush] has been unduly ignored. El Hodiri (1967) rediscovered Karush and put it in a better perspective."

El Hodiri [41] in turn wrote that he became acquainted with Karush's thesis by reading Pennisi [104] and that he got the reference to Pennisi from the book of Saaty and Bram [108], see also Cottle [24]. Even more curious is the fact that the thesis of Karush is mentioned also by Fiacco and McCormick [47], in the "Additional References" to Chap. 2. After its publication, their book immediately became one of the most quoted books on nonlinear programming.

Karush's thesis has the title "Minima of Functions of Several Variables with Inequalities as Side Conditions", and certainly, a modern reader familiar with mathematical programming would recognize this as a paper belonging to nonlinear programming. But the term nonlinear programming did not exist when Karush wrote his thesis, so in order to understand the mathematical context of his work we need to go back to the Department of Mathematics at Chicago University in the 1930s. This was at a time where calculus of variations was one of the high points of the department's research—the so-called Chicago School of the Calculus of Variations, a school that in those years counted mathematicians of the caliber of G.A. Bliss, L.M. Graves, F.A. Valentine and M.R. Hestenes. When Karush wrote his thesis, calculus of variation problems with inequalities as side conditions were under investigation, and his thesis can be seen as a finite-dimensional version of such problems.

Karush began his thesis with a reference to a paper by Bliss [14] in which Bliss treated the equality constrained version of the optimization problem to gain insights into questions about normality and abnormality in the calculus of variations, explaining that his, i.e. Karush's thesis "proposes to take up the corresponding problem" for inequality constrained problems. Karush only referred to works of Bliss, Dines and Farkas but his thesis was most likely also inspired by Valentine's paper from 1937 "The Problem of Lagrange with Differential Inequalities as added Side Conditions" [118]. Indeed, Karush's thesis resembles Valentine's paper both in its title, its structure, its notation and its approach, so it is very likely, that Karush's thesis was inspired by the work of Valentine. A summary of the thesis of Karush was published, for the first time, as an appendix to Kuhn [82]. The whole thesis is published for the first time in the present collection.

Karush began his thesis with a clear presentation of the main problem: to determine necessary and sufficient conditions for a relative minimum of a function of n variables in the class of points satisfying m inequality conditions. He then listed three results of Bliss [14] on the classical Lagrange optimization problem. One of this results had already been given by Caratheodory [18] and is the "Fritz John version" of the multipliers rule for this problem (see section "Basic Results"). We have to say that the English translation of Caratheodory [18] appeared only in 1965 (vol. I) and 1967 (vol. II). Karush obtained necessary and sufficient optimality conditions which involve, at first, first-order partial derivatives and subsequently second-order partial derivatives. He recalled that the classical Lagrange problem of constrained optimization had, by that time, received a full satisfactory treatment, at least for C^2-class functions.

The problem treated by Karush is what is now known as a typical nonlinear programming problem. The constraints are written in the form $g_i(x) \geq 0,\ i = 1, \ldots, m$, and are considered to be *all active at* x° (feasible point), as, owing to continuity, a constraint of the form $g_i(x^\circ) > 0$ does not impose, locally, any restriction.

The first necessary optimality condition reported by Karush, Proposition 3.1 of his thesis, if read hastily, may give the impression that the author described a Fritz John type result:

> if x° is a local solution of the problem
>
> $$\text{Min } f(x)$$
>
> with $x \in S = \{x \in A, g_i(x) \geq 0,\ i = 1, \ldots m\}$, A open set of $\mathbb{R}^n$ and every function at least differentiable on A, then there exists a nonzero vector $(\lambda_0, \lambda_1, \ldots, \lambda_m)$ such that x° is a stationary point of the following Lagrangian function
>
> $$L(x, \lambda) = \lambda_0 f(x) + \sum_{i=1}^{m} \lambda_i g_i(x).$$

But this, indeed, is *not* the Fritz John theorem, since Proposition 3.1 of Karush gives no information on the sign of the multipliers. Giving so "large" results, the proof of this proposition is quite immediate. For $m < n$ its proof becomes even "unnecessary", as in this case the result is a direct consequence of the necessary optimality conditions of Bliss–Caratheodory. On the other hand, it is always possible to make reference to this theorem, also for the general case, by means of the device, used by Karush, of adding to, or subtracting from each inequality the square of a real number (this was motivated by a similar procedure, used by Valentine [118] in the Calculus of Variations). For example, the constraints $g_i(x) \geq 0,\ i = 1, \ldots m$, can be converted into the following equality constraints

$$g_i(x) - \alpha_i^2 = 0,\ i = 1, 2, \ldots m.$$

Karush's version of what later became known as the Kuhn–Tucker theorem, and what we, in section "Basic Results", have called the Karush–Kuhn–Tucker theorem, appears as Theorem 3.2 in his thesis. Karush introduced the *linearizing cone at* x° which he called "the admissible directions", that is all directions $dx \neq 0$ such that $\nabla g_i(x^\circ)dx \geq 0$. Then he proved the following necessary optimality conditions:

> "Suppose that for each admissible direction ... $[dx]$ there is an admissible arc issuing from x° in the direction dx. Then a first necessary condition for $f(x^\circ)$ to be a minimum is that there exist multipliers $\lambda_i \leq 0$ such that the derivatives L_{x_i} of the function $L = f + \sum \lambda_i g_i$ all vanish at x°" [75, p. 13] (Karush used the symbol F instead of L for the Lagrangian.)

The condition on the linearizing cone is just what is now called "the Kuhn–Tucker constraint qualification", even if it would be more philologically correct to call it "the Karush–Kuhn–Tucker constraint qualification". The reader will note that the sign condition on the multipliers, as it was given by Karush, is opposed to the one given in section "Basic Results", since Karush considered a minimum problem and a Lagrangian function $L = f + \lambda g$.

Thanks to the Farkas lemma, the proof given by Karush is relatively simple. Karush was aware of the role played by the constraint qualification, which he referred to as "property Q", in proving this optimality condition, and he wondered "what the probability is, roughly, that the functions $g_i(x)$ will satisfy property Q". In geometrical terms the Karush–Kuhn–Tucker constraint qualification assumes that each direction vector of the linearizing cone (roughly speaking, for each "tangent direction"), there exists a regular arc of the feasible region which approximates the said "tangent" direction. The conclusion of Karush is quite reassuring: "if the functions $g_i(x)$ are regular enough, it seems that the satisfaction of property Q is not a great restriction" [75, p. 14].

It is now well known (see, e.g., [11, 59, 105]), that property Q can be substituted by other assumptions, varying in generality. Karush considered only one other constraint qualification:

There exists an admissible direction dx such that $\nabla g_i(x^\circ)dx > 0, \ \forall i$.

This condition is now known as *Cottle constraint qualification* [21] or also as *Arrow–Hurwicz–Uzawa constraint qualification* [5]. The Cottle constraint qualification implies the Karush–Kuhn–Tucker constraint qualification, but it is not implied by it. Karush did not prove this statement but built a two-dimensional numerical example which satisfies his property Q, but not his second constraint qualification.

Karush also offered some sufficient optimality conditions, of the first order and of the second order in his thesis. For example, Theorem 4.1 says that x° is a (strict) local minimum point if $m \geq n$, the Jacobian matrix $Jg(x^\circ)$ has rank n and there exists a vector of multipliers $(\lambda_1, \lambda_2, \ldots, \lambda_m)$, with all negative elements, such that $\nabla f(x^\circ) + \sum_{i=1}^m \lambda_i \nabla g_i(x^\circ) = 0$.

In the two last sections of his thesis, Karush gave some second-order optimality conditions; we quote only the Corollary to Theorem 5.1 and to Theorem 6.1. The Corollary asserts that if x° is a local minimum point for the usual problem, where now the functions involved are C^2, and if the Jacobian matrix $Jg(x^\circ)$ has rank m, then a necessary optimality condition is:

$$\nabla f(x^\circ) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^\circ) = 0,$$

$$\lambda_i \leq 0, \ \forall i$$

$$(dx)^T HL(x^\circ)dx \geq 0,$$

for each admissible vector dx satisfying the system $\nabla g_i(x^\circ)dx = 0$, where $L = f + \sum \lambda_i g_i$ and HL is the Hessian matrix of L.

Theorem 6.1 assures that x° is a (strict local) solution of the problem if there exists a vector of multipliers $(\lambda_1, \lambda_2, \ldots, \lambda_m)$ with all negative components, such that

$$\nabla f(x^\circ) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^\circ) = 0,$$

$$(dx)^T HL(x^\circ)dx > 0,$$

for each admissible vector $dx \neq 0$ such that $\nabla g_i(x^\circ)dx = 0$.

We take the opportunity to correct the second-order sufficient optimality conditions given by El Hodiri [39] and reported also in El Hodiri [40,41] and in Takayama [115] for the problem

$$\text{Max}\, f(x) \text{ subject to } g_i(x) \geq 0,\ i = 1, \dots, m.$$

The "generalization" provided by El Hodiri is correct if the active constraints are all associated with *positive* multipliers, with reference to a Lagrangian function of the form $L = f + \lambda g$. See, e.g., Giorgi [58]; for second-order optimality conditions in mathematical programming problems the reader may consult Avriel [9], Ben-Tal [13], Hestenes [68], Fiacco and McCormick [47] and McCormick [94]; this last paper is republished in the present collection. We add that the classical second-order sufficient optimality conditions for a nonlinear programming problem, due to McCormick, can be derived from a paper of Pennisi [104] on Calculus of Variations, here republished.

We note also that Hestenes [67] gave more sophisticated second-order necessary and sufficient optimality conditions, for a general nonlinear programming problem, by means of the so-called "contingent cone" or "Bouligand tangent cone" (see, e.g., [59]).

The Contribution of Fritz John

Fritz John (1910–1994) was a student of Richard Courant at Göttingen University, where he received a Ph.D. in 1933. He was compelled (like Courant) to leave Germany for racial reasons. After a short period at Cambridge University (England), he moved to the United States, where he collaborated from 1943 to 1945 with the U.S. War Department, and then taught at the Universities of Kentucky, of New York and at last at the Courant Institute at New York University.

Fritz John was a first-rate mathematician and his scientific production is vast and generally of a high level. He was mainly concerned with convex geometry, partial differential equations, elasticity theory and numerical analysis. His papers are published by Birkhäuser (see [99]).

Kuhn [82] properly delineates the geometric motivations of the paper of John [71]. The history of this paper is well known, at least among researchers in optimization theory: it was first rejected by the *Duke Mathematical Journal* and later appeared in the Courant anniversary volume of 1948.

The paper of John is quoted by Kuhn and Tucker in their 1951 paper, but they surely worked independently of John, as Kuhn [84, p. 133] writes:

> "Tucker and I were made aware of the work of John when our paper was in galley proofs; the evidence of this fact is that when we inserted a reference to this paper in the references, we did not renumber the bibliography correctly".

John's approach is quite general. The problem (P_3), described in section "Basic Results" of this chapter, is only a "finite version" of the problem considered by John. John divided his paper into two parts: the first is concerned with finding necessary and sufficient conditions for the existence of a minimum to a function subject to inequality constraints, the second part is devoted to two geometrical applications of the conditions found in the first part. We quote directly from John's paper [71, pp. 187–188]:

> "Let R be a set of points x in a space E, 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° of R' has to satisfy in order that
>
> $$M = F(x^\circ) = \min_{x \in R'} F(x).''$$

John then relaxed the generality of the approach by assuming that $E = R^n$ and S is a compact subset of a metric space H. Moreover, John required that F and G are C^1, although with an opening towards what, at the end of the 1970s, would be called "nonsmooth analysis":

> "from the point of view of applications it would seem desirable to extend the methods used here to cases, where the functions involved are not necessarily differentiable" [71, p. 187].

Again we quote directly the paper of Fritz John [71, pp. 188–189]:

> "Theorem 1. Let x° be an interior point of R, and belong to the set R' of all points x of R, which satisfy (1) [the constraints $G(x, y) \geq 0$] for all $y \in S$. Let
>
> $$F(x^\circ) = \min_{x \in R'} F(x).$$
>
> Then there exists a finite set of points $y^1, \ldots, y^s$ in S and numbers $\lambda_0, \lambda_1, \ldots, \lambda_s$ which do not all vanish, such that
>
> $$G(x^\circ, y^r) = 0 \text{ for } r = 1, \ldots, s,$$
> $$\lambda_0 \geq 0,\ \lambda_1 > 0, \ldots, \lambda_s > 0,$$
> $$0 \leq s \quad \leq n,$$

the function

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

has a critical point at x°, i.e. [the partial derivatives are zero at x°]

$$\phi_i(x^\circ) = 0, \text{for } i = 1, \ldots, n.''$$

We note first that the version of this theorem for the usual nonlinear programming problem, like the problem (P_3) of section "Basic Results", was first taken into consideration, at least as far as we are aware, by Cottle [21]. Then, again we point out that, with relation to a finite-dimensional problem with only equality constraints, the Fritz John conditions had already been anticipated by Caratheodory [18] and by Bliss [14]. Always for the case of equality constraints, but for a *vector objective function* (i.e. for a Pareto optimization problem) the John conditions had been anticipated by de Finetti [29, 30]. See Giorgi and Guerraggio [60].

The proof given by John, similarly to the proof given by Karush (also John surely did not know of Karush's thesis), makes use of the separation theorems of the Convex Analysis and of their "algebraic versions", i.e. theorems of the alternative for linear systems. But, instead of using Farkas' lemma as Karush had done, John made references to recent works of Dines [37] and Stokes [112].

John also gave a first-order sufficient condition for local optimality. If we "translate" his sufficient conditions for the usual nonlinear programming problem (P_3) of section "Basic Results", i.e. the problem

$$\text{Max} f(x) \text{ subject to } g_i(x) \leq 0, \ i = 1, 2, \ldots, m,$$

we have the following result (see also Stoer and Witzgall [111]):

If relations i) and ii) of point 8) of "Basic Results" are satisfied by a nonnegative and nonzero vector $(y_0, y_1, \ldots, y_m)$ *at a feasible point* x° *and if the rank of the following matrix, evaluated at* x°

$$\begin{bmatrix} y_0 \frac{\partial f}{\partial x_1} & \cdots & y_0 \frac{\partial f}{\partial x_n} \\ y_1 \frac{\partial g_1}{\partial x_1} & \cdots & y_1 \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ y_m \frac{\partial g_m}{\partial x_1} & \cdots & y_m \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

is n, then x° *is a point of local minimum for* (P_3).

The reader will note the similarity with the first-order sufficient optimality conditions found by Karush.

In the second part of his paper John was concerned with two geometrical applications of the results of part one of his paper. Indeed, these applications seem to be the true motivations that led Fritz John towards a modern mathematical programming problem. Concerning this point the following words of Kuhn [82, p. 15] are sufficiently clear:

> "The main impulse came from trying to prove the theorem (which forms the main application in John (1948) [71]) that asserts that the boundary of a compact convex set S in $\mathbb{R}^n$ lies between two homothetic ellipsoids of ratio $\leq n$, and that the outer ellipsoid can be taken to be the ellipsoid of least volume containing S. The case $n = 2$ had been settled by F. Behrend (1938) with whom John had become acquainted in 1934 in Cambridge, England. A student of John's, O. B. Ader, dealt with the case $n = 3$ in 1938 (Ader (1938)). By that time, John had become deeply interested in convex sets and in the inequalities connected with them".

This problem of the ellipsoid of least volume containing a set S of $\mathbb{R}^n$ is the second geometrical application of John's paper. The first one is concerned with finding a sphere with minimum radius containing a given bounded set S of $\mathbb{R}^n$. This problem goes back to Sylvester [114], who in 1857 published a one sentence note: "It is required to find the least circle which shall contain a given set of points in the plane". The history of this problem is expounded by Kuhn [82]. Finally, we note, following Kjeldsen [76, pp. 341–342], that

> "in John's work, ..., 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. ...reading the second part of the paper,...,which is concerned with the two geometrical applications, ...also explains why John did not touch upon the problem of abnormality [i.e. the first multiplier $\lambda_0 = 0$] and thereby did not consider the problem of constraint qualifications ...[which can] 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".

The Contribution of Kuhn and Tucker

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 he began his Ph.D. study at Princeton University. In 1932 he received the Ph.D. with a thesis in the field of topology, and 2 years later he was appointed assistant professor. In 1938 he became associate professor, and then full professor in 1946.

Tucker was a leading figure in the American mathematical community and he was one of those who made in those years Princeton University a world famous center of mathematical research. In particular, due to the ONR logistic project he made Princeton the center of mathematical programming and game theory in the 1950s, while chairing the Mathematics Department (1953–1963). Research in these areas also took place at the RAND Corporation. Tucker had a tremendous influence on the students who came in contact with him: among his Ph.D. students were

Michael Balinski, John Nash Jr., David Gale, Alan Goldman, John Isbell, Stephen Maurer, Marvin Minsky, Lloyd Shapley.

Harold W. Kuhn was born in 1925 in Santa Monica, California. He received a bachelor's degree in science from the California Institute of Technology in 1947, and then moved to Princeton, where he wrote, in 1950, his Ph.D. thesis, entitled "Subgroup Theorems for Groups Presented by Generators and Relations", under the supervision of Ralph Fox. After some travelling and a 7-year appointment at Bryn Mawr College, Kuhn returned to Princeton as associate professor. Later he declared [84] that it was his good fortune to be in the right place at the right time to work with an exceptional man as A.W. Tucker.

The famous paper of Kuhn and Tucker of 1951 was a continuation of the work they had done in the summer of 1948 on linear programming in the trial project on linear programming and game theory financed by ONR of which we spoke previously. Indeed, the paper begins with a formulation of a linear programming problem and its equivalence with a "saddle-value problem":

> x° is a (global) minimum point for the linear function f, under the restrictions
>
> $$g_i(x) = b_i - \sum_{j=1}^{n} a_{ij} x_j \geq 0,\ x_j \geq 0,\ j = 1, \ldots, n,$$
>
> if and only if there exists a multipliers vector λ°, with nonnegative components and such that (x°, λ°) is a saddle point for the Lagrangian function $L = f + \lambda g$, i.e. the following inequalities hold:
>
> $$L(x, \lambda^\circ) \leq L(x^\circ, \lambda^\circ) \leq L(x^\circ, \lambda),$$
>
> for all nonnegative vectors x and λ.

Kuhn and Tucker then made the connection between the saddle point problem and game theory, since a saddle point for the Lagrangian provides a solution for a related two-person zero-sum game. It also, as they wrote in the introduction, "yields the characteristic duality of linear programming" [85, p. 481]. The declared purpose of the paper is the extension of the above result to a nonlinear programming problem, with nonnegativity conditions on the decision variables. More precisely, they consider the problem to find the maximum of a differentiable function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ subject to the constraints $g_1(x) \geq 0, \ldots, g_m(x) \geq 0,\ x_1 \geq 0, \ldots, x_n \geq 0$. Also the constraints $g_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, $i = 1, \ldots, m$, are assumed to be differentiable.

Their first step is to find necessary and sufficient conditions for a generic differentiable function $\varphi(x, \lambda)$ to have a nonnegative saddle point at (x°, λ°) :

$$\varphi(x, \lambda^\circ) \leq \varphi(x^\circ, \lambda^\circ) \leq \varphi(x^\circ, \lambda),$$

for all $x \geq 0,\ \lambda \geq 0$. They denote ${\varphi^\circ}_x, {\varphi^\circ}_\lambda$ the partial derivatives of φ evaluated at x° and λ° and they proved that the conditions

$$\varphi^{\circ}{}_{x} \leq 0,\ \varphi^{\circ}{}_{x} x^{\circ} = 0,\ x^{\circ} \geq 0 \tag{1}$$

$$\varphi^{\circ}{}_{\lambda} \geq 0,\ \varphi^{\circ}{}_{\lambda} \lambda^{\circ} = 0,\ \lambda^{\circ} \geq 0 \tag{2}$$

are necessary for $(x^{\circ}, \lambda^{\circ})$ to be a nonnegative saddle point for $\varphi(x, \lambda)$. The same conditions (1) and (2) are sufficient for the same problem, if the following conditions are added:

$$\varphi(x, \lambda^{\circ}) \leq \varphi(x^{\circ},\ \lambda^{\circ}) + \varphi^{\circ}{}_{x}(x - x^{\circ}) \tag{3}$$

$$\varphi(x^{\circ},\ \lambda) \geq \varphi(x^{\circ},\ \lambda^{\circ}) + \varphi^{\circ}{}_{\lambda}(\lambda - \lambda^{\circ}) \tag{4}$$

for any $x \geq 0,\ \ \lambda \geq 0$.

Kuhn and Tucker felt compelled to, justify relations (3) and (4), so that they do not seem "as artificial as may appear at first sight", but that today—after more than half a century—we recognize as characterizations of concavity (convexity) of the differentiable function φ_{λ} with respect to x (with respect to λ). Conditions (1) and (2) later become known as "the Kuhn–Tucker conditions" (for the problem considered) and now are also known as "the Karush–Kuhn–Tucker conditions".

As a second step they introduced, just with this name, a "constraint qualification", which is the same regularity assumption on the constraints we have already met in the thesis of Karush. This constraint qualification is needed to get rid of pathological situations, on the boundary of the feasible region, "such as an outward pointing cusp". We know also that a constraint qualification makes $y_0 > 0$ in the first Fritz John necessary optimality conditions. Let us consider, e.g., the following example, provided by Kuhn and Tucker:

> the feasible region in $\mathbb{R}^2$ is given by $g_1(x) = x_1 \geq 0$; $g_2(x) = x_2 \geq 0$; $g_3(x) = (1 - x_1)^3 - x_2 \geq 0$, and the boundary point $x^{\circ} = (1, 0)$. The constraint qualification is not satisfied, since the feasible region does not contain an arc issuing from this point in the direction $dx_1 = 1, dx_2 = 0$. If the problem is to maximize $f(x) = x_1$, we see that $x^{\circ} = (1, 0)$ is the optimal solution, but the Kuhn-Tucker conditions (1)-(2) are not satisfied at the optimal point.

Equipped with all these conditions and assumptions, Kuhn and Tucker stated and proved the first two of their theorems:

Theorem 1. *In order that x° be a solution of the maximum problem, where the constraint qualification holds, it is necessary that x° and some λ° satisfy conditions (1) and (2) for some $\varphi(x, \lambda) = f(x) + \lambda g(x)$.*

Theorem 2. *In order that x° be a solution of the maximum problem, is sufficient that x° and some λ° satisfy conditions (1)–(3) for $\varphi(x, \lambda) = f(x) + \lambda g(x)$.*

It is worthwhile to examine briefly the proof given by Kuhn and Tucker on the necessary conditions: it appears at once that the relevant steps are the same ones used by Karush in his proof of Theorem 3.2 of his thesis. First of all, thanks to the constraint qualification, every direction dx of the linearizing cone verifies the inequality $\nabla f(x^\circ)dx \leq 0$, and at this point Kuhn and Tucker invoked Farkas' lemma to ensure the existence of nonnegative multipliers.

Kuhn and Tucker's approach to prove the sufficiency part is different from the one of Karush. Kuhn and Tucker choose the saddle point formulation and then establish a sort of "loop" between their maximization problem, the nonnegative saddle point problem for the Lagrangian function $L(x, \lambda) = f + \lambda g$, and the conditions (1)–(2) and (3); (condition (4), concerning the convexity of L with respect to λ is obviously always satisfied).

The third and last step performed by Kuhn and Tucker in order to reach the purpose of the paper as they had declared it in section "Introduction", was to place restrictions on the functions to ensure the equivalence of solutions of the maximum problem:

$$\text{Max} f(x) \text{ subject to } g_i(x) \geq 0,\ i = 1, \ldots, m,\ x \geq 0$$

and the saddle value problem for the Lagrangian. Indeed, after their Theorem 2, they proved an "equivalence theorem", i.e.

Theorem 3. *Let the functions* $f(x),\ g_1(x), \ldots, g_m(x)$ *be concave as well as differentiable for* $x \geq 0$. *Let, as before, the constraint qualification hold. Then,* x° *is a solution of the maximum problem if and only if* x° *and some* λ° *give a solution of the saddle value problem for* $\varphi(x, \lambda) = f(x) + \lambda g(x)$.

Section 6 of the paper of Kuhn and Tucker is concerned with an extension to a vector maximum problem, Sect. 7 treats the so-called "minimum component maximum problem" and the final Sect. 8 treats other types of constraints, i.e. without $x \geq 0$ and equality constraints (with and without $x \geq 0$).

As for what concerns Sect. 6 on vector optimization, we may say that it is one of the first accurate mathematical treatments of this kind of problems (see also de Finetti [29, 30] and, for an historical analysis of the contributions of this author to vector optimization problems, Giorgi and Guerraggio [60]). We have already said of the difficulty one has in the definition itself of a maximum point ("efficient point" in the literature) for a vector-valued function. Usually, a feasible point x° is "Pareto efficient" or simply "efficient" for a vector maximization problem, when, for each feasible vector x, there exists an index i such that $f_i(x) < f_i(x^\circ)$. Kuhn and Tucker were the first authors to present a more restrictive definition of efficiency, a definition which rules out some anomalies occurring in the usual definition of efficiency (given above). A feasible point x° is said to be *properly efficient* (in the sense of Kuhn and Tucker) when it is efficient and for no direction dx of the linearizing cone at x° we have $\nabla f_i(x^\circ)dx \geq 0$, with at least one index j for which $\nabla f_j(x^\circ)dx > 0$ holds.

In the next decades, after this first definition of Kuhn and Tucker, numerous other definitions of proper efficiency were proposed, almost all more restrictive than the one of Kuhn and Tucker. Some of these definitions are again based on the notion of "trade-off", whereas others are of a more geometrical nature and refer to the image space (for the various definitions of proper efficiency and for their comparison, the reader may consult Guerraggio et al. [63]).

Also the treatment given by Kuhn and Tucker of the vector optimization problem, follows closely the scheme adopted for the scalar problem. A necessary condition for x° to be a properly efficient point (in a maximum vector problem) for $f = (f_1, f_2, \dots, f_p)$, under the restrictions $g_i(x) \geq 0, i = 1, \dots, m;\ x \geq 0$, is that there exists a positive vector $\theta^\circ \in \mathbb{R}^p$ and a non negative vector $\lambda^\circ \in \mathbb{R}^m$, such that the function $L = \theta f + \lambda g$ satisfies conditions (1) and (2) at $(x^\circ, \theta^\circ, \lambda^\circ)$. The same conditions become sufficient if, besides (1) and (2), also (3) is satisfied. Finally, under concavity assumptions on each f_s and each g_i, x° is a properly efficient (maximum) point if and only if (x°, λ°) is a saddle point for the Lagrangian function $L = \theta^\circ f + \lambda g$. About these subjects, we may note that the constraint qualification is no longer needed, as in the vector case we always obtain that $\theta^\circ \neq 0$, because now we require a stronger definition of a maximum point: the point x° must be properly efficient. Kuhn and Tucker also provided a useful numerical example which justifies their definition of proper efficiency.

If we compare the theorems of Karush, and Kuhn and Tucker with Fritz John's theorem one gets the impression that the two statements are quite similar. Indeed, if we add to the assumptions of the Fritz John theorem a constraint qualification, we are sure that the John multiplier y_0, associated with the objective function, is different from zero, i.e. positive: therefore we obtain the Karush–Kuhn–Tucker theorem.

However, none of these mathematicians moved further in this direction. Karush made no remarks on what would happen if his "property Q" does not hold. As we have seen, due to the nature of the two applications to convex sets, Fritz John did not even consider the fact that his first multiplier could be zero resulting in a Lagrangian function without the objective function. Also Kuhn and Tucker made no comments on the non validity of any constraint qualification and on the consequences of this fact. Indeed, the studies on "non regular mathematical programming problems", i.e. problems where no constraint qualification holds, are recent: see, for example, the book of Arutyunov [8].

Some Further Considerations and Conclusions

Another important contribution to the theory of nonlinear programming at its beginning is the non published paper of Slater [109]; in this paper the findings of Kuhn and Tucker are quoted, 1 year before its publication. Slater reconsidered the relationship between a nonlinear programming problem and a saddle value problem for a suitable Lagrangian function. The "novelty" is that Slater did not assume any differentiability of the functions involved, which are only required to be concave (or convex, according to the type of problem). Moreover, Slater introduced a new constraint qualification, which is more easily checked than the one of Kuhn and Tucker; it is now universally known as the "Slater condition".

The proof of Slater is quite intricate and makes use of the Kakutani fixed point theorem. The results of Slater have been proved in a more elegant and elementary way by Uzawa [117] and by Karlin [74] by means of classical separation theorems of Convex Analysis. The constraint qualification used by Karlin is formally different from the one of Slater, but it can be shown that the two conditions are equivalent: see Hurwicz and Uzawa [70]. For some comments and corrections of the paper of Uzawa [117] see Moore [97].

Another important paper from the middle of the 1950s on nonlinear programming is "Reduction of Constrained Maxima to Saddle-Point Problems" by Arrow and Hurwicz [3]. As we have seen, Kuhn and Tucker related, under concavity assumptions, the solution of a nonlinear programming problem (research of constrained maxima) to the solution of a saddle point problem. In this paper Arrow and Hurwicz proved that the mentioned result of Kuhn and Tucker can be extended by relaxing the concavity assumptions, at the expense of obtaining the results only locally. This is perhaps the first paper where a "modified Lagrangian function" is introduced in the study of saddle points problems. Subsequently K.J. Arrow reconsidered these questions in 1958 (see Solow and Arrow [113]) and, within a more general framework, in 1973 (see Arrow et al. [7]).

The question of duality was apparently the initial motivation for Tucker to generalize the results on linear programming obtained by himself, Kuhn, and Gale in the summer of 1948. Even so, Kuhn and Tucker did not prove a duality result in their "Nonlinear programming" paper from 1951, but they did prove, as we have seen, the equivalence between the nonlinear programming problem with concavity and differentiability conditions of the involved functions and the saddle value problem for the corresponding Langragian function, which indeed does suggest the existence of a duality result. The first such result for nonlinear programming was developed at Princeton University, not by Tucker's group but by the Danish expert on the theory of convexity, Werner Fenchel, who was visiting in Princeton in the spring of 1951. Tucker invited Fechel to give a series of lectures on convexity at the mathematics department; Fenchel developed the first duality theorem for nonlinear programming during the course of preparing these lectures (see [80]).

Fenchel used a result from a paper he had published in 1949 in which he introduced the concept of conjugate convex functions. Here he had shown that to each convex function $f(x_1, \ldots, x_n)$, defined in a convex subset G, of $\mathbb{R}^n$ and satisfying some conditions of continuity, there corresponds in a unique way a convex subset, Γ, of $\mathbb{R}^n$ and a convex function $\phi(\xi_1, \ldots, \xi_n)$, defined in Γ and with the same properties as f, such that the inequality

$$x_1\xi_1 + \ldots + x_n\xi_n \leq f(x_1, \ldots, x_n) + \phi(\xi_1, \ldots, \xi_n),$$

is fulfilled for all points $x = (x_1, \ldots, x_n)$ in G and all points $\xi = (\xi_1, \ldots, \xi_n)$ in Γ. The correspondence between G, f and Γ, ϕ is symmetric, and Fenchel called the functions f and ϕ for conjugate functions [45, pp. 73–75].

Fenchel defined Γ to be the set of all points ξ, for which the function

$$\sum_{i=1}^{n} x_i \xi_i - f(x)$$

is bounded from above in G, and defined

$$\phi(\xi) = \sup_{x \in G}(\sum_{i=1}^{n} x_i \xi_i - f(x)).$$

The definition of ϕ then ensures that the above inequality makes sense.

He used this concept of conjugate functions to develop his duality result for nonlinear programming. His notes from the lectures were published by ONR in 1953, and here Fenchel's duality result for "A generalized Programming Problem" as he called it, appears in the last section. Here Fenchel argued that also to each concave function there corresponds a conjugate function, and he considered a closed convex function $f(x)$, defined on a convex set C in $\mathbb{R}^n$, and a closed concave function $g(x)$, defined on a convex set D. He let $\phi : \Gamma \longrightarrow \Re$ and $\psi : \Delta \longrightarrow \Re$ denote the conjugates of f and g respectively, which allowed him to formulate the following two (dual) problems [46, p. 105].

> "PROBLEM I: To find a point x^0 in $C \cap D$, such that $g(x) - f(x)$ as a function in $C \cap D$ has a maximum at x^0.
> PROBLEM II: To find a point ξ^0 in $\Gamma \cap \Delta$, such that $\phi(\xi) - \psi(\xi)$ as a function in $\Gamma \cap \Delta$ has a minimum at ξ^0."

From this he derived the Fenchel duality theorem:

> If the sets $C \cap D$ and $\Gamma \cap \Delta$ are non-empty then $g(x) - f(x)$ is bounded above, $\phi(\xi) - \psi(\xi)$ is bounded below and under some further conditions on the origin in relation to the sets C, D, Γ and Δ, then [46, pp. s.105–106]:
>
> $$\sup_{x \in C \cap D}(g(x) - f(x)) = \inf_{\xi \in \Gamma \cap \Delta}(\phi(\xi) - \psi(\xi)).$$

Fenchel did not explore this further, but the lecture notes from his course became a huge source of inspiration. They had quite an influence on the following development of the theory of convexity in the USA and in the development of convex programming.

The book of 1958 edited by Arrow et al. [4] is perhaps the first important collection of papers concerned with nonlinear programming (and also with linear programming). In this collection we find, among others, the paper of L. Hurwicz "Programming in Linear Spaces" (reproduced in the present collection), which is one of the first works on optimization problems (both scalar and vector) defined in topological spaces: here the Karush–Kuhn–Tucker theorem is generalized in various forms, with and without the use of differentials. This book collects also the studies of the Stanford mathematical economics community of those years on the gradient methods in mathematical programming, which go back to a seminal paper of Arrow and Hurwicz of 1950 (reproduced in the present collection). We find also some first interesting economic applications of nonlinear programming: a good motivation from economics is always given (we point out that K.J. Arrow and L. Hurwicz have been awarded the Nobel Prize in Economics), but predominantly the book is on mathematical results.

Other classical papers in nonlinear programming at the beginning of the 1960s are Arrow et al. [5], who weakened the Kuhn–Tucker constraint qualification and provided an analysis of the various constraint qualifications proposed since then, and Arrow and Enthoven [6], who first generalized the sufficient optimality conditions of Kuhn and Tucker, originally established under the assumption of concavity, to quasiconcave functions. Since then the literature on nonlinear programming has grown at an exponential rate; and an entire volume would be necessary just to list all the contributions that have appeared to date.

At this point it is possible to draw some conclusions on the originality of the papers of Karush [75] and of Kuhn and Tucker [85] (see also the works of Kjeldsen [76, 77, 80]). Kuhn and Tucker also obtained "their" conditions for the case considered by Karush, i.e. when the condition $x \geq 0$ is absent (see Sect. 8 of Kuhn and Tucker [85]). Therefore, the main findings of Karush on one side and of Kuhn and Tucker on the other side, are similar, very similar, almost equal and also the methods of proof are similar, both based on the same constraint qualification and on the use of Farkas' lemma.

What is different and allows us to speak of "independent" contributions is the context in which their findings took place, both scientifically and socially. The way followed by Karush is perhaps more "abstract" and, in a sense, more modern, because it underlines the analogy between nonlinear programming problems and the classical unconstrained optimization problems. Indeed Karush gave necessary optimality conditions of the first and of the second order and sufficient optimality conditions of the first and of the second order. In this sense the contribution of Karush is closer to several more recent presentations of the theory of nonlinear programming. On the other hand, Kuhn and Tucker turned their attention to the saddle value problem and to the consequent generalization of what had already been proved for a linear programming problem. Therefore, their approach was more "productive" for practical applications, particulary in Economics, even though their work as such was not motivated by application, but by the desire for generalization for the sake of gaining insights into the mathematical structure of inequality constrained optimization problems.

Indeed, the paper of Kuhn and Tucker almost at once gave rise to a myriad of contributions, whereas the papers of Karush and of Fritz John did not have any influence on the emergence and early development of nonlinear programming. The trivial reasons for Karush's work are that his thesis was never published, and he was an unknown graduate student, but none of these reasons apply to the case of Fritz John. First, his paper was published in 1948, but the paper became interesting only after nonlinear programming had become an autonomous field of research. Second, Fritz John was a well known mathematician at the time, and he was linked with the famous school of Richard Courant. A more reasonable explanation is the fact, that W. Karush and F. John were moved by theoretical aims, in scientific environments of the calculus of variations and the theory of convexity. Kuhn and Tucker, even though their work with nonlinear programming was not motivated directly by applications, but by "pure" mathematical investigations of generalizations, they nevertheless operated in an environment where practical applications to Economics, Management Science, Operations Research, Logistics, Engineering, etc., were a source of new researches.

Kuhn [84, p. 134] recognized three factors which he found were significant for the rapid development of nonlinear programming after the 1951 paper of Kuhn and Tucker:

> "First, the model of nonlinear programming was flexible enough to encompass a large class of real-life problems that had not been adequately treated by the techniques then available. In social terms, after the successes of operations research in the Second World War, a number of major industries were willing to try out this new model. Second, the necessary conditions established by Karush, Kuhn and Tucker formed the starting point for a large number of algorithms to solve nonlinear programs. Third, and perhaps the most necessary factor, the first half of the 1950s saw the development and rapid expansion of computers that could be programmed to solve this sort of problem".

And we might want to add at least one more factor, namely the enormous amount of research money the government gave to academic research in the post war period in the USA.

We have already mentioned two great pioneers of mathematical programming, T.C. Koopmans and L.V. Kantorovich who in 1975 received the Nobel Prize in Economics. We can add also the names of K.J. Arrow, who was awarded the Nobel Prize in Economics in 1972, of H.M. Markowitz, who was awarded the same prize in 1990, for his contributions to quadratic programming, within his famous model of "portfolio selection", of L. Hurwicz, who was awarded the Nobel Prize in economics in 2007, and of L. S. Shapley and A. E. Roth who were awarded the prize in 2012.

With Kuhn's and Tucker's paper the theory of nonlinear programming became an autonomous research field. In a very short time it reached a development (also towards algorithmic aspects) which is by now impossible to describe in a single book let alone in a single paper, if one wants to present its present status in a sufficient analytic way.

We add at this point some brief final conclusions. We think that the history of the emergence of mathematical programming (linear and nonlinear) is surely a "great history", both for the relevance of its contents (melting pot of several research sectors and stimulating observatory for the didactics), and for the highly esteemed names of great mathematicians and Nobel prize winners we meet in this history.

In describing the emergence of mathematical programming, we have been compelled to stress some typical aspects of this birth: the influence of some mathematical tools, such as Linear Algebra and Convex Analysis, which proved crucial for its take-off; the convergence of various mathematical motivations (calculus of variations, geometrical problems, theory of games, problems of operations research, etc.) which have led to a "mix", that has at once proved to be a new and original field of research, and finally, the influence of the historical circumstances of the Second World War and the civilian mobilization of U.S. scientists and mathematicians for the war effort along with the military support of science in the U.S.A in the post war period, see also Kjeldsen [79].

From this point of view the emergence of mathematical programming can be considered as an expression of that great analytical tradition which started from the key concepts of "function" and differential calculus (with all their modern developments) to offer the main tool in supporting the study of various problems arising from real life (especially from social and economic sciences).

Like every history of quite recent facts, also the history of nonlinear programming covers perhaps a too brief period, in order to allow the researcher to see in a right perspective all its fruits and to distinguish between important results and short-lived researches. The history of mathematical programming is a history of a topic "in fieri", but just for this reason it appears more stimulating, as its topic is still in progress and not yet embalmed.

Finally, another exciting aspect for those concerned with the history of a contemporary field of research, is the possibility to try to contribute to its construction, and not to be only a passive beneficiary of the results of other researchers.

For the present collection we have chosen the following contributions, as representative of the emergence process and first developments of nonlinear programming: (in alphabetical order of the authors)

1. K. J. ARROW and L. HURWICZ, A gradient method for approximating saddle points and constrained maxima, RAND Corporation, paper P-223, June 1951.
2. K. J. ARROW and L. HURWICZ, Reduction of constrained maxima to saddle-point problems; in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Vol. 5, 1955, 1–20.
3. K. J. ARROW, L. HURWICZ and H. UZAWA, Constraint qualifications in maximization problems, Naval Res. Logist. Quart., 8, 1961, 175–191.
4. G. BLISS, Normality and abnormality in the calculus of variations, Trans. Amer. Math. Soc., 43, 1938, 365–376.
5. R. W. COTTLE, A theorem of Fritz John in mathematical programming, RAND Corporation Memorandum RM-3858-PR, 1963.
6. W. FENCHEL, On conjugate convex functions, Canad. J. Math., 1, 1949, 73–77.
7. L. HURWICZ, Programming in linear spaces; in K. J. Arrow, L. Hurwicz and H. Uzawa (Eds.), Studies in Linear and Nonlinear Programming, Stanford University Press, Stanford, 1958, 38–102.
8. F. JOHN, Extremum problems with inequalities as subsidiary conditions; in K. O. Friedrichs, O. E. Neugebauer and J. J. Stoker (Eds.), Studies and Essays - Presented to R. Courant on his 60th Birthday, January 8, 1948, Wiley Interscience, New York, 1948, 187–204.

9. W. KARUSH, Minima of Functions of Several Variables with Inequalities as Side Conditions, MSc Thesis, Department of Mathematics, University of Chicago, 1939.
10. H. W. KUHN and A. W. TUCKER, Nonlinear programming; in J. Neyman (Ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 1951, 481–492.
11. G. P. McCORMICK, Second-order conditions for constrained minima, SIAM J. on Applied Mathematics, 15, 1967, 641–652.
12. L. L. PENNISI, An indirect sufficiency proof for the problem of Lagrange with differential inequalities as added side conditions, Trans. Amer. Math. Soc., 74, 1953, 177–198.
13. M. SLATER, Lagrange multipliers revisited, Cowles Commission Discussion Paper 403, 1950.
14. H. UZAWA, The Kuhn-Tucker theorem in concave programming; in K. J. Arrow, L. Hurwicz and H. Uzawa (Eds.), Studies in Linear and Nonlinear Programming, Stanford University Press, Stanford, 1958, 32–37.
15. F. A. VALENTINE, The problem of Lagrange with differential inequalities as added side conditions; in Contributions to the Calculus of Variations (1933–1937), University of Chicago Press, Chicago, 1937, 407–448.

The book is completed with an Appendix containing the following contributions:

(A) L. L. DINES and N. H. McCOY, On linear inequalities, Trans. Roy. Soc. Canada, 27, 1933, 37–70 [This paper was quoted by Karush in his thesis].

(B) H. W. KUHN, Nonlinear programming: a historical view; in R. W. Cottle and C. E. Lemke (Eds.), Nonlinear Programming, SIAM-AMS Proceedings, Vol. XI, American Mathematical Society, Providence, 1976, 1–26. [In this paper, for the first time, Kuhn admits the priority of the results of Karush and his ignorance, since then, of these results. In the Appendix of this paper there is a summary of the thesis of Karush: it is the first time that the main results of Karush are published. In the present book we omit this Appendix of Kuhn's paper].

(C) J.-L. LAGRANGE, Recherches sur la Méthode de Maximis et Minimis, Miscellanea Taurinesia, Tom 1, 1762, pages 173–195.

References

1. O. B. ADER (1938), An affine invariant of convex regions, Duke Math. J., 4, 291-299.
2. D. J. ALBERS and G. L. ALEXANDERSON (1985) (Eds.): *Mathematical People, Profiles and Interviews*. Boston: Birkhäuser, 1985
3. K. J. ARROW and L. HURWICZ (1955), Reduction of constrained maxima to saddle-point problems; in Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, Vol. 5, 1-20.

4. K. J. ARROW, L. HURWICZ and H. UZAWA (Eds.) (1958), Studies in Linear and Nonlinear Programming, Stanford Univ. Press, Stanford.
5. K. J. ARROW, L. HURWICZ and H. UZAWA (1961), Constraint qualifications in maximization problems, Naval Res. Logist. Quart., 8, 175-191.
6. K. J. ARROW and A. C. ENTHOVEN (1961), Quasi-concave programming, Econometrica, 29, 779-800.
7. K. J. ARROW, F. J. GOULD and S. M. HOWE (1973), A General Saddle Point Result for Constrained Optimization. *Mathematical Programming* 5:225–234.
8. A. V. ARUTYUNOV (2000), Optimality Conditions: Abnormal and Degenerate Problems, Kluwer Acad. Publishers, Dordrecht.
9. M. AVRIEL (1976), Nonlinear Programming, Prentice Hall, Englewood Cliffs, N. J.
10. M. AVRIEL, W. E. DIEWERT, S. SCHAIBLE and I. ZANG (Eds.) (1988), Generalized Concavity, Plenum Press, New York.
11. M. BAZARAA and C. M. SHETTY (1976), Foundations of Optimization, Springer Verlag, Berlin.
12. F. BEHREND (1938), Ueber die kleinste umbeschriebene und die grösste einbeschriebene Ellipse eines konvexen Bereiches, Math. Ann., 115, 379-411.
13. A. BEN-TAL (1980), Second-order and related optimality conditions in nonlinear programming, Journal of Optimization Theory and Applications, 31, 143-165.
14. G. BLISS (1938), Normality and abnormality in the calculus of variations, Trans. Amer. Math. Soc., 43, 365-376.
15. G. BOOLE (1854), On the conditions by which the solutions of questions in the theory of probabilities are limited, Philosophical Magazine, 8, 91-98.
16. R. J. BOSCOVICH (1757), De litteraria expeditione per Pontificiam ditionem, Bononiensi Scientiarum et Artium Instituto atque Academia Commentarii, 4, 353-396.
17. A. CAMBINI and L. MARTEIN (2009), Generalized Convexity and Optimization. Theory and Applications, Springer, Berlin.
18. C. CARATHEODORY (1935), Variationsrechnung und Partielle Differential Gleichungen Erster Ordnung, B. G. Tubner, Berlin. English translations: Calculus of variations and Partial Differential Equations of the First-Order, Holden day, San Francisco, Cal., Vol. I (1965), Vol. II (1967).
19. W. B. CARVER (1922), Systems of linear inequalities, Annals of Mathematics, 23, 212-220.
20. A. CHARNES and W. W. COOPER (1961): "On some Works of Kantorovich, Koopmans and Others." *Management Science*, 8, 1961, pp. 246–263.
21. R. W. COTTLE (1963), A theorem of Fritz John in mathematical programming, RAND Corporation Memorandum RM-3858-PR.
22. R. W. COTTLE (ed.) (2003), The Basic George B. Dantzig, Stanford University Press, Stanford, California.
23. R. W. COTTLE (2010), A brief history of the International Symposia on Mathematical Programming, Mathematical Programming, Series B, 125, 207–233.
24. R. W. COTTLE (2012), William Karush and the KKT Theorem, in M. GROTSCHEL (ed.) Optimization Stories, Documenta Mathematica, Bielefeld, Germany, pp. 255–269.
25. A. A. COURNOT (1826), Sur le calcul des conditions d'inégalité annoncé par M. Fourier, Bulletin Universel des Sciences et de l'Industrie, Sciences Mathématiques, 6, 1-8.
26. A. A. COURNOT (1827), Extension du principe des vitesses virtuelles au cas où les conditions de liaison du système sont exprimées par des inégalités, Bulletin Universel des Sciences et de l'Industrie, Sciences Mathématiques, 8, 165-170.
27. G. B. DANTZIG (1991), Linear programming - The story how it began: some legends, a little about its historical significance, and comments about where its many mathematical programming extensions may be headed. In J. K. Lenstra and others (1991), 19-31.
28. G. DARBOUX (1890): "Avertissement." In J. B. J. FOURIER: *Oeuvres*, Vol. 2, (ed.): G. DARBOUX, Paris: Gauthier-Villars.
29. B. DE FINETTI (1937a), Problemi di "optimum", Giornale dell'Istituto Italiano degli Attuari, 6, 48-67.

30. B. DE FINETTI (1937b), Problemi di “optimum” vincolato, Giornale dell’istituto Italiano degli Attuari, 6, 112-126.
31. B. DE FINETTI (1949), Sulle stratificazioni convesse, Ann. Mat. pura e Appl., 30, 173-183.
32. C. DE LA VALLÉE POUSSIN (1910–1911), Sur la méthode de l’approximation minimum, Annales de la Socété Scientifique de Bruxelles, 35, 1-16.
33. G. C. F. M. DE PRONY (1804), Recherches physico-mathématiques sur la théorie des eaux courantes, Imprimerie Impériale, Paris.
34. L. L. DINES (1918–1919), Systems of linear inequalities, Annals of Mathematics, 20, 191-199.
35. L. L. DINES (1925), Definite linear inequalities, Annals of Mathematics, 27, 57-64.
36. L. L. DINES (1927), Sets of functions of a general variable, Transactions of the American Math. Society, 29, 463-470.
37. L. L. DINES (1936), Convex extension and linear inequalities, Bull. Amer. Math. Soc., 42, 353-365.
38. L. L. DINES and N. H. McCOY (1933), On linear inequalities, Trans. Roy. Soc. Canada, 27, 37-80.
39. M. A. EL-HODIRI (1967), The Karush characterization of constrained extrema of functions of a finite number of variables, Research memorandum A.3, UAR Ministry of Treasury, Cairo.
40. M. A. EL-HODIRI (1971), Constrained Extrema: Introduction to the Differentiable Case with Economic Applications, Springer, Berlin (Originally: M. A. El-Hodiri (1966), Constrained Extrema of Functions of a Finite Number of variables: Review and Generalizations, Krannert Institute Paper 141, Purdue University, Purdue).
41. M. A. EL-HODIRI (1991), Extrema of Smooth Functions, Springer Verlag, Berlin.
42. M. ELLMAN (1973), Planning Problems in the USSR, Cambridge Univ. Press, Cambridge.
43. J. FARKAS (1895) Über dieAnwendung des Mechanischen Princips von Fourier. *Mathematische und Naturwissenschaftliche Berichte aus Ungarn*, 12, 263–281.
44. J. FARKAS (1901), Theorie der einfachen Ungleichungen, Journal für die reine und angewandte Mathematik, 124, 1-27.
45. W. FENCHEL (1949), On conjugate convex functions, Canad. J. Math., 1, 73-77.
46. W. FENCHEL (1953), Convex Cones, Sets and Functions, Lecture Notes by D. Blackett, Princeton University, Princeton (from lectures delivered in 1951).
47. A. V. FIACCO and G. P. McCORMICK (1968), Nonlinear Programming: Sequential Unconstrained Minimization Techniques, J. Wiley, New York.
48. A. V. FIACCO (1983), Introduction to Sensitivity and Stability Analysis in Nonlinear programming, Academic Press, New York.
49. J. B. J. FOURIER (1798), Mémoire sur la statique, contenant la démonstration du principe de vitesses virtuelles et la théorie des moments, Journal de l’Ecole Politechnique, 2, cah. 5, 20-60. Also in : Oeuvres, Vol. 2, G. Darboux Ed., Gauthier Villars, Paris, 1890.
50. J. B. J. FOURIER (1823). Extract from reference 5 for 1823. *Histoire de l’Académie Royale des Sciences*, 6, 1827, pp. xxix–xli.
51. J. B. J. FOURIER (1824): (Extract from) Analyse des travaux de l’Académie Royale des Sciences, pendant l’année 1824. Partie mathématque. *Histoire de l’Académie Royale des Sciences*, 7, 1827, pp. xlvii–lv.
52. J. B. J. FOURIER (1890), *Oeuvres*, Vol. 2, (Ed.): G. Darboux, Paris: Gauthier-Villars.
53. J. B. J. FOURIER (1902), Auflösung der Bestimmten Gleichungen. (Ed.): Wilhelm Engelsmann, Leipzig, 1902.
54. O. L. FRANKSEN (1985a), Irreversibility by Inequality Constraints I: On Fourier’s Inequality. *Systems Analysis, Modelling, Simulation*, 2, 137–149.
55. O. L. FRANKSEN (1985b), Irreversibility by Inequality Constraints II: The Second Law of Thermodynamics. *Systems Analysis, Modelling, Simulation*, 3, 251–273.
56. O. L. FRANKSEN (1985c), Irreversibility by Inequality Constraints III: Towards Mathematical Programming. *Systems Analysis, Modelling, Simulation*, 4, 337–359.

57. J. W. GIBBS (1879), On the fundamental formulae of dynamics, American journal of Mathematics, 2, 49-64.
58. G. GIORGI (2003), Some remarks on second-order optimality conditions, International Journal of Mathematical Sciences, 2, 97-102.
59. G. GIORGI, A. GUERRAGGIO and J. THIERFELDER (2004), Mathematics of Optimization: Smooth and Nonsmooth Case, American Elsevier, Amsterdam.
60. G. GIORGI and A. GUERRAGGIO (2010), De Finetti's contributions to Pareto optimization problems and to generalized convexity. In: Bruno de Finetti, Istituto Lombardo Accademia di Scienze e Lettere. Incontri di Studio, 45, Edizioni Universitarie di Lettere Economia Diritto, Milan, 73-92.
61. I. GRATTAN-GUINNESS (1976), Joseph Fourier's anticipation in nonlinear programming, Operational Research Quarterly, 21, 361-364.
62. I. GRATTAN-GUINNESS (1994), "A new type of question": On the prehistory of linear and nonlinear programming, 1770-1940. In E. Knobloch and D. E. Rowe (Eds.), The History of Modern Mathematics, Vol.III, Academic Press, Boston, 43-89.
63. A. GUERRAGGIO, E. MOLHO and A. ZAFFARONI (1994), On the notion of proper efficiency in vector optimization, Journal of Optimization Theory and Applications, 82, 1-21.
64. A. GUERRAGGIO and E. MOLHO (2004), The origins of quasi-concavity: a development between mathematics and economics, Historia Mathematica, 31, 62-75.
65. A. HAAR (1924), Ueber lineare Ungleichungen, Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio Scientiarum Mathematicarum, 2, (1924-1926), 1-14.
66. N. HADJISAVVAS, S. KOMLOSI and S. SCHAIBLE (Eds.) (2005), Handbook of Generalized Convexity and Generalized Monotonicity, Springer, New York.
67. M. R. HESTENES (1966), Calculus of Variations and Optimal Control Theory, J. Wiley, New York.
68. M. R. HESTENES (1975), Optimization Theory. The Finite Dimensional Case, J. Wiley & Sons, New York.
69. F. L. HITCHCOCK (1941), Distribution of a product from several sources to numerous localities, J. Math. Phys., 20, 224-230.
70. L. HURWICZ and H. UZAWA (1958), A note on the Lagrangian saddle points; in K. J. Arrow, L. Hurwicz and H. Uzawa (Eds.), Studies in Linear and Nonlinear Programming, Stanford Univ. Press, Stanford, 103-113.
71. F. JOHN (1948), Extremum problems with inequalities as subsidiary conditions; in K. O. FRIEDRICHS, O. E. NEUGEBAUER and J. J. STOKER (Eds), *Studies and Essays - Presented to R. Courant on his 60th Birthday January 8*. New York: Interscience, 187–204.
72. L. V. KANTOROVICH (1960), Mathematical methods of organizing and planning production, Management Science, 6, 366-422. (English translation of "Mathematicheskie metody organizatriya planirovanija proizvosdka", Leningrad, 1939).
73. L. V. KANTOROVICH (1965), The Best Use of Economic Resources, Oxford University Press, Oxford. (Russian original publication: Academy of Sciences, Moscow, 1959. French translation: Calcul Economique et Utilisation des Ressources, Dunod, Paris, 1963).
74. S. KARLIN (1959), Mathematical Methods and Theory in Games, Programming, and Economics. 2 vols, Addison-Wesley, Reading, Mass.
75. W. KARUSH (1939), Minima of Functions of Several Variables with Inequalities as Side Conditions, MSc Thesis, Department of Mathematics, University of Chicago, Chicago.
76. T. H. KJELDSEN (2000a), A contextualized historical analysis of the Kuhn-Tucker theorem in nonlinear programming: the impact of World War II, Historia Mathematica, 27, 331-361.
77. T. H. KJELDSEN (2000b), The emergence of nonlinear programming: interactions between practical mathematics and mathematics proper, The Mathematical Intelligencer, 22, 50-54.
78. T. H. KJELDSEN (2002), Different motivations and goals in the historical development of the theory of systems of linear inequalities, Archive for History of Exact Sciences, 56, 469-538.

79. T. H. KJELDSEN (2003), New Mathematical Disciplines and Research in the Wake of World War II; in B. BOOSS-BAVNBEK and J. HØYRUP (Eds.): *Mathematics and War*. Birkhäuser Verlag: Basel-Boston-Berlin, 126–152.
80. T. H. KJELDSEN (2006), The Development of Nonlinear Programming in Post War USA: Origin, Motivation, and Expansion. In H. B. Andersen, F. V. Christiansen, K. F. Jrgensen, and V. Hendricks (eds): *The Way Through Science and Philosophy: Essays in Honour of Stig Andur Pedersen.* College Publications, London, 2006, 31-50.
81. T. C. KOOPMANS (ED.) (1951), Activity Analysis of Production and Allocation, Cowles Commission Monograph N. 13, J. Wiley, New York.
82. H. W. KUHN (1976), Nonlinear programming: a historical view; in R. W. Cottle and C. E. Lemke (Eds.), Nonlinear Programming, SIAM-AMS Proceedings, Vol. IX, American Mathematical Society, Providence, N. J., 1-26.
83. H. W. KUHN (1991), Nonlinear programming: a historical note; in J. K. Lenstra and others, 82-96.
84. H. W. KUHN (2002), Being in the right place at the right time, Operations Research, 50, 132-134.
85. H. W. KUHN and A. W. TUCKER (1951), Nonlinear programming. In J. Neyman (Ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, 481-492.
86. J. L. LAGRANGE (1788), Mécanique Analitique, V.ve Desaint, Paris. Then in : Oeuvres de Lagrange, Vol. XI and Vol. XII (J.-A. Serret and G. Darboux, Eds.), Gauthier-Villars, Paris, 1888 and 1889.
87. J. L. LAGRANGE (1797), Théorie des Fonctions Analytiques, Impr. de la République, Paris. Then in: Oeuvres de Lagrange, Vol. IX (J.-A. Serret, Ed.), Gauthier-Villars, Paris, 1881.
88. P. S. LAPLACE (1793), Sur quelques points du système du monde, Mémoire de l'Académie Royale des Sciences, 1-87. Also in : Oeuvres Complètes, Vol. XI, Gauthier-Villars, Paris, 1910, 477-558.
89. P. S. LAPLACE (1799), Traité de Mécanique Céleste, Vol. 2, Duprat, Paris. Also in: Oeuvres Complètes, Vol. II, Gauthier-Villars, Paris, 1880.
90. L. J. LEIFMAN (ED.) (1990), Functional Analysis, Optimization , and Mathematical Economics: A Collection of Papers Dedicated to the Memory of Leonid Vital'evich Kantorovich, Oxford University Press, Oxford and New York.
91. J. K. LENSTRA, A. H. G. RINNOOY KAN and A. SCRIJVER (Eds.) (1991), History of Mathematical Programming - A Collection of Personal Reminiscences, CWI and North Holland, Amsterdam.
92. G. LORIA (1950), Storia delle Matematiche. Dall'Alba della Civiltà al Tramonto del Secolo XIX, Hoepli, Milan.
93. W. V. LOVITT (1916), Preferential Voting. *American Mathematical Monthly*, 23, 363–366.
94. G. P. McCORMICK (1967), Second-order conditions for constrained minima, SIAM J. on Applied mathematics, 15, 641-652.
95. H. MINKOWSKI (1896), Geometrie der Zahlen, Teubner, Leipzig and Berlin.
96. S. K. MISHRA and G. GIORGI (2008), Invexity and Optimization, Springer Verlag, Berlin.
97. J. C. MOORE (1968), Some extensions of the Kuhn-Tucker results in concave programming. In J. P. Quirk and A. M. Zarley (Eds.), Papers in Quantitative Economics, The University Press of Kansas, Lawrence, Kansas, 31-71.
98. O. MORGENSTERN (ED.) (1954), Economic Activity Analysis, J.Wiley, New York.
99. J. MOSER (ED.) (1985), Fritz John Collected papers, Vol. I, Birkhäuser, Boston.
100. T. MOTZKIN (1936), Beiträge zur Theorie der linearen Ungleichungen, Azriel, jerusalem. English translation: RAND Corporation Report T-22, 1952.
101. C. L. M. H. NAVIER (1825), Sur le calcul des conditions d'inégalité, Nouveau Bulletin des Sciences, par la Société Philomatique de Paris, 66-68; 81-84.
102. M. V. OSTROGRADSKY (1838a), Considérations générales sur le moments des forces, Mémoires de l'Académie Impériale des Sciences de St. Pétersbourg, Sixième Série, 1, 129-150.

103. M. V. OSTROGRADSKY (1838b), Mémoire sur les déplacements instantanés des systèmes assujettis à des conditions variables, Mémoiresde l'Académie Impériale des Sciences de St. Pétersbourg, Sixième Série, 1, 565-600.
104. L. L. PENNISI (1953), An indirect sufficiency proof for the problem of Lagrange with differential inequalities as added side conditions, Trans. Amer. Math. Soc., 74, 177-198.
105. D. W. PETERSON (1973), A review of constraint qualifications in finite-dimensional spaces, SIAM Rev., 15, 639-654.
106. B. T. POLYAK (2002), History of mathematical programming in the USSR: analyzing the phenomenon, Math. Programming, 91, Ser. B, 401-416.
107. A. PREKOPA (1980), On the development of optimization theory, Amer. Math. Monthly, 87, 527-542.
108. T. L. SAATY and J. BRAM (1964), Nonlinear Mathematics, McGraw-Hill, New York.
109. M. SLATER (1950), Lagrange multipliers revisited, Cowles Commission Discussion Paper 403.
110. G. J. STIGLER (1945), The cost of subsistence, Journal of Farm Economics, 27, 303-314.
111. J. STOER and C. WITZGALL (1970), Convexity and Optimization in Finite Dimensions - I -, Springer Verlag, Berlin.
112. R. W. STOKES (1931), A geometric theory of solution of linear inequalities, Transactions of the American Math. Soc., 33, 782-805.
113. R. M. SOLOW and K. J. ARROW (1958), Gradient Methods for Constrained Maxima with Weakened Assumptions; in Studies in Linear and Nonlinear Programming, K. J. Arrow et al., Stanford University Press, 166–176.
114. J. J. SYLVESTER (1857), A question in the geometry of situation, Quarterly J. Pure and Appl. Math., 1, 79.
115. A. TAKAYAMA (1974), Mathematical Economics, Dryden Press, Hinsdale, Il. Second edition: Cambridge University Press, Cambridge, 1985.
116. V. M. TIKHOMIROV (1996), The evolution of methods of convex optimization, American Math. Monthly, 103, 65-71.
117. H. UZAWA (1958), The Kuhn-Tucker theorem in concave programming. In K. J. Arrow, L. Hurwicz and H. Uzawa (Eds.), Studies in Linear and Nonlinear Programming, Stanford Univ. Press, Stanford, 1958, 32-37.
118. F. A: VALENTINE (1937), The problem of Lagrange with differential inequalities as added side conditions. In Contributions to the Calculus of Variations (1933-1937), University of Chicago Press, Chicago, 407-448.
119. C. VAN DE PANNE and F. RAHNAMA (1985), The first algorithm for linear programming: an analysis of Kantorovich's method, Economics of Planning, 19, 76-91.
120. J. VON NEUMANN and O. MORGENSTERN (1944), Theory of Games and Economic Behavior, Princeton Univ. Press, Princeton.

A GRADIENT METHOD FOR APPROXIMATING SADDLE POINTS
AND CONSTRAINED MAXIMA

Kenneth J. Arrow and Leonid Hurwicz

P-223

13 June 1951

The RAND Corporation
SANTA MONICA • CALIFORNIA

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*,
DOI 10.1007/978-3-0348-0439-4_2, © Springer Basel 2014

P-223
6-13-51

A GRADIENT METHOD FOR APPROXIMATING SADDLE POINTS AND CONSTRAINED MAXIMA

Kenneth J. Arrow and Leonid Hurwicz

1. Introduction.

In the following, X and Y will be vectors with components X_i, Y_j. By $X \geq 0$ will be meant $X_i \geq 0$ for all i. Let $g(X)$, $f_j(X)$ $(j = 1, \cdots, m)$ be functions with suitable differentiability properties, where $f_j(X) \geq 0$ for all X, and define

$$(1) \qquad F(X, Y) = g(X) + \sum_{j=1}^{m} Y_j \left\{ 1 - \left[f_j(X)\right]^{1+\gamma} \right\}.$$

Let $(\overline{X}, \overline{Y})$ be a saddle-point of (1) subject to the conditions $X \geq 0$, $Y \geq 0$; assume it unique in X. The function $F(X, \overline{Y})$ attains its maximum for variation in X subject to the condition $X \geq 0$ at the point $X = \overline{X}$. Since F is a maximum for variation in each component X_i separately, it follows that

$$(2) \qquad \overline{F}_{X_i} \leq 0 \quad \text{for all } i, \text{ and}$$

$$(3) \qquad \overline{X}_i = 0 \quad \text{if } \overline{F}_{X_i} < 0.$$

We will refer to those subscripts for which (3) holds as <u>corner</u> indices and the remainder as <u>interior</u> indices. Let X^1 be the vector of components of X with corner indices, and X^2 the vector of interior components. Since F is also a maximum for variations in X^2 alone (holding X^1 at 0 and Y at $\overline{Y}$), and the first-order terms vanish by (2) and (3), it follows, under the usual differentiability assumptions, that the matrix,

$$(4) \qquad \overline{F}_{X^2X^2} \text{ is negative semi-definite,}$$

where $\bar{F}_{X^2X^2}$ is the matrix of elements $\partial^2 F/\partial X_i \partial X_j$, with i and j ranging over interior indices, evaluated at $(\bar{X}, \bar{Y})$.

It is shown in another paper, now in preparation, that for all γ sufficiently large, $\bar{X}$ maximizes $g(X)$ subject to the restraints $f_j(X) \leq 1$, $X \geqq 0$, and

(5) $\bar{F}_{X^2X^2}$ is negative definite.

Hence the determination of the constrained maximum is equivalent to finding the saddle-point of a function $F(X, Y)$ which is linear in Y and satisfies (5). We seek here a convergent process for approximating such a saddle-point. The intuitively natural method, in terms of the motivations of the two players (interpreting F as the pay-off of a game in which player I chooses X and player II chooses Y), is for the player who chooses X to move "uphill" with regard to variation in that variable, while the other player moves against the gradient with respect to Y. Such processes have been investigated by Brown and von Neumann [1][1] for the case where F is linear in both X and Y. In that case, the "naive" gradient method just described leads to an oscillatory behavior (see Samuelson [2], pp. 17-22) and must be modified. In the present case, even if the functions g, f_j were linear to begin with, the introduction of the power γ creates a nonlinear system satisfying (5); as will be seen, this implies that the naive gradient method will be at least locally stable.

[1]Numbers in brackets refer to the bibliography at the end of the paper.

2. Description of the Gradient Method.

It must be recalled that the variables X and Y are constrained to be non-negative, so that the movements of the players with and against the gradients of X and Y, respectively, cannot carry the variables into areas of negativity. The gradient method for finding a saddle-point then is the following system of differential equations:

(1) $\dot{X}_i = 0$ if $F_{X_i} < 0$ and $X_i = 0$,

(2) $= F_{X_i}$ otherwise;

(3) $\dot{Y}_j = 0$ if $F_{Y_j} > 0$ and $Y_j = 0$,

(4) $= - F_{Y_j}$ otherwise;

the dot denotes differentiation with respect to time. In this system the derivatives are discontinuous functions of the variables. The usual existence theorems for nonlinear differential equations assume continuity (see [3], Chapter II). If it could be shown that equations (2.1-4) have a unique solution for any initial position continuous with respect to variations in the starting-point, considerably stronger statements could be made about the convergence of the system.

3. Theorem.

Let F(X, Y) be linear in Y, possess a saddle-point $(\bar{X}, \bar{Y})$ under the constraint $X \geq 0$, $Y \geq 0$, and be analytic in some neighborhood of $(\bar{X}, \bar{Y})$. Suppose further that (a) condition (1.5) holds and (b) $\bar{X}_i > 0$ and $\bar{Y}_j > 0$ for every interior index i or j.[2]

[2] Analogously to (1.3), j is a corner index for Y if $\bar{Y}_j = 0$, $\bar{F}_{Y_j} < 0$; an interior index for Y is any subscript which is not a corner index.

Then for every initial position in a sufficiently small neighborhood of $(\overline{X}, \overline{Y})$, there is a unique solution $X(t)$, $Y(t)$ of the equations (2.1-4), such that $\lim_{t\to\infty} X(t) = \overline{X}$ and, for every limit-point Y^* of $Y(t)$, $(\overline{X}, Y^*)$ is a saddle-point of $F(X, Y)$.

4. Proof.

If (X^*, Y^*) were another saddle-point of $F(X, Y)$, then $(X^*, \overline{Y})$ would be still another. That is, X^* would maximize $F(X, \overline{Y})$ for variation in X. Then (1.5) implies that if X^* is in a sufficiently small neighborhood of $\overline{X}$, then $\overline{X} = X^*$, so that $\overline{X}$ is at least locally unique.

In what follows, let $x = X - \overline{X}$, $y = Y - \overline{Y}$, expanding the derivatives of F into power series. Then

$$F_{X^1} = \overline{F}_{X^1} + a(x, y), \tag{1}$$

where $a(x, y)$ is a continuous vector function with $a(0, 0) = 0$.

In the expansion of F_{X^2}, there are no constant terms by definition. Divide the terms of the expansion into four types: those containing components of x^1; the terms linear in x^2; the terms linear in y; and the terms in x^2 and y of degree higher than the first. Define the (variable) matrix A as follows: for any interior index i and corner index j, let $A_{ij}\, x_j$ be the sum of all terms in the expansion of F_{X_i} which have x_j as a factor but do not have x_k as a factor for any corner index $k < j$. Then, clearly, $\sum_{j1} A_{ij}\, x_j$ is the sum of all terms in the expansion of F_{X_i} which contain corner components of x, the summation extending only over corner indices. The matrix A is a function of x and y. Now consider the fourth type of term in the expansion of F_{X^2}, the non-linear terms involving x^2 and y only. Since F, and therefore F_{X^2}, is linear in y, each such term must involve a component of x^2. Define the matrix B so that, for every pair of interior indices i and j, $B_{ij}\,x_j$ is the sum of all non-linear terms x in the expansion of F_{X_i} which have x_j for a factor but do not have x_k as a factor for any corner index k or for any interior index $k < j$. Then, $\sum_{j^2} B_{ij}\,x_j$ is the sum of all non-linear terms in the expansion of F_{X_i} which contain no corner components of x as factors. B therefore is a function of x^2 and y; further, since each component of Bx^2 is non-linear, B vanishes if both x^2 and y do.

$$(2)\qquad F_{X^2} = A(x,\,y)\,x^1 + \bar{F}_{X^2X^2}\,x^2 + \bar{F}_{X^2Y}\,y + B(x^2,\,y)\,x^2,$$

where A and B are continuous matrix functions, and $B(0,\,0) = 0$.

Since F is linear in y, F_Y is independent of y. By a discussion similar to the preceding, it follows that

$$(3)\qquad F_Y = \bar{F}_Y + C(x)\,x^1 + \bar{F}'_{X^2Y}\,x^2 + b(x^2),$$

where C is a continuous matrix function and the vector $b(x^2)$ is of the second order with respect to components of x^2.

Now define

$$(4)\qquad D = (1/2)\,(x'x + y'y).$$

D is proportional to the distance in the $(X,\,Y)$ space to the saddle-point $(\bar{X},\,\bar{Y})$. Differentiate (4) with respect to time.

$$(5)\qquad DD = x'x + y'y.$$

First suppose that for each i either $X_i > 0$ or $F_{X_i} \geq 0$ and that for each j either $Y_j > 0$ or $F_{Y_j} \leq 0$. Then from (2.2), (2.4) and (5),

$$(6)\qquad DD = x'F_X - y'F_Y.$$

Substitute from (1-3) into (6)

(7) $$DD = (x^1)'F_{X^1} + (x^2)'F_{X^2} - y'F_Y$$

$$= (x^1)'\overline{F}_{X^1} + (x^1)'a(x, y) + (x^2)'A(x, y)(x^1) + (x^2)'\overline{F}_{X^2X^2}\, x^2 + (x^2)'\overline{F}_{X^2Y}\, y$$

$$+ (x^2)'B(x^2, y)\, x^2 - y'\overline{F}_Y - y'C(x)\, x^1 - y'\overline{F}'_{X^2Y}\, x^2 - y'\, b(x^2).$$

The last term is homogeneous linear in y and of the second order in x^2. Hence, it can be written in the form,

(8) $$y'\, b(x^2) = (x^2)'\, E(x^2, y)\, (x^2),$$

where E is a continuous matrix function and $E(x^2, 0) = 0$.

Each term in (7) is a scalar and therefore equal to its transpose. In particular, $(x^2)'\overline{F}_{X^2Y}\, y = y'\overline{F}'_{X^2Y}\, x^2$. Let

(9) $$c(x, y) = a(x, y) + A'(x, y)\, x^2 - C'(x)\, y,$$

(10) $$G(x^2, y) = B(x^2, y) - E(x^2, y).$$

In view of (8-10) and the preceding remarks, (7) can be simplified to the following expression:

(11) $$DD = (x^1)'\overline{F}_{X^1} + (x^1)'\, c(x, y) + (x^2)'\overline{F}_{X^2X^2}\, x^2 + (x^2)'\, G(x^2, y)\, (x^2) - y'\overline{F}_Y.$$

From (1) and (9),

(12) $$c(0, 0) = 0.$$

Let m_1 be the minimum of $\left|\bar{F}_{X_i}\right|$ over the corner indices i; by definition, $m_1 > 0$. By (12), we can choose ϵ_1 so that every component of c is less than m_1 whenever $D < \epsilon_1$. Let $\sum_1$ denote summation over the corner indices only; then, since $x_i \geq 0$ for all corner indices,

$$(x^1)' c(x, y) < m_1 \sum_{i^1} |x_i| \leq -(x^1)' \bar{F}_{X^1}, \quad \text{if } x^1 \neq 0 \text{ and } D < \epsilon_1, \text{ or}$$

$$(13) \qquad (x^1)' \bar{F}_{X^1} + (x^1)' c(x, y) < 0 \quad \text{if } x^1 \neq 0 \text{ and } D < \epsilon_1.$$

From (1.5),

$$(x^2)' \bar{F}_{X^2X^2}\, x^2 < 0 \quad \text{unless } x^2 = 0.$$

Let m_2 be the maximum of $(x^2)' \bar{F}_{X^2X^2}\, x^2$ subject to the condition $D = 1$, m_3 the maximum of $\sum_{i^2} |x_i|$ subject to the same condition. Then,

$$(14) \qquad m_2 < 0, \; (x^2)' \bar{F}_{X^2X^2}\, x^2 \leq m_2 D^2, \; \sum_{i^2} |x_i| \leq m_3 D.$$

From (2), (8) and (10), $G(0, 0) = 0$. If D is sufficiently small, (x^2, y) will be sufficiently close to $(0, 0)$ to insure that the largest of the components g_{ij} of G is less than $-m_2/(m_3)^2$ in absolute value. Then, from (14),

$$(x^2)' G(x^2, y)\, x^2 \leq \left|\sum_{i^2}\sum_{j^2} g_{ij}\, x_i x_j\right| \leq \sum_{i^2}\sum_{j^2} |g_{ij}|\,|x_i|\,|x_j|$$

$$< (-m_2/m_3^2)\left(\sum_{i^2} |x_i|\right)^2 \leq -m_2 D^2 \leq -(x^2)' \bar{F}_{X^2X^2}\, x^2,$$

the strict inequality holding provided that $\sum_{i^2} |x_i| > 0$, which is equivalent to $x^2 \neq 0$.

$$(15) \qquad (x^2)' \bar{F}_{X^2X^2}\, x^2 + (x^2)' G(x^2, y)\, x^2 < 0 \quad \text{if } D < \epsilon_2, \; x^2 \neq 0.$$

For each corner index j, $y_j \geq 0$ always, while $\bar{F}_{Y_j} > 0$; for interior indices j, $\bar{F}_{Y_j} = 0$. Hence,

(16) $y'\bar{F}_Y \geq 0.$

By (11), (13), (15) and (16),

(17) $D\dot{D} < 0$ if $D < \epsilon$, $x \neq 0$; $D\dot{D} \leq 0$ if $D < \epsilon$,

where ϵ is chosen smaller than ϵ_1 or ϵ_2, and also sufficiently small so that,

(18) $F_{X_i} < 0$, $F_{Y_j} > 0$ when $D \leq \epsilon$, for all corner indices i and j;

(19) $\epsilon < \min_i 2\,\bar{X}_i$, $\epsilon < \min_j 2\,\bar{Y}_j$, the minima being taken over all interior indices;

(20) $F_{X^2X^2}$ is negative definite when $X = \bar{X}$ and $y'y/2 < \epsilon$.

By assumption (b) of the theorem, (19) is possible with positive ϵ. By (1.5), $F_{X^2X^2}$ is negative definite when $X = \bar{X}$ and $y = 0$; since F is certainly continuous in Y, (20) can hold for sufficiently small ϵ.

We will now show that there does in fact exist a unique solution of (2.1-4) continuous in the initial position and in time if, at the initial position $[X(0), Y(0)]$, $D < \epsilon$. Let S_0 be the set of all indices for which $X_i(0) = 0$, By (19), any index in S_0 must be a corner index for X. Similarly, let T_0 be the set of all indices for which $Y_j(0) = 0$. By (18), $F_{X_i} < 0$, $F_{Y_j} > 0$ for all indices in S_0 and T_0, respectively. By the differential equation system (S_0, T_0), we shall mean

(21) $\dot{X}_i = F_{X_i}$ for i not in S_0, $\dot{Y}_j = -F_{Y_j}$ for j not in T_0,
$X_i = Y_j = 0$ for i in S_0, j in T_0.

In this system, the derivatives are continuous functions of the variables. By the Cauchy-Lipschitz Existence Theorem (see [3], Theorem (4.1), p. 23), the system (S_0, T_0) has a solution uniquely defined by the initial conditions. Let $Z = (X, Y)$, and let a given solution be $Z[t, Z(0)]$, where $Z(0)$ is the initial position. Then it is further known ([3], (7.3), p.30) that

(22) $Z[t, Z(0)]$ is a continuous function of $Z(0)$ and of t.

For every i not in S_0, $X_i[t, Z(0)] > 0$ in some interval of time; similarly, $Y_j[t, Z(0)] > 0$ in some interval for every j not in T_0. Since F_{X_i} and F_{Y_j} are continuous functions of Z, which is in turn continuous in t there is an interval in which $F_{X_i} < 0$, $F_{Y_j} > 0$ for i in S_0, and j in T_0. The solution to system (S_0, T_0) is then a solution to the system (2.1-4), and further it is clearly the only one.

Since S_0 and T_0 contain only corner indices, $x_i = y_j = 0$ for all i in S_0 and j in T_0. If we fix these variables at 0, F, considered as a function of the remaining variables, has the same properties as assumed to begin with. Hence, (17) is valid; since $\dot{D} \leq 0$, $D[t, Z(0)]$ (the value of D for the point $Z[t, Z(0)]$) is non-increasing. Since $D[0, Z(0)] < \epsilon$, $D[t, Z(0)] < \epsilon$ for all t. Hence, $F_{X_i} < 0$, $F_{Y_j} > 0$ for all i in S_0 and j in T_0 for all points of the solution $Z[t, Z(0)]$. The solution for (S_0, T_0) therefore ceases to be a solution for (2.1-4) only when $X_i[t, Z(0)] = 0$ for some i not in S_0 or $Y_j[t, Z(0)] = 0$ for some j not in T_0. Let this occur at time t_0. Since $D[t_0, Z(0)] < \epsilon$, $X_i[t_0, Z(0)] > 0$, $Y_j[t_0, Z(0)] > 0$ for all interior indices by (19); hence, i or j must be a corner index by (18). Let S_1 be now the set of all indices for which $X_i[t_0, Z(0)] = 0$, T_1 the set of all indices for which $Y_j[t_0, Z(0)] = 0$. Clearly, S_1 includes S_0, T_1 includes T_0. Again, the solution of the system (S_1, T_1) is the unique solution of (2.1-4) in some interval of time beginning with t_0. The argument can be repeated; since the sets S_i, T_i are increasing and there are only a finite number of indices, only a finite number of systems are involved. It then follows easily that the system (2.1-4) has a unique solution $Z[t, Z(0)]$ continuous in t and in $Z(0)$.

By (18), for each corner index i, there is a number $m_i < 0$ such that $F_{X_i} \leq m_i$ whenever $D \leq \epsilon$. As $\dot{D}[t, Z(0)] \leq 0$ for all t, $D[t, Z(0)] < \epsilon$. So long as $X_i[t, Z(0)] > 0$, $\dot{X}_i[t, Z(0)] \leq m_i$, so that $X_i[t, Z(0)]$ reaches 0 in finite time. Since $F_{X_i} < 0$ for all t, $X_i[t, Z(0)] = 0$ for all t from then on. The same argument holds for corner indices of Y.

(23) $$X^1[t, Z(0)] = Y^1[t, Z(0)] = 0 \text{ for all } t \text{ sufficiently large.}$$

As $\dot{D}[t, Z(0)] \leq 0$ for all t, $D[t, Z(0)]$ converges to a limit. Let

(24) $$\lim_{t \to \infty} D[t, Z(0)] = D^*.$$

Let $Z^* = (X^*, Y^*)$ be any limit point of $Z[t, Z(0)]$. There is a sequence $\{t_n\}$ such that

(25) $$\lim_{n \to \infty} t_n = \infty, \lim_{n \to \infty} Z[t_n, Z(0)] = Z^*.$$

Let $Z_n = Z[t_n, Z(0)]$. Then, by (22),

(26) $$Z(t, Z^*) = \lim_{n \to \infty} Z(t, Z_n) = \lim_{n \to \infty} Z[t + t_n, Z(0)].$$

Since D is a continuous function of Z, it follows from (26) and (24) that

(27) $\quad D(t, Z^*) = \lim_{n \to \infty} D\left[t + t_n, Z(0)\right] = D^*,$

a constant. That is, $\dot{D}(t, Z^*) = 0$ for all t. By (17), $x(t, Z^*) = 0$ for all t, or

(28) $\quad X(t, Z^*) = \overline{X}$ for all t.

In particular, $X(0, Z^*) = X^* = \overline{X}$. Since Z^* was any limit-point of $Z\left[t, Z(0)\right]$,

(29) $\quad \lim_{t \to \infty} X\left[t, Z(0)\right] = \overline{X}.$

Let an asterisk denote evaluation at $Z^* = (\overline{X}, Y^*)$. By (23) and (18),

(30) $\quad \overline{X}^1 = 0, \; F^*_{X^1} < 0,$

(31) $\quad Y^{*1} = 0, \; F^*_{Y^1} > 0.$

By (28), $\dot{X}^2(t, Z^*) = 0$; since $\overline{X}^2 > 0$ by hypothesis, it follows from (2.2) that

(32) $\quad F^*_{X^2} = 0.$

By (20), $F^*_{X^2X^2}$ is negative definite. In conjunction with (30) and (32), this shows that

(33) $\quad F(X, Y^*)$ has a maximum at $\overline{X}$ for variation in X subject to $X \geq 0$.

Since F is linear in Y, F_Y is independent of Y, so that $F^*_{Y^2} = \overline{F}_{Y^2} = 0$. That is, $F(\overline{X}, Y)$ is independent of Y^2. From (31), then,

(34) $F(\overline{X}, Y)$ has a minimum at Y^* for variation in Y subject to $Y \geq 0$.

(28), (33) and (34) complete the proof of the theorem.

5. A Remark on the Hypotheses of the Theorem.

Condition (b) of the theorem, that no component of $\overline{X}$ or $\overline{Y}$ is at the boundary of the domain of variation unless it is actually a corner extremum in the proper sense, is inserted to avoid the possibility that at some point $X_i = 0$ and $F_{X_i} = 0$ for some i. We have been unable to show, in this situation, either that there exists a solution of (2.1-4) with such an initial position or that, if it exists, it is unique. Some experiments with simple systems suggest that in fact there is a unique solution beginning at such a point; if so, condition (b) could be dropped.

6. Economic Interpretation.

Let X_i $(i = 1, \cdots, n)$ be activity levels of the n different possible production activities (measured, e. g., by the outputs of one of the products). Let $g(X)$ be the social utility derived from activities, $F_{ij}(X_i)$ the quantity of input j needed to carry on activity i at level X_i, α_j the stock of input j available to begin with, and $f_j(X) = \sum_i f_{ij}(X_i) - \alpha_j + 1$. The unit of measurement of commodity j should be chosen sufficiently large that $f_j(X)$ (which is excess demand plus one) will be positive throughout the adjustment process. Note that production of an output j by means of process i would be represented by a negative value for the function f_{ij}; also, $\alpha_j = 0$ for intermediate products. Hence, it is desired to choose a set of activity levels $\overline{X}$ which will maximize $g(X)$ subject to the constraints that the excess demand of the productive system for any input does not exceed the initial supply, i.e., $\sum_i f_{ij}(X) \leq \alpha_j$, or, $f_j(X) \leq 1$ for all j. By definition, $X_i \geq 0$ for all i. As noted in section 1, if $\overline{X}$ is the optimum set of activity levels, then, there is some $\overline{Y}$ such that $(\overline{X}, \overline{Y})$ is the saddle-point of the function $F(X, Y)$ defined in (1.1).

It follows then, by the Theorem, that the X-components of the solution of the system of differential equations (2.1-4) will approach $\overline{X}$. The equations (2.1-2) can be written,

(1) $$\dot{X}_i = (\partial g/\partial X_i) - \sum_j Y_j (1+\eta)(f_j)^{\eta} (df_{ij}/dX_i),$$

unless the right-hand side is negative when $X_i = 0$, in which case the right-hand side is replaced by 0. Let

(2) $$q_i = \partial g/\partial X_i,$$

(3) $$p_j = Y_j (1+\eta)(f_j)^{\eta}.$$

Then, from(1-3),

(4) $$\dot{X}_i = q_i - \sum_j p_j (df_{ij}/dX_i) \quad \text{if} \quad X_i > 0,$$

$$= \max\left[0,\ q_i - \sum_j p_j (df_{ij}/dX_i)\right] \quad \text{if} \quad X_i = 0.$$

By (2.3-4), p_j is determined by (3) in conjunction with the equations,

(5) $$\dot{Y}_j = f_j^{1+\eta} - 1 \quad \text{if} \quad Y_j > 0,$$

$$= \max\left[f_j^{1+\eta} - 1,\ 0\right] \quad \text{if} \quad Y_j = 0.$$

Note that $\dot{Y}_j > 0$ if $f > 1$, i. e., if there is excess demand, and $\dot{Y}_j < 0$ if there is excess supply (except for free goods).

Institutionally, the process can be visualized as follows: there is a central board which evaluates the social worth of a given constellation of activity levels, and therefore the marginal social valuation q_i of each; for each activity, there is a plant manager who determines the activity level X_i; for each primary or intermediate product, there is a price-fixing authority who determines p_j. The central board announces the marginal social valuations q_i, and each price-fixing authority announces a price p_j. then, each plant manager expands or contracts at a rate equal to the difference between the marginal social valuation of the activity, q_i, and the marginal cost of increasing the activity, $\sum_j p_j\,(df_{ij}/dX_i)$ (apart from the corner case of unused activities). At the same time, the price-fixing authority adjusts Y_j in accordance with the excess demand, as given in (5) and then arrives at p_j.

It is important to observe that these rules of decision-making are highly decentralized. Once the prices are announced, the individual activity managers need know only their own technologies to determine their rate of expansion. Similarly, the price-fixers need know only the excess demands on their own markets.

Even the decisions of the central board in regard to the marginal social valuations of the commodities can be simplified. Actually, the social valuation depends on the outputs of final products. Let $g_{ik}(X_i)$ be the output of final product k if activity i is operated at level X_i, $g_k(X) = \sum_i g_{ik}(X_i)$ be the total output of final product k, and $U(g_1, \cdots, g_m)$ the social utility derived from having total outputs $g_1, \cdots, g_m$ of the final products $1, \cdots, m$, respectively. Then $g(X) = U[g_1(X), \cdots, g_m(X)]$, so that

(6) $$\partial g/\partial X_i = \sum_k (\partial U/\partial g_k)(dg_{ik}/dX_i).$$

If the central board announces merely the marginal social valuations of the various final products, $r_k = \partial U/\partial g_k$, the firm can compute its marginal social valuation,

(7) $$q_i = \sum_k r_k (dg_{ik}/dX_i),$$

by the knowledge of its own technology.

Bibliography

1. G. W. Brown and J. von Neumann, "Solutions of Games by Differential Equations," P-142, 19 April 1950, RAND.

2. P. A. Samuelson, "Market Mechanisms and Maximizations," P-69, 28 March 1949, RAND.

3. S. Lefschetz, Lectures on Differential Equations, Princeton: Princeton University Press, 1946, 209 pp.

/ebc

REDUCTION OF CONSTRAINED MAXIMA TO SADDLE-POINT PROBLEMS

KENNETH J. ARROW AND LEONID HURWICZ
STANFORD UNIVERSITY, UNIVERSITY OF MINNESOTA

1. Introduction

1.1. The usual applications of the method of Lagrangian multipliers, used in locating constrained extrema (say maxima), involve the setting up of the *Lagrangian expression*,

$$\phi(x, y) = f(x) + y'g(x) \,, \tag{1}$$

where $f(x)$ is being (say) maximized with respect to the (vector) variable $x = \{x_1, \cdots, x_N\}$, subject to the constraint $g(x) = 0$, where $g(x)$ maps the points of the N-dimensional x-space into an M-dimensional space, and $y = \{y_1, \cdots, y_M\}$ is the Lagrange multiplier (vector). Here, $\{\quad\}$ indicates a column vector; the prime indicates transposition, so that y' is a row vector.

The essential step of the customary procedure is the solution for x, as well as y, of the pair of (vector) equations,

$$\phi_x(x, y) = 0, \; g(x) = 0 \,, \tag{2}$$

where $\phi_x(x, y) = \{\partial\phi(x, y)/\partial x_1, \cdots, \partial\phi(x, y)/\partial x_N\}$. Let $(\bar{x}, \bar{y})$ be the solutions of equations (2), while $\hat{x}$ maximizes $f(x)$ subject to $g(x) = 0$. Then, under suitable restrictions,

$$\bar{x} = \hat{x} \,. \tag{3}$$

1.2. In [1] Kuhn and Tucker treat the related problem of maximizing $f(x)$ subject to the constraints[1] $g(x) \geqq 0$, $x \geqq 0$, where, for an arbitrary K-dimensional vector $a = \{a_1, \cdots, a_K\}$, the relation $a \geqq 0$ is here defined to mean $a_k \geqq 0$ for $k = 1, \cdots, K$. Another definition of vectorial inequalities, permitting greater generality of treatment, will be used in later sections of this paper. There we shall treat directly the class of situations where $f(x)$ is to be maximized subject to $g^{(1)}(x) \geqq 0$, $g^{(2)}(x) = 0$, $x^{[1]} \geqq 0$, $x^{[2]}$ not restricted as to sign, $x = \{x^{[1]}, x^{[2]}\}$.

Denote by C_g the set of all x satisfying the constraints $g(x) \geqq 0$, $x \geqq 0$. The two results stated below are of fundamental importance for the problem considered.

(A) (See theorem 1 [1].) Let g satisfy the following condition (called Constraint

Most of the work of this paper was done under the auspices of The RAND Corporation, with additional support and assistance from the Cowles Commission for Research in Economics and the Office of Naval Research.

[1] In [1] our f and g are respectively written as g and F. The symbol in [1] for the Lagrange multiplier (our y) is u.

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_3, © Springer Basel 2014

Qualification, here abbreviated as C.Q.).[2] If $\tilde{x}$ is a boundary point of C_g and x satisfies the relations,

$$\tilde{g}_x^a(x - \tilde{x}) \geqq 0\,, \tag{4}$$

$$x^b - \tilde{x}^b \geqq 0\,, \tag{5}$$

where "~" over a symbol denotes its evaluation at $x = \tilde{x}$, $g = \{g^a, g^\beta\}$, $\tilde{g}^a = 0$, $\tilde{g}^\beta > 0$, $x = \{x^a, x^b\}$, $\tilde{x}^a > 0$, and $\tilde{x}^b = 0$, then there exists a differentiable vector-valued function ψ of the real variable θ whose domain is the closed interval (0, 1) and the range is in C_g; that is, $x = \psi(\theta)$, such that $\psi(0) = \tilde{x}$ and $\psi'(0) = \lambda(x - x)$ for some positive scalar λ.

Under this condition, if all derivatives used below exist and if $\bar{x}$ maximizes $f(x)$ for $x \in C_g$, there exists y satisfying the conditions

$$\bar{x} \geqq 0, \quad \bar{\phi}_x \leqq 0, \quad \bar{x}'\bar{\phi}_x = 0\,, \tag{6}$$

$$\bar{y} \geqq 0, \quad \bar{\phi}_y \geqq 0, \quad \bar{y}'\bar{\phi}_y = 0\,, \tag{7}$$

where $\bar{\phi}_x$ and $\bar{\phi}_y$ are partial (vector) derivatives of the Lagrangian expression (1) evaluated at $(\bar{x}, \bar{y})$.

(B). (See theorem 3 [1].) If the hypotheses specified in (A) hold and, in addition, the functions $f(x)$, $g_m(x)$, $m = 1, \cdots, M$ are *concave*,[3] there exists a pair $(\bar{x}, \bar{y})$, satisfying conditions (6) and (7), such that (x, y) is a *nonnegative saddle-point* (NNSP) of $\phi(x, y)$, that is,

$$\phi(x, \bar{y}) \leqq \phi(\bar{x}, \bar{y}) \leqq \phi(\bar{x}, y) \quad \text{for all} \quad x \geqq 0, \quad y \geqq 0\,; \tag{8}$$

furthermore, any NNSP $(\bar{x}, \bar{y})$ of $\phi(x, y)$ has the property that x maximizes $f(x)$ in C_g. According to lemma 1 [1], conditions (6), (7) are implied by (8) regardless of the nature of $\phi(x, y)$, that is, even if $\phi(x, y)$ is not given by (1).

2. A modified Lagrangian approach

2.1. Because of the interesting game theoretical and economic implications of the theorem in (B), section 1.2 (which the authors will study elsewhere), the question arises as to the possibility of similar results when some of the conditions of the theorem are relaxed.

It turns out that results of such nature can be obtained, though not without some sacrifices. The relaxation is primarily with regard to the convexity assumptions which fail to hold in some important economic applications (the case of "increasing returns"). The main sacrifices are (1) the Lagrangian expression is modified, and (2) the results are proved only locally.

The results are presented below in the form of three theorems. Theorem 1 is auxiliary in nature; theorems 2 and 3 together imply the existence of a local nonnegative saddle-

[2] This restriction "is designed to rule out singularities on the boundary of the constraint set, such as an outward-pointing 'cusp'" (see p. 483 in [1]). It should be noted, however, that because of (4), C.Q. is a property of g, not merely of C_g. Thus $g(x) \equiv -(x - 1)^3$, x one-dimensional, lacks C.Q., while $g(x) \equiv -(x - 1)$, with the same C_g, does have it.

[3] A function $f(x)$ is said to be concave if

$$(1 - \theta)f(x^0) + \theta f(x) \leqq f[(1 - \theta)x^0 + \theta x]$$

for all $0 \leqq \theta \leqq 1$ and all x^0 and x in the region where $f(x)$ is defined (see [1], pp. 10–11).

point for the modified Lagrangian expression. Theorem 3 shows this saddle-point to be of the type leading to convergence in gradient procedures described by the authors in [3].

The notation differs in some detail from that introduced in section 1. To facilitate reading, some notational principles are stated in 2.2.1; the main symbols used are listed in sections 2.2.2 and 2.3.4.

2.2.1. *Some principles of notation.* A K-dimensional column vector $\{a_1, a_2, \cdots, a_K\}$ is denoted by a; dim a denotes the number of components in a. If A is a matrix, A' is its transpose. Hence, in particular, a' is a row vector and $a'b \equiv \sum_{k=1}^{K} a_k b_k$ is the inner product of the vectors a and b; $a \cdot b$ is an alternative, and sometimes more convenient, notation for $a'b$.

$[a_1, a_2, \cdots, a_K]$ is the finite (unordered) set whose elements are $a_1, a_2, \cdots, a_K$. $A \sim B$ is the set of all elements in A but not in B (the set-theoretic difference).

$\{x \mid p_x\}$ denotes the set of all x possessing the property p_x.

If

(9) $$c(a) = \{c_1(a), c_2(a), \cdots, c_P(a)\},$$

(10) $$a = \{a_1, a_2, \cdots, a_K\},$$

then

(11) $$c_a \equiv c_a(a) = \left\| \frac{\partial c_T}{\partial a_k} \right\|, \qquad p = 1, 2, \cdots, P; \qquad k = 1, 2, \cdots, K.$$

Further, $\bar{c}$, $\bar{c}_a$ denote, respectively, $c(a)$ and $c_a(a) \equiv c_a$ evaluated at $a = \bar{a}$.

If $\psi(a, b)$ is a real-valued (scalar) function of the vectors $a = \{a_1, a_2, \cdots, a_K\}$, $b = \{b_1, b_2, \cdots, b_R\}$, then

(12) $$\psi_{ab} = \left\| \frac{\partial^2 \psi}{\partial a_k \partial b_r} \right\|, \qquad k = 1, 2, \cdots, K; \qquad r = 1, 2 \cdots, R,$$

where $\bar{\psi}_{ab}$ denotes ψ_{ab} evaluated at $(\bar{a}, \bar{b})$.

$S_\rho(x^0) = \{x \mid d(x, x^0) \leqq \rho\}$ where $d(x', x'')$ denotes the Euclidean distance between x' and x''.

2.2.2. *Some symbols used.*

(N.1.1) $$x = \{x_1, x_2, \cdots, x_N\}.$$

X is the Euclidean N-space of the x's.

$$\mathcal{N} = [1, 2, \cdots, N].$$

$\mathcal{N}'$ is a fixed (possibly empty, not necessarily proper) subset of $\mathcal{N}$. As will be seen in (N.1.4), the elements of $\mathcal{N}'$ are the indices of the components of $x^{[1]}$ as defined in the first paragraph of section 1.2.

(N.1.2) $$z = \{z_1, z_2, \cdots, z_M\}.$$

Z is the Euclidean M-space of the z's.

$$\mathcal{M} = [1, 2, \cdots, M].$$

$\mathcal{M}'$ is a fixed (possibly empty, not necessarily proper) subset of $\mathcal{M}$. As will be seen from (N.1.4), (N.2), (N.3), the elements of $\mathcal{M}'$ are the indices of the components of $g^{(1)}$ as defined in the first paragraph of section 1.2; the elements of $\mathcal{M} \sim \mathcal{M}'$ are the indices of $g^{(2)}$ (see same paragraph); g will be defined as $\{g^{(1)}, g^{(2)}\}$.

(N.1.3) $$y = \{y_1, y_2, \cdots, y_M\} .$$

Y is the Euclidean M-space of the y's. Here Y is the space of the real-valued linear functions on Z. Even in the Euclidean case it is convenient to distinguish between the two, since our definitions of nonnegativity in the two spaces differ.

(N.1.4) $$x \geqq 0 \text{ means } \begin{cases} x_n \geqq 0 \text{ for } n \in \mathcal{N}' . \\ x_n \text{ unrestricted as to sign for } n \notin \mathcal{N}' . \end{cases}$$

X^+ is the set of all $x \geqq 0$.

$$z \geqq 0 \text{ means } \begin{cases} z_m \geqq 0 \text{ for } m \in \mathcal{M}' . \\ z_m = 0 \text{ for } m \notin \mathcal{M}' . \end{cases}$$

$$y \geqq 0 \text{ means } \begin{cases} y_m \geqq 0 \text{ for } m \in \mathcal{M}' . \\ y_m \text{ unrestricted as to sign for } m \notin \mathcal{M}' . \end{cases}$$

For any vector $a = \{a_1, a_2, \cdots, a_k\}$,

$$a = 0 \text{ means } a_1 = 0, a_2 = 0, \cdots, a_K = 0 ;$$

$$a > 0 \text{ means } a_1 > 0, a_2 > 0, \cdots, a_K > 0 ;$$

$$a < 0 \text{ means } -a > 0 .$$

(N.2.1) $'g$ is a function on X^+ to Z. Hence $'g(x) = \{'g_1(x), 'g_2(x), \cdots, 'g_M(x)\}$ where the $'g_m$, $m \in \mathcal{M}$ are real-valued functions.

(N.2.2) We shall find it convenient to work with some of the $'g_m$, $m \in \mathcal{M}'$ replaced by their negatives. More precisely, we write

$$g_m = \begin{cases} 'g_m \text{ if } m \in \mathcal{M} \sim \mathcal{M}^- \\ -'g_m \text{ if } m \in \mathcal{M}^- , \end{cases}$$

where $\mathcal{M}^- \subseteq \mathcal{M} \sim \mathcal{M}'$ will be defined in section 2.3.4.

$$g = \{g_1, g_2, \cdots, g_M\} .$$

Note. Since $\mathcal{M}^- \subseteq \mathcal{M} \sim \mathcal{M}'$, it is seen that the conditions

$$'g(x) \geqq 0, \quad g(x) \geqq 0$$

are equivalent. For practical purposes, one could consider the problem as given directly in terms of g, rather than $'g$. We start with $'g$, however, in order to avoid the impression of a loss of generality in connection with the assumptions of section 2.3.4.

(N.3) $$C_g = \{x \,|\, 'g(x) \geqq 0, \ x \geqq 0\} \equiv \{x \,|\, g(x) \geqq 0, \ x \geqq 0\}$$

(the "constraint set").

(N.4) f is a real-valued function on X^+ (the "maximand").

(N.5) $$O_{fg} = \{x' \,|\, x' \in C_g \text{ and } f(x) \leqq f(x') \text{ for all } x \in C_g\}$$

(the "optimal set").

(N.6) $x = \{x^{(1)}, x^{(2)}\}$ where

$\mathcal{N}^{(i)}$ = the set of indices of the components of $x^{(i)}$, $i = 1, 2$

$n \in \mathcal{N}^{(1)}$ if $n \notin \mathcal{N}'$ or $n \in \mathcal{N}'$ and $\bar{x}_n > 0$

$n \in \mathcal{N}^{(2)}$ if $n \in \mathcal{N}'$ and $\bar{x}_n = 0$

for a given $\bar{x} \in O_{f_0}$ and either component may be empty.

Note 1. When a vector a is partitioned into two subvectors, say

$$a = \{a^*, a^{**}\}$$

and we say that a^* (or a^{**}) is empty, this means that $a = a^{**}$ (or $a = a^*$).

Note 2. The above partitioning of the vector x obviously depends on the point $\bar{x}$ in O_{f_0} chosen. The same is true of the partitioning in (N.7) below and of various subsequent partitionings of x and g. It is understood that all these partitionings refer to the same choice of $\bar{x}$, and that $\bar{x}$, once chosen, remains fixed.

(N.7) $$g = \{g^{[1]}, g^{[2]}\}$$

where

$$g^{[1]}(\bar{x}) = 0, \quad g^{[2]}(\bar{x}) > 0$$

and either component may be empty.

(N.8) $$h(x) = 1 - g(x)$$

where 1 denotes the M-dimensional vector with 1's as components; $h^{[i]} = 1 - g^{[i]}$, $i = 1, 2$.

(N.9) $_{\eta_m}p_m(x) = 1 - [h_m(x)]^{1+\eta_m}, \quad m \in \mathcal{M}.$

(N.10) $\eta = \{\eta_1, \eta_2, \cdots, \eta_M\}$.

(N.11) $_\eta p(x) = \{_{\eta_1}p_1(x), {}_{\eta_2}p_2(x), \cdots, {}_\eta p_M(x)\}$.

(N.12) $_\eta\phi(x, y) = f(x) + y'[_\eta p(x)]$ (the "modified Lagrangian expression") .

2.3.1. *A reformulation of Kuhn-Tucker theorem* 1. This slight generalization of theorem 1 (see [1], p. 484) is needed here because of the meaning of inequalities given in (N.1.4). [The possibility of this type of generalization is indicated in [1] (see pp. 491–492).][4]

We shall say that g satisfies the Constraint Qualification (C.Q.) at x, if the requirements of the definition in (A) of section 1.2 are satisfied with the inequalities (4), (5) in the same section interpreted in the sense of (N.1.4). $\phi(x, y)$ is given by (1) in 1.1. (It is immaterial whether g or $'g$ is used.)

THEOREM. *If f and g are differentiable, $\bar{x} \in O_{f_0}$ and g satisfy* C.Q. *at $\bar{x}$, then there exists a $\bar{y} \in Y$ such that*

$$\bar{y} \geqq 0; \ \bar{\phi}_y \cdot \bar{y} = 0; \ \bar{\phi}_y \cdot y \geqq 0 \textit{ for all } y \geqq 0 ;$$

$$\bar{x} \geqq 0; \ \bar{\phi}_x \cdot \bar{x} = 0; \ \bar{\phi}_x \cdot x \leqq 0 \textit{ for all } x \geqq 0 .$$

[Note that, by virtue of the definitions in 2.2.2, this means that $\bar{\phi}_{y_m} \geqq 0$ if $m \in \mathcal{M}'$, $\bar{\phi}_{y_m} = 0$ if $m \notin \mathcal{M}'$, $\bar{\phi}_{x_n} \leqq 0$ if $n \in \mathcal{N}'$, $\bar{\phi}_{x_n} = 0$ if $n \notin \mathcal{N}'$. The other inequalities of the theorem are also to be interpreted in the sense of (N.1.4).]

[4] See also Hurwicz [9], pp. VIII – 2–6.

2.3.2. Theorem 1.

DEFINITION.[5] *An M-dimensional vector* $\eta = \{\eta_1, \eta_2, \cdots, \eta_M\}$ *is said to be acceptable if, for each* $m \in \mathcal{M}$, (1) $\eta_m \geqq 0$, *and* (2) η_m *is an even integer if* $h_m(\bar{x}) < -1$.

THEOREM 1. *If, for some* $\rho > 0$, $x \in S_\rho(\bar{x})$, $\bar{x} \in O_{f_0}$, f *and* g *are differentiable, and* g *satisfies* C.Q. *at* $\bar{x}$, *then, for any acceptable* η, *there exists a vector* $\bar{y} = \bar{y}(\eta)$ *such that*

(13) $$_\eta\bar{\phi}_x \cdot x \leqq 0 \text{ for all } x \geqq 0 ;$$

(14) $$_\eta\bar{\phi}_x \cdot \bar{x} = 0 ;$$

(15) $$\bar{x} \geqq 0 ;$$

(16) $$_\eta\bar{\phi}_y \cdot y \geqq 0 \text{ for all } y \geqq 0 ;$$

(17) $$_\eta\bar{\phi}_y \cdot \bar{y} = 0 ;$$

(18) $$\bar{y} \geqq 0 .$$

The bar over ϕ denotes evaluation at $x = \bar{x}$, $y = \bar{y}(\eta)$.

Note that the relations (13)–(18) are necessary conditions for a nonnegative, in the sense of (N.1.4), saddle-point of $_\eta\phi(x, y)$ at $(\bar{x}, \bar{y})$. In particular, the relations (13)–(18) are satisfied if one selects $\bar{y} = \bar{y}(\eta)$ such that[6]

(19) $$(1 + \eta_m)\bar{y}_m(\eta) = \bar{y}_m(0) \text{ for all } m \in \mathcal{M} .$$

If the selection is made in accordance with (19), the equality

(20) $$_0\bar{\phi}_x = {}_\eta\bar{\phi}_x$$

will hold. Here $_0\phi(x, y)$ is $_\eta\phi(x, y)$ with $\eta = 0$; this is obviously the same as $\phi(x, y)$ in (1) of 1.1.

PROOF. For $\eta = 0$, the preceding theorem follows directly from the reformulated version of the Kuhn-Tucker theorem 1 given in 2.3.1. Thus there exists a vector

(21) $$\bar{y}(0) = \{\bar{y}_1(0), \bar{y}_2(0), \cdots, \bar{y}_M(0)\}$$

with the required properties.

Consider now the case $\eta \neq 0$. We shall show that $\bar{y}(\eta)$ defined by (19), that is, explicitly, by

(22) $$y_m(\eta) = \frac{1}{1 + \eta_m} \bar{y}_m(0) , \qquad m \in \mathcal{M}$$

[where $\bar{y}_m(0)$ is that of (21)], satisfies the relations (13)–(20).

We first observe that (22) yields

(23) $$(1 + \eta_m)\bar{y}_m(\eta)[h_m(\bar{x})]^{\eta_m} = \bar{y}_m(0) , \qquad m \in \mathcal{M} .$$

[When $h_m(\bar{x}) = 1$, (23) follows directly from (22). When $h_m(\bar{x}) \neq 1$, we have $_0\bar{\phi}_{y_m} = g_m(\bar{x}) > 0$, and hence, by (16)–(18), $\bar{y}_m(0) = 0$; (22) then yields $\bar{y}_m(\eta) = 0$ and (23) follows.]

Since

(24) $$_\eta\phi_{x_n} = f_{x_n} + \sum_{m=1}^{M} (1 + \eta_m) y_m(\eta) [h_m(x)]^{\eta_m} \frac{\partial g_m(x)}{\partial x_n} , \qquad n \in \mathcal{N} ,$$

[5] In many applied problems, $h_m(x) \geqq 0$ for all m and all $x \geqq 0$. It was pointed out by Dr. Masao Fukuoka that, in the absence of such an assumption, the requirement of nonnegativity of the components of η is insufficient for the proof of the theorem.

[6] $\bar{y}_m(0) = \bar{y}$ in Kuhn-Tucker theorem 1 (see 2.3.1).

formula (23) implies

$$ {}_\eta\bar{\phi}_{x_n} = f_{x_n} + \sum_{m=1}^{M} \bar{y}_m(0) \frac{\partial g_m(\bar{x})}{\partial x_n}, \qquad n \in \mathcal{N}. \tag{25} $$

Noting that the right member of (25) is identical with ${}_0\bar{\phi}_{x_n}$, we conclude that the relations (13)–(15) hold for all η with nonnegative components, since they are known to hold for $\eta = 0$.

Relation (16) is established by the fact that the right member of

$$ {}_\eta\bar{\phi}_{y_m} = {}_\eta p_m(\bar{x}) = 1 - [h_m(\bar{x})]^{1+\eta_m}, \qquad m \in \mathcal{M}, \tag{26} $$

is nonnegative for $m \in \mathcal{M}'$, zero for $m \notin \mathcal{M}'$ when η is acceptable (see the definition above) since, for any $m \in \mathcal{M}$, $h_m(\bar{x}) \leqq 1$, and, furthermore, ${}_\eta\bar{\phi}_{y_m} = 0$ if $m \notin \mathcal{M}'$, in which case $h_m(\bar{x}) = 1$.

Now suppose that, for some $m_0 \in \mathcal{M}$, ${}_\eta\bar{\phi}_{y_{m_0}} > 0$, that is, $h_{m_0}(\bar{x}) < 1$; then, by (16)–(18) for $\eta = 0$, $\bar{y}_{m_0}(0) = 0$; hence $\bar{y}_{m_0}(\eta) = 0$, and, therefore,

$$ \bar{\phi}_{y_{m_0}} \cdot \bar{y}_{m_0}(\eta) = 0. \tag{27} $$

Since (27) clearly holds in the alternative case ${}_\eta\phi_{m_0} = 0$, (17) follows.

Finally, (18) holds because $\bar{y}_m(\eta)$ has the same sign as $\bar{y}_m(0)$ and the latter, by (18) for $\eta = 0$, is nonnegative if $m \in \mathcal{M}'$.

2.3.3. THEOREM 2. *Let, for some* $\rho > 0$, $x \in S_\rho(\bar{x})$, $\bar{x} \in O_{f_0}$, *such that* (13)–(20) *are satisfied. Then*

$$ {}_\eta\phi(\bar{x}, \bar{y}) \leqq {}_\eta\phi(\bar{x}, y) \text{ for all } y \geqq 0. \tag{28} $$

For we have

$$ {}_\eta\phi(\bar{x}, y) - {}_\eta\phi(\bar{x}, \bar{y}) = (y - \bar{y}) \cdot {}_\eta\bar{p} = y \cdot {}_\eta\bar{p} \geqq 0 \text{ for } y \geqq 0 \tag{29} $$

where, since

$$ {}_\eta\bar{\phi}_y = {}_\eta\bar{p}, \tag{30} $$

the second equality follows from (17) and the inequality from (16).

2.3.4. *Notation.*

$$ x^{(2)} = \{x^{(21)}, x^{(22)}\} \tag{N.13} $$

where

$$ {}_0\bar{\phi}_{x^{(21)}} = 0, \quad {}_0\bar{\phi}_{x^{(22)}} < 0 $$

and either component may be empty.

$$ x = \{x^I, x^{II}\} \tag{N.14} $$

where

$$ \begin{aligned} x^I &= \{x^{(1)}, x^{(21)}\} \\ x^{II} &= x^{(22)}. \end{aligned} \tag{N.14.1} $$

(Either x^I or x^{II} may be empty.)

It should be noted that, by (13)–(15) and (N.13),

$$ \begin{aligned} {}_0\bar{\phi}_{x^I} &= 0, \\ {}_0\bar{\phi}_{x^{II}} &< 0. \end{aligned} \tag{N.14.2} $$

2.3.5. *Definition of a regular constrained maximum.* In theorem 3 below we use the concept of a *regular* constrained maximum. The definition of such a maximum is given in the last part of this section. To state it, we must first formulate three regularity conditions denoted by R_1, R_2, R_3.

The first regularity condition R_1. Let $\bar{x}$ be a value maximizing the function $f(x)$ subject to $'g(x) \geqq 0$, $x \geqq 0$, and hence also subject to

$$
\begin{aligned} g(x) &\geqq 0 \\ x &\geqq 0 \end{aligned} \tag{31}
$$

where the inequalities are to be interpreted in the sense of (N.1.4).

From (N.6) and (N.7) it is clear that, for sufficiently small variations of x, the constraints

$$
\begin{aligned} g^{[2]}(x) &\geqq 0 \\ x^{(1)} &\geqq 0 , \end{aligned} \tag{32}
$$

which are a part of (31), can be disregarded. Hence, at $\bar{x}, f(x)$ possesses a *local* maximum subject to

$$
\begin{aligned} g^{[1]}(x) &\geqq 0 , \\ x^{(2)} &\geqq 0 . \end{aligned} \tag{33}
$$

Let $g^{\dagger}$ be a subvector of $g^{[1]}$ such that $C_g = C_{\{g^{\dagger}, g^{[1]}\}}$ and write

$$
g^{[1]} = \{g^{\dagger}, g^{\dagger\dagger}\} \tag{34}
$$

The components of $g^{\dagger\dagger}$ can be disregarded in the 'process of maximization, that is, $O_{f,g} = O_{f,\{g^{\dagger}, g^{[1]}\}}$. If the Lagrangian multiplier vector $\bar{y}^{[1]}$ (corresponding to the constraints $g^{[1]}(x) \geqq 0$) is partitioned according to

$$
\bar{y}^{[1]} = \{\bar{y}^{\dagger}, \bar{y}^{\dagger\dagger}\} \tag{35}
$$

it is always possible to put

$$
\bar{y}^{\dagger\dagger} = 0 , \tag{36}
$$

and this will be done in what follows.

Assuming that the constraints (33) are consistent, we may replace them by

$$
\begin{aligned} g^{\dagger}(x) &= 0 \\ x^{(2)} &\geqq 0 . \end{aligned} \tag{37}
$$

The *first regularity condition* is

$$
\text{rank}\ (\bar{g}^{\dagger}_{x^{(1)}}) = \dim g^{\dagger} = M^{\dagger} , \tag{R_1}
$$

say.

Note 1. R_1 corresponds to the requirement of nondegeneracy in linear programming (see [4], p. 340).

Note 2. R_1 implies C.Q. (see appendix I).

The second regularity condition R_2. Since, by (N.7), (N.6), (N.14.1), and (34),

$$
\begin{aligned} g^{\dagger}(\bar{x}) &= 0 \\ \bar{x}^{II} &= 0 , \end{aligned} \tag{38}
$$

it follows that, as a function of x^I, $f(x^I, \bar{x}^{II}) \equiv f(x^I, 0)$ has at $\bar{x}^I$ a local maximum subject to the constraints

$$(39) \qquad \begin{aligned} g^\dagger(x^I, \bar{x}^{II}) &\equiv g^\dagger(x^I, 0) = 0 \\ x^{(21)} &\geqq 0 . \end{aligned}$$

The corresponding Lagrangian expression becomes

$$(40) \qquad {}_0\phi^I(x^I, y^\dagger) = f(x^I, 0) + y^\dagger \cdot g^\dagger(x^I, 0) .$$

Using the reformulation of Kuhn-Tucker theorem 1, given in 2.3.1, we may assert the existence of a $\bar{y}^\dagger$ such that

$$(41) \qquad \bar{x}^I \geqq 0; \; {}_0\bar{\phi}^I_{x^I} = 0 ;$$

$$(42) \qquad \bar{y}^\dagger \geqq 0; \; {}_0\bar{\phi}^I_{y^\dagger} = 0 .$$

It might happen that some components of $\bar{y}^\dagger$ vanish. Write $y^\dagger = \{y^*, y^0\}$ where every component of $\bar{y}^*$ is different from zero and

$$(43) \qquad \bar{y}^0 = 0 .$$

Let $g^\dagger$ be correspondingly partitioned as

$$(44) \qquad g^\dagger = \{g^*, g^0\} .$$

Now suppose that ${}_0\phi^I(x^I, y^\dagger)$ has a nonnegative saddle-point at $(\bar{x}^I, \bar{y}^\dagger)$. By theorem 3 in Kuhn-Tucker, a sufficient condition for this is that f and g be both concave. One can then easily verify that

$$(45) \qquad {}_0\phi^I(x^I, y^*) \equiv f(x^I, 0) + y^* \cdot g^*(x^I, 0)$$

has a nonnegative saddle-point at $(\bar{x}^I, \bar{y}^*)$.

But then $\bar{x}^I$ maximizes $f(x^I, 0)$ subject to $g^*(x^I, 0) \geqq 0$ and $x^{(21)} \geqq 0$. Hence in this case the components of g^0 could have been disregarded in the original maximation problem ($O_{f,g} = O_{f,\{g^*, g^{[2]}\}}$).

However, complications might arise if ${}_0\phi^I(x^I, y^\dagger)$ did not have a nonnegative saddle-point at $(\bar{x}^I, \bar{y}^\dagger)$. To take care of this case, one might require that

(46) g^0 is empty unless ${}_0\phi^I(x^I, y^\dagger)$ has a local nonnegative saddle-point at $(\bar{x}^I, \bar{y}^\dagger)$.

However, to simplify matters we shall impose the seemingly[7] stronger condition

(47) g^0 is empty .

It follows that

$$(48) \qquad M^* \equiv \dim g^* = \dim g^\dagger \equiv M^\dagger .$$

Let $\mathcal{M}^*$ [$= \mathcal{M}^\dagger$ by (47)] denote the set of indices of g^*. Clearly, for $m \in \mathcal{M}^* \cap (\mathcal{M} \sim \mathcal{M}')$, we may have $\bar{y}_m < 0$.

Now suppose the preceding reasoning had been carried out in terms of $'g$ instead of g. Nothing would be changed, except, possibly, the signs of some components of the Lagrangian multiplier, to be denoted by $'\bar{y}$.

That is, we would have $'\bar{y}_m > 0$ for $m \in \mathcal{M}^* \cap \mathcal{M}'$ and $'\bar{y}_m > 0$ or $'\bar{y}_m < 0$ for

[7] See section 2.3.7.

$m \in \mathcal{M}^* \cap (\mathcal{M} \sim \mathcal{M}')$. Let $\mathcal{M}^\cdot$ be defined by the relation $m \in \mathcal{M}^\cdot$ if and only if $m \in \mathcal{M}^* \cap (\mathcal{M} \sim \mathcal{M}')$ and $'\bar{y}_m < 0$. Then, it is clear from (N.2.2) that we may put

$$\begin{aligned} \bar{y}_m &= {}'\bar{y}_m \text{ for } m \in \mathcal{M} \sim \mathcal{M}^\cdot \\ \bar{y}_m &= -{}'\bar{y}_m \text{ for } m \in \mathcal{M}^\cdot , \end{aligned} \tag{49}$$

so that $\bar{y}_m > 0$ for all $m \in \mathcal{M}^*$.

Hence, *without loss of generality* [as compared with (47)] condition (47) may be restated as the *second regularity condition*,

$$\begin{aligned} &g^0 \text{ is empty and} \\ &\bar{y}_m > 0 \text{ for all } M \in \mathcal{M}^* . \end{aligned} \tag{R$_2$}$$

The first regularity condition then implies

$$\text{rank } (\bar{g}^*_{x^{(1)}}) = M^* \tag{50}$$

where

$$M^* = \dim g^* . \tag{51}$$

The third regularity condition R_3. When the first two regularity conditions are satisfied, second derivatives are continuous, and x^I is nonempty, it is possible to show (see appendix II) that a certain quadratic form is nonpositive when some of the variables are restricted in sign. The third regularity condition is a strengthening of (71) requiring that the quadratic form in question be negative under the same restrictions. This condition, analogous to that used by Samuelson (see [5], p. 358) makes it possible to avoid going beyond second order terms in the expansions used.

The third regularity condition is formulated in terms of a function $q(t)$ of a new variable vector

$$t = \{t^*, t^{**}\} \tag{52}$$

which is obtained by a transformation of coordinates from x^I after the latter has been partitioned so that

$$x^I = \{x^*, x^{**}\} , \tag{53}$$

where x^* is a subvector of $x^{(1)}$.

We shall (a) define x^* and x^{**}; (b) write down the transformation defining $\{t^*, t^{**}\}$ in terms of $\{x^*, x^{**}\}$; (c) define $q(t)$; and (d) formulate the third regularity condition.

In the remainder of this section it is assumed that R_1 holds; it is also assumed that x^I is not empty.

First case: $M^* = 0$. Write

$$t = t^{**} = x^{**} = x^I , \tag{54}$$

so that, by (52) and (53), x^* and t^* are empty, and define

$$q(t) = f(x^I, \bar{x}^{II}) = f(t^{**}, 0) . \tag{55}$$

The third regularity condition for this case is formulated in R_3 below.

Second case: $M^* > 0$. (a) *The definition of* x^*. From R_1 it follows that there exists a (nonempty) M^*-dimensional subvector x^* of $x^{(1)}$ such that

$$\bar{g}^*_{x^*} \text{ is an } M^* \text{ by } M^* \ (M^* \geqq 1) \text{ nonsingular matrix} . \tag{56}$$

We then define x^{**} by (53) and $x^{(12)}$ by

(57) $$x^{(1)} = \{x^*, x^{(12)}\}$$

Clearly

(58) $$x^{**} = \{x^{(12)}, x^{(21)}\}$$

(b) *The transformation from x^I to t.* Let

(59) $$h^* = 1 - g^*$$

where 1 is the M^*-dimensional vector with (scalar) 1's as components. $t = \{t^*, t^{**}\}$ is then defined by the transformation

(60) $$t^* = h^*(x^*, x^{**}, \bar{x}^{II})$$

(61) $$t^{**} = x^{**}$$

We also partition t^{**} by

(62) $$t^{**} = \{t^{(12)}, t^{(21)}\}$$

where

(63) $$t^{(12)} = x^{(12)}, \quad t^{(21)} = x^{(21)}$$

This is obviously consistent with (57) and (61).

(c) *The definition of $q(t)$.* By (59), the Jacobian H of the transformation (58)–(59) is

(64) $$H = \begin{pmatrix} h^*_{x^*} & h^*_{x^{**}} \\ 0 & I \end{pmatrix} = -\begin{pmatrix} g^*_{x^*} & g^*_{x^{**}} \\ 0 & -I \end{pmatrix}$$

so that, by (56),

(65) $$|\bar{H}| = -|-\bar{g}^*_{x^*}| \neq 0,$$

that is,

(66) $$\bar{H} \text{ is nonsingular}.$$

Hence, locally, (60)–(61) can be solved for x^I in terms of t; we may write this solution as

(67) $$x^I = r(t)$$

where

(68) $$r = \{r^*, r^{**}\}$$

and

(69) $$x^* = r^*(t), \ x^{**} = r^{**}(t) = t^{**}.$$

The function $q(t)$ is now defined as $f(x)$ evaluated at $x^{II} = \bar{x}^{II}$ and with x^I expressed in terms of t, that is,

(70) $$q(t) = f[r(t), \bar{x}^{II}] \equiv f[r^*(t^*, t^{**}), t^{**}, 0].$$

The statement of the third regularity condition. We have now defined $q(t)$ for all M^* provided the first regularity condition R_1 is satisfied and x^I is nonempty. It is shown in appendix II that, assuming R_1, R_2, and the continuity of the second derivatives, unless x^{**} is empty, there exists $\rho > 0$ such that, for all $t^{**} \in S_\rho(\bar{x}^{**})$,

$$(t^{**} - \bar{x}^{**})' \bar{q}_{t^{**}t^{**}} (t^{**} - \bar{x}^{**}) \leqq 0 , \qquad \text{if } t^{(21)} \geqq 0. \tag{71}$$

The *third regularity condition* is a strengthening of the preceding inequality. It states that

(R_3) (a) x^{**} is empty or

(b) there exists $\rho > 0$ such that, for all $t^{**} \in S_\rho(\bar{x}^{**})$, $(t^{**} - \bar{x}^{**})' \bar{q}_{t^{**}t^{**}}(t^{**} - \bar{x}^{**}) < 0$ if $t^{(21)} \geqq 0$ and $t^{**} \neq \bar{x}^{**}$.

Note. The situation covered by (a) of R_3 is of importance since it permits the treatment of a large class of cases where f and g are linear.

DEFINITION. *$f(x)$ is said to have a regular maximum at $\bar{x}$ subject to $g(x) \geqq 0$, $x \geqq 0$, if the three regularity conditions R_1, R_2, R_3 are satisfied at $\bar{x}$ and $\bar{x} \in O_{fg}$.*

2.3.6. THEOREM 3. *If, for some $\rho > 0$, $x \in S_\rho(\bar{x})$, $\bar{x}$ a regular maximum*[8] *of $f(x)$ subject to $g(x) \geqq 0$ and $x \geqq 0$, f and g are differentiable (with regard to x), and furthermore, when x^I is nonempty, have continuous second order derivatives with regard to x^I, then, for all acceptable*[9] *η sufficiently large in each component,*

$$x^I \text{ is empty} , \tag{72}$$

or

$$(x^I - \bar{x}^I)'{}_\eta\bar{\phi}_{x^Ix^I}(x^I - \bar{x}^I) < 0 \textit{ if } x^{(21)} \geqq 0,\ x^I \neq \bar{x}^I , \tag{73}$$

and for some $\rho' > 0$, and all $x \in S_{\rho'}(\bar{x})$ such that $x \geqq 0$, $x \neq \bar{x}$,

$$_\eta\phi[x, \bar{y}(\eta)] < {}_\eta\phi[\bar{x}, \bar{y}(\eta)] \tag{74}$$

where $_\eta\phi$ and $\bar{y}(\eta)$ are defined as in theorem 1.

Note.[10] Theorem 3 is valid for f, g linear if x^{**} is empty (regardless of whether x^* is empty), provided the first two regularity conditions hold. However, if both x^* and x^{**} are empty, x^I is empty, and the theorem follows from the first case considered below. If x^{**} is empty while x^* is nonempty, use the first two cases below together with (90) (since g^* is nonempty and t^{**} is empty). Note that x^{**} is empty at the basic solutions of a linear programming problem.

2.3.7. *Proof of theorem* 3. First it is shown that (72) or (73) implies (74). Then it is shown that (72) or (73) is true.

It can be seen that if theorem 3 is established for the case of $\{g^{\dagger\dagger}, g^0\}$ empty, then theorem 3 is also true if (i) $g^{\dagger\dagger}$ is not empty, and/or (ii) g^0 is not empty but $_0\phi^I(x^I, y^\dagger)$ has a nonnegative saddle-point at $(\bar{x}^I, \bar{y}^\dagger)$, since in either case $\bar{x}$ remains unchanged and the additional terms in the modified Lagrangian expression vanish at $\bar{y}$ [compare equations (36) and (43)].

Hence, with no loss of generality, we may henceforth assume $\{g^{\dagger\dagger}, g^0\}$ to be empty, that is,

$$g^{[1]} = g^* . \tag{75}$$

We now show that (72) or (73) implies (74), that is, that in a sufficiently small neighborhood, if (72) or (73) is assumed to be valid and the inequalities $x \geqq 0$, $x \neq \bar{x}$, hold,

[8] The term "regular maximum" is defined at the end of section 2.3.5.

[9] The term "acceptable" is defined at the beginning of section 2.3.2.

[10] The desirability of explicit treatment of the linear case was emphasized by Dr. Masao Fukuoka.

the conclusion of (74) follows. We write ϕ instead of ${}_\eta\phi$ throughout. Also (72) or (73), $x \geqq 0$, $x \neq \bar{x}$, is assumed.

Let

$$\xi = x - \bar{x} \tag{76}$$

$$\xi^I = x^i - \bar{x}^i\,, \qquad i = I, II. \tag{77}$$

First case: $\xi^{II} \neq 0$. By (20) and (N.14.2),

$$\bar{\phi}_x \cdot \xi = \bar{\phi}_{x^I} \cdot \xi^I + \bar{\phi}_{x^{II}} \cdot \xi^{II} < 0\,. \tag{78}$$

But then the conclusion of (74) follows from the well-known "Fréchet" property of differentials[11] which, as applied to the present case, states that, given any $\sigma > 0$, there exists an $\epsilon > 0$ such that,

$$\left| \frac{1}{|\xi|} [\phi(x, \bar{y}) - \phi(\bar{x}, \bar{y}) - \bar{\phi}_x \cdot \xi] \right| < \sigma \tag{79}$$

if $|\xi| < \epsilon$.

Choose

$$\sigma = - \frac{1}{|\xi|} \bar{\phi}_x \cdot \xi \tag{80}$$

which is positive by (78). Then, for a sufficiently small $|\xi|$, we have by (79)

$$\left| \frac{1}{|\xi|} [\phi(x, \bar{y}) - \phi(\bar{x}, \bar{y}) + \sigma] \right| < \sigma \tag{81}$$

which implies

$$\frac{1}{|\xi|} [\phi(x, \bar{y}) - \phi(\bar{x}, \bar{y})] < 0 \tag{82}$$

and hence the conclusion of (74).

If x^I is empty, this completes the proof of the theorem 3, since $x \neq \bar{x}$ then implies $\xi^{II} \neq 0$. If x^I is not empty, we must consider the

Second case: $\xi^{II} = 0$. Since it is assumed that $x \neq \bar{x}$, $\xi^{II} = 0$ implies

$$\xi^I \neq 0\,. \tag{83}$$

In virtue of the existence of the second derivatives of ϕ with regard to x^I (by definition of ϕ, and the assumptions concerning the second derivatives of f and g with regard to x^I) we have, by Taylor's theorem,

$$\phi(x, \bar{y}) - \phi(\bar{x}, \bar{y}) = \bar{\phi}_{x^I} \cdot \xi^I + \tfrac{1}{2}(\xi^I)' \tilde{\phi}_{x^I x^I} \xi^I\,, \tag{84}$$

where $\tilde{\phi}_{x^I x^I}$ denotes $\phi_{x^I x^I}$ evaluated at $x = \tilde{x}$, $\tilde{x} = \bar{x} + \theta\xi$, $0 < \theta < 1$. It now suffices to note that $(\xi^I)' \phi_{x^I x^I} \xi^I$ is negative at $\bar{x}$ [since (72) or (73) is assumed to hold and its hypotheses are satisfied] and continuous in the neighborhood of $\bar{x}$ (by the hypotheses of the theorem concerning the second derivatives of f and g), so that, for a sufficiently small $|\xi^I|$, $(\xi^I)' \tilde{\phi}_{x^I x^I} \xi^I < 0$. Since $\bar{\phi}_{x^I} \cdot \xi^I = 0$ by (N.14.2), (74) follows.

We now show that (72) holds if x^I is nonempty.

First case: g^* empty. By equation (75), $g^{[1]}$ is also empty. Hence, by (13)–(15) in theorem 1,

$$\bar{y}(\eta) = \bar{y}^{[2]}(0) = 0 \tag{85}$$

[11] See Hille [10], p. 72, definition 4.3.4, equation (iii).

and, using (N.12),

(86) $$_\eta\phi[x, \bar{y}(\eta)] = f(x) .$$

Since g^* is empty, we have $M^* = 0$, and, therefore, the definition (55) of q applies, so that (since x^* is empty but x^I is not) t^{**} is not empty and

(87) $$\bar{q}_{t^{**}t^{**}} = \bar{f}_{x^{**}x^{**}} = {}_\eta\bar{\phi}_{x^Ix^I} .$$

Equations (86) and (87), together with the third regularity condition R_3, yield (73) for a sufficiently small neighborhood of $\bar{x}$.

Second case: g^* nonempty. Write

(88) $$\psi(t, y) = \phi[r(t), \bar{x}^{II}, y]$$

where $r(t)$ is defined in (67). (Where it is desired to indicate the dependence of ψ on η, we may write $_\eta\psi$ instead of ψ.)

Then, by (66), that is, R_1, we have

(89) $$\bar{\psi}_{tt} = \psi_{tt}\big|_{t=\bar{t}} = (\bar{H}^{-1})'{}_\eta\bar{\phi}_{x^Ix^I}\bar{H}^{-1}, \ \bar{t} = \{h^*(\bar{x}), \bar{x}^{**}\} ,$$

since $_\eta\bar{\phi}_{x^I} = {}_0\bar{\phi}_{x^I} = 0$ by (20) and (N.14.2).

We shall now show that (73) is implied by

(90) $$\tau'\bar{\psi}_{tt}\tau < 0, \text{ if } \tau^{(21)} \geqq 0, \text{ and } \tau \neq 0$$

where the partitioning of τ corresponds to that of t. We show later that (90) holds.

To see that (90) implies (73), let x^I satisfy the inequalities $x^{(21)} \geqq 0$, $x^I \neq \bar{x}^I$. Choose

(91) $$\begin{pmatrix} \tau^* \\ \tau^{**} \end{pmatrix} = \tau = \bar{H}\,(x^I - \bar{x}^I) = \begin{pmatrix} \bar{h}^*_{x^*} & \bar{h}^*_{x^{**}} \\ 0' & I \end{pmatrix} \begin{pmatrix} x^* - \bar{x}^* \\ x^{**} - \bar{x}^{**} \end{pmatrix} .$$

Since, by (66), $\bar{H}$ is nonsingular, $x^I \neq \bar{x}^I$ implies $\tau \neq 0$. Also, (91) yields

(92) $$\tau^{**} = x^{**} - \bar{x}^{**} ,$$

hence, in particular,

(93) $$\tau^{(21)} = x^{(21)} - \bar{x}^{(21)} .$$

But

(94) $$\bar{x}^{(21)} = 0 ,$$

since $x^{(21)}$ is a component of $x^{(2)}$ by (N.13), and $\bar{x}^{(2)} = 0$ by (N.6). Hence

(95) $$\tau^{(21)} = x^{(21)}$$

and thus $x^{(21)} \geqq 0$ implies $\tau^{(21)} \geqq 0$.

Having shown that the hypotheses of (73) imply those of (90), we see that the hypotheses of (73), together with the validity of the assertion in (90), yield

(96) $$\tau'\bar{\psi}_{tt}\tau < 0 .$$

But, using in succession (91), (89), and simplifying, we have

(97) $$\begin{aligned} \tau'\psi_{tt}\tau &= (x^I - \bar{x}^I)'\bar{H}'\bar{\psi}_{tt}\bar{H}(x^I - \bar{x}^I) \\ &= (x^I - \bar{x}^I)'\bar{H}'(\bar{H}^{-1})'{}_\eta\phi_{x^Ix^I}\bar{H}^{-1}\bar{H}(x^I - \bar{x}^I) \\ &= (x^I - \bar{x}^I)'{}_\eta\bar{\phi}_{x^Ix^I}(x^I - \bar{x}^I) . \end{aligned}$$

Formulas (96) and (97) yield the conclusion of (73). Thus it has been established that (90) implies (73). It remains to be shown that (90) is valid. It is convenient to write $\bar{\psi}_{tt}$ in the partitioned form

$$\bar{\psi}_{tt} = \begin{pmatrix} \bar{\psi}_{t^*t^*} & \bar{\psi}_{t^*t^{**}} \\ \bar{\psi}_{t^{**}t^*} & \bar{\psi}_{t^{**}t^{**}} \end{pmatrix} \equiv \begin{pmatrix} A & B \\ B' & C \end{pmatrix} \tag{98}$$

where t^{**} may be empty; t^* is assumed nonempty, since the case of t^* empty was treated earlier.

It will now be shown that A, that is, $\bar{\psi}_{t^*t^*}$ [compare (98)], which depends on η, can be made negative definite by a suitable choice of η.

Recalling that $\mathcal{M}^*$ denotes the set of indices of the components of g^*, and using (N.9) and (60), we see that, for $m \in \mathcal{M}^*$,

$$_{\eta_m}p_m(x) = 1 - t_m^{1+\eta_m} \tag{99}$$

where t_m is a component of t^*.

Since, by theorem 1 and equation (75),

$$\bar{y}_m(\eta) = 0 \text{ for } m \in \mathcal{M} \sim \mathcal{M}^*, \tag{100}$$

we have, from the definitions of ψ, q, and $_\eta\phi$ [equations (88), (70), and (N.12), respectively], and the preceding relations (99) and (100), the equality

$$\psi[t, \bar{y}(\eta)] = q(t) + \sum_{m \in \mathcal{M}^*} [\bar{y}_m(\eta)](1 - t_m^{1+\eta_m}). \tag{101}$$

Writing

$$F = \bar{q}_{t^*t^*}, \tag{102}$$

we have, from (101) and the definition of A that

$$A = F - D, \tag{103}$$

where $D = ||d_{m,m'}||$, $m \in \mathcal{M}^*$, $m' \in \mathcal{M}^*$, is a diagonal matrix [that is, $d_{m,m'} = 0$ for $m \neq m'$] with

$$d_{m,m} = [\bar{y}_m(\eta)](1 + \eta_m)\eta_m = \bar{y}_m(0)\eta_m, \qquad m \in \mathcal{M}^*, \tag{104}$$

where the second equality follows from (19).

Let λ denote the largest characteristic root of F. Since, by the second regularity condition R_2, $\bar{y}_m(0) > 0$ if $m \in \mathcal{M}^*$, we may choose η_m^0, for each $m \in \mathcal{M}^*$, to be a positive even integer satisfying

$$\eta_m^0 > \lambda/\bar{y}_m(0), \tag{105}$$

so that

$$\min_{m \in \mathcal{M}^*} d_{m,m} > \lambda \tag{106}$$

for all acceptable $\eta_m \geqq \eta_m^0$.

Then, for any $t^* \neq 0$, and each acceptable $\eta_m \geqq \eta_m^0$, we have

$$\begin{aligned} t^{*\prime}Ft^* \leqq \lambda t^{*\prime}t^* \equiv \lambda \sum_{m \in \mathcal{M}^*} t_m^2 < \sum_{m \in \mathcal{M}^*} d_{m,m}\, t_m^2 \\ \equiv t^{*\prime}Dt^*, \end{aligned} \tag{107}$$

that is, $t^* \neq 0$ implies $t^{*\prime}(F - D)t^* < 0$ for all sufficiently large acceptable η, or A is negative definite for all sufficiently large acceptable η.

This suffices to establish (90) and, therefore, (73) if t^{**} is empty.

Now assume t^{**} not empty. Write

$$P = \begin{pmatrix} I & -A^{-1}B \\ 0 & I \end{pmatrix} \tag{108}$$

and

$$_\eta\Omega = P'{}_\eta\bar{\psi}_{tt}P \,. \tag{109}$$

Then methods used to show that (90) implies (73) can be used to show that

$$w'{}_\eta\Omega w < 0 \text{ for } w \neq 0,\ w^{(21)} \geqq 0 \tag{110}$$

implies (90). This is because

$$P^{-1} = \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix} \tag{111}$$

and, like its analogue $\bar{H}$, performs an identity transformation on t^{**}, so that the condition $t^{(21)} \geqq 0$ is transformed into the condition $w^{(21)} \geqq 0$. It remains to establish (110). Now from (109), (108), and (98), we have

$$_\eta\Omega = \begin{pmatrix} A & 0 \\ 0 & C - B'A^{-1}B \end{pmatrix}, \tag{112}$$

so that $w'_\eta\Omega w = w^{*\prime}Aw^* + w^{**\prime}\,(C - B'A^{-1}B)w^{**}$.

Now, we may take A as negative definite, and hence, to establish (110), it will suffice to show that

$$Q \equiv w^{**\prime}(C - B'A^{-1}B)w^{**} < 0, \text{ if } w^{**} \neq 0,\ w^{(21)} \geqq 0 \,. \tag{113}$$

Before doing so, we shall obtain an auxiliary result.

It will now be shown that the norm of A^{-1} can be made arbitrarily small by choosing η sufficiently large. It does not matter which of the many norms is used (see Bowker [6]). Note that, denoting by $N(X)$ the norm of the matrix X, we have $N(A + B) \leqq N(A) + N(B)$, $N(AB) \leqq N(A)N(B)$; if all the elements of a matrix approach 0, so does its norm. If I denotes the identity matrix, $N(I) = 1$.

D^{-1} is a diagonal matrix whose nonzero elements approach zero for η large; hence, the same is true of $D^{-1}F$. Therefore, η can be chosen sufficiently large so that,

$$I - D^{-1}F \text{ is nonsingular}\,, \tag{114}$$

and

$$N(D^{-1}F) < 1\,. \tag{115}$$

Following Waugh (see p. 148, [7]), we use the identity, valid because of (114),

$$(I - D^{-1}F)^{-1} = I + (I - D^{-1}F)^{-1}D^{-1}F\,, \tag{116}$$

and the properties of the norm to derive the relation,

$$N[(I - D^{-1}F)^{-1}] \leqq 1 + N[(I - D^{-1}F)^{-1}]N(D^{-1}F)\,. \tag{117}$$

From (117) and (115), it follows that,

$$N\,[\,(D^{-1}F - I)^{-1}] \leqq \frac{1}{1 - N\,(D^{-1}F)}\,. \tag{118}$$

Since $A = F - D = D(D^{-1}F - I)$, it follows that $A^{-1} = (D^{-1}F - I)^{-1}D^{-1}$, and hence

$$N(A^{-1}) \leqq N(D^{-1})\, N[(D^{-1}F - I)^{-1}] \leqq \frac{N(D^{-1})}{1 - N(D^{-1}F)} \tag{119}$$

which can be made arbitrarily small for η large.

Consider now the quadratic form Q in (113). We have shown, using (101), that

$$C = \bar{\psi}_{t^{**}t^{**}} = \bar{q}_{t^{**}t^{**}}\,. \tag{120}$$

Hence the third regularity condition, R_3, implies

$$w^{**\prime}Cw^{**} < 0 \text{ if } w^{**} \neq 0,\ w^{(21)} \geqq 0\,. \tag{121}$$

As shown earlier $N(B'A^{-1}B) \leqq N(B')N(A^{-1})N(B) = N(A^{-1})[N(B)]^2$ can be made arbitrarily small by choosing a large enough η. Now

$$|w^{**\prime}B'A^{-1}Bw^{**}| \leqq N(B'A^{-1}B)w^{**\prime}w^{**}\,, \tag{122}$$

since the characteristic roots of a matrix are bounded in absolute value by its norm.

Also, denoting by μ the maximum of $w^{**\prime}Cw^{**}$ subject to $w^{**\prime}w^{**} = 1$, $w^{(21)} \geqq 0$, we have

$$w^{**\prime}Cw^{**} \leqq \mu w^{**\prime}w^{**} \tag{123}$$

and, by (121), $\mu < 0$. With the aid of (122),

$$Q < [\mu + N(B'A^{-1}B)]w^{**\prime}w^{**} \text{ if } w^{**} \neq 0,\ w^{(21)} \geqq 0\,. \tag{124}$$

By choosing η sufficiently large, so that

$$\mu + N(B'A^{-1}B) < 0\,, \tag{125}$$

we establish (113), which, in turn, yields (110), (90), (73), and hence theorem 3.

APPENDIX I[12]

Let the first regularity condition R_1 hold. Consider $\bar{x}$ such that,

$$g^{[1]}(\bar{x}) = 0,\ g^{[2]}(\bar{x}) > 0,\ \bar{x} \geqq 0\,, \tag{126}$$

and x such that,

$$\bar{g}_x^{[1]}(x - \bar{x}) \geqq 0,\ x^{(2)} - \bar{x}^{(2)} \geqq 0\,, \tag{127}$$

where all inequalities are to be interpreted in the sense of (N.1.4). Define now the function $g^{\sharp}$ of $\bar{x}$ by

$$g^{\sharp}(x) = \{g^{\dagger}(x),\ x^{**},\ x^{II}\} \tag{128}$$

where x^{**} is defined by (58). Notice that assuming g^0 to be empty as in (47), $g^{\sharp}$, like x, has N dimensions.

It follows that

$$g_x^{\sharp} = \begin{pmatrix} g_{x^*}^{\dagger} & g_{x^{**}}^{\dagger} & g_{x^{II}}^{\dagger} \\ 0 & I & 0 \\ 0 & 0 & I \end{pmatrix} \tag{129}$$

and hence

$$|\bar{g}_x^{\sharp}| = |\bar{g}_{x^*}^{\dagger}| \neq 0\,. \tag{130}$$

[12] This appendix parallels lemma 76.1 in Bliss [11].

Consider now the relation which associates with a real number α the values $\bar{\bar{x}}$ of x for which the equation

$$g^{\#}(\bar{\bar{x}}) = g^{\#}(\bar{x}) + \alpha \bar{g}^{\#}_{x}(x - \bar{x}) \tag{131}$$

is satisfied. In virtue of the implicit function theorem, for sufficiently small values of α(131) defines $\bar{\bar{x}}$ as a (single-valued) differentiable function of α, say

$$\bar{\bar{x}} = \psi_1(\alpha)\,, \tag{132}$$

such that

$$\psi_1(0) = \bar{x}\,. \tag{133}$$

Differentiating (131) with respect to α and setting $\alpha = 0$, we have

$$\bar{g}^{\#}_{x}\psi_1'(0) = \bar{g}^{\#}_{x}(x - \bar{x}) \tag{134}$$

and hence, because of (130),

$$\psi_1'(0) = x - \bar{x}\,. \tag{135}$$

We shall now show that

$$\psi_1(\alpha) \in C_g \text{ for } \alpha \geqq 0, \ \alpha \text{ sufficiently small}\,. \tag{136}$$

By (131), (126), and (127)

$$g^{\dagger}(\bar{\bar{x}}) = \alpha \bar{g}^{\dagger}_{x}(x - \bar{x}) \geqq 0 \text{ for } \alpha \geqq 0\,. \tag{137}$$

It follows that

$$g^{[1]}(\bar{\bar{x}}) \geqq 0 \text{ for } \alpha \geqq 0\,, \tag{138}$$

which together with

$$g^{[2]}[\psi_1(\alpha)] \geqq 0 \text{ for } \alpha \text{ sufficiently small}\,, \tag{139}$$

yields

$$g[\psi_1(\alpha)] \geqq 0 \text{ for } \alpha \geqq 0 \text{ sufficiently small}\,. \tag{140}$$

Now, since x^* is a subvector of $x^{(1)}$, $x^{(2)}$ is a subvector of $\{x^{**}, x^{II}\}$, hence (127) and (131) imply

$$\bar{\bar{x}}^{(2)} = \psi_1^{(2)}(\alpha) = \bar{x}^{(2)} + \alpha(x^{(2)} - \bar{x}^{(2)}) \geqq 0 \text{ for } \alpha \geqq 0 \tag{141}$$

which, together with

$$\bar{\bar{x}}^{(1)} = \psi_1^{(1)}(\alpha) \geqq 0 \text{ for } \alpha \text{ sufficiently small, yields} \tag{142}$$

$$\psi_1(\alpha) \geqq 0 \text{ for } \alpha \geqq 0, \ \alpha \text{ sufficiently small}\,. \tag{143}$$

In turn, (140) and (143) yield (136).

Now let us interpret "α sufficiently small" as $0 \leqq \alpha \leqq \lambda$ where $\lambda > 0$ and define the function ψ by

$$\psi(\theta) = \psi_1(\lambda\theta) \text{ for all } 0 \leqq \theta \leqq 1\,. \tag{144}$$

Then

$$\begin{aligned} &\psi(0) = \bar{x}\,, \\ &\psi'(0) = \lambda\psi_1'(0) = \lambda(x - \bar{x})\,, \qquad && \lambda > 0\,, \\ &\psi(\theta) \in C_g\,, && 0 \leqq \theta \leqq 1\,. \end{aligned} \tag{145}$$

Since (145) are precisely the requirements of C.Q., it has been shown that R_1 implies C.Q.

APPENDIX II

We shall now show that, if the first two regularity conditions hold and if in a neighborhood of $\bar{x}$, f and g are assumed to possess continuous derivatives of second order with regard to x^I, then (71) is valid.

Let x^{**} be nonempty. Then, writing

$$\bar{t}^* = h^*(\bar{x}) = 1 \qquad \text{(a vector of 1's)}, \tag{146}$$

$$\bar{t}^{**} = \bar{x}^{**}, \tag{147}$$

we have, using Taylor's theorem,

$$q(\bar{t}^*, t^{**}) - q(\bar{t}^*, \bar{t}^{**}) = \bar{q}_{t^{**}} \cdot (t^{**} - \bar{t}^{**}) + \tfrac{1}{2}(t^{**} - \bar{t}^{**})' \tilde{q}_{t^{**}t^{**}}(t^{**} - \bar{t}^{**}), \tag{148}$$

where "$-$" over a symbol denotes the evaluation at $t = \bar{t}$, while "$\sim$" over a symbol denotes evaluation at $t = \tilde{t}$, where $\tilde{t} = \bar{t} + \theta\,(t^{**} - \bar{t}^{**})$, $0 < \theta < 1$. Now suppose it has been shown that (a) $q(\bar{t}^*, t^{**})$ has, as a function of t^{**}, subject to the constraint $t^{(21)} \geqq 0$, a local maximum at $t^{**} = \bar{t}^{**}$, and (b) $\bar{q}_{t^{**}} = 0$. From (a) it follows that, in a sufficiently small neighborhood, the left member of (148) is nonpositive if $t^{(21)} \geqq 0$. But then, using (b), we see that the quadratic form in the right member of (148) is nonpositive. Since, by hypothesis, $q_{t^{**}t^{**}}$ is a continuous function of t^{**}, we have, for $t^{(21)} \geqq 0$, and in a sufficiently small neighborhood of $\bar{t}$,

$$(t^{**} - \bar{t}^{**})\bar{q}_{t^{**}t^{**}}(t^{**} - \bar{t}^{**}) \geqq 0 \tag{149}$$

which is the desired result (71). Hence it remains to prove (a) and (b).

(a) *$q(\bar{t}^*, t^{**})$ has, as a function of t^{**}, subject to $t^{(21)} \geqq 0$, a local maximum at $t^{**} = \bar{t}^{**}$.*

It follows from the remarks at the beginning of the discussion of the second regularity condition that $f(x^I, 0)$, as a function of x^I, has a local maximum at $x^I = \bar{x}^I$, subject to the constraints

$$g^\dagger(x^I, 0) = 0, \ x^{(21)} \geqq 0. \tag{150}$$

Hence, subject to the same constraints, $q(t)$ has a local maximum at $\bar{t}$. Now we must distinguish the two ways in which the "milder" (46) second regularity condition R_2 may be satisfied.

First way: ${}_0\phi^I(x^I, y^\dagger)$ has a nonnegative saddle-point at $(\bar{x}^I, \bar{y}^\dagger)$, that is, locally, since $\bar{y}^0 = 0$ by (43),

$$f(x^I, 0) + \bar{y}^* \cdot g^*(x^I, 0) \leqq f(\bar{x}^I, 0) + \bar{y}^* \cdot g^*(\bar{x}^I, 0) \tag{151}$$

for all x^I such that $x^{(21)} \geqq 0$.

But $g^*(\bar{x}^I, 0) = 0$ because of (150), and $g^*(x^I, 0)$ in the left member of (151) vanishes for $t^* = \bar{t}^*$. Hence (151) yields, locally and for $t^{(21)} \geqq 0$,

$$f[r^*(\bar{t}^*, t^{**}),\ t^{**}, 0] \leqq f[r^*(\bar{t}^*, \bar{t}^{**}),\ \bar{t}^{**}, 0] \tag{152}$$

which means precisely that $q(\bar{t}^*, t^{**})$ has a local maximum at $\bar{t}^{**}$ subject only to $t^{(21)} \geqq 0$.

Second way: g^0 is empty. In this case (150) is equivalent to

$$g^*(x^I, 0) = 0, \tag{153}$$

$$x^{(21)} \geqq 0. \tag{154}$$

But (153) is necessarily satisfied if $t^* = \bar{t}^*$ and hence can be disregarded. Since $q(t)$

was seen to have a local maximum at $\bar{t}$ subject to (150), it follows that $q(\bar{t}^*, t^{**})$ will have a local maximum at $\bar{t}^{**}$ subject only to $t^{(21)} \geqq 0$.

(b) $$\bar{q}_{t^{**}} = 0\,.$$

We have

$$\bar{q}_{t^{**}} = \bar{f}_{x^*}\bar{r}^*_{t^{**}} + \bar{f}_{x^{**}} \tag{155}$$

We now evaluate the three expressions on the right-hand side of (155). We start with $\bar{r}^*_{t^{**}}$. Noting that

$$f^*\{[r^*(\bar{t}^*, t^{**}),\ t^{**}],\ 0\} = 0 \text{ for all } t^{**}\,, \tag{156}$$

we obtain by differentiation with respect to t^{**}, using (60) and (69), and evaluating at $t = \bar{t}$,

$$\bar{g}^*_{x^*}\bar{r}^*_{t^{**}} + \bar{g}^*_{t^{**}} = 0\,; \tag{157}$$

in virtue of R_1 this can be solved yielding

$$\bar{r}^*_{t^{**}} = -(\bar{g}^*_{x^*})^{-1}\bar{g}^*_{x^{**}}\,. \tag{158}$$

To find $\bar{f}_{x^*}$, $\bar{f}_{x^{**}}$, we write the condition that ${}_0\bar{\phi}^I{}_{x^I} = 0$, using equation (41) in the form

$$\begin{aligned} \bar{f}_{x^*} + \bar{g}^*_{x^*}\bar{y}^* &= 0\,, \\ \bar{f}_{x^{**}} + \bar{g}^*_{x^{**}}\bar{y}^* &= 0\,. \end{aligned} \tag{159}$$

The terms involving g^0 vanish, of course.

Substituting (159) and (158) into (155), we have

$$\bar{q}_{t^{**}} = (-\bar{y}^*\cdot\bar{g}^*_{x^{**}}) + (-\bar{y}^*\cdot\bar{g}^*_{x^*})[-(\bar{g}^*_{x^*})^{-1}\bar{g}^*_{x^{**}}] = (-\bar{y}^*\cdot\bar{g}^*_{x^{**}}) + (\bar{y}^*\cdot\bar{g}^*_{x^{**}}) = 0\,. \tag{160}$$

This completes the proof of (71).

REFERENCES

[1] H. W. KUHN and A. W. TUCKER, "Nonlinear programming," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley and Los Angeles, University of California Press, 1951, pp. 481–492.

[2] T. C. KOOPMANS, "Analysis of production as an efficient combination of activities," edited by T. C. Koopmans, *Activity Analysis of Production and Allocation*, Cowles Commission Monograph No. 13, New York, John Wiley and Sons, 1951, pp. 33–97.

[3] K. J. ARROW and L. HURWICZ, "A gradient method for approximating saddle-points and constrained maxima," P-223, The RAND Corporation, 1951 (hectographed).

[4] G. B. DANTZIG, "Maximization of a linear function of variables subject to linear inequalities," *Activity Analysis of Production and Allocation*," edited by T. C. Koopmans, Cowles Commission Monograph No. 13, New York, John Wiley and Sons, 1951, pp. 339–347.

[5] P. A. SAMUELSON, *Foundations of Economic Analysis*, Cambridge, Harvard University Press, 1948.

[6] A. H. BOWKER, "On the norm of a matrix," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 285–288.

[7] F. V. WAUGH, "Inversion of the Leontief matrix by power series," *Econometrica*, Vol. 18 (1950), pp. 142–154.

[8] P. A. SAMUELSON, "Market mechanisms and maximization," P-69, The RAND Corporation, 1949 (hectographed).

[9] L. HURWICZ, "Programming in general spaces," Cowles Commission Discussion Paper, Economics No. 2109, 1954 (hectographed).

[10] E. HILLE, *Functional Analysis and Semi-Groups*, New York, American Mathematical Society Colloquium Publications, Vol. 31, 1948.

[11] G. A. BLISS, *Lectures on the Calculus of Variations*, Chicago, The University of Chicago Press, 1946.

CONSTRAINT QUALIFICATIONS IN MAXIMIZATION PROBLEMS*†

Kenneth J. Arrow, Leonid Hurwicz, and Hirofumi Uzawa

Many problems arising in logistics and in the application of mathematics to industrial planning are in the form of constrained maximizations with nonlinear maximands or constraint functions or both. Thus a depot facing random demands for several items may wish to place orders for each in such a way as to maximize the expected number of demands which are fulfilled; the total of orders placed is limited by a budget constraint. In this case, the maximand is certainly nonlinear. The constraint would also be nonlinear if, for example, the marginal cost of storage of the goods were increasing. Practical methods for solving such problems in nonlinear programming almost invariably depends on some use of Lagrange multipliers, either by direct solution of the resulting system of equations or by a gradient method of successive approximations (see [5], Part II). This article discusses a part of the sufficient conditions for the validity of the multiplier method.

INTRODUCTION

This article covers an examination of the interrelationships of the additional assumptions under which the following two propositions, both extensions of the classical Lagrange multiplier method ([1], p. 153), are valid.

Quasi-Saddle-Point Condition

If $\bar{x}$ maximizes $f(x)$, subject to the constraints $g(x) \geqq 0$, and $f(x)$ and $g(x)$ are differentiable, then there exists $\bar{y} \geqq 0$, such that $\bar{f}_x + \bar{y}\,\bar{g}_x = 0$, $\bar{y}\, g(\bar{x}) = 0$.[1]

*Manuscript received January 28, 1959.
†This work was supported in part by the Office of Naval Research (Task NR-047-004) at Stanford University. Hurwicz's participation was made possible by a Rockefeller Foundation grant to Stanford for mathematical research in the social sciences.

[1] x is a column vector with components $x_1, \ldots, x_n$, y a row vector with m components, $f(x)$ is a real-valued function of x, $g(x)$ a column vector with real-valued components $g^j(x)$, $(j = 1, \ldots, m)$, f_x a row vector with components $f_{x_i} = \partial f/\partial x_i$, g_x a matrix with components $g^j_{x_i} = \partial g^j/\partial x_i$, where i varies over columns and j over rows. Bars over f and g or their derivatives denote evaluation at $\bar{x}$. If v is a vector, $v \geqq 0$ means that each component of v is nonnegative; $v > 0$ means that each component is positive.

A function $f(x)$ of a vector variable is said to be differentiable at $\bar{x}$, if there is a row vector a such that

$$\lim_{h \to 0} \; [f(\bar{x} + h) - f(\bar{x}) - ah]/|h| = 0 .$$

(Continued)

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_4, © Springer Basel 2014

Saddle-Point Criterion

If $f(x)$ and $g(x)$ are concave, then a necessary and sufficient condition that $\bar{x}$ maximize $f(x)$, subject to the constraints, $g(x) \geqq 0$ is that there exist a $\bar{y} \geqq 0$, such that $(\bar{x}, \bar{y})$ is a saddle-point subject to $y \geqq 0$, of the Lagrangian function $f(x) + y\, g(x)$.

It is known that neither proposition is valid without additional assumptions.[2] Kuhn and Tucker [11] showed that both propositions are valid if f and g are differentiable and the following condition is satisfied.

Constraint Qualification KT[3]

For all $\bar{x}$ in the constraint set C (defined by the conditions $g(x) \geqq 0$) and all $\bar{\bar{x}}$ such that $\bar{g}^k_x \cdot (\bar{\bar{x}} - \bar{x}) \geqq 0$ for each component $\underline{\underline{k}}$ of $g(x)$ for which $g^k(\bar{x}) = 0$, there exists a differentiable vector-valued function $\psi(\theta)$ such that $\psi(0) = \bar{x}$, $\psi(\theta)$ belongs to C for all positive θ sufficiently small, and $\psi'(0) = \bar{\bar{x}} - \bar{x}$.

Also in this article, further results on the subject are discussed and simplified proofs are given. First, Constraint Qualification KT are slightly weakened so that the meaning of the qualification becomes more transparent. Theorem 1 shows that the Lagrangian method can be applied to those constrained maxima for which the weaker version of the Constraint Qualification is satisfied. Next this article shows that the Constraint Qualification in the present formulation is the weakest requirement for the Lagrange method to be applicable; namely, in Theorem 2 below, it is proved that if the Lagrange method is justified for all differentiable maximands (or even all linear maximands), then the constraint function satisfies the Constraint Qualification provided the constraint set is convex.

The direct verification of the Constraint Qualification in specific cases is difficult, and it is useful to find simpler hypotheses which imply it. Several apparently new conditions implying the Constraint Qualification and, therefore, the validity of the Saddle-Point Criterion and the Quasi-Saddle-Point Condition are proved in a later section (A Sufficient Condition for the Constraint Qualification). Note that for differentiable functions $f(x)$ and $g(x)$, the Quasi-Saddle-Point Condition implies the Saddle-Point Criterion. For if $f(x)$ and $g(x)$ are concave, the Lagrangian function, $f(x) + y\, g(x)$, is concave in x for any given $y \geqq 0$; if $\bar{x}$ maximizes

---

[1](Continued)

If this condition holds, then the partial derivatives of $f(x)$ all exist at $\bar{x}$ and $\bar{f}_x = a$, but the condition of differentiability at a point is stronger than the existence of partial derivatives.

The function f is said to be concave, if $f[tx + (1 - t)\bar{x}] \geqq tf(x) + (1 - t) f(\bar{x})$ for any pair of vectors, x, $\bar{x}$, and any real number t, $0 \leqq t \leqq 1$.

A vector function f is said to be differentiable (or concave), if each component is.

A saddle-point, subject to $y \geqq 0$ of a function $L(x, y)$ of two vectors, is defined by the properties that $\bar{x}$ maximizes $L(x, \bar{y})$ and $\bar{y}$ minimizes $L(\bar{x}, y)$ for nonnegative y. The concept is due to Kuhn and Tucker [11]; see their definition of the Saddle Value Problem on p. 482 of Ref. [11]. A quasi-saddle-point is a point where the first-order (necessary) conditions for a saddle-point are satisfied.

In this article, the variables are not necessarily restricted to be nonnegative. Any such restrictions are therefore assumed to be included among the conditions $g(x) \geqq 0$. The formulation of the following conditions therefore differs in detail but not in essence from that of Kuhn and Tucker [11].

[2]See, for example, Courant, [7], pp. 189-190 and 192-93. For the case of inequalities, the following example is due to Slater [12]: $f(x) = x$, $g(x) = -(1 - x)^2$; here both f and g are concave and differentiable, yet there is no saddle-point and hence no quasi-saddle-point. See also the example of Kuhn and Tucker, Ref. [11], pp. 483-84.

[3]See Ref. [11], p. 483. For a corresponding condition in the context of equality constraints, see Bliss, Ref. [6], p. 210, conclusion of Lemma 76.1.

$f(x)$ subject to $g(x) \geqq 0$, the Quasi-Saddle-Point Condition implies that the Lagrangian function has a zero derivative at $\overline{x}$ and, therefore, as a concave function, must have a maximum there. Since the Lagrangian is linear in y and $g(\overline{x}) \geqq 0$, it is obvious that $\overline{y}\, g(\overline{x}) = 0$ implies that $\overline{y}$ minimizes $f(\overline{x}) + y\, g(\overline{x})$ subject to $y \geqq 0$. The converse part of the Saddle-Point Criterion, that if $(\overline{x}, \overline{y})$ is a saddle-point of the Lagrangian for some $\overline{y}$, subject to $y \geqq 0$, then $\overline{x}$ is a constrained maximum of $f(x)$ subject to $g(x) \geqq 0$, holds without any assumptions on $g(x)$. (See Kuhn and Tucker, [11], Theorem 2.)

Still another constraint qualification has been given by Hurwicz ([5], Chapter 4, Section V.3.3.2). In another section (Equivalence of Constraint Qualifications KT and H) it is shown to be equivalent to Constraint Qualification KT, at least for finite-dimensional spaces.

To state these results more precisely and to relate them to other work, we will introduce some notation and definitions. In the first place, we define the constraint set,

$$C = \{x\colon\ g(x) \geqq 0\}\,. \tag{1}$$

In the second place, we will denote $\overline{\overline{x}} - \overline{x}$ by ξ. Finally, to simplify the statement of the conclusion of Constraint Qualification KT which will appear frequently in the following discussion, we will find it convenient to introduce the following definitions:

Definition 1. A contained path (with origin $\overline{x}$ and direction ξ) is an n-vector-valued function $\psi(\theta)$ of a real variable which satisfies:

(2) $\psi(\theta)$ is defined for all $0 \leqq \theta \leqq \overline{\theta}$ for some $\overline{\theta} > 0$;

(3) $\psi(0) = \overline{x}$, $\psi(\theta) \in C$ for all $0 \leqq \theta \leqq \overline{\theta}$;

(4) $\psi(\theta)$ has a right-hand derivative at $\theta = 0$ such that $\psi'(0) = \xi$.

Definition 2. An n-vector ξ such that there is a contained path with origin $\overline{x}$ and direction ξ will be referred to as an attainable direction at $\overline{x}$. The set of attainable directions (at any given $\overline{x}$) will be denoted by $\underline{\underline{A}}$.

The set of indices $\{1, \ldots, m\}$ is divided into two parts, E and F. E is the set of all indices effective at $\overline{x}$, namely,

$$E = \{j\colon\ g^j(\overline{x}) = 0\}, \tag{5}$$

and F is the set of all indices ineffective at $\overline{x}$, namely,

$$F = \{j\colon\ g^j(\overline{x}) > 0\}. \tag{6}$$

Definition 3. An n-vector ξ is termed a locally constrained direction if $\overline{g}^E_x\, \xi \geqq 0$ (i.e., $\overline{g}^j_x\, \xi \geqq 0$ for all $j \in E$). The set of locally constrained directions will be denoted by $\underline{\underline{L}}$.

With these definitions the Kuhn-Tucker Constraint Qualification can be written.

Constraint Qualification KT

Every locally constrained direction is attainable, i.e., $L \subset A$.

(Since the definition of a contained path does not require it to be differentiable throughout, this formulation is apparently weaker than Kuhn and Tucker's [11]. We do not know if the weakening is more than apparent when $g(x)$ is differentiable.)

We now observe that the set L of locally constrained directions is a closed convex cone. (By a cone is meant a set which, if it contains any point x, also contains λx for every scalar $\lambda \geqq 0$.) The set A of attainable directions is a cone but is not necessarily convex.

Definition 4. Let W be the closure of the convex cone spanned by A, the set of attainable directions (i.e., the smallest closed convex cone containing A). The elements of W will be termed weakly attainable directions.

A weakly attainable direction is, then, the limit of a sequence of nonnegative linear combinations of attainable directions. We now introduce a weaker constraint qualification:

Constraint Qualification W

Every locally constrained direction is weakly attainable, i.e., $L \subset W$.[4]

It will be shown (Theorem 1) that Constraint Qualification W is sufficient for the validity of the Quasi-Saddle-Point Condition and therefore for that of the Saddle-Point Criterion.

---

[4]To see that A is not necessarily convex and Constraint Qualification W is truly weaker than Constraint Qualification K T, consider the constraints $x_1 \geqq 0$, $x_2 \geqq 0$, $-x_1x_2 \geqq 0$. The constraint set C consists of the origin and the two positive half-axes. If $\overline{x}$ is taken to be the origin, the set of attainable directions A is the same as C. The set of weakly attainable directions, W, is the convex cone spanned by this set; that is, the nonnegative quadrant. All constraints are effective, and

$$\overline{g}_x^E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

so that the set of locally constrained directions L is defined by $\xi_1 \geqq 0$, $\xi_2 \geqq 0$, and thus is again the nonnegative orthant. L is therefore contained in W but not in A.

In general, if A were a closed set, then the convex cone spanned by it would also be closed, and the words, "the closure of," in Definition 4 could be deleted without loss of generality. It can be shown fairly easily that it is closed for convex constraint sets. The referee has supplied the following example, which shows that A is not necessarily closed for differentiable $g(x)$ in general:

$$g(x, 0) = -x^2,$$

$$g(x, y) = -r^2 \{\theta^4 \sin^2(1/\theta) + [\max(r - \sec\theta \tan\theta, 0)]^4\} \text{ for } y > 0,$$

where

$$r = (x^2 + y^2)^{1/2}, \quad 0 \leqq \theta = \arctan(y/x) \leqq \pi/2,$$

and the domain of definition is the nonnegative quadrant.

Since $g(x, y) \leqq 0$ everywhere in the domain of definition, the constraint set

$$C = \{(x, y): g(x, y) \geqq 0\} \text{ is the union of the segments,}$$

$$\{(x, y): y = x \arctan(1/n\pi), \ x \geqq 0, \ x^2 \leqq y\},$$

for all positive integers n. Thus, the attainable directions from $(0, 0)$ include $(1, \arctan(1/n\pi))$ for all n, but not $(1, 0)$, so that A is not closed.

We now turn to some other conditions in the literature which have been found to be sufficient for the Quasi-Saddle-Point Condition. In the usual treatment of the Lagrange multiplier method in the case of equality constraints (Courant [7], p. 198), it is required that the matrix $\overline{g}_x$ of the (partial) derivatives of the constraint functions with respect to the variables have a rank equal to the number of constraints. This condition has been extended to the case of inequalities in [3], p. 8. We write, in the present notation, the

Nondegeneracy Condition

The rank of $\overline{g}_x^E$ equals the number of effective constraints.

(The condition given in [3] is actually slightly weaker.) It was shown in [3], Appendix I, that the Nondegeneracy Condition implies Constraint Qualification KT. In a later section (A Sufficient Condition for the Constraint Qualification), we will deduce the Nondegeneracy Condition from a more general sufficient condition for Constraint Qualification W.

In concave programming—that is, where the functions $f(x)$ and $g(x)$ are assumed to be concave—there are theorems which state conditions under which the Saddle-Point Criterion is valid but which do not involve Constraint Qualification KT. The simplest case is that of linear programming where the functions $f(x)$ and $g(x)$ are assumed to be linear. In this case, the Saddle-Point Criterion always holds, as is well known ([10], Chapters XIX and XX). Since the constraint qualifications are always assumptions about the constraint functions only and do not involve $f(x)$, the question arises whether or not it is the linearity of $g(x)$ that is vital. The answer is in the affirmative, in the sense that the linearity of $g(x)$ is sufficient for the Quasi-Saddle-Point Condition; see Corollary 2 of Theorem 3.

Another constraint qualification for concave programming has been proposed by Slater [12]:

Constraint Qualification S

The function $g(x)$ is concave and, for some x^*, $g(x^*) > 0$.

Slater showed that if the function $f(x)$ is concave, his constraint qualification implied the Saddle-Point Criterion. A simplified proof was given by Uzawa ([5], Chapter 3). Slater's theorem was extended to more general spaces by Hurwicz ([5], Chapter 4, Theorem V.3.1). (Slater's assumption of continuity is dispensable.)

Karlin ([9], Chapter 7, Theorem 7.1.1) suggested still another constraint qualification:

Constraint Qualification K

The function $g(x)$ is concave and, for every $y \geq 0$, there is an x such that $y\,g(x) > 0$.

This condition is clearly implied by Constraint Qualification S; Hurwicz and Uzawa ([5], Chapter 5) have shown that in spaces of considerable generality the two conditions are in fact equivalent.

It is natural to investigate the relation between Constraint Qualification S (or K) and the more general Constraint Qualifications, such as KT or W. Obviously, conditions S or K do not require differentiability of $g(x)$, so that they cannot be completely subsumed under conditions KT or W, which do. We may, however, ask whether or not the former do imply the latter conditions under the additional assumption that $g(x)$ is differentiable. In the section devoted to A Sufficient Condition for the Constraint Qualification, it is shown that such is indeed the case; in fact, a single condition (Theorem 3) is given under which all previous conditions can be subsumed as special cases, as well as some conditions not previously given in the literature.

A still weaker version of this condition is presented in Theorem 4. In one of its corollaries, the hypothesis refers to functions which are simultaneously concave and quasi-convex. A characterization of such functions and also of functions which are simultaneously quasi-concave and quasi-convex is presented in the section, A Characterization of Functions Simultaneously Concave and Quasi-Convex; this result may be of interest in other contexts.

In [5], Chapter 4, Section V.3.3.2, Hurwicz introduces the following constraint qualification:

Constraint Qualification H

For all $\overline{x} \epsilon C$ and all ξ such that $\overline{g}_x \xi + g(\overline{x}) \geqq 0$, ξ is attainable.

Since Constraint Qualification H does not identify separate coordinates, it is meaningful in all linear topological spaces in which a differentiation operation can be defined.[5] It has been shown ([5], Chapter 4, Theorems V.3.3.2 and V.3.3.3) to be a sufficient condition for both the Saddle-Point Criterion and the Quasi-Saddle-Point Condition in spaces of considerable generality. We will show in the section, Equivalence of Constraint Qualifications KT and H, that in finite-dimensional spaces, Constraint Qualifications KT and H are equivalent.

PRELIMINARY LEMMAS AND REMARKS

LEMMA 1: Every weakly attainable direction is locally constrained.

PROOF: Let ξ be an attainable direction, $\psi(\theta)$ a contained path with origin $\overline{x}$ and direction ξ. Then for some $\overline{\theta} > 0$ and every $j \epsilon E$,

$$g^j[\psi(0)] = 0 \text{ , and } g^j[\psi(\theta)] \geqq 0 \qquad (0 \leqq \theta \leqq \overline{\theta});$$

hence, for $\theta = 0$, $dg^j[\psi(\theta)]/d\theta = \overline{g}_x^j \psi'(0) = \overline{g}_x^j \xi \geqq 0$ for every $j \epsilon E$, so that ξ is locally constrained.[6] That is, A is included in L. Since L is a convex cone, the convex cone spanned by A must also be included in L; and since L is closed, W, the closure of the convex cone spanned by A, must also be included in L. (Q.E.D.)

Let us define, for $\overline{x} \epsilon C$,

(7) $$K = \text{closure of the set } \{\lambda(x - \overline{x}): \lambda \geqq 0 \text{ , } x \epsilon C\}.$$

K is the union of all half-lines from $\overline{x}$ through elements of C, together with the boundary of the union. K is clearly a closed cone. That the set in braces is not necessarily closed is illustrated by the case $g(x, y) = y - x^2$, $\overline{x} = (0, 0)$, where the x-axis belongs to K but not to the set in braces.

LEMMA 2: If the constraint set C is convex, then K is a closed convex cone, and $K \subset W$.

PROOF: The convexity of K follows immediately from that of C.

[5]It is possible, although we have not investigated this point, that Constraint Qualification KT can be extended in a natural form to infinite-dimensional (function) spaces.

[6]Note that from the differentiability of $g^j(x)$ and from that of $\psi(\theta)$ at $\theta = 0$, the chain rule for differentiation is valid. Theorem 6.14, p. 113 in [1] may be easily extended to the present case of one-sided differentiation.

If $x \epsilon C$, then by the convexity of the set C,

$$\overline{x} + \theta(x - \overline{x}) \epsilon C \quad \text{for all} \quad 0 \leqq \theta \leqq 1 .$$

Hence, $x - \overline{x}$ is attainable and therefore weakly attainable. Since W is a cone, $\lambda(x - \overline{x}) \epsilon W$ for all $\lambda \geqq 0$. (Q.E.D.)

Let B be any set of vectors. The negative polar cone, to be denoted by B', is defined by

$$B' = \{u: \ ux \leqq 0 \ \text{for all} \ x \epsilon B\} .$$

We have (Fenchel [8], pp. 8-10),

(8) B' is a closed convex cone;

(9) $B_1 \subset B_2$ implies that $B_1' \supset B_2'$;

(10) if B is a closed convex cone, $B'' = B$.

LAGRANGE REGULARITY AND THE CONSTRAINT QUALIFICATION

Definition 5. An m-vector-valued function $g(x)$ will be termed Lagrange regular if, for any differentiable function $f(x)$, the Quasi-Saddle-Point Condition holds.

LEMMA 3: If $\overline{x}$ maximizes $f(x)$ subject to $x \epsilon C$, then

$$\overline{f}_x \epsilon W'$$

where W' is the negative polar cone of W.

PROOF: Let $\psi(\theta)$ be a contained path with origin $\overline{x}$ and direction ξ ; then

$$f[\psi(\theta)] \leqq f[\psi(0)] = f(\overline{x}) \quad \text{for all} \quad 0 \leqq \theta \leqq \overline{\theta} .$$

Then

$$\overline{f}_x \xi = \overline{f}_x \psi'(0) \leqq 0 ,$$

for any ξ in A and, by continuity and convexity, for any $\xi \epsilon W$ (Definitions 2 and 4). (Q.E.D.)

THEOREM 1: If $g(x)$ satisfies Constraint Qualification W, then $g(x)$ is Lagrange regular.

PROOF: Let $f(x)$ be a differentiable function and $\overline{x}$ maximize $f(x)$ subject to $x \epsilon C$. Then, by Lemma 3, $\overline{f}_x \epsilon W'$. On the other hand, Constraint Qualification W states that $L \subset W$ and therefore implies, from (9), that

$$W' \subset L' .$$

Hence, we have

(11) $$\overline{f}_x \epsilon L' .$$

If B is the closed convex cone consisting of all vectors $y^E(-\bar{g}^E_x)$ with $y^E \geqq 0$, then Definition 3 and (11) show that $\bar{f}_x \in B''$, and by (10),

$$-\bar{f}_x = \bar{y}^E \bar{g}^E_x \quad \text{for some} \quad \bar{y}^E \geqq 0 .$$

Define

$$\bar{y} = (\bar{y}^E, \bar{y}^F) \quad \text{with} \quad \bar{y}^F = 0 .$$

Then $\bar{f}_x + \bar{y}\,\bar{g}_x = 0$, $\bar{y}\,g(\bar{x}) = \bar{y}^E g^E(\bar{x}) = 0$, from (5), so that the Quasi-Saddle-Point Condition is satisfied. (Q.E.D.)

Theorem 1 is the basic necessity theorem for nonlinear programming ([11], Theorem 1) extended to the weaker Constraint Qualification W of this paper.

THEOREM 2: If $g(x)$ is Lagrange regular and if the constraint set C defined by it is a convex set, then $g(x)$ satisfies the Constraint Qualification W.

PROOF: It will be shown first that

$$K' \subset L' . \tag{12}$$

Let $a \in K'$; then from (7), for $\lambda = 1$,

$$a(x - \bar{x}) \leqq 0 \quad \text{for all} \quad x \in C . \tag{13}$$

Then $\bar{x}$ maximizes the function $f(x) = ax$ subject to $x \in C$. By the Lagrange regularity of $g(x)$, there is an m-vector $\bar{y}$ such that

$$a + \bar{y}\,\bar{g}_x = 0 ; \quad \bar{y} \geqq 0 , \tag{14}$$

and

$$\bar{y}\,g(\bar{x}) = 0 . \tag{15}$$

The conditions (14) and (15) imply that $\bar{y}^F = 0$ and, thus, that

$$a + \bar{y}^E \bar{g}^E_x = 0 , \quad \bar{y}^E \geqq 0 . \tag{16}$$

The condition (16) implies that

$$a\xi \leqq 0 \quad \text{for all } \xi \quad \text{such that} \quad \bar{g}^E_x \xi \geqq 0 ;$$

i.e.,

$$a \in L' .$$

Hence, we have the relation (12). Then, by Eqs. (9) and (10), $K \supset L$. Applying Lemma 2,

$$W \supset K \supset L . \quad \text{(Q.E.D.)}$$

A SUFFICIENT CONDITION FOR THE CONSTRAINT QUALIFICATION

THEOREM 3: If E' is the set of effective constraints which are convex functions, E'' is the set of all other effective constraints, and if there exists ξ^* such that $\bar{g}_x^{E'} \xi^* \geqq 0$, $\bar{g}_x^{E''} \xi^* > 0$, then Constraint Qualification W holds.

PROOF: Let ξ be any element of L, α any positive real number, and

$$\psi(\theta) = \bar{x} + (\xi + \alpha \xi^*) \theta, \qquad \text{for } \theta \geqq 0 .$$

We will show that $\psi(\theta)$ is a contained path for θ sufficiently small and, hence, $\xi + \alpha \xi^*$ is attainable. If we let α approach zero, it will follow from Definition 4 that ξ belongs to W, in fact to the closure of A, so that we will have shown that $L \subset W$, which is Constraint Qualification W.

For any $j \in E$, at $\theta = 0$,

$$dg^j [\psi(\theta)]/d\theta = \bar{g}_x^j(\xi + \alpha \xi^*) = \bar{g}_x^j \xi + \alpha \bar{g}_x^j \xi^* \geqq \alpha \bar{g}_x^j \xi^* ,$$

since $\bar{g}_x^j \xi \geqq 0$ by definition of L. If $j \in E'$, it follows from the hypothesis of the theorem that

$$dg^j [\psi(\theta)]/d\theta \geqq 0 \quad \text{at} \quad \theta = 0 ,$$

which, for a convex function, implies that $g^j [\psi(\theta)]$ has its minimum for $\theta \geqq 0$ at $\theta = 0$. If $j \in E''$, then

$$dg^j [\psi(\theta)]/d\theta > 0 \quad \text{at} \quad \theta = 0 ,$$

so that $g^j [\psi(\theta)]$ has a local right-hand minimum at $\theta = 0$. It follows that

$$g^E [\psi(\theta)] \geqq g^E [\psi(0)] = g^E(\bar{x}) = 0 \quad \text{for} \quad \theta \quad \text{sufficiently small.}$$

Since, by definition, $g^F[\psi(0)] > 0$, $g^F[\psi(\theta)] \geqq 0$ for θ sufficiently small, so that $\psi(\theta) \in C$ for θ sufficiently small, from which the theorem follows, as has previously been shown.

COROLLARY 1: If $g(x)$ is convex, then $g(x)$ is Lagrange regular.

PROOF: In this case, E'' is the null set, and it suffices to set $\xi^* = 0$. The conclusion follows from Theorem 1.

As a special case, we state

COROLLARY 2: If $g(x)$ is linear, it is Lagrange regular.

COROLLARY 3: If $g(x)$ is concave, E' the set of effective constraints which are linear, E'' the set of all other effective constraints, and there exists x^* such that $g^{E'}(x^*) \geqq 0$, $g^{E''}(x^*) > 0$, then $g(x)$ is Lagrange regular.

PROOF: For any concave function,

$$\bar{g}^j_x(x^* - \bar{x}) \geqq g^j(x^*) - g^j(\bar{x}) = g^j(x^*) \quad \text{for} \quad j \in E .$$

Since the only functions which are both concave and convex are linear, the results follow from Theorem 3.

The following special case is precisely Constraint Qualification S.

COROLLARY 4: If $g(x)$ is concave and $g(x^*) > 0$ for some x^*, then $g(x)$ is Lagrange regular.

Corollary 4 was originally stated as Theorem 3 in Ref. [4]. The following corollary generalizes Theorem 2 in Ref. [2] which, in turn, generalized Corollary 4.

COROLLARY 5: If the constraint set C is convex and possesses an interior, and $\bar{g}^j_x \neq 0$ for each $j \in E$, then $g(x)$ is Lagrange regular.

PROOF: By Lemma 2, $x - \bar{x}$ is weakly attainable for all $x \in C$ and therefore belongs to L by Lemma 1. Since C possesses an interior, L must possess one also and therefore has the full dimensionality of the entire space. If, for some $j \in E$, $\bar{g}^j_x \xi = 0$ for all ξ in L, it follows that $\bar{g}^j_x \xi = 0$ for all ξ, which means that $\bar{g}^j_x = 0$, contrary to hypothesis. Hence, for each $j \in E$, there exists $\xi^j \in L$ such that $\bar{g}^j_x \xi^j \neq 0$; since $\bar{g}^j_x \xi \geqq 0$ for all ξ in L by definition, we must have

$$\bar{g}^j_x \xi^j > 0 , \quad \bar{g}^j_x \xi^k \geqq 0 \quad \text{for} \quad j, k \in E .$$

If we let

$$\xi^* = \sum_{j \in E} \xi^j ,$$

we see that

$$\bar{g}^E_x \xi^* > 0 ,$$

and the conclusion follows trivially from Theorem 3.

REMARK: To establish that C has an interior, it is sufficient that $g(x^*) > 0$ for some x^*; to establish that C is convex, it is sufficient that $g(x)$ be quasi-concave.[7]

We can also relate the Nondegeneracy Condition to this analysis.

COROLLARY 6: If the rank of $\bar{g}^E_x$ equals the number of effective constraints, then $g(x)$ is Lagrange regular.

PROOF: Let u be an arbitrary positive column vector; then from the hypothesis, we can find ξ^* such that $\bar{g}^E_x \xi^* = u > 0$.

[7] A function $f(x)$ is quasi-concave if, for all c, the set $\{x: f(x) \geqq c\}$ is convex. A function is quasi-convex if it is the negative of a quasi-concave function, so that the sets $\{x: f(x) \leqq c\}$ are all convex.

If we reconsider the proof of Theorem 3, we see that the convexity assumption for the elements of E' is only needed to insure that $g^j[\psi(\theta)]$ has a minimum at $\theta = 0$ for $j \epsilon E'$. For the purposes of the proof, a local minimum is sufficient. It would therefore suffice to define E' as the set of effective constraints which are locally convex (i.e., which are convex over some neighborhood of $\bar{x}$), which, for example, would be implied by well-known conditions on the matrix of second partial derivatives of the $g^j(x)$ evaluated at $x = \bar{x}$. It is clear that the larger E' is, the weaker the hypothesis of the theorem.

A somewhat different weakening of Theorem 3 is suggested by observing that if δ^j is the j^{th} unit row vector, then

$$(17) \qquad -\bar{g}_x^j + \delta^j \bar{g}_x^E = 0 \quad \text{for any} \quad j \epsilon E .$$

Since $\delta^j \bar{g}_x^E(0) = 0$, we see that the Lagrangian conditions for a constrained maximum of $-g^j(\bar{x} + \xi)$ subject to $\bar{g}_x^E \xi \geqq 0$ are satisfied at $\xi = 0$. Suppose for the moment that the Lagrangian conditions (i.e., those in the Quasi-Saddle-Point Condition) were sufficient to insure a constrained maximum. Then $g^j(\bar{x} + \xi)$ would have a minimum at $\xi = 0$ for $\xi \epsilon L$. Since $(\xi + \alpha . \xi^*)\theta \epsilon L$ for all $\theta \geqq 0$, this implies that $g^j[\psi(\theta)] \geqq g^j[\psi(0)] = 0$, and so the argument of Theorem 3 is still valid if E' is defined to contain all effective constraints for which the Lagrangian conditions are sufficient to insure a constrained maximum for $-g^j(x)$.

A set of hypotheses, under which the Lagrangian conditions are sufficient for a constrained maximum, is presented in [2], Theorem 1. We state the relevant results as a lemma (the statement below is slightly more general in that the domain of definition is not restricted to the nonnegative orthant; the previous proof extends easily).

LEMMA 4: Let $f(x)$ be a differentiable quasi-concave function and $g(x)$ an m-vector valued differentiable quasi-concave function defined over some convex domain D. Let $\bar{x}$ and $\bar{y} \geqq 0$ satisfy the Lagrangian conditions, $\bar{f}_x + \bar{y}\,\bar{g}_x = 0$, $\bar{y}\, g(\bar{x}) = 0$, and let one of the following conditions be satisfied:

(a) $\bar{f}_x x^1 > \bar{f}_x x^2$ for some $x^1 \epsilon C$, $x^2 \epsilon D$;

(b) $\bar{f}_x \neq 0$ and $f(x)$ is twice differentiable in a neighborhood of $\bar{x}$;

(c) $f(x)$ is concave.

Then $\bar{x}$ maximizes $f(x)$ subject to the constraints $g(x) \geqq 0$.

Lemma 4 can be applied to the preceding argument, with x replaced by ξ, $f(x)$ by $-g^j(\bar{x} + \xi)$, and $g(x)$ by $\bar{g}_x^E \xi$. Since the latter is linear and therefore necessarily quasi-concave, the hypotheses of the lemma are equivalent to requiring that one of (a), (b) or (c) hold for the function $-g^j(\bar{x} + \xi)$. The application of (b) and (c) is straightforward. Condition (a) becomes

$$(18) \qquad \bar{g}_x^j \xi^1 < \bar{g}_x^j \xi^2 ,$$

for some ξ^1 in the constraint set $\bar{g}_x^E \xi \geqq 0$ and some ξ^2 for which $\bar{x} + \xi^2$ is in the domain of definition of $g^j(x)$. Since $\bar{g}_x^j \xi \geqq 0$ for all $\xi \epsilon L$, while $0 \epsilon L$, (18) is equivalent to

$$\bar{g}_x^j (x - \bar{x}) > 0 ,$$

for some x in the domain of definition of $g^j(x)$. We can thus state the following generalization of Theorem 3:

THEOREM 4: Let E' be the set of effective constraints $g^j(x)$ which are quasi-convex functions and which satisfy one of the following three conditions:

(a) $\bar{g}^j_x x > \bar{g}^j_x \bar{x}$ for some x for which $g^j(x)$ is defined;

(b) $\bar{g}^j_x \neq 0$ and $g^j(x)$ is twice differentiable in a neighborhood of $x = \bar{x}$;

(c) $g^j(x)$ is convex.

If E'' is the set of all other effective constraints and if there exists ξ^* such that $\bar{g}^{E'}_x \xi^* \geqq 0$, $\bar{g}^{E''}_x \xi^* > 0$, then Constraint Qualification W holds.

From this theorem can be deduced generalizations of some of the corollaries to Theorem 3. Corresponding to Corollary 1 of the latter, we have

COROLLARY 1. If $g(x)$ is quasi-convex and each effective component satisfies one of the conditions (a), (b), or (c) of Theorem 4, then $g(x)$ is Lagrange regular.

Corollary 3 of Theorem 3 can be generalized to

COROLLARY 2: Suppose $g(x)$ is concave. Let E' be the set of effective constraints which are also quasi-convex and which satisfy one of the following conditions:

(a) $g^j(x) > 0$ for some x ;

(b) $\bar{g}^j_x \neq 0$ and $g^j(x)$ is twice differentiable in a neighborhood of $x = \bar{x}$;

(c) $g^j(x)$ is linear.

If E'' is the set of all other effective constraints and if there exists ξ^* such that $g^{E'}(x^*) \geqq 0$, $g^{E''}(x^*) > 0$, then $g(x)$ is Lagrange regular.

PROOF: If (a) holds, then, since $g^j(x)$ is concave,

$$\bar{g}^j_x(x - \bar{x}) \geqq g^j(x) - g^j(\bar{x}) = g^j(x) > 0 ,$$

so that (a) in Theorem 4 holds. Since (b) is the same in the two statements and (c) here implies (c) in the statement of Theorem 4, E' as defined here satisfies the conditions of Theorem 4. The proof is then the same as that of Corollary 3, Theorem 3.[8]

A CHARACTERIZATION OF FUNCTIONS SIMULTANEOUSLY CONCAVE AND QUASI-CONVEX

To better understand the domain of applicability of Corollary 2 of Theorem 4, it is useful to give a characterization in Lemma 6 of functions which are simultaneously concave and quasi-convex. First, we characterize functions which are simultaneously quasi-concave

[8] It should be stated that the case of nonlinear equality constraints is not well handled by these theorems and corollaries; these all depend on the construction of a linear contained path, but for nonlinear equality constraints none may exist. A form of Corollary 6 to Theorem 3 does remain valid in this case ([3], Appendix 1).

and quasi-convex. In what follows, an indifference set is a set $\{x: f(x) = c\}$; a maximal (minimal) indifference set is the set on which $f(x)$ attains its maximum (minimum). A set S will be said to be bounded by two noncrossing hyperplanes in D if there exist linear functions, $L_1(x)$, $L_2(x)$, not identically constant in D, such that

(19) $$S = \{x \in D: L_1(x) \geqq 0 , L_2(x) \leqq 0\} ,$$

and

(20) $$L_1(x) < 0 , L_2(x) > 0 \text{ for no } x \in D .$$

A diagram will show that (19) and (20) are algebraic transcriptions of the geometric concept they define.

LEMMA 5: A function is both quasi-concave and quasi-convex over a convex domain D if and only if every indifference set not minimal or maximal is bounded by two noncrossing hyperplanes in D.

PROOF: Without loss of generality, assume that D is the domain of definition of $f(x)$. We first note that quasi-convexity implies that $\{x: f(x) < c\}$ is convex for all c. Suppose $f(x^o) < c$, $f(x^1) < c$, x^2 a convex combination of x^o and x^1. If $c' = \max[f(x^o), f(x^1)]$, we have x^o, x^1 belonging to the set $\{x: f(x) \leqq c'\}$, which is convex by definition of quasi-convexity, and hence $f(x^2) \leqq c' < c$.

We suppose, without loss of generality, that the linear space is the smallest containing D. Consider any value, c, of $f(x)$, which is neither the maximum nor the minimum. If the set $\{x: f(x) \geqq c\}$ did not have the full dimensionality of the space, there would exist a linear function $L(x)$ not identically zero, such that $L(x) = 0$ whenever $f(x) \geqq c$. Choose y so that $L(y) \neq 0$, and let $z = y - x$. Then, for any x, $L(x + tz)$ is a linear function of t which is nonzero for $t = 1$ and therefore takes on the value 0 for at most one value of t; hence we can find ϵ arbitrarily small for which $L(x + \epsilon z) \neq 0$, $L(x - \epsilon z) \neq 0$. Suppose x is an interior point of D. Choose ϵ sufficiently small so that $x \pm \epsilon z \in D$. Since $L(x \pm \epsilon z) \neq 0$, $f(x \pm \epsilon z) < c$; by the convexity of $\{x: f(x) < c\}$, $f(x) < c$. Since D is convex, it follows by continuity that $f(x) \leqq c$ for all $x \in D$, so that c would be the maximum value of $f(x)$, contrary to assumption.

The set $\{x: f(x) \geqq c\}$ is convex and has the full dimensionality of the space; the set $\{x: f(x) < c\}$ has been shown to be convex. Hence, there is a separating hyperplane, i.e., a linear function, not constant over the entire space and hence not over D, $L_1(x)$ such that $L_1(x) \leqq 0$ for $f(x) < c$, $L_1(x) \geqq 0$ if $f(x) \geqq c$ ([8], Theorem 28, p. 48). If $L_1(x) = 0$ whenever $f(x) \geqq c$, the set $\{x: f(x) \geqq c\}$ would lie in a hyperplane and therefore not have the full dimensionality of the space, a contradiction. Hence, $L_1(x^o) > 0$ for some x^o for which $f(x^o) \geqq c$. Suppose $L_1(x) = 0$ for some x for which $f(x) < c$. Then for y a convex combination of x^o and x sufficiently close to x, we have $L_1(y) > 0$, $f(y) < c$, contrary to the separation result. Thus,

$$L_1(x) < 0 \text{ if } f(x) < c , \quad L_1(x) \geqq 0 \text{ if } f(x) \geqq c .$$

Since the sets $\{x: f(x) < c\}$ and $\{x: f(x) \geqq c\}$ together exhaust D, we have

$$(21) \qquad \{x\colon\ f(x) < c\} = \{x \in D\colon\ L_1(x) < 0\},$$

$$(22) \qquad \{x\colon\ f(x) \geqq c\} = \{x \in D\colon\ L_1(x) \geqq 0\}.$$

Similarly, from the quasi-concavity of $f(x)$, we find there exists a linear function, $L_2(x)$, such that

$$(23) \qquad \{x\colon\ f(x) \leqq c\} = \{x \in D\colon\ L_2(x) \leqq 0\},$$

$$(24) \qquad \{x\colon\ f(x) > c\} = \{x \in D\colon\ L_2(x) > 0\}.$$

From (22) and (23), (19) holds for the set $S = \{x\colon\ f(x) = c\}$. Since the set (24) is included in the set (22), (20) holds.

We now prove the converse theorem. First, consider any c which lies strictly between the maximum and minimum values of $f(x)$. By assumption, the set $S = \{x\colon\ f(x) = c\}$ satisfies (19) and (20) for some linear functions, $L_1(x)$, $L_2(x)$. Let $\underline{S} = \{x \in D\colon\ L_1(x) < 0\}$, $\overline{S} = \{x \in D\colon\ L_2(x) > 0\}$. We will first show that the set $\{x\colon\ f(x) \leqq c\}$ is convex. Since $c < \max_x f(x)$, $f(x^o) - c > 0$ for some x^o.

By (19) we must have $f(x) - c \neq 0$ for all $x \in \underline{S}$, and all $x \in \overline{S}$. Further, $f(x) - c$ cannot assume both signs in $\underline{S}$, for, by the convexity of the set and the continuity of $f(x)$, we would have $f(x) - c = 0$ in that set for some x, which has just been shown impossible. Similarly, $f(x) - c$ must have a single sign in $\overline{S}$.

Suppose $f(x^o) - c > 0$ for some x^o in $\underline{S}$. We will show that it is impossible that $f(x^1) - c > 0$ for some x^1 in $\overline{S}$. For then, $f(x) - c > 0$ for all $x \in \underline{S}$ and all $x \in \overline{S}$, while $f(x) = c$ for all $x \in S$, and therefore $f(x) \geqq c$ for all $x \in D$, contrary to assumption.

Hence, if $f(x^o) > c$ for some x^o in $\underline{S}$, $f(x) > c$ for all $x \in \underline{S}$, and only for such x. The set $\{x\colon\ f(x) \leqq c\}$ is then precisely the set $\{x \in D\colon\ L_1(x) \geqq 0\}$, which is certainly convex.

Alternatively, we might have $f(x^o) - c > 0$ for some $x^o \in \overline{S}$. The argument is completely parallel.

We have shown that $\{x\colon\ f(x) \leqq c\}$ is convex for any c, neither maximal nor minimal. The proof that $\{x\colon\ f(x) \geqq c\}$ is convex for such c is completely parallel. There remain the cases where $c = \max_x f(x)$ or $\min_x f(x)$. In the first case, the set $\{x\colon\ f(x) \leqq c\}$ is the entire set D and is certainly convex. The set $\{x\colon\ f(x) \geqq c\}$ is the intersection of all the sets $\{x\colon\ f(x) \geqq c'\}$ for $c' < c$; it is an intersection of convex sets and therefore convex. The case where $c = \min_x f(x)$ is handled by the same argument.

LEMMA 6: A concave function is also quasi-convex over a convex domain D if and only if every indifference set not maximal or minimal is the intersection of D with a hyperplane.

PROOF: Let S, as before, be an indifference set, $\{x\colon\ f(x) = c\}$. We consider five cases:

(a) For some $x^o, x^1 \in \underline{S}$, $L_1(x^o) > 0$, $L_2(x^1) < 0$. From (19), $L_2(x^o) \leqq 0$, $L_1(x^1) \geqq 0$. In this case, let $x^2 = (x^o + x^1)/2$, which belongs to S by convexity,

$$L_1(x^2) > 0 , \quad L_2(x^2) < 0 .$$

For any $x \in D$, let $x(\theta) = (1 - \theta) x + \theta x^2$, $g(\theta) = f[x(\theta)]$. Clearly, $x(\theta) \in S$ for θ in the neighborhood of 1, so that $g(\theta) = c$ for all θ in the neighborhood of 1, and $g'(1) = 0$. Since $g(\theta)$ is concave, it has its maximum at $\theta = 1$, and, in particular, $f(x) = g(0) \leqq g(1) = c$, so that $c = \max_x f(x)$.

(b) $L_1(x) = 0$ for all $x \in S$, $L_2(x) \leqq 0$ for all $x \in D$ for which $L_1(x) = 0$. In this case, $x \in S$, if and only if $x \in D$, $L_1(x) = 0$. (Q.E.D.)

(c) $L_2(x) = 0$ for all $x \in S$, $L_1(x) \geqq 0$ for all $x \in D$ for which $L_2(x) = 0$. In this case, $S = \{x \in D\colon\ L_2(x) = 0\}$.

(d) $L_1(x) = 0$ for all $x \in S$, $L_2(x^o) > 0$, $L_1(x^o) = 0$, for some $x^o \in D$. First suppose $L_1(x) < 0$ for some $x \in D$. Then we could find a convex combination of x^o and x for which L_1 is negative, L_2 positive, contrary to (20).

(25) $$L_1(x) \geqq 0 \quad \text{for all} \quad x \in D .$$

Since $L_1(x)$ is not identically 0 in D, we can find x^1 so that $L_1(x^1) > 0$. If $L_2(x) < 0$ for some $x \in S$, we can find a convex combination of $\underline{x}$, x^1 for which L_1 is positive, L_2 negative, contrary to the assumption that L_1 is zero for all $x \in S$.

(26) $$L_2(x) = 0 \quad \text{for all} \quad x \in S.$$

From (25) and (26), the hypotheses of (c) are satisfied.

(e) $L_2(x) = 0$ for all $x \in S$, $L_1(x^o) < 0$, $L_2(x^o) = 0$ for some $x^o \in D$. This case is completely parallel to (d).

Thus, except in case (a), the indifference sets are the intersections of hyperplanes with D.

The converse follows trivially from Lemma 5; by assumption, for every c not maximal or minimal, there exists a linear function $L(x)$ such that

$$\{x\colon\ f(x) = c\} = \{x \in D\colon\ L(x) = 0\} ,$$

so that (19) and (20) are satisfied with $L_1(x) = L_2(x) = L(x)$.

EQUIVALENCE OF CONSTRAINT QUALIFICATIONS KT AND H

THEOREM 5: In finite-dimensional spaces, Constraint Qualifications KT and H are equivalent.

PROOF: Clearly, the hypothesis of Constraint Qualification KT can be inferred from that of condition H by considering only those components j for which $g^j(\overline{x}) = 0$. Since the conclusions of the two Constraint Qualifications are the same, the KT condition implies condition H.

To establish the converse, suppose Constraint Qualification H and the hypothesis of Constraint Qualification KT hold. Then for any $\varepsilon > 0$,

$$\text{if } g^j(\bar{x}) = 0\,,\ g^j_x \cdot (\varepsilon\ \xi) \geqq 0\,. \tag{28}$$

On the other hand, clearly

$$\text{if } g^j(\bar{x}) > 0\,,\ g^j(\bar{x}) + g^j_x \cdot (\varepsilon\ \xi) \geqq 0 \quad \text{for } \varepsilon \text{ sufficiently small.} \tag{29}$$

Since $\bar{x} \in C$, $g^j(\bar{x}) \geqq 0$ for all j. From (28) and (29), $g(\bar{x}) + g_x(\varepsilon\ \xi) \geqq 0$. By Constraint Qualification H, there exists a path $\psi_\varepsilon(\theta)$ such that $\psi_\varepsilon(0) = \bar{x}$, $\psi_\varepsilon(\theta) \in C$ for θ sufficiently small, $\psi'_\varepsilon(0) = \varepsilon\ \xi$. If we now define $\psi(\theta) = \psi_\varepsilon(\theta/\varepsilon)$, we have a contained path at $\bar{x}$ in direction ξ.

BIBLIOGRAPHY

[1] T. M. Apostol, Mathematical Analysis: A Modern Approach to Advanced Calculus, Reading, Mass.: Addison-Wesley, 1957.

[2] K. J. Arrow and A. C. Enthoven, "Quasi-Concave Programming," The RAND Corporation, P-1847, December 16, 1959. (To be published in Econometrica.)

[3] K. J. Arrow and L. Hurwicz, "Reduction of Constrained Maxima to Saddle-Point Problems," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles: University of California Press, 1956, Vol. V, pp. 1-20.

[4] K. J. Arrow, L. Hurwicz, and H. Uzawa, "Constraint Qualification in Maximization Problems," Office of Naval Research Technical Report No. 64, Department of Economics, Stanford University, 1958.

[5] K. J. Arrow, L. Hurwicz, and H. Uzawa, Studies in Linear and Nonlinear Programming, Stanford: Stanford University Press, 1958.

[6] G. A. Bliss, Lectures on the Calculus of Variations, Chicago: University of Chicago Press, 1946.

[7] R. Courant, Differential and Integral Calculus (E. J. McShane, tr.), Vol. II, New York: Interscience Publishers, 1936.

[8] W. Fenchel, Convex Cones, Sets and Functions. Princeton University, 1953 (hectographed).

[9] S. Karlin, Mathematical Methods and Theory in Games, Linear Programming, and Economics, Vol. I, Cambridge, Mass.: Addison-Wesley, 1960.

[10] T. C. Koopmans, Editor, Activity Analysis of Production and Allocation, Cowles Commission Monograph No. 13, New York: John Wiley and Sons, 1951.

[11] H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in J. Neyman, Editor, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Berkeley and Los Angeles: University of California Press, 1951, pp. 481-492.

[12] M. Slater, "Lagrange Multipliers Revisited: A Contribution to Nonlinear Programming," Cowles Commission Discussion Paper, Math. 403, November 1950.

NORMALITY AND ABNORMALITY IN THE CALCULUS OF VARIATIONS*

BY
GILBERT A. BLISS

Within the past few years a number of papers concerning the problem of Bolza in the calculus of variations have been published which make it possible to carry through the theory of this problem with much simplified assumptions concerning what is called the normality of the minimizing arc. I refer especially to papers by Graves [8],† Hestenes [11, 14, 16], Reid [15], and Morse [13]. These papers and others are also important because they bring the theory of problems of the calculus of variations with variable end points to a stage comparable with that already attained for the more special case in which the end points are fixed.

In the theories of Bolza [1, chap. 11, 12] and Bliss [2] for the problem of Lagrange with fixed end points it was assumed that the minimizing arc considered, extended slightly at both ends, was normal on every sub-interval. Morse [4] showed that the theory could be carried through on the assumption that the arc itself, without extensions, was normal on every sub-interval. The most important case, however, turns out to be the one for which the arc as a whole is normal relative to the problem considered, but not necessarily normal on sub-intervals. Graves proved the necessary condition of Weierstrass for such a normal minimizing arc, and Hestenes deduced further necessary conditions and gave sufficiency proofs for a minimum. The importance of these results is emphasized by the fact that for the very general problem of Mayer, which may be regarded as a sub-case of the problem of Bolza, every minimizing arc is abnormal on every sub-interval, even though the arc as a whole is normal relative to the problem. Thus the problem of Mayer needs a separate treatment, such as was given by Bliss and Hestenes [9, 10], unless one has at his command results equivalent to the recent extensions of the theory of the problem of Bolza mentioned above.

In this paper I am attempting 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. In §2 analogous notions are discussed for problems of the cal-

* Presented to the Society, April 20, 1935; received by the editors April 1, 1937.

† The numbers in brackets here and elsewhere refer to the bibliography at the end of this paper.

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_5, © Springer Basel 2014

culus of variations. From this discussion it will be clear that a normal arc for the problem of Bolza is a non-singular arc of the class in which a minimizing one is sought. The singular arcs of the class are the abnormal ones. They have an enormous variety of types. It is not likely that a general theory can be formulated which would apply to all of them, though one might characterize and study successfully some very general cases.

In the papers of Graves and Hestenes mentioned above there is no explicit assumption concerning normality. The arc studied is assumed only to have a set of multipliers like those which it would have if it were normal for the problem of Bolza considered. In the following pages it will be seen that, though such an arc may be abnormal for the problem originally considered, it is nevertheless normal for a second problem of Bolza obtained from the first by suitably extending the class of arcs in which a minimizing one is sought. Furthermore the properties characterizing a minimizing arc for the original problem are effective for the second, so that the sufficiency theorems of Hestenes for arcs which are normal have as easy consequences those for the abnormal arcs permitted by his hypotheses. This makes possible a number of simplifications in the details of the proofs. It is not to be expected, of course, that new necessary conditions on a minimizing arc can be secured by extending the class of arcs in which a minimizing one is sought. The paper of Graves, therefore, seems to contain results not attainable by considering only normal arcs.

In the introduction to his paper [13] Morse makes a statement concerning priority for the proofs of sufficiency theorems without assumptions of normality which might easily be misunderstood and about which I should like to make the following comments. Hestenes had previously proved, in his paper [11], three sufficiency theorems (Theorems 9:1, 9:3, 9:5) without explicit assumptions of normality, and also a fourth theorem (Theorem 9:4) with normality assumptions still undesirably strong, but weaker than those which had before been used. Reid [15] and Morse [13] showed independently that by means of a further lemma, but aided still essentially by the results of Hestenes, this fourth theorem can be brought to a par with the others. The condition VI′ [11, p. 811] of Theorem 9:4 is analogous to one which I used in the paper [5], and which was originally due to A. Mayer. Its statement involves the notion of conjugate points and is therefore more closely related to the classical conditions of Jacobi for simpler problems than the corresponding conditions of the other theorems. I think it should be understood that the priority comment of Morse is applicable to Theorem 9:4 of Hestenes, but not to the other three theorems of his paper, which are equally important. I may add that the theorems of Hestenes were proved with great originality and ingenuity while he was my research assistant at the University of Chicago in 1933 [16, p. 543]. When he went away he left a manuscript with me in which the theorems were, at my suggestion, deduced only for normal arcs, the ones which then, as well as now, seemed to me the most important, even though the justification of the arguments of the present paper was at that time missing. This manuscript has since appeared in much modified form in my mimeographed lectures on the problem of Bolza [12]. In his paper [11] Hestenes showed that his methods are also effective for the problem of Bolza in the form adopted by Morse.

1. **Abnormality for minima of functions of a finite number of variables.** The significance of the notion of abnormality in the calculus of variations can be indicated by a study of the theory of the simpler problem of finding, in the set of points $y=(y_1, \cdots, y_n)$ satisfying a system of equations of the form

$$\phi_\beta(y)=0 \quad (\beta=1, \cdots, m<n),$$

one which minimizes a function $f(y)$. For a point $y^0=(y_1^0, \cdots, y_n^0)$ near which the functions f and ϕ_β have continuous partial derivatives of at least the second order, and which satisfies the equations $\phi_\beta=0$, we have the following theorems, some of which are, of course, well known.

THEOREM 1:1. *A first necessary condition for $f(y^0)$ to be a minimum is that there exist constants l_0, l_β not all zero such that the derivatives F_{y_i} of the function*

$$F=l_0f+l_\beta\phi_\beta$$

all vanish at y^0.

To prove this we have only to note that the determinants of the matrix

$$\left\| \begin{matrix} f_{y_i}(y^0) \\ \phi_{\beta y_i}(y^0) \end{matrix} \right\|$$

must all vanish. Otherwise, according to well known implicit function theorems, the equations $f(y)=f(y^0)+u$, $\phi_\beta(y)=0$ would have solutions y for negative values of u, and $f(y^0)$ could not be a minimum.

A point y^0 has by definition *order of abnormality equal to q* if there exist q linearly independent sets of multipliers of the form $l_0=0$, l_β having the property of the theorem. When $q=0$ the point y^0 is said to be *normal.* A necessary and sufficient condition for abnormality of order q is evidently that the matrix $\|\phi_{\beta y_i}(y^0)\|$ have rank $m-q$. At a normal point y^0 the multipliers l_0, l_β of the theorem can be divided by l_0 and put into the form $l_0=1$, l_β. In this form they are unique, since the non-vanishing difference of two such sets would be a set of multipliers implying abnormality.

LEMMA 1:1. *If a point y^0 is normal, then for every set of constants η_i $(i=1, \cdots, n)$ satisfying the equations*

$$\phi_{\beta y_i}(y^0)\eta_i = 0 \tag{1:1}$$

there exists a set of functions $y_i(b)$ having continuous second derivatives near $b=0$, satisfying the equations $\phi_\beta=0$, and such that

$$y_i(0)=y_i^0,\ y_i'(0)=\eta_i.$$

The proof can be made by considering the equations

$$\phi_\beta(y) = 0,\ \phi_\gamma(y) = \phi_\gamma(y^0) + b\zeta_\gamma \quad (\beta = 1, \cdots, m;\ \gamma = m+1, \cdots, n) \tag{1:2}$$

in which the auxiliary functions $\phi_\gamma(y)$ are selected so that they have continuous second derivatives near y^0 and make the functional determinant $|\phi_{iy_k}(y^0)|$ different from zero, and in which the constants ζ_γ are defined by the equations

$$\phi_{\gamma y_i}(y^0)\eta_i = \zeta_\gamma. \tag{1:3}$$

Equations (1:2) then have solutions $y_i(b)$ with continuous derivatives of at least the second order near $b=0$, and such that $y_i(0)=y_i^0$. By differentiating with respect to b the equations (1:2) with these solutions substituted, we find the equations

$$\phi_{\beta y_i}(y^0)y_i'(0) = 0,$$

$$\phi_{\gamma y_i}(y^0)y_i'(0) = \zeta_\gamma.$$

With equations (1:1) and (1:3) these show that $y_i'(0)=\eta_i$.

THEOREM 1:2. *If y^0 is a normal point and $f(y^0)$ a minimum then the condition*

$$F_{y_iy_k}(y^0)\eta_i\eta_k \geqq 0$$

must hold for every set η_i satisfying the equations (1:1), where $F=f+l_\beta\phi_\beta$ is the function formed with the unique set of multipliers $l_0=1$, l_β belonging to y^0.

The conclusion of the theorem is due to the fact that the function $g(b)=f[y(b)]$, formed with the functions $y_i(b)$ of the lemma, must have a minimum at $b=0$. Since

$$\phi_{\beta y_i}[y(b)]y_i'(b) = 0$$

the derivatives of $g(b)$ are seen to have the values

$$g'(b) = f_{y_i}[y(b)]y_i'(b) = F_{y_i}[y(b)]y'(b),$$
$$g''(0) = F_{y_iy_k}(y^0)\eta_i\eta_k,$$

and for $g(0)$ to be a minimum we must have $g''(0) \geqq 0$.

THEOREM 1:3. *If a point y^0 has a set of multipliers $l_0=1$, l_β for which the function $F=f+l_\beta\phi_\beta$ satisfies the conditions*

$$(1{:}4) \qquad F_{y_i}(y^0) = 0, \qquad F_{y_iy_k}(y^0)\eta_i\eta_k > 0$$

for all sets η_i satisfying the equations

$$(1{:}5) \qquad \phi_{\beta y_i}(y^0)\eta_i = 0,$$

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

This can be proved with the help of Taylor's formula with integral form of remainder. For every point y near y^0 satisfying the equations $\phi_\beta=0$ we have the equations

$$(1{:}6) \qquad \begin{aligned} f(y) - f(y^0) &= f_{y_i}(y^0)\eta_i + \int_0^1 (1-\theta)f_{y_iy_k}(y')\eta_i\eta_k d\theta, \\ 0 &= \phi_{\beta y_i}(y^0)\eta_i + \int_0^1 (1-\theta)\phi_{\beta y_iy_k}(y')\eta_i\eta_k d\theta, \\ 0 &= \int_0^1 \phi_{\beta y_i}(y_i')\eta_i d\theta, \end{aligned}$$

where $y_i' = y_i^0 + \theta(y_i - y_i^0)$, $\eta_i = y_i - y_i^0$. From these we find readily

$$f(y) - f(y^0) = \int_0^1 (1-\theta)F_{y_iy_k}(y')\eta_i\eta_k d\theta.$$

Since the quadratic form in the integrand of the last integral, thought of as a function of independent variables y' and η, is positive for $y'=y^0$ and all sets η satisfying the equations (1:5), it will remain positive for $y'=y^0+\theta(y-y^0)$ and all sets η, including $\eta=y-y^0$, satisfying equations (1:6), provided that y lies in a sufficiently small neighborhood N of the point y^0. Thus we see that for all points y in N satisfying the equations $\phi_\beta=0$ the difference $f(y)-f(y^0)$ is positive.

The last theorem is analogous to the sufficiency theorems of Hestenes in the calculus of variations. In it there is no explicit assumption concerning the normality or abnormality of the point y^0. If y^0 has abnormality of order q, however, let ν be a variable which ranges over a subset of $m-q$ of the numbers $1, \cdots, m$ such that the matrix $\|\phi_{\nu y_i}(y^0)\|$ has rank $m-q$, and let ρ range over the complementary subset. Then we have the following theorem:

THEOREM 1:4. *Let y^0 be a point which satisfies the hypotheses of Theorem 1:3 with a set of multipliers $l_0=1$, l_β, and let ν and ρ be variables having the ranges described in the last paragraph. Then y^0 is normal for the modified problem of minimizing the function $g=f+l_\rho\phi_\rho$ in the class of points y satisfying the restricted system of equations $\phi_\nu=0$, and y^0 satisfies the hypotheses of Theorem 1:3 for the modified problem with the multipliers $l_0=1$, l_ν. Furthermore if $g(y^0)$ is a minimum for the modified problem, then $f(y^0)$ is a minimum for the original one.*

We see that the point y^0 is normal for the modified problem, since the matrix $\|\phi_{\nu y_i}(y^0)\|$ has rank $m-q$. For the function $F=g+l_\nu\phi_\nu=f+l_\beta\phi_\beta$ of the modified problem the conditions (1:4) are satisfied for all sets η satisfying the equations

$$\phi_{\nu y_i}(y^0)\eta_i = 0, \tag{1:7}$$

since equations (1:5) are linear and have a matrix of coefficients of rank $m-q$ and hence are consequences of equations (1:7). The set of points y satisfying the equations $\phi_\nu=0$ includes the points satisfying the complete system $\phi_\beta=0$ as a subclass in which $g=f$. Hence if $g(y^0)$ is a minimum for the modified problem, the value $f(y^0)=g(y^0)$ must have the same property for the original problem.

From the last theorem it is evident that generality is not lost by proving Theorem 1:3 only for points y^0 which are normal. Such points are, in fact, the non-singular points of the class which satisfy the equations $\phi_\beta=0$. Near each of them there are infinitely many points of the class, as is shown by Lemma 1.1, and the minimum problem near one of them is therefore never trivial. Abnormal points, on the other hand, are the singular points of the class, and may occur in a wide variety of types. For some of these points the minimum problem is trivial, as, for example, in the case of a point y^0, for which $\phi_1=0$, which minimizes the function ϕ_1 in the class of points y satisfying the equations $\phi_2= \cdots =\phi_m=0$. Near such a point y^0 there is no other point satisfying all of the equations $\phi_\beta=0$.

An idea of the great variety of types of abnormal points may be gained by considering the problem of minimizing a function $f(y_1, y_2)$ of two variables in the class of points (y_1, y_2) satisfying a single equation $\phi(y_1, y_2)=0$. The variety of abnormal points possible in this case is at least as great as the variety of singular points of an algebraic curve. The particular example $f=2y_1^2-y_2^2$, $\phi=y_1^2y_2-y_2^3=0$, with minimizing point $(0, 0)$, shows that the condition involving the quadratic form in Theorem 1:3 is not in general necessary for a minimum.

2. **Abnormality in the calculus of variations.** The problem to be considered in this section [12, p. 4] is that of finding in a class of arcs

$$(2:1) \qquad y_i(x) \qquad (i = 1, \cdots, n;\ x_1 \leqq x \leqq x_2)$$

satisfying conditions of the form

$$\phi_\beta(x, y, y') = 0 \qquad (\beta = 1, \cdots, m < n),$$
$$\psi_\mu[x_1, y(x_1), x_2, y(x_2)] = 0 \qquad (\mu = 1, \cdots, p \leqq 2n + 2)$$

one which minimizes a sum

$$J = g[x_1, y(x_1), x_2, y(x_2)] + \int_{x_1}^{x_2} f(x, y, y')dx.$$

A set of values (x, y, y') and end values $[x_s, y_{is}] = [x_s, y_i(x_s)]$ $(s = 1, 2)$ is said to be *admissible* if it lies interior to a region of such values in which the functions f, g, ϕ_β, ψ_μ have continuous derivatives of at least the fourth order, and in which the matrix $\|\phi_{\beta y_i'}\|$ and the matrix of first derivatives of the functions ψ_μ have ranks m and p, respectively. An *admissible arc* C defined by functions of the form (2:1) is one which is continuous and consists of a finite number of sub-arcs with continuously turning tangents, and whose elements (x, y, y') and end values are admissible. When convenient we may represent by $J(C)$, $g(C)$, $\psi_\mu(C)$ the values of these functions determined by the arc C.

The conditions involved in the sufficiency theorems for this problem are the following, the numbering being that which I have often used [see, e.g., 12, chap. 3]:

I. THE MULTIPLIER RULE. A set of multipliers l_0, $l_\beta(x)$, e_μ for an admissible arc E is a set for which the l_0, e_μ are constants and the functions $l_\beta(x)$, defined on the interval x_1x_2 belonging to E, are continuous except possibly at values of x defining corners of E at which they nevertheless have well-defined forward and backward limits. The arc E satisfies the multiplier rule if there exist constants c_i and multipliers l_0, $l_\beta(x)$, e_μ such that for $F = l_0 f + l_\beta(x)\phi_\beta$ the equations

$$F_{y_i'} = \int_{x_1}^{x} F_{y_i}dx + c_i, \qquad \phi_\beta = 0$$

are satisfied along E, and furthermore such that the end values of E satisfy the equations

$$(2:2) \qquad \left[(F - y_i'F_{y_i'})dx + F_{y_i'}dy_i\right]_1^2 + l_0dg + e_\mu d\psi_\mu = 0, \qquad \psi_\mu = 0$$

identically in the differentials dx_s, dy_{is}.

It has been proved [12, p. 27] that the identically vanishing set of multipliers is the only set having constants l_0, e_μ all zero, or having functions l_0, $l_\beta(x)$ which vanish simultaneously at some value x on the interval x_1x_2.

II'_N. An admissible arc E satisfies the strengthened condition of Weierstrass if for every set of the type (x, y, y', l) in a neighborhood N of those belonging to E the inequality

$$E(x, y, y', l, Y') > 0$$

is satisfied for all admissible sets $(x, y, Y') \neq (x, y, y')$, where

$$E = F(x, y, Y', l) - F(x, y, y', l) - (Y'_i - y'_i)F_{y_i'}(x, y, y', l).$$

III'. An admissible arc E satisfies the strengthened condition of Clebsch if at every element (x, y, y', l) belonging to E, the inequality

$$F_{y_i'y_k'}(x, y, y', l)\pi_i\pi_k > 0$$

is satisfied for all non-vanishing sets π_i satisfying the equations

$$\phi_{\beta y_i'}(x, y, y')\pi_i = 0.$$

If we represent by q, q_μ the quadratic forms in dx_s, dy_{is} whose coefficients are the second derivatives of the functions g, ψ_μ, respectively, the second variation of J for an extremal arc E with multipliers $l_0 = 1$, $l_\beta(x)$, e_μ has the value

$$J_2(\xi, \eta) = 2\gamma[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')dx$$

in which

$$2\omega = F_{y_iy_k}\eta_i\eta_k + 2F_{y_iy_k'}\eta_i\eta'_k + F_{y_i'y_k'}\eta'_i\eta'_k ,$$

$$2\gamma = \left[(F_x - y'_iF_{y_i})dx + 2F_{y_i}dy_idx\right]_1^2 + 2q + 2e_\mu q_\mu$$

with dx, dy_i replaced by ξ, $y'_i\xi + \eta_i$ [12, p. 71]. The equations of variation along E are the equations

$$\Phi_\beta(x, \eta, \eta') = 0, \qquad \Psi_\mu[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] = 0 \tag{2:3}$$

in which

$$\Phi_\beta = \phi_{\beta y_i}\eta_i + \phi_{\beta y_i'}\eta'_i ,$$

and Ψ_μ is $d\psi_\mu$ with dx, dy_i replaced as above by ξ, $y'_i\xi + \eta_i$ [12, p. 14]. An *admissible set* ξ_1, ξ_2, $\eta_i(x)$ is one for which ξ_1, ξ_2 are constants and the functions $\eta_i(x)$ have on x_1x_2 the continuity properties of an admissible arc $y_i(x)$. The second variation $J_2(\xi, \eta)$ for E is by definition positive definite if it is positive for all non-vanishing admissible sets $\xi_1, \xi_2, \eta_i(x)$ satisfying the equations (2:3).

IV′. An extremal arc E satisfies the condition IV′ if its second variation is positive definite.

The condition IV′ is applicable to an admissible arc which has no corners and satisfies conditions I and III′, since such an arc is necessarily non-singular and an extremal [12, pp. 112, 117].

The sufficiency theorem of Hestenes to be considered here is now the following one:

THEOREM 2:1. *If an admissible arc E has no corners and satisfies the conditions* I, II_N', III′, IV′ *with a set of multipliers $l_0=1$, $l_\beta(x)$, e_μ then $J(E)$ is a strong relative minimum.*

Every admissible arc E satisfies the multiplier rule with none or a limited number of linearly independent non-vanishing sets of multipliers having $l_0=0$, It is said to have *order of abnormality equal to q* if it satisfies I with q and only q such sets $l_{0\sigma}=0$, $l_{\beta\sigma}(x)$, $e_{\mu\sigma}$ $(\sigma=1, \cdots, q)$. When $q=0$ it is said to be *normal.* A set of non-vanishing multipliers with $l_0=0$ will be called an *abnormal set of multipliers.*

For an admissible arc with order of abnormality equal to q the equation

$$(2:4)\qquad \Big[F_{\sigma y_i'}\eta_i\Big]_1^2 + e_{\mu\sigma}\Psi_\mu = 0$$

with $F_\sigma = l_{\beta\sigma}(x)\phi_\beta$ is for each σ an identity in the variables ξ_s, $\eta_{is}=\eta_i(x_s)$, since this is what the first equation (2:2) becomes for the multipliers $l_{0\sigma}=0$, $l_{\beta\sigma}(x)$, $e_{\mu\sigma}$ when the end values of dx, dy_i are replaced by those of ξ, $y_i'\xi+\eta_i$. The usual integration by parts applied to the sum

$$l_{\beta\sigma}(x)\Phi_\beta = F_{\sigma y_i}\eta_i + F_{\sigma y_i'}\eta_i'$$

gives the equation

$$(2:5)\qquad \int_{x_1}^{x_2} l_{\beta\sigma}\Phi_\sigma dx = \Big[F_{\sigma y_i'}\eta_i\Big]_1^2,$$

so that for every admissible set of variations satisfying the equations $\Phi_\beta=0$ we find with the help of equations (2:4) and (2:5) the relations

$$(2:6)\qquad \Big[F_{\sigma y_i'}\eta_i\Big]_1^2 = 0, \qquad e_{\mu\sigma}\Psi_\mu = 0.$$

The matrix of the q sets of values $e_{\mu\sigma}$ $(\sigma=1, \cdots, q)$ is necessarily of rank

q. Otherwise there would be a linear combination of these sets vanishing identically, and, according to a remark made above, the same combination of the linearly independent complete sets $l_{0\sigma}$, $l_{\beta\sigma}(x)$, $e_{\mu\sigma}$ would then also vanish identically, which is impossible. In the following paragraphs the variable ρ is understood to have as its range a subset of the numbers $\mu=1, \cdots, p$ such that the determinant $|e_{\rho\sigma}|$ is different from zero, and the variable ν will have the range complementary to that of ρ. The second equation (2:6) then shows that for an admissible set ξ_1, ξ_2, $\eta_i(x)$ the equations $\Psi_\rho=0$ are consequences of the equations $\Phi_\beta=\Psi_\nu=0$.

THEOREM 2:2. *Let E be an admissible arc without corners which satisfies the hypotheses of Theorem 2:1 with a set of multipliers $l_0=1$, $l_\beta(x)$, e_μ, and let ρ and ν be variables whose ranges are determined by the linearly independent abnormal sets of multipliers of E as described in the last paragraph. Then the arc E is normal for the modified problem of minimizing the functional $J(C)+e_\rho\psi_\rho(C)$ in the class of admissible arcs C satisfying the reduced system of equations $\phi_\beta=\psi_\nu=0$, and the arc E with the multipliers $l_0=1$, $l_\beta(x)$, e_ν satisfies the hypotheses of Theorem 2:1 for the modified problem. Furthermore if $J(E)+e_\rho\psi_\rho(E)$ is a strong relative minimum for the modified problem, then $J(E)$ is a similar minimum for the original problem.*

It is easy to see that the arc E is normal for the modified problem. For if E had for that problem a set of non-vanishing multipliers of the form $l_0=0$, $l_\beta(x)$, e_ν, the set $l_0=0$, $l_\beta(x)$, $e_\rho=0$, e_ν would be multipliers for E and the original problem, necessarily linearly expressible in terms of the q sets $l_{0\sigma}=0$, $l_{\beta\sigma}(x)$, $e_{\mu\sigma}$ $(\sigma=1, \cdots, q)$. This is, however, impossible on account of the fact that the determinant $|e_{\rho\sigma}|$ is not zero.

The arc E satisfies the hypotheses of Theorem 2:1 for the modified problem with the multipliers $l_0=1$, $l_\beta(x)$, e_ν, as one readily sees by an examination of the conditions I, II'_N, III$'$, IV$'$. For the condition IV$'$ one needs to note that on account of the second equation (2:6) the restricted system $\Phi_\beta=\Psi_\nu=0$ implies the complete system $\Phi_\beta=\Psi_\mu=0$.

Since the class of arcs in which a minimizing one is sought for the modified problem includes as a subclass those among which a minimizing arc is sought for the original problem, and since on the subclass the values of the functionals $J(C)+e_\rho\psi_\rho(C)$ and $J(C)$ are equal, the last statement of the theorem is evidently true.

The remarks made at the end of §1 are now applicable for the most part to the problem of Bolza also. As a result of Theorem 2:2 it is clear that no generality is lost by proving Theorem 2:1 for normal arcs only, and the proof for such arcs turns out to be in some respects simpler than for the abnormal arcs included in the proof of Hestenes. A normal arc is a non-singular arc of the class in which a minimizing arc is sought in the sense that near every normal arc there are an infinity of other arcs of the class [12, pp. 49, 51]. The minimum problem near such an arc is therefore never trivial. Near an abnormal arc E, on the other hand, there may be no other arc of the class in which a minimizing one is sought, as in the case when $\psi_1(E)$ vanishes and is a strong relative minimum or maximum in the class of admissible arcs satisfying the conditions $\phi_\beta = \psi_2 = \cdots = \psi_p = 0$. In this case the minimum problem near E is trivial. The variety of types of abnormal arcs is evidently very great. Those included in the sufficiency theorems of Hestenes are of a special type closely related to normal arcs. Other important special types can doubtless be described and discussed, and it might be useful to have results of this kind. But it seems likely that a comprehensive theory would at this time be exceedingly elaborate and difficult, and perhaps impossible.

When the number of the end conditions $\psi_\mu = 0$ is equal to the number $2n+2$ of end values x_s, y_{is} $(s=1, 2)$ the problem is said to have fixed end points. An admissible arc E is by definition normal on a sub-interval $x'x''$ if its corresponding sub-arc is normal relative to the problem with fixed end points on that interval. The assumption that an arc E is normal on every sub-interval is evidently undesirable, for the same reason that it would be undesirable to assume for the problem of §1 that the determinants of order m of some particular set belonging to the matrix $\|\phi_{\beta y_i'}\|$ are all different from zero. For the problem of Mayer, which is the problem of Bolza with integrand function f identically zero, every minimizing arc is abnormal on every sub-interval, as has been pointed out by Carathéodory [6, 7] and others. No theory based upon the assumption of normality on sub-intervals can therefore be effective in this important case.

Bibliography

1. Bolza, *Vorlesungen über Variationsrechnung*, 1st edition, 1909; 2d edition, 1933.

2. Bliss, *The problem of Lagrange in the calculus of variations*, American Journal of Mathematics, vol. 52 (1930), pp. 673–744.

3. Morse, *Sufficient conditions in the problem of Lagrange with variable end conditions*, American Journal of Mathematics, vol. 53 (1931), pp. 517–546; see also Morse, *The problems of Lagrange and Mayer under general end conditions*, Proceedings of the National Academy of Sciences, vol. 16 (1930), pp. 229–233.

4. Morse, *Sufficient conditions in the problem of Lagrange with fixed end points*, Annals of Mathematics, vol. 32 (1931), pp. 567–577.

5. Bliss, *The problem of Bolza in the calculus of variations*, Annals of Mathematics, vol. 33 (1932), pp. 261–274.

6. Carathéodory, *Die Theorie der zweiten Variation beim Problem von Lagrange*, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, 1932, pp. 99–114.

7. Carathéodory, *Über die Einteilung der Lagrangeschen Variationsprobleme nach Klassen*, Commentarii Mathematici Helvetici, vol. 5 (1933), pp. 1–10.

8. Graves, *On the Weierstrass condition for the problem of Bolza in the calculus of variations*, Annals of Mathematics, vol. 33 (1932), pp. 747–752.

9. Bliss and Hestenes, *Sufficient conditions for a problem of Mayer in the calculus of variations*, these Transactions, vol. 35 (1933), pp. 305–326.

10. Hestenes, *Sufficient conditions for the general problem of Mayer with variable end points*, these Transactions, vol. 35 (1933), pp. 479–490; also *Contributions to the Calculus of Variations* 1931–1932, The University of Chicago.

11. Hestenes, *Sufficient conditions for the problem of Bolza in the calculus of variations*, these Transactions, vol. 36 (1934), pp. 793–818.

12. Bliss, *The problem of Bolza*, Mimeographed lecture notes, University of Chicago, Winter Quarter, 1935.

13. Morse, *Sufficient conditions in the problem of Lagrange without assumptions of normalcy*, these Transactions, vol. 37 (1935), pp. 147–160.

14. Hestenes, *The problem of Bolza in the calculus of variations in parametric form*, American Journal of Mathematics, vol. 58 (1936), pp. 391–406.

15. Reid, *The theory of the second variation for the non-parametric problem of Bolza*, American Journal of Mathematics, vol. 57 (1935), pp. 573–586.

16. Hestenes, *On sufficient conditions in the problems of Lagrange and Bolza*, Annals of Mathematics, vol. 37 (1936), pp. 543–551.

University of Chicago,
Chicago, Ill.

MEMORANDUM
RM-3858-PR
OCTOBER 1963

A THEOREM OF FRITZ JOHN IN MATHEMATICAL PROGRAMMING

Richard W. Cottle

This research is sponsored by the United States Air Force under Project RAND—contract No. AF 49(638)-700 monitored by the Directorate of Development Planning, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

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

DOI 10.1007/978-3-0348-0439-4_6, © Springer Basel 2014

PREFACE

This Memorandum contributes to an aspect of the research program of The RAND Corporation consisting of basic supporting studies in mathematics. Considered is the application of a theorem of Fritz John to mathematical programming.

The author, a consultant to The RAND Corporation, is a Research Assistant at the Operations Research Center, Richmond, California, where some of the earlier research was conducted, partially supported by the Office of Naval Research, Contract No. Nonr-222(83), with the University of California. Preparation of the Memorandum was undertaken under U. S. Air Force Project RAND.

SUMMARY

In this Memorandum, the author specializes a theorem of Fritz John to the case of mathematical programming. It is shown that when a certain multiplier is positive, the well-known Kuhn-Tucker conditions obtain. A sufficient condition for the positivity of this multiplier is proposed.

A THEOREM OF FRITZ JOHN IN MATHEMATICAL PROGRAMMING

1. INTRODUCTION

A 1948 article [1] by F. John, entitled "Extremum Problems with Inequalities as Subsidiary Conditions," appears to be the first paper in which the classical theory of equality-constrained extremization is extended to deal with inequality-constrained extremization. John establishes a theorem on necessary conditions for a minimum and another theorem on sufficient conditions for a relative minimum. The remainder of the paper is devoted to applications of the results. Only the first of these two theorems will be mentioned here.

A more widely known paper is "Nonlinear Programming" by H. W. Kuhn and A. W. Tucker [2] in which John's article is referred to but not discussed in detail. The Kuhn-Tucker paper treats necessary and sufficient conditions for an inequality-constrained maximum.

The purpose of this Memorandum is to point out how the addition of a suitable regularity condition in John's theorem enables one to deduce the "Kuhn-Tucker conditions."

2. THE MATHEMATICAL PROGRAMMING PROBLEM

A typical formulation of the mathematical programming problem is

$$\text{Maximize } f(x) \text{ subject to } x \in E_n \text{ , } g(x) \geq 0 \text{ .} \tag{2.1}$$

Further assumptions on the function f and the mapping g yield special types of programming, such as linear, quadratic, concave, etc.

The classical problem referred to earlier is

$$\text{Maximize } f(x) \text{ subject to } x \in E_n \text{ , } g(x) = 0 \tag{2.2}$$

where f is a differentiable function and

$$g(x) = [g_1(x), \ldots, g_m(x)]^T$$

is a vector-valued differentiable mapping on E_n $(m < n)$. In the method of Lagrange (or undetermined) multipliers, one forms the Lagrangian function

$$L(x,u) = f(x) + u^T g(x) \tag{2.3}$$

If at a maximum, x^0 , the Jacobian matrix $(\partial g_i / \partial x_j)$ has rank m , the following conditions must hold:

$$L_x(x^0, u^0) = 0 \quad \text{some } u^0 \tag{2.4}$$

$$g(x^0) = 0 \text{ .} \tag{2.5}$$

One seeks the solutions of the problem (2.2) among the extrema of the unconstrained Lagrangian $L(x,u)$.*
The conditions (2.4) and (2.5) are necessary, though not sufficient, for an extremum.

We shall state John's theorem below, but with certain notational changes and maximization replacing minimization. The intention is to maintain consistency in problem statements.

Let R be a set of points in E_n and f a real-valued function on R . Let S be a compact metric space. Let $g(x,\sigma) = g_\sigma(x)$ be a real-valued function on $R \times S$. Now define the set

$$R' = \{x \in R \mid g_\sigma(x) \geq 0 \quad \text{all} \quad \sigma \in S\} . \tag{2.6}$$

We seek $x^0 \in R'$ such that

$$f(x^0) = \max_{x \in R'} f(x) . \tag{2.7}$$

Assume that f and $\partial f/\partial x_j$ are continuous on R $(j = 1,\ldots,n)$ and that g and $\partial g/x_j$ are continuous on $R \times S$ $(j = 1,\ldots,n)$. Notice that $R \times S$ can be given a metric space structure.†

---

*See Ref. 3.

†See, for example, Ref. 4, p. 91.

With these notations we are prepared to state

Theorem A:* Let $x^0 \varepsilon R'$ be an interior point of R, and let

$$f(x^0) = \max_{x \varepsilon R'} f(x) .$$

Then there exists a finite set of points, $\sigma_1,\ldots,\sigma_s \varepsilon S$, and numbers, $u_0, u_1,\ldots,u_s$, not all zero, such that

$$g_{\sigma_r}(x^0) = 0 \quad r = 1,\ldots,s \tag{2.8a}$$

$$u_0 \geqq 0 \ , \ u_1 > 0,\ldots,u_s > 0 \tag{2.8b}$$

$$0 \leqq s \leqq n \tag{2.8c}$$

the function (2.8d)

$$\Phi(x) = u_0 f(x) + \sum_{r=1}^{s} u_r g_{\sigma_r}(x)$$

has a critical point at x^0 ; i.e., $\Phi_x(x^0) = 0$.

It is important to notice that the "multiplier" u_0 could be zero and that there are no regularity conditions imposed on the constraint set R' .

We shall specialize John's theorem to the case where S is the set $\{1,\ldots,m\}$--which is trivially a compact metric space--and the variables x_j are non-negative.

---

*John, Ref. 1, p. 188.

Theorem B: Let x_0 maximize $f(x)$ constrained by $g_i(x) \geqq 0$, $i = 1,\ldots,m$, and $x_j \geqq 0$, $j = 1,\ldots,n$. Then there exists a semi-positive (i.e., non-negative and non-zero) vector $(u_0,u_1,\ldots,u_m,v_1,\ldots,v_n)^T$ such that

$$u_i g_i(x^0) = 0 \quad i = 1,\ldots,m \tag{2.9a}$$

$$v_j x_j^0 = 0 \quad j = 1,\ldots,n \tag{2.9b}$$

the function (2.9c)

$$\Phi(x) = u_0 f(x) + \sum_{i=1}^{m} u_i g_i(x) + \sum_{j=1}^{n} v_j x_j$$

has a critical point at x^0 .

In order to state the analogous theorem of Kuhn and Tucker, we must first recall the constraint qualification.*

Let x^0 belong to the boundary of the constraint set

$$R' = \{x \epsilon E_n | g_i(x) \geqq 0 \quad i = 1,\ldots,m \text{ , } x_j \geqq 0 \text{ , } j = 1,\ldots,n\}$$

Let $g^{[1]}(x)$ be the mapping defined by all those component functions of g which vanish at x^0 . Let I_1 consist of those rows of the $n x n$ identity matrix corresponding to components of x^0 which are zero. The Kuhn-Tucker constraint qualification is satisfied at x^0

*Ref. 2, p. 483.

if every vector differential dx satisfying the homogeneous linear inequalities

$$g_x^{[1]}(x^0)dx \geqq 0 \ , \ I_1 dx \geqq 0 \tag{2.10}$$

is tangent to an arc contained in the set R'. This means that to any dx satisfying (2.10) there corresponds a differentiable arc $x = \alpha(\theta)$, $0 \leq \theta \leq 1$, contained in R' such that $x^0 = \alpha(0)$ and some positive scalar λ such that $\alpha'(0) = \lambda dx$.

Theorem C:* Let R' satisfy the constraint qualification. In order that x^0 maximize $f(x)$ subject to $x \varepsilon R'$, it is necessary that x^0 and some $u = (u_1, \ldots, u_m)^T$ satisfy the following conditions:

$$f_x(x^0) + [g_x(x^0)]^T u \leqq 0 \tag{2.11a}$$

$$(x^0)^T \{f_x(x^0) + [g_x(x^0)]^T u\} = 0 \tag{2.11b}$$

$$x^0 \geqq 0 \tag{2.11c}$$

$$g(x^0) \geqq 0 \tag{2.11d}$$

$$u^T g(x^0) = 0 \tag{2.11e}$$

$$u \geqq 0 \ . \tag{2.11f}$$

These relations have also been called the quasi-saddle point conditions [5]. They are necessary conditions of optimality in the program

$$\text{maximize} \quad f(x) \quad \text{subject to} \quad x \varepsilon R' \tag{2.12}$$

---

*Kuhn-Tucker conditions, Ref. 2, p. 484.

when R' satisfies the constraint qualification. The program (2.12) is called the maximum problem [2].

Theorem 1: Let x^0 solve the problem in Theorem B. If the multiplier u_0 is positive, then the Kuhn-Tucker conditions hold.*

Proof. If $u_0 > 0$ we may assume $u_0 = 1$. Let $u = (u_1,\ldots,u_m)^T$ and $v = (v_1,\ldots,v_n)^T$. Then

$$f_x(x^0) + [g_x(x^0)]^T u = -v \leq 0 \tag{2.13}$$

which is (2.11a). From (2.13) and (2.9b) we get

$$\begin{aligned}(x^0)^T\{f_x(x^0) + [g_x(x^0)]^T u + v\} \\ = (x_0)^T\{f_x(x^0) + [g_x(x^0)]^T u\} = 0\end{aligned}$$

which is (2.11b). The remainder of the conditions (2.11) are even more obvious.

3. A SUFFICIENT CONDITION FOR POSITIVE u_0

With all notations as above, let x^0 solve the problem of Theorem B, and let $(u_0,u_1,\ldots,u_m,v_1,\ldots,v_n)^T$ be the associated semi-positive vector of multipliers. $g^{[1]}$ is the mapping composed of components of g which vanish at x^0 . Let $\bar{x}$ be the vector of components x_j of x such that $x_j^0 > 0$.

---

*A similar result may be found in Ref. 6, p. 227, the English translation of Ref. 4; see also Ref. 2, p. 489.

The regularity condition we shall impose is that the equation

$$y^T[g_{\overline{x}}^{[1]}(x^0)] = 0 \tag{2.14}$$

have no semipositive solution. This condition is slightly more general than the nondegeneracy condition of Ref. 5, which is that $g_{\overline{x}}^{[1]}(x^0)$ be of full rank.

Theorem 2: If g satisfies the regularity condition (2.14), the multiplier u_0 in Theorem B is positive.

Proof. With $\overline{x}$ "evaluated" at x^0, we get $(\overline{x})^0 > 0$, and consequently $\overline{v} = 0$; that is, the corresponding vector of multipliers is zero. Suppose $u_0 = 0$. Then

$$[g_x(x^0)]^T u + v = 0 \tag{2.15}$$

and in particular

$$[g_{\overline{x}}(x^0)]^T u + \overline{v} = [g_{\overline{x}}(x^0)]^T u = 0 . \tag{2.16}$$

Let $u^{[1]}$ be the vector of multipliers corresponding to $g^{[1]}$. Any components in u but not in $u^{[1]}$ must be zero. Therefore, we conclude

$$(u^{[1]})^T[g_{\overline{x}}^{[1]}(x^0)] = 0 . \tag{2.17}$$

Now $u^{[1]}$ is non-negative and cannot be zero, for otherwise u in (2.15) is zero and then so is v.

But this contradicts the semi-positivity of $(u_0, u_1, \ldots, u_m, v_1, \ldots, v_n)^T$. Hence, $u^{[1]}$ is semi-positive. However, (2.17) contradicts our regularity assumption. Therefore, $u_0 > 0$.

REFERENCES

1. John, F., "Extremum Problems with Inequalities as Subsidiary Conditions," Studies and Essays, Courant Anniversary Volume, Interscience, New York, 1948.

2. Kuhn, H. W., and A. W. Tucker, "Nonlinear Programming," Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman, ed., University of California Press, Berkeley, 1951, pp. 481-492.

3. Courant, R., Differential and Integral Calculus, Interscience, New York, 1936.

4. Berge, C., Espaces Topologiques, Dunod, Paris, 1959.

5. Arrow, K. J., L. Hurwicz, and H. Uzawa, "Constraint Qualifications in Maximization Problems," Naval Res. Logist. Quart., Vol. 8, No. 2, June 1961, pp. 175-191.

6. Berge, C., Topological Spaces, trans. by E. M. Patterson, MacMillan, New York, 1963.

ON CONJUGATE CONVEX FUNCTIONS

WERNER FENCHEL

1. Since the classical work of Minkowski and Jensen it is well known that many of the inequalities used in analysis may be considered as consequences of the convexity of certain functions. In several of these inequalities pairs of "conjugate" functions occur, for instance pairs of powers with exponents a and α related by $1/a + 1/\alpha = 1$. A more general example is the pair of positively homogeneous convex functions defined by Minkowski and known as the distance (or gauge) function and the function of support of a convex body. The purpose of the present paper is to explain the general (by the way rather elementary) idea underlying this correspondence. Subjected to a more precise formulation the result is the following:

To each convex function $f(x_1, \ldots, x_n)$ defined in a convex region G and satisfying certain conditions of continuity there corresponds in a unique way a convex region Γ and a convex function $\phi(\xi_1, \ldots, \xi_n)$ defined in Γ and with the same properties such that

$$x_1\xi_1 + \ldots + x_n\xi_n \leqq f(x_1, \ldots, x_n) + \phi(\xi_1, \ldots, \xi_n), \tag{1}$$

for all points $(x_1, \ldots, x_n)$ in G and all points $(\xi_1, \ldots, \xi_n)$ in Γ. The inequality is exact in a sense explained below. The correspondence between G, f and Γ, ϕ is symmetric, and the functions f and ϕ are called conjugate.[1]

The hypersurfaces $y = f(x_1, \ldots, x_n)$ and $\eta = \phi(\xi_1, \ldots, \xi_n)$ correspond to each other in the polarity with respect to the paraboloid

$$2y = x_1^2 + \ldots + x_n^2 .$$

Let $F(x)$ be strictly increasing for $x \geqq 0$. Then $f(x) = \int_0^x F(x)dx$ is convex, and its conjugate function is $\phi(\xi) = \int_0^\xi \Phi(\xi)d\xi$ where $\Phi(\xi)$ is the inverse function of $F(x)$. The inequality (1) for $n = 1$ therefore yields the well-known inequality of W. H. Young[2]

$$x\xi \leqq \int_0^x F(x)dx + \int_0^\xi \Phi(\xi)d\xi .$$

(1) may thus be considered as a generalization of this inequality.

Received March 24, 1948.

[1]The case $n = 1$ has been considered by S. Mandelbrojt [3] under the assumption that the ranges G and Γ are identical with the entire axis $-\infty < x < \infty$. This, however, is incompatible with the complete reciprocity between f and ϕ which will appear from an example given below. Mandelbrojt's formulation of the theorem is thus not quite correct due to the fact that the least upper bounds occurring in it may be infinite.

[2]See e.g. [2] p. 111.

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, 125
DOI 10.1007/978-3-0348-0439-4_7,

If $f(x_1, \ldots, x_n)$ is positively homogeneous of degree one, then G is the entire space $x_1, \ldots, x_n$ while Γ is closed and bounded, and $\phi(\xi_1, \ldots, \xi_n)$ is identically zero. In this case (1) expresses that $f(x_1, \ldots, x_n)$ is the function of support of the convex body Γ.[3]

2. The euclidean spaces with coordinates $x_1, \ldots, x_n$ and $x_1, \ldots, x_n, y$ will be denoted by R^n and R^{n+1} respectively, points and vectors in these spaces by x and x, y respectively. Furthermore we write

$$x' + x'' = (x'_1 + x''_1, \ldots, x'_n + x''_n), \quad \lambda x = (\lambda x_1, \ldots, \lambda x_n),$$

$$\Sigma x\xi = x_1\xi_1 + \ldots + x_n\xi_n.$$

θ will always denote a number in the interval $0 < \theta < 1$.

The point set G of R^n is supposed to be convex, i.e. if x' and x'' belong to G, the whole segment $(1 - \theta)x' + \theta x''$ belongs to G. But G need neither be closed nor open nor bounded. The interior points of segments belonging to G are shortly called the interior points of G. All other points of accumulation of G, belonging to G or not, will be called the boundary or extreme points of G.

A function $f(x)$ defined in G is called convex if

$$f\big((1 - \theta)x' + \theta x''\big) \leqq (1 - \theta)f(x') + \theta f(x'') \tag{2}$$

for any two points x' and x'' of G and all θ. It is well known that this implies that $f(x)$ is continuous at the interior points of G. For our purpose we have also to consider the behaviour of $f(x)$ at the boundary points. Let x^* be a boundary point of G. For functions of one variable $\lim\limits_{x \to x^*} f(x)$ exists or is ∞. But this is not necessarily the case for functions of several variables. If x^* belongs to G the only general conclusion to be drawn from (2) is that

$$\varliminf_{x \to x^*} f(x) \leqq f(x^*); \tag{3}$$

for, from

$$f\big((1 - \theta)x + \theta x^*\big) \leqq (1 - \theta)f(x) + \theta f(x^*)$$

it follows that

$$\varliminf_{x \to x^*} f(x) \leqq \varliminf_{\theta \to 1} f\big((1 - \theta)x + \theta x^*\big) \leqq f(x^*),$$

and (2) remains valid if $f(x^*)$ is replaced by any other value satisfying (3).

If necessary, we now change G and f by adding to G all those boundary points x^* not yet belonging to G for which $\varliminf\limits_{x \to x^*} f(x)$ is finite and by defining f at these and at the boundary points previously belonging to G by

$$f(x^*) = \varliminf_{x \to x^*} f(x). \tag{4}$$

The new G and the function f obtained in this way are obviously again convex; for, let x' and x'' be arbitrary points of the new G and $x'_{(\nu)}$ and $x''_{(\nu)}$, $\nu = 1, 2, \ldots$,

---

[3]See e.g. [1] p. 23-24.

sequences of interior points of G such that

$$x'_{(\nu)} \to x',\ x''_{(\nu)} \to x'',\ f(x'_{(\nu)}) \to f(x'),\ f(x''_{(\nu)}) \to f(x''),$$

then we get from

$$f\big((1-\theta)x'_{(\nu)} + \theta x''_{(\nu)}\big) \leqq (1-\theta)f(x'_{(\nu)}) + \theta f(x''_{(\nu)})$$

for $\nu \to \infty$

$$\underline{\lim}_{x \to (1-\theta)x' + \theta x''} f(x) \leqq \underline{\lim}_{\nu \to \infty} f\big((1-\theta)x'_{(\nu)} + \theta x''_{(\nu)}\big) \leqq (1-\theta)f(x') + \theta f(x''),$$

which shows that $(1-\theta)x' + \theta x''$ belongs to G and that (2) is valid, as the left-hand side is $f\big((1-\theta)x' + \theta x''\big)$.

With (3) in mind we may say that (4) expresses that the functions which will be considered in the following are convex and semi-continuous from below, and G is "closed relative to f," i.e. all boundary points at which $\underline{\lim}\, f(x)$ is finite belong to G, or in other words, at each boundary point which does not belong to G we have $\lim f(x) = \infty$.

3. The theorem to be proved may now be formulated thus:

Let G be a convex point set in R^n and $f(x)$ a function defined in G convex and semi-continuous from below and such that $\lim_{x \to x^*} f(x) = \infty$ for each boundary point x^* of G which does not belong to G. Then there exists one and only one point set Γ in R^n and one and only one function $\phi(\xi)$ defined in Γ with exactly the same properties as G and $f(x)$ such that

$$\Sigma x\xi \leqq f(x) + \phi(\xi), \tag{5}$$

where to every interior point x of G there corresponds at least one point ξ of Γ for which equality holds.

In the same way $G, f(x)$ correspond to $\Gamma, \phi(\xi)$.

We define Γ as the set of all points ξ with the property that the function $\Sigma x\xi - f(x)$ is bounded from above in G, and we define $\phi(\xi)$ in Γ as the least upper bound of this function:

$$\phi(\xi) = \underset{x \in G}{\text{l.u.b.}} \big(\Sigma x\xi - f(x)\big).$$

Then (5) is valid. The inequality $\Sigma x\xi - f(x) \leqq z$ or

$$f(x) \geqq \Sigma x\xi - z$$

means that the hyperplane $y = \Sigma x\xi - z$ in R^{n+1} with the normal vector $\xi, -1$ lies nowhere above the hypersurface $y = f(x)$, and $-z$ is the intercept of this hyperplane on the y-axis. It is a well-known fact that there exists at least one hyperplane of support of the convex hypersurface, i.e. a hyperplane which contains at least one point of the hypersurface and lies nowhere above it. This shows that Γ is not empty. Further we see that if there exists a hyperplane of support with the normal vector $\xi, -1$ and if $x^o, f(x^o)$ is a point of contact, then we have

$$\phi(\xi) = \Sigma x^o\xi - f(x^o),$$

and $-\phi(\xi)$ is the y-intercept of this hyperplane. If x^o is an arbitrary interior point of G, a hyperplane of support through $x^o, f(x^o)$ exists, and this proves the assertion on the equality sign in (5).

It is evident that Γ and $\phi(\xi)$ are convex. In fact, let ξ' and ξ'' be arbitrary points of Γ, then we have for $x \epsilon G$,

$$\Sigma x\xi' - f(x) \leqq \phi(\xi'), \ \Sigma x\xi'' - f(x) \leqq \phi(\xi''),$$

hence
$$\Sigma x((1-\theta)\xi' + \theta\xi'') - f(x) \leqq (1-\theta)\phi(\xi') + \theta\phi(\xi'')$$

which shows that $(1-\theta)\xi' + \theta\xi''$ belongs to Γ and that

$$\phi((1-\theta)\xi' + \theta\xi'') \leqq (1-\theta)\phi(\xi') + \theta\phi(\xi'').$$

Let now ξ^* be a boundary point of Γ and $\xi\epsilon\Gamma$, $x\epsilon G$. Then it follows from (5) that

$$\varliminf_{\xi \to \xi^*} \phi(\xi) \geqq \Sigma x\xi^* - f(x)$$

and this shows on the one hand that $\xi^* \epsilon \Gamma$ if $\varliminf \phi(\xi)$ is finite, i.e. that Γ is closed relative to $\phi(\xi)$, and on the other hand that

$$\varliminf_{\xi \to \xi^*} \phi(\xi) \geqq \phi(\xi^*),$$

i.e. that $\phi(\xi)$ is semi-continuous from below. Hence Γ and ϕ have the same properties as G and f.

4. It remains to be proved that if we start with Γ and $\phi(\xi)$ the same procedure gives G and $f(x)$ again. We have to consider the set G^* of all points x for which $\Sigma\xi x - \phi(\xi)$ is bounded from above in Γ, together with the function

$$f^*(x) = \underset{\xi \epsilon \Gamma}{\text{l.u.b.}} \left(\Sigma\xi x - \phi(\xi)\right)$$

defined in G^*.

If $x \epsilon G$ we get from (5)

$$\Sigma\xi x - \phi(\xi) \leqq f(x) \tag{6}$$

for all $\xi\epsilon\Gamma$, hence $G \subset G^*$ and $f^*(x) \leqq f(x)$ in G. But to an interior point x of G there corresponds a ξ such that equality is valid in (6), which implies $f^*(x) \geqq f(x)$. Hence $f^*(x) = f(x)$ at the interior points of G and, as both functions are convex and semi-continuous from below, also at the boundary points of G.

Let now x^o be a point of R^n not in G. We have to prove that it does not belong to G^*, i.e. that

$$\underset{\xi \epsilon \Gamma}{\text{l.u.b.}} \left(\Sigma\xi x^o - \phi(\xi)\right) = \infty. \tag{7}$$

Since the quantity $\Sigma\xi x^o - \phi(\xi)$ is the y-coordinate of the point at which the hyperplane

$$y = \Sigma\xi x - \phi(\xi)$$

of R^{n+1} intersects the line $x = x^o$ parallel to the y-axis, we have to show that there are hyperplanes below the hypersurface $y = f(x)$ which have arbitrary large intercepts on the line $x = x^o$. Suppose first that x^o is an exterior point of G. Then there exists a hyperplane H parallel to the y-axis which separates the line $x = x^o$ from G and $y = f(x)$. Consider any hyperplane of support S of $y = f(x)$. Let S turn around the intersection of H and S so that the part lying below $y = f(x)$ moves downwards. Then the point at which S intersects the line $x = x^o$ moves upwards and tends to infinity. Suppose next that x^o is a boundary point of G but not belonging to G. Then we have $f(x) \to \infty$ for $x \to x^o$. Consider any segment belonging to G and having x^o as one of its end points. Let x' be a fixed point and x'' a variable point of the segment between

x' and x^o such that $f(x'') > f(x')$. A plane of support through x'', $f(x'')$ then intersects the line $x = x^o$ at a point the y-coordinate of which is greater than $f(x'')$ and therefore tends to infinity if $x'' \to x^o$. This completes the proof of the theorem.

5. In section **1** it has been asserted that the hypersurfaces $y = f(x)$ and $\eta = \phi(\xi)$ correspond to each other in the polarity with respect to $2y = \Sigma x^2$. This is obviously true in the sense that each of the hypersurfaces is the envelope of the polar hyperplanes of the points of the other. For $y = f(x)$ may be considered as the envelope of the hyperplanes

$$y = \Sigma x\xi - \phi(\xi),$$

where $\xi \epsilon \Gamma$ is the parameter, and the poles of these hyperplanes are the points ξ, $\phi(\xi)$.

6. Suppose now that $y = f(x)$ is strictly convex, i.e. each hyperplane of support contains only one point of $y = f(x)$. Let further $\eta = \phi(\xi)$ satisfy the same condition; for $y = f(x)$ this means that there passes at most one hyperplane of support through a point of $y = f(x)$. Then $f(x)$ has continuous derivatives[4]

$$f_i(x) = \frac{\partial f}{\partial x_i}$$

and we have

$$\xi_i = f_i(x).$$

These relations establish a continuous one to one correspondence between the interior points of G and those of Γ. Solving them with respect to the x we get

$$x_i = \phi_i(\xi)$$

where, for reasons of symmetry, the ϕ_i must be the derivatives of ϕ. From this it is seen that in the case of $n = 1$ the derivatives of two conjugate convex functions are mutually inverse functions. This proves the assertion of section **1** on the inequality of Young. Furthermore we get an explicit expression for $\phi(\xi)$ if $f(x)$ is given, viz.

$$\phi(\xi) = \sum_{i=1}^{n} \xi_i\phi_i(\xi) - f(\phi_i(\xi))$$

valid in the interior of Γ. Hence, our correspondence between f and ϕ is the Legendre transformation of the theory of differential equations.

References

[1] T. Bonnesen, W. Fenchel, *Theorie der konvexen Körper* (Berlin, 1934).

[2] G. H. Hardy, J. E. Littlewood, G. Pólya, *Inequalities* (Cambridge, 1934).

[3] S. Mandelbrojt, "Sur les fonctions convexes," *C.R. Acad. Sci. Paris*, vol. **209** (1939), 977-978.

The Technical University of Denmark
Copenhagen

[4]See [1] p. 23, 26. The argument used there in the case of positively homogeneous convex functions may easily be generalized to the case considered here.

4

PROGRAMMING IN LINEAR SPACES[1]

LEONID HURWICZ

I. Introduction

I.1. The present study has its origin in problems of optimal resource allocation, especially those related to the possibilities of a price mechanism. While for some purposes Pareto-optimality might be the more relevant concept, we have confined ourselves here to the case where by "optimal" is meant "efficient" resource allocation.[2]

The main result of the present chapter is an extension of the Kuhn-Tucker results [31] on "non-linear programming" to more general linear topological spaces.[3] (Numbers in square brackets indicate the references). The initial stimulus toward this type of generalization was the paper by Rosenbloom [39]. The need for it became apparent in the course of a larger study on problems of decentralization in resource allocation mechanisms.

The remainder of the Introduction is devoted to a brief statement of the nature of the problem. I.2 is primarily directed at the reader interested in the relevance of the study from the viewpoint of economics. I.3 provides a summary of some of the results being generalized and some of the mathematical issues arising.

II is devoted to introducing some of the basic concepts and notations. III and IV are devoted to the derivation of certain theorems on linear inequalities in linear topological spaces, among them the "Minkowski-Farkas Lemma," fundamental in the sequel, and another result of

[1] This study might not have been undertaken, and almost certainly would not have been completed, without persevering encouragement by Tjalling C. Koopmans of the Cowles Commission for Research in Economics. The author is greatly indebted to Professor Paul Rosenbloom, of the University of Minnesota, for having made him aware of the potentialities of the Banach space theory for the problems here treated. He also wishes to acknowledge valuable advice received from Professors Bernard R. Gelbaum and Gerhard K. Kalisch, also of the University of Minnesota, as well as from Kenneth J. Arrow and Hirofumi Uzawa of Stanford University. Thanks are due a number of readers, and to K. S. Kretschmer in particular, for pointing out errors in the first printing; see especially the note on p. 74.

[2] Cf. Koopmans [27].

[3] Some of the results, as for instance the generalized "Minkowski-Farkas Lemma," may be of independent interest.

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_8, © Springer Basel 2014

importance in relating the theory of programming to that of games of strategy. An appendix to III relates the results of III to the theory of linear equations in Banach spaces, as formulated in a paper by Hausdorff [18].[4] (See the NOTE in brackets on page 74.) V.1 states conditions under which a Lagrangian saddle-point implies maximality (efficiency). V.2 deals with the problem of scalarization, i.e., of reducing a vectorial maximization problem to one of scalar maximization. V.3 contains the main results concerning the existence of saddle-points and "quasi-saddle-points." The third section, V.3.3 treats situations where the differential ("marginal") first-order conditions for the saddle-points are satisfied, while in the first, V.3.1, the differentiability is not assumed. (From the economist's viewpoint, the existence of a saddle-point corresponds to the existence of a price-vector equilibrating the market.) V.3.2 is devoted to the special ("linear") case which, in view of the interest in "linear programming" models, seemed worthy of separate direct treatment. The author has not completely avoided repetition where he feared that brevity might cause ambiguity. Also, many "obvious" and "trivial" proofs are spelled out in detail.

I.2. In problems of efficient resource allocation we deal with a model where commodities are classified into *resources*, typically available in limited amounts, and *desirables* in terms of which efficiency is defined. The amounts of resources used up and of the desirables produced are determined by the decision as to *activity levels*. Thus the model is of the type treated in activity analysis,[5] although not necessarily under the assumptions of additivity and linearity.

Some of the mathematical problems arising in models where linearity and additivity are not assumed have been explored by Kuhn and Tucker [31], and Slater [41]. The treatment in both papers is confined to the case where there is a finite number of commodities and a finite number of activities. This, of course, limits their domain of application. In economics there are many problems where, for instance, an infinity of commodities could be more naturally postulated, as in the case of problems involving time or location. But there is another reason why an economist may be interested in having a theory of resource allocation in which the commodity space or the activity space is of a more general nature: the logical structure of treatment of the more general situations often reveals the "deeper" or "intuitive" bases of important propositions and helps focus attention on the more fundamental features of the problem.

The economic interpretation of the Kuhn-Tucker and Slater results (discussed by Kuhn and Tucker) has to do with the possibility of reaching positions of efficient resource allocation through the price mechanism.

[4] On a number of occasions we explore the question of the necessity of the underlying hypotheses and several theorems are devoted to this.

[5] Cf. [27].

Roughly speaking,[6] when suitable conditions are satisfied [in the economist's language the main ones could be described as "perfect divisibility" (all positive multiples permissible), and absence of "external (dis-) economies of scale" and of "increasing returns"], (a) "competitive equilibrium" implies that an efficient point has been reached and (b) any efficient point can be one of "competitive equilibrium" (provided prices are properly selected). When the absence of "increasing returns" is not assumed (while the assumptions of "perfect divisibility" and absence of "external (dis-) economies of scale" are retained), it is still possible to obtain criteria[7] permitting classification of certain situations as non-efficient.

Results of the latter type are of considerable importance, for they serve as a basis for the development of a theory of resource allocation applicable to a class of situations not excluding "increasing returns."[8]

In attempting to generalize results of this type, the writer was guided by his interest in cases where "increasing returns" might prevail and hence "marginal" type phenomena would have to be considered. Mathematically, this meant working in a space with an operation of differentiation possessing most of its usual properties. The Banach spaces form the most general class of spaces with which the writer was familiar at the time this was written, although a more general theory of differentiation does exist.[9] However, in theorems where the differential operations were not used, an attempt was made to obtain proofs valid for a more general class of linear spaces. In V.3.1, the author has treated the case of Lagrangian saddle-points by methods of the type used by Slater, i.e., relying on convexity but not using differentiability.

If one is to treat phenomena of "indivisibilities," one must go beyond linear spaces. But since one knows that most of the important results valid in linear spaces cannot be expected to hold when "indivisibilities" appear, it becomes desirable to reappraise the objectives of the theory of resource-allocating mechanisms, especially in their "decentralization" aspects. This is done to some extent in another paper now being completed by the writer.

I.3. In this section we give a brief description of some of the results being generalized in the present chapter.

[6] For a more precise statement the reader is referred to Koopmans [27], Kuhn and Tucker [31], and Arrow and Hurwicz [2], and Chapter 3 of the present book.

[7] These criteria are of differential ("marginal") first-order nature; they involve prices, but not the full conditions of "competitive equilibrium." Cf. Theorem 1 in Kuhn and Tucker [31].

[8] Cf. the work of Hotelling [21], Lange and Taylor [33], and Lerner [35], especially as it involves "marginal cost" pricing. Cf. also Arrow and Hurwicz [2], and Chapter 6 of the present book.

[9] Cf. Hyers [22], and its Bibliography. See V. 3.3.8, where results involving differential operations are extended to a wider class of linear spaces.

Let $\mathscr{X}$, $\mathscr{Y}$, and $\mathscr{Z}$ be finite-dimensional Euclidean spaces,[10] their dimensionalities being respectively n_x, n_y, n_z.

In each space we define certain ordering relations. If $v' = (v'_1, \cdots, v'_{n_v})$, $v'' = (v''_1, \cdots, v''_{n_v})$ are two vectors in the space $\mathscr{V}$ (where $\mathscr{V}$ may be $\mathscr{X}$, $\mathscr{Y}$, or $\mathscr{Z}$), we write

$$v' \geqq v'' \text{ to mean } v'_i \geqq v''_i \text{ for } i = 1, 2, \cdots, n_v \,,$$
$$v' \geq v'' \text{ to mean } v' \geqq v'' \text{ and } v'' \ngeqq v' \text{ (i.e., } v' \neq v'') \,,$$
$$v' > v'' \text{ to mean } v'_i > v''_i \text{ for } i = 1, 2, \cdots, n_v \,.$$

The origins of the three spaces treated are denoted by $0_x, 0_y, 0_z$, but the subscripts are omitted where no danger of confusion seems to exist.

Given a set Y in the space $\mathscr{Y}$ of desirables (Y is the "attainable" set, cf. [27], p. 47), we define its *maximal* (= "efficient") subset $\hat{Y}$ by the condition

$$\hat{Y} = \{y \in Y : \; y' \in Y, y' \geqq y \text{ imply } y' \leqq y\} \,.$$

I.e., y is a maximal (= efficient) element of Y if, and only if, $y' \geq y$ does not hold for any element y' of Y.

The set Y, however, is given indirectly, since our decision variable is x, not y. Thus there are given two (single-valued) functional relations[11] f, g with a common domain in $\mathscr{X}$ and ranges in $\mathscr{Y}$ and $\mathscr{Z}$ respectively and

$$Y = f[P_x \cap g^{-1}(P_z)]$$

where P_x and P_z are the respective non-negative orthants of $\mathscr{X}$ and $\mathscr{Z}$. I.e., $y \in Y$ if, and only if, $y = f(x)$ with $x \geqq 0$ and $g(x) \geqq 0$.

Somewhat inaccurately we shall call an element x *maximal* when $f(x)$ is maximal and we write $\hat{X} = f^{-1}(\hat{Y})$.

The problem of finding necessary and sufficient conditions for the maximality of a point x in $\mathscr{X}$ is usually called the problem of *vectorial* maximization of $f(x)$ subject to the *constraints* $x \geqq 0$, $g(x) \geqq 0$. When $n_y = 1$, we are of course dealing with scalar maximization. Corresponding to a given maximization problem one may construct its *Lagrangian* (*expression*) given by

$$\Phi(x, z^* \,; y^* \,; f, g) = y^*[f(x)] + z^*[g(x)]$$

where

$$y^*(y) = y^{*\prime} y = \sum_{i=1}^{n_y} y_i^* y_i$$

[10] In the economic interpretation, $\mathscr{X}$ is the space of activity level vectors, $\mathscr{Y}$ the space of the vectors of desirables, $\mathscr{Z}$ that of resources.

[11] In line with the prevailing practice, we use f where Kuhn and Tucker use g and vice versa.

and

$$z^*(z) = z^{*\prime}z = \sum_{i=1}^{n_z} z_i^* z_i .$$

(In matrix and vector notation A' is the transpose of A. Depending on the content, we omit some or all of the detail following the symbol Φ.)

We say that Φ has a *non-negative saddle-point* at $(x_0, z_0^* ; y_0^*)$ if and only if

$$x_0 \geqq 0, \qquad z_0^* \geqq 0, \qquad y_0^* > 0,$$

and

$$\Phi(x, z_0^* ; y_0^*) \leqq \Phi(x_0, z_0^* ; y_0^*) \leqq \Phi(x_0, z^* ; y_0^*) ,$$

for all $x \geqq 0$ and all $z^* \geqq 0$.

Φ is said to have a *non-negative quasi-saddle-point* at $(x_0, z_0^* ; y_0^*)$ if and only if

$$x_0 \geqq 0, \qquad z_0^* \geqq 0, \qquad y_0^* > 0 ,$$

and

$$\begin{cases} \Phi_x^0 \leqq 0, & \Phi_x^{0\prime} x_0 = 0, \\ \Phi_{z^*}^0 \geqq 0, & \Phi_{z^*}^{0\prime} z_0^* = 0 . \end{cases}$$

[Here $\Phi_x^0 = <\partial\Phi/\partial x_1, \cdots, \partial\Phi/\partial x_{n_x}>$ (see fn. 1, p. 2, for this notation) with all derivatives evaluated at $(x_0, z_0^* ; y_0^*)$; $\Phi_z^{0*} = <\partial\Phi/\partial z_1^*, \cdots, \partial\Phi/\partial z_{n_z}^*>$ with all derivatives evaluated at $(x_0, z_0^* ; y_0^*)$.] It was shown by Kuhn and Tucker ([31], Lemma 1) that a non-negative saddle-point is always a non-negative quasi-saddle-point. The converse is false.

In order to state the main results of Kuhn and Tucker we need three concepts:

a) a function f, with convex domain D in $\mathscr{X}$ and range in $\mathscr{Y}$, is said to be *concave* if, and only if, for any x', x'' in D and any $0 < \theta < 1$ the inequality $(1-\theta)f(x') + \theta f(x'') \leqq f[(1-\theta)x' + \theta x'']$ holds;

b) the function g, with domain in $\mathscr{X}$ and range in $\mathscr{Y}$, is said to be *regular* if and only if the "constraint qualification" (cf. Kuhn and Tucker [31], p. 483) is satisfied;

c) x_0 is *properly maximal* if it is a proper solution of the vector maximum problem in the sense of Kuhn and Tucker ([31], p. 488).[12]

The following are results of interest:

1. (Kuhn and Tucker [31], Theorem 4.) Let x_0 be properly maximal, f and g differentiable, and g regular for $x \geqq 0$. Then there exist y_0^*, z_0^* such that $\Phi(x, z^* ; y^*)$ has a non-negative quasi-saddle-point at $(x_0, z_0^* ; y_0^*)$. (Note: For $n_y = 1$, the term "properly" may be omitted and Theorem 4 becomes Theorem 1 of Kuhn and Tucker.)

[12] Let $f(x_0)$ be properly maximal whenever x_0 is. Then the result of Arrow, Barankin, and Blackwell [1] seems to show that, at least when $Y = f[P_x \cap g^{-1}(P_z)]$ is closed and convex, the set of properly maximal y's is dense in the set of maximal points.

2. (Kuhn and Tucker [31], Theorem 6.) Let both f and g be differentiable and concave, and g regular for $x \geqq 0$. Then x_0 is properly maximal if and only if there exist y_0^*, z_0^* such that $\Phi(x, z^* ; y^*)$ has a non-negative saddle-point at $(x_0, z_0^* ; y_0^*)$. (Note : For $n_y = 1$, the term "properly" may be omitted and Theorem 6 becomes Theorem 3 of Kuhn and Tucker.)

The following comments may be found helpful in following the later sections of this paper.

1. The "if" part of Theorem 6 fails to hold when $y^* \geq 0$ instead of the stronger $y^* > 0$ which is postulated. This raises a difficulty in generalizing to linear (or even Banach) spaces, since in some of them a $y^* > 0$ may not exist.

2. The "if" part of Theorem 6 remains valid when the assumptions of differentiability and concavity of f and g and regularity of g are abandoned.[13]

3. The "only if" part of Theorem 6 depends on Theorem 4 and the concavity of f and g.

4. The proof of Theorem 4 consists in "scalarizing" the problem by means of an appropriate $y^* > 0$ (which will exist if x_0 is properly maximal) and then using Kuhn-Tucker Theorem 1 covering the case $n_y = 1$.

5. The crucial step in the proof of the Kuhn-Tucker Theorem 1 involves the use of the Minkowski-Farkas Lemma which states that if, A being an $m \times n$ matrix,

$$Ax \geqq 0 \quad \text{implies } b'x \geqq 0 \text{ for all } x\,,$$

then there exists $t \geqq 0$ such that $b = A't$. (Cf. [31], p. 484.) Thus in attempting to generalize the results of Kuhn and Tucker the success hinges on finding the conditions under which the linear topological space counterpart of the Minkowski-Farkas proposition is valid.

6. The relationship of the present chapter to the results of Kuhn and Tucker is similar to that of Goldstine's paper [15] to, say, Bliss's discussion in [4], p. 210 ff. Goldstine treats the case of constraints in the form of equalities and imposes requirements strict enough to imply the existence of unique Lagrangian functionals ("multipliers"). Some of these results (in the "relaxed" form where uniqueness need not be present) are special cases of the theorems obtained in the present chapter.

7. Slater [41], assumes f and g to be continuous and postulates that they have a property (which we shall call "almost concavity")[14] implied by (but not implying) the concavity of both f and g; neither f nor g

[13] This suggests itself in reading Slater [41], p. 11

[14] Suppose that, for some $y^* \geqq 0$, $z^* \geqq 0$, $y^*[f(x^1)] + z^*[g(x^1)] = y^*[f(x^2)] + z^*[g(x^2)]$. Then "almost concavity" of (f, g) requires that $y^*[f(x)] + z^*[g(x)] \geqq y^*[f(x^i)] + z^*[g(x^i)]$ for all x on the segment joining x^1, x^2.

is assumed differentiable; instead of requiring that g be regular, it is required that, for some $x_1 \geqq 0$, $g(x_1) > 0$. (When this is so, we shall call g *Slater-regular.*) If by a *Slater-maximal* element of $Y = f[P_x \cap g^{-1}(P_z)]$ is meant a $y_0 \in Y$ such that $y' > y_0$ for no $y' \in Y$, and if x_0 is called Slater-maximal when $f(x_0)$ is Slater-maximal, then Slater's main result (Slater [41], Theorem 3) may be stated as follows:

Let f and g be continuous and almost concave and let g be Slater-regular. Then x_0 is Slater-maximal if, and only if, there exist

$$y_0^* \geq 0, z_0^* \geqq 0 \text{ such that}$$

$$\Phi(x, z_0^* ; y_0^*) \leq \Phi(x_0, z_0^* ; y_0^*) \leq \Phi(x_0, z^* ; y_0^*)$$

for all $x \geqq 0$ and all $z^* \geqq 0$.

It may be noted that the concept of Slater-maximality is weaker than that of maximality (as previously defined) and that it makes the "if" part of the theorem valid even though $y_0^* > 0$ is not required.

If one wanted to substitute "maximal" for "Slater-maximal" in Slater's Theorem 3, it is clear from known examples that one would have to require x_0 to be properly maximal and not merely maximal, as well as to specify that $y_0^* > 0$.

Slater's Theorem 3 is, of course, a counterpart of the Kuhn and Tucker Theorem 6. In the special case $n_y = 1$ the two concepts of maximality coincide; also $y_0^* \geq 0$ becomes equivalent to $y_0^* > 0$. Hence in this case the Slater result differs from that of Kuhn and Tucker only with regard to the hypotheses, since the assertion is precisely the same.

II. Notation, Terminology, and Some Fundamental Lemmas

II.1.0. This chapter deals with problems arising in spaces here called *linear topological spaces.* These spaces have both an algebraic structure (they are linear systems, i.e., sets of vectors, with vector addition and scalar multiplication) and a topological structure (they are topological spaces), and, furthermore, the two structures are related by the requirement that each of the algebraic operations be a (jointly) continuous function of its two arguments.[15]

The concept of a linear topological space, to be introduced more formally below, is a natural generalization of the properties of the finite-dimensional space (the real line being the simplest case) in its

[15] The definition of a linear topological space used in this paper is exactly the same as that used in Bourbaki [7], p. 1, for the term "espace vectoriel topologique." Our concept of a linear topological space is, therefore, broader than, e.g., that used by Bourgin [9], where the additional assumption is made that the space satisfies the Hausdorff separation axiom (i.e., that it is a T_2 space).

On the other hand, there are authors (e.g., Hille [20]) who use a concept broader than ours by relaxing slightly the nature of the continuity requirement for the algebraic operations, the continuity being required in each argument separately, but not necessarily jointly. Many of our results remain valid for this broader class of spaces.

customary Euclidean distance (metric) topology. Since any linear system must contain all scalar multiples of all its elements, the scalars used in a linear system being real numbers, the set of all integers (or even the set of all rational numbers) is not a linear system, hence not a linear topological space. From the economist's viewpoint this rules out applications involving indivisibilities.

II.1.1. *Linear topological spaces.* A linear topological space is both a linear system and a topological space. To avoid ambiguities, and for the sake of completeness, we supply some of the standard information concerning these concepts.

II.1.1.1. *Linear systems.* What we call a linear system is a purely algebraic concept. A fuller label would be "real linear system" since the scalars used are the reals. (Banach uses the term "linear space," Bourbaki "vector space"; our usage of the term "linear system" agrees with Hille's.) We shall find it convenient to refer to the elements of a linear system as *vectors.*

Since a linear system is an *additive group,* we start by defining the latter. A set $\mathscr{X}$ is called an additive group if it satisfies the following conditions:

1. With each pair (x', x'') of elements of $\mathscr{X}$ is associated a unique element x of $\mathscr{X}$; x is called the *sum* of x' with x'' and this is written as $x = x' + x''$.

2. Addition is associative; i.e., given any three elements x', x'', x''' of $\mathscr{X}$, $x' + (x'' + x''') = (x' + x'') + x'''$.

3. There is in $\mathscr{X}$ an element (the identity element of addition, later called the origin) denoted by 0_x (or, more simply, by 0) such that $x + 0_x = 0_x + x = x$ for every element x of $\mathscr{X}$.

4. To each element x of $\mathscr{X}$ corresponds uniquely an element $-x$ (called the negative of x) such that $x + (-x) = 0_x$. [Subtraction is defined by the relation $x' - x'' = x' + (-x'')$.] The foregoing conditions imply that the law of cancellation holds, i.e., that

$$x' + x = x'' + x \text{ implies } x' = x''$$

for any three elements x', x'', x of the group.

An additive group is called *commutative* (*Abelian*) if it satisfies the following additional condition:

5. $x' + x'' = x'' + x'$ for any two elements of the group.

A linear system is a commutative additive group in which there is further an operation of *scalar multiplication* (by reals, which we shall often call scalars). I.e.,

6. With each pair (α, x) where α is a scalar (real) and x a vector (an element of $\mathscr{X}$), there is associated a unique vector x', called their scalar product; this is written as $x' = \alpha \cdot x$. [Scalar multiplication is

commutative, i.e., $\alpha \cdot x = x \cdot \alpha$; the multiplication symbol $(\cdot)$ is often omitted.]

7. Scalar multiplication is distributive with regard to both scalars and vectors, i.e.,

$$(\alpha' + \alpha'')x = \alpha' x + \alpha'' x$$

and

$$\alpha(x' + x'') = \alpha x' + \alpha x''$$

for all selections of the scalars and vectors.

8. Scalar multiplication is associative, i.e., $\alpha'(\alpha'' x) = (\alpha' \alpha'')x$ for all selections of scalars and vectors.

9. The number one is the identity element of scalar multiplication, i.e., $1 \cdot x = x$ for all vectors x.

The preceding conditions imply that $(-1) \cdot x = -x$ and $0 \cdot x = 0_x$. [The last equation is an example of a situation where both the number 0 (zero) and the vector 0_x (origin) appear together. This is sometimes written simply as $0 \cdot x = 0$ and one must infer from the context that 0 denotes a scalar on the left and a vector on the right.]

Definition. A *linear system* is a set satisfying condition 1–9 above, i.e., an additive commutative group with scalar multiplication which is commutative, distributive, and associative, with reals as scalars and 1 as the identity element of scalar multiplication.

Algebraic set operations. Let X, X', X'' be subsets of a linear system and α a scalar (a real number). We write

$$\alpha X = \{\alpha x : x \in X\} ,$$
$$X' + X'' = \{x' + x'' : x' \in X', x'' \in X''\} ,$$
$$X' - X'' = \{x' - x'' : x' \in X', x'' \in X''\} .$$

Also,

$$-X = (-1)X = \{-x : x \in X\} .$$

These algebraic operations must be distinguished from the set-theoretic operations of union and difference. The union of two sets X' and X'' is written as $X' \cup X''$; the set-theoretic difference (i.e., the set of all elements that are in X' but not in X'') is written as $X' \sim X''$. The complement of X (with respect to X') is the difference $X' \sim X$.

We should also note that the algebraic operations do not have some of the properties suggested by the symbolism; e.g., it need not be true that $X + X = 2X$.

Some geometric terms. Given two vectors x', x'', the set $\{\lambda x' + (1 - \lambda)x'' : 0 \leq \lambda \leq 1\}$ is called the *segment joining* x' and x''. A set is called *convex* if with any two points x', x'' it also contains all points of the segment joining them. If $-X = X$, the set X is called *symmetric* (with respect to the origin). X is said to be *star-shaped from* the point x if, with any point x', it also contains the segment joining x and x'.

A subset X of the linear system $\mathscr{X}$ is called *absorbing* if, given any point x in the system $\mathscr{X}$, there is a point x' in the set X and a positive real number λ such that $x = \lambda x'$.[16]

II.1.1.2. *Topological spaces.* To define a topological space[17] it is convenient to start by introducing the concept of "a topology." A collection S of subsets of a given set A is called *a topology for* A if it satisfies the following conditions: (1) A is an element of S and so is the empty set ϕ; (2) the intersection of any two sets belonging to S belongs to S; (3) the union of the members of any (possibly infinite) sub-collection of S belongs to S. The subsets of A which belong to S are called open (relative to S, or in S). The union of all open sets contained in a given set is called its *interior*.

We *topologize* a set by selecting a topology for it. Any set can be topologized, for the two-element collection $\{\phi, A\}$ is a topology for A, i.e., it satisfies the above three conditions; such a two-element topology will be referred to as the *coarse* topology for A; it is sometimes called in the literature the indiscrete or trivial topology. On the other hand, the power set (sometimes written 2^A) of A, i.e., the set of all subsets of A, is also a topology for A, to be called the *fine* (often called discrete) topology for A. When A has two or more elements, the two topologies differ; for instance, one-element sets are open in the fine topology, but not in the coarse topology. Given two topologies for a set A, we call S' *finer than* S'' (and S'' *coarser than* S') if S'' is a proper subset of S', i.e., if every set open in S'' is also open in S' and there are some sets open in S' that are not open in S''. (Two topologies are non-comparable with respect to fineness when neither is a subset of the other.) Clearly, the fine topology is the finest topology possible, while the coarse topology is the coarsest topology possible. In most cases of applied interest, we deal with topologies that are somewhere between the fine and the coarse topologies.

Denote by $R^\sharp$ the linear system whose elements are all real numbers, i.e., the "real line." Its so-called "natural" topology is defined as consisting of all subsets B of $R^\sharp$ characterized by the following property: each element of B must belong to an "open interval" which is a subset of B. (An open interval is defined as the set of all numbers greater than some fixed number and less than another fixed number; an open interval is an open set in the natural topology, but there are open sets which are not open intervals, e.g., the set of all numbers other than zero.) A set is *closed* (in a specified topology) if its complement (with

---

[16] This usage of the term *absorbing*, as well as some of the subsequent formulation, is due to the author's exposure to lectures by Professor Hans Radstrom of the Royal Institute of Technology in Stockholm, to whom the author is also indebted for clarification on certain properties of linear spaces.

[17] See, for instance, Kelley [23].

respect to A) is open. A set may be both open and closed (e.g., the empty set and A), or it may be neither open nor closed (e.g., one-element sets in the coarse topology when A has two or more elements). What is ordinarily called a closed interval (i.e., one including its end-points) is a closed set in the natural topology of the real line. An interval including only one of its end-points is neither open nor closed in the natural topology. The *closure* of a set is the intersection of all closed sets containing it.

A *topological space* is defined formally as an ordered pair (A, S) where S is a topology for A. Often, when the topologization of A is understood, we refer to A itself as a topological space.

Let (A, S) be a topological space, B a subset of A, x an element of B. B is called a *neighborhood of* x (with respect to the topology S) if it contains a subset C which is open (with respect to the topology S). Obviously, any open set containing x is a neighborhood of x, but a neighborhood need not be open. (Some authors use a narrower concept of a neighborhood and require that it be an open set.) The collection of all neighborhoods of a given point x is called the *complete neighborhood system* for the point x. For instance, in the fine topology all sets of which x is an element constitute a complete neighborhood system for x; in particular, the one-element set consisting of x alone is a neighborhood of x. In the coarse topology, on the other hand, a point has only one neighborhood, namely, the set A. A topological space (A, S) is called a *Hausdorff* (topological) *space* if any two distinct points of the space have disjoint neighborhoods. Thus the fine topology is Hausdorff, but the coarse topology (when there are two or more elements in the space) is not. $R^\#$ in its natural topology is Hausdorff, for we can use as disjoint neighborhoods open intervals centered on the two points, the width of the intervals being less than half the distance of the two points. Most spaces of applied interest are Hausdorff.

Sometimes we are interested in certain subsets of the complete neighborhood system of a point. A subset F of the complete neighborhood system of a point is called a *fundamental system* of neighborhoods of the point if every neighborhood of the point contains a neighborhood belonging to the set F; if we call the neighborhoods belonging to F fundamental, we can say that every neighborhood of a point must contain a fundamental neighborhood of that point.

It is often convenient to define a topology indirectly, viz., by assigning to each point a of a set A a (non-empty) collection F_a and declaring it to be a fundamental neighborhood system of a. The complete neighborhood system of a is then defined as the collection G_a of subsets of A, each of which contains a fundamental set (i.e., a set belonging to F_a); finally a subset A' of A is declared as open if and only if it is a neighborhood of all of its points.

In order for such a procedure to result in a topology for A, the collection F_a must, of course, satisfy certain conditions. First, naturally, each set belonging to F_a must contain a, otherwise it would not qualify as a neighborhood of a; hence every set belonging to F_a is non-empty. Second, the collection F_a must satisfy the following finite intersection requirement: the intersection of any two sets belonging to F_a must contain a set belonging to F_a. (The intersection itself need not belong to F_a.) A non-empty collection F_a of sets each of which contains a and satisfying the preceding finite intersection requirement will be called a *neighborhood base* at a. We shall see that it is convenient to discuss the properties of linear topological spaces in terms of fundamental neighborhood systems and neighborhood bases.

It was mentioned earlier that a linear topological space is a set which is both a linear system and a topological space, with certain continuity conditions imposed on the algebraic operations of addition and scalar multiplication. To be able to state these conditions, we must introduce the concept of continuity.

Let A and B be two sets and let f denote a functional relation whose domain is A and whose range is B, i.e., which associates with each element a in A a unique element $b = f(a)$ in B. Given a subset A' of A, we define the *image* of A' by f as the set $\{f(a) \in B : a \in A'\}$. Given a subset B' in B, we define as the *inverse image* of B' by f the set $\{a \in A : f(a) \in B'\}$. The image of A' by f is denoted by $f(A')$; the inverse image of B' by f is denoted by $f^{-1}(B')$.

Now let us topologize A and B, with S denoting the topology for A, T the topology for B. The function f is said to be *continuous* if the inverse image $f^{-1}(B')$ of every set B' open in T is itself open (in S). It is important to realize that continuity depends not only on the nature of the function, but also on the manner in which the two spaces have been topologized. Thus if S is fine, any function on A is continuous. Similarly, the constant function (which has the same value for all elements of A) is continuous for any topology, since the inverse image of the one-element set (consisting of the constant f value) is the whole space A. Now suppose $A = B$ and f is the identity function, i.e., $f(a) = a$ for all a in A. (When $A = B = R^{\sharp}$, the identity function is represented by the positively inclined 45° straight line through the origin.) Whether f is continuous depends on the topologization of A and B. If A and B are given the same topologies (i.e., $S = T$), then f is continuous, since $f^{-1}(B') = B'$ for all B'. But, even though the sets A and B are the same, their topologies may differ. For instance, let $A = B = R^{\sharp}$, with f still the identity function, and let B have the natural topology while to A we give the coarse topology. Let B' be a finite open interval on the B-axis which is an open set in the natural topology. The inverse image of B' is the same interval, taken on the

A-axis ; but, in the coarse topology of the real line, a non-empty proper subset of the line is not open ; hence $f^{-1}(B')$ is not open (in S) ; hence with this topologization the identity function is not continuous.

One more topological concept is essential in discussing the properties of linear topological spaces. As was indicated earlier, the continuity of the operations in such a space is joint continuity in the two arguments. To clarify the point, consider the operation of scalar multiplication. We may write, for a scalar (real) α and a vector x, $\alpha x = g(\alpha, x)$, so that the scalar product may be viewed as a function of two variables α and x. In order to explain what is meant by the *joint* continuity of g in the two variables, we restate the situation as follows : First, we write $\alpha x = f((\alpha, x))$, i.e., we now view the scalar product as a function whose domain is the set of ordered pairs (α, x), i.e., the Cartesian product $R^\sharp \times \mathscr{X}$, while the range, of course, is the set $\mathscr{X}$. To apply the above definition of continuity, we must topologize the product set $R^\sharp \times \mathscr{X}$. Similarly, addition may be viewed as a function on the Cartesian product $\mathscr{X} \times \mathscr{X}$ with the range in $\mathscr{X}$. Here, again, the product set must be topologized.

In both cases, the appropriate topologization (i.e., the one implicit in the definition of a linear topological space) is the so-called product topology which we shall now define.

Let there be two topological spaces (A, S) and (B, T) and let $C = A \times B$. The *product topology*, about to be defined, will be denoted by $P[S, T]$; hence the topological product space is written as $(C, P[S, T])$. To define the product topology, it is enough to characterize the open sets of C. A set C' is open in the $P[S, T]$ topology if and only if every point c' of C' is a member of a set of the form $A' \times B'$ where A' is open (in S), B' is open (in T), and $A' \times B' \subseteq C'$. As an illustration, if $A = B = R^\sharp$ (the real line), so that C is the Cartesian plane (the set of all ordered pairs of real numbers), and $S = T =$ the natural topology for the reals, then the product topology $P[S, T]$ is the usual Euclidean topology of the plane where a set is open if every one of its points can be enclosed in a disk of positive radius wholly belonging to the set.

Another example is obtained if C is again $R^\sharp \times R^\sharp$ but $S = T =$ coarse topology ; here the product topology is the coarse topology of the plane. Similarly, the topological product of fine topological spaces is fine. An interesting case is obtained if we take $A = B = R^\sharp$ but with different topologies on the two spaces, viz., A coarse and B natural. In the product topology a one-element set (a point) is not closed and the topology is not Hausdorff : every open set must contain an infinite "strip" (of positive width) parallel to the A-axis, and any neighborhood containing $(a', 0)$ must contain all other points of the form $(a, 0)$.

II.1.1.3. *Linear topological spaces.* We now have a sufficient vocabulary to provide a precise definition of a linear topological space.

Definition. Let X be a linear system and S a topology for X. (X, S) is said to be a *linear topological space* if (1) addition is continuous in the product topology $P[S, S]$, and (2) scalar multiplication is continuous in the product topology $P[N, S]$, where N denotes the natural topology of the real line.

To illustrate, let X be the real line $R^{\sharp}$. As might have been expected, $(R^{\sharp}, N)$, i.e., the real line in its natural topology, is a linear topological space ; the real line in its coarse topology also turns out to be a linear topological space ; the real line in its fine (discrete) topology is not a linear topological space, although it is a linear system and a topological space. Hence the continuity conditions in the above definition are not automatically satisfied for every linear system which is also a topological space.

The verification of the continuity properties of the algebraic operations directly from the topology of the space can be quite awkward ; the situation becomes much more transparent when the properties of linear topological spaces are stated in terms of neighborhood systems. Furthermore, we may confine ourselves to the discussion of the neighborhood system of the origin 0_x ; if G_{0_x} is a collection of neighborhoods of the origin, the corresponding collection of neighborhoods of any point x is given by $\{x\} + G_{0_x} = G_x$. Thus the topology of a linear topological space may be defined in terms of a fundamental neighborhood system of the origin, the corresponding complete system of neighborhoods then being defined as the collection of sets containing a fundamental set, and finally, an open set being defined as a set which is a neighborhood of each of its elements. But to follow such procedures we must know what types of neighborhoods one may encounter in linear topological spaces.

The answer to this question is contained in a theorem we shall state in a moment. To simplify this statement, we shall coin an *ad hoc* term ; we shall call a non-empty family G of sets *acceptable* if it satisfies the following conditions : (1) if V is in the family G, then the family must also contain a set W such that $W + W \subseteq V$; (2) every set in the family is symmetric, i.e., $V = -V$ for all V in G ; (3) every set of the family contains the origin, i.e., $0_x \in V$ for each V in G ; (4) every set of the family is star-shaped from the origin, i.e., if a point x is in V, then so is the whole segment joining x to the origin ; (5) every set in the family is absorbing, i.e., if x is any point of the space X and V is a set in the family G, then V has an element x' such that $x = \lambda x'$ for some positive number λ ; (6) the family G is invariant under homotheties (from the origin), i.e., if V is a set in the family and α a real number different from zero, then the set αV is also in the family.

THEOREM (Bourbaki [7], Prop. 5, p. 7).

A. *If (X, S) is a linear topological space, then there exists an accepta-*

ble (i.e., satisfying conditions 1–6 above) fundamental neighborhood system of the origin.

B. *In a linear system X, let F be a neighborhood base at 0_x (i.e., a non-empty collection of sets each containing the origin and such that an intersection of any two members of the collection contains a member of the collection) and suppose that F is acceptable (i.e., satisfies conditions 1–6 above). Then there exists a topology (and only one topology) such that F is the fundamental neighborhood system of the origin in that topology. In this topology, X is a linear topological space.*

[The above six conditions are somewhat redundant, since 3 follows from the others. We have chosen this form, however, partly in order to show that a linear topological space is a linear topological group, i.e., an additive Abelian group with a topology in which addition and subtraction are both continuous (jointly in the two arguments). Conditions 1–3 above are precisely those characterizing a fundamental neighborhood system of the origin (identity element of addition) of a topological group. Cf. Bourbaki [6], p. 6.]

We can now verify our statements about the various topologizations of the real line. Thus for its coarse topology the neighborhood base at the origin consists of the single one-element set $\{R^\#\}$. It may be seen that this base is acceptable (i.e., satisfies conditions 1–6). For the natural topology of the real line we use the family of all open intervals centered on 0; again, the family is acceptable. Hence $R^\#$ is indeed a linear topological space in both the coarse and the natural topology. But the situation is different when $R^\#$ is given its fine (discrete) topology. Since the one-element set consisting of the origin is open in this topology, it is a neighborhood and hence any fundamental neighborhood system of the origin must contain $\{0\}$. However, $\{0\}$ is not an absorbing set (i.e., condition 6 of acceptability is violated) and hence $R^\#$ in its fine topology does not have an acceptable fundamental system; hence it is not a linear topological space.

A linear topological space satisfying the Hausdorff separation axiom (distinct points have disjoint neighborhoods) is called a *Hausdorff linear* (topological) *space*. It will be noted that the real line, depending on its topologization, may fail to be a linear topological space (in the fine topology), it may be a Hausdorff linear space (in the natural topology), or it may be a non-Hausdorff linear topological space (in the coarse topology). Euclidean finite-dimensional spaces are all Hausdorff linear.

A linear topological space may or may not possess a fundamental neighborhood system of the origin consisting of convex neighborhoods. If, in a linear topological space, there exists such a fundamental system consisting of convex neighborhoods (i.e., every fundamental neighborhood is convex), the space is called a *locally convex* (linear topological) *space*. We may note that the real line forms a locally convex space in both its natural and its coarse topology. This is not accidental: according to a theorem due to Tychonoff (cf. [43], p. 769) every finite-dimensional linear topological space is locally convex. [A linear system is finite-dimensional, say n-dimensional, if there exists a finite set of elements $x_1, x_2, \cdots, x_n$ such that every element x of X can be written in the form $x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n$ where the α_i are scalars (reals).] Furthermore, if a space is finite-dimensional and Hausdorff linear, then its topology is Euclidean.

Most linear topological spaces occurring in applications are locally convex, but there do exist linear topological spaces that are not locally convex. Tychonoff's example (*loc. cit.*, p. 768) is the space denoted by $l_{1/2}$ consisting of all the infinite sequences $x = (x_1, x_2, \cdots)$ of numbers x_i such that

$$\sum_{i=1}^{\infty} |x_i|^{1/2} < \infty .$$

The space $l_{1/2}$ is topologized in the following manner. We construct a fundamental neighborhood system of the origin consisting of sets of the form

$$\{x : (\sum_{i=1}^{\infty} |x_i|^{1/2})^2 < \rho\}$$

with ρ varying over positive reals. It was shown by Tychonoff that (a) this space is a linear topological space, but (b) there does not exist a fundamental neighborhood system of the origin consisting of convex sets. (The fact that the fundamental system as given does not consist of convex sets is by itself inconclusive, since there might exist another fundamental system consisting of convex sets and yielding the same topology.) Hence $l_{1/2}$ is a non-locally convex linear topological space. On the other hand, the space is Hausdorff linear; this can be shown by utilizing the fact that the function

$$q(x) = (\sum_{i=1}^{\infty} |x_i|^{1/2})^2$$

satisfies the inequality $q(x' + x'') \leqq 2[q(x') + q(x'')]$. Let x, y be two distinct elements of the space such that $q(x - y) = a$, where a is necessarily positive. Now select a neighborhood of x to be the set of points such that $q(x - z')$ is less than $a/6$; similarly, select a neighborhood of y consisting of points such that $q(y - z'')$ is less than $a/6$. Suppose that there is a point z belonging to both neighborhoods. Then, in virtue of the inequality, $q(x - y) \leqq 2[a/6 + a/6]$ which contradicts the assumption made. Hence the space does satisfy the Hausdroff separation axiom.

We have had examples of spaces that are both Hausdorff and locally convex (real line in its natural topology), Hausdorff but not locally convex (the space $l_{1/2}$), locally convex but not Hausdorff (real line in its coarse topology). To complete the picture let us point out that the topological product of $l_{1/2}$ with the real line in its coarse topology is neither Hausdorff nor locally convex, although it is a linear topological space.

A fundamental neighborhood system of the origin in a locally convex space can always be defined (Bourbaki [7], pp. 95–96) by means of a set of functions called *semi-norms*. A semi-norm is defined as a finite real-valued function p on a linear system satisfying the following two requirements: (1) for any scalar α and any vector x, $p(\alpha x) = |\alpha| \cdot p(x)$; (2) for any two vectors, $p(x' + x'') \leqq p(x') + p(x'')$. It follows that $p(0_x) = 0$ and $p(x)$ is always non-negative. A semi-norm is called a *norm* if it has the further property (3) $p(x) = 0$ only if $x = 0_x$. Hence, for a norm, $p(x) = 0$ if, and only if, $x = 0_x$. The norm of x is usually written $\|x\|$. For instance, let X be the linear system consisting of all ordered pairs (x_1, x_2) of real numbers (the Cartesian plane). Then $p^*(x) = |x_1| + |x_2|$ is a norm, while $p^{**}(x) = |x_1|$ is a semi-norm, but not a norm.

If p is a semi-norm, the set $\{x : p(x) < \lambda\} = [p\,;\lambda]$ is called an *open strip* (of width 2λ). Denote by $[p]$ the set of all open strips $[p\,;\lambda]$, obtained by keeping p fixed while λ varies over the positive reals. Given a set P of semi-norms, we shall denote by $[P]$ the set of all open strips, obtained by taking all the sets $[p\,;\lambda]$ with λ varying over the positive reals and p over P. Finally, let $F(P)$ denote the set of all finite intersections of members of $[P]$. The set $F(P)$ is a fundamental neighborhood system of the origin, as can be verified from the "acceptability" conditions 1–6 above; also, the elements of $F(P)$ are convex sets, since all strips are convex and so are their intersections. Hence $F(P)$ defines a locally convex topology for the linear system on which the semi-norms are defined. Conversely (Bourbaki [7], p. 96, Prop. 4), every locally convex topology can be defined by a fundamental set $F(P)$ for a suitably chosen set of semi-norms P.

A *normed space* is a linear system where the fundamental neighborhood system of the origin consists of sets (open spheres) $S(\rho) = \{x : \|x\| < \rho\}$ where ρ is the *radius* of the *sphere*. (The spheres are centered at the origin.) The system consists of spheres with the radius varying over the positive reals, although a smaller system (e.g., with rational radii) would be sufficient. That a normed space is a locally convex linear space follows from the fact that the spheres constitute an "acceptable" family (i.e., satisfy conditions 1–6 above) and are convex sets. Also, a normed space is Hausdorff. The proof proceeds exactly as in the case of the space $l_{1/2}$ above, except that the relevant inequality does not have the factor 2 on the right-hand side.

A linear topological space is called *normable* if its topology can be defined by a norm as just indicated. From what has just been said it follows that a normable space must be locally convex Hausdorff. However, not every locally convex Hausdorff linear space is normable. Because of the convenience in dealing with normed spaces, it is of interest to know under what conditions a space is normable. In order to do so, we must introduce a new concept, that of a *bounded* subset of a linear topological space. A subset B is said to be bounded if, given *any* neighborhood V of the origin, there is a positive scalar λ such that $B \subseteq \lambda V$; this is expressed by saying that a bounded set is *absorbed* by every neighborhood.

We may now state Kolmogoroff's theorem on normability of linear topological spaces: a linear topological space is normable if and only if it is locally convex Hausdorff and there exists a bounded neighborhood of the origin.

The following is an example of a non-normable locally convex Hausdorff space. Its elements are all the infinite numerical sequences $x = (x_1, x_2, \cdots)$. The space is topologized by the set $P = \{p^1, p^2, \cdots\}$ of semi-norms where $p^d(x) = \max(|x_1|, |x_2|, \cdots, |x_d|)$. Its topology, being based on the family $F(P)$ as the fundamental neighborhood system of the origin, is necessarily a locally convex linear space. It is also Hausdorff because for each element x other than 0_x of the space there exists a norm p in P such that $p(x) \neq 0$. (Cf. Bourbaki [7], p. 97, Prop. 5.) Now if this space were normable, there would exist a bounded neighborhood of the origin; hence, by definition of a fundamental system, there would exist a bounded set of the family $F(P)$, since a subset of a bounded set is bounded. Hence to establish the non-normability, it is enough to show that no member of the family $F(P)$ is bounded. Now the members of the family $F(P)$ are formed by finite intersections of the open strips defined by the norms p^d. Hence it is true for each member of $F(P)$ that, starting with, say, the k-th component, the values of the components with subscripts $\geqq k$ are completely unrestricted. Now let V_k be a member of $F(P)$, with the components whose subscripts $\geqq k$ are unrestricted, while the components 1 through $k-1$ cannot exceed M $(0 < M < \infty)$ in absolute value. We show that V_k is not absorbed by a neighborhood V_{k+1}. This follows from the fact that in V_{k+1} the $(k+1)$th component is restricted, while in V_k it is not. Hence, no matter what $\lambda > 0$ we choose, there will be elements in V_k that are not in λV_{k+1}. Hence V_k is not bounded; but since V_k is a typical member of $F(P)$, no set in $F(P)$ is bounded. By the previous argument it follows that there is no bounded neighborhood of the origin, and hence the space is not normable.

Many spaces we deal with are normed; in particular, the finite-dimensional Euclidean spaces are normed. The norm of a point x in a Euclidean n-dimensional space can be defined in various ways. The *Euclidean* norm of x is defined as

$$\left(\sum_{i=1}^{n} x_i^2\right)^{1/2};$$

another norm (which results in the same topology) can be defined as $\max(|x_1|, |x_2|, \cdots, |x_n|)$.

The space of all infinite sequences $x = (x_1, x_2, \cdots)$ of numbers x_i with only finitely many components different from zero can be normed by defining

$$||x|| = \left(\sum_{i=1}^{\infty} x_i^2\right)^{1/2}.$$

In a normed space it is possible to define a *distance* function $d(x', x'') = ||x' - x''||$. In any space on which a distance function has been defined one can introduce a "metric" topology, by using as the fundamental neighborhood system for x_0 the (metric) spheres, i.e., the sets $\{x : d(x, x^0) < \rho\}$ with the radius ρ varying over positive reals. Because of the triangle inequality satisfied by the distance function a metric space is always a Hausdorff topological space, the proof being analogous to that sketched above for the normed and $l_{1/2}$ spaces.

Let $(x^1, x^2, \cdots)$ be an infinite sequence of points x^i in a metric space with a distance function d. The sequence is said to be a *Cauchy sequence* if, given $\varepsilon > 0$, there exists a positive integer N such that $d(x^m, x^n) < \varepsilon$ provided both m and n are greater than N. A sequence $(x^1, x^2, \cdots)$ is said to converge to x^0 if, for any $\varepsilon > 0$, there exists a positive integer N such that $d(x^n, x^0) < \varepsilon$ provided n is greater than N. A sequence is said to be *convergent* if it converges to some element x^0 of the space. It is known that every convergent sequence is a Cauchy sequence. On the other hand, there are spaces with non-convergent Cauchy sequences. One example of such a space is that of infinite sequences with only finitely many components different from zero. A space where every Cauchy sequence is convergent is called *complete*. The reals are complete in their natural topology, while the rationals with the same topology (i.e., defined by the Euclidean distance or norm) form a space that is not complete because there are sequences of reals converging to an irrational number. A normed space which is also complete is called a *Banach space*. Thus the reals (as well as all finite-dimensional Euclidean spaces) are Banach, but the above space of infinite sequences with only finitely many non-zero components is not Banach, though normed. The classic example of an infinite-dimensional Banach space is the space l_2 of all infinite numerical sequences $x = (x_1, x_2, \cdots)$ such that

$$\sum_{i=1}^{\infty} x_i^2 < \infty,$$

the norm being defined as the square root of the preceding infinite sum;

l_2 belongs to a sub-class of Banach spaces known as *Hilbert spaces*, where with each pair of elements it is possible to associate a number called their *inner product* $x' \cdot x''$, with $x' \cdot x''$ linear in each of its arguments, $x' \cdot x'' = x'' \cdot x'$, $x \cdot x$ always non-negative, and $x \cdot x = 0$ if and only if $x = 0_x$. In a space with such an inner product it is possible to define the norm of a vector as $\|x\| = (x \cdot x)^{1/2}$, and the resulting normed space is called Hilbert if it is Banach, i.e., if it is complete. (Some authors use somewhat different definitions of a Hilbert space.) According to this definition the Euclidean spaces are Hilbert, with the inner product defined as

$$\sum_{i=1}^{n} x_i' x_i'' .$$

The resulting norm is, of course, that corresponding to the Euclidean distance.

As an example of a Banach space which is not a Hilbert space we may take the space of all infinite bounded numerical sequences $x = (x_1, x_2, \cdots)$. The norm of this space is defined by $\|x\| = \sup(|x_1|, |x_2|, \cdots)$.

II.1.2. *Linear transformations.* A function T whose domain is a linear system $\mathscr{X}$ and the range a subset of a linear system $\mathscr{Y}$ is called a *linear transformation* on $\mathscr{X}$ into $\mathscr{Y}$ if it is additive and homogeneous, i.e., if

$$T(x' + x'') = T(x') + T(x'') \qquad \text{for all } x', x'' \text{ in } \mathscr{X} ,$$

and

$$T(\alpha x) = \alpha T(x) \qquad \text{for all real } \alpha \text{ and all } x \text{ in } \mathscr{X} .$$

If both spaces are linear topological, an additive continuous function is homogeneous (Hille [20], Theorem 2.6.1, p. 16). The converse, however, is not true; i.e., there are (in infinite dimensional spaces) linear transformations which are not continuous at any point (see Bourbaki [7], p. 93).

If both spaces are Banach, an additive function is continuous if and only if it is bounded (i.e., carries bounded sets into bounded sets). Hence, in Banach spaces, "linear bounded" as applied to transformations is synonymous with "linear continuous."

A linear transformation on $\mathscr{X}$ whose range is a subset of the reals (i.e., a homogeneous additive real-valued function on $\mathscr{X}$) is called a *linear functional* on $\mathscr{X}$. There are linear functionals on locally convex spaces that are not continuous at any point (Bourbaki [7], p. 93). Since a linear functional is both convex and concave, it follows that, even in locally convex spaces, a convex or concave function need not be continuous. This is of interest in connection with results of this chapter where only concavity, but not continuity, of a function is assumed, since it proves that the concavity assumption is less restrictive; the same remark applies to results where only linearity, but not continuity, of transformations is assumed.

II.1.3. *The conjugate space.* Let $\mathscr{X}$ be a linear topological space. The set $\mathscr{X}^*$ of all linear *continuous* (with respect to the natural topology of the reals) functionals on $\mathscr{X}$ is called the *conjugate* (adjoint, dual) space of $\mathscr{X}$. $\mathscr{X}^*$ is a linear system whose typical element will be written as x^*; the null element (origin) of $\mathscr{X}^*$ (i.e., the real-valued function on $\mathscr{X}$ whose value is zero for each element of $\mathscr{X}$) is denoted by 0_x^* or 0, as the occasion demands.

Two ways of topologizing the conjugate space are of particular interest. They are respectively labeled "strong" and "weak star," the latter usually being written "weak*."

In each case the topology is defined through a fundamental neighborhood system.

In the *strong* topology a fundamental neighborhood of 0_x^* is of the form

$$U(\varepsilon, B) = \{x^* \in \mathscr{X}^* : |x^*(x)| < \varepsilon \qquad \text{for all } x \in B\}$$

where B is a *bounded* set, taking all neighborhoods $U(\varepsilon, B)$ with ε varying over the positive reals and B over the class of *all bounded* sets in $\mathscr{X}$.

In the *weak* topology* a fundamental neighborhood of 0_x^* is also of the form

$$U(\varepsilon, B) = \{x^* \in \mathscr{X}^* : |x^*(x)| < \varepsilon \qquad \text{for all } x \in B\} ,$$

but B here is required to be a *finite* set; the fundamental system is again obtained by letting ε vary over positive reals and B over the class of *all finite* sets in $\mathscr{X}$.

Since every finite set is bounded, it follows that every weak* neighborhood is also a strong neighborhood, but there may be strong neighborhoods that are not weak* neighborhoods. It follows that the strong topology is at least as fine as, and possibly finer than, the weak* topology. I.e., every set open (resp. closed) in the weak* topology is also open (resp. closed) in the strong topology, but the converse need not be true.

In both topologies the conjugate space is a Hausdorff locally convex linear topological space (Bourbaki [8], pp. 16–19). Moreover, when the space $\mathscr{X}$ is normed, the conjugate space is *normable* in its *strong* topology, the norm of an element x^* of the conjugate space being defined by

$$\|x^*\| = \sup_{\|x\| \leq 1} |x^*(x)| .$$

In its strong (norm) topology, the conjugate of any normed space is complete, hence it is a Banach space.

In finite-dimensional Euclidean spaces, the strong and weak* topologies coincide. But in infinite-dimensional spaces the strong topology is, in most cases likely to be considered, actually finer than the weak* topology.

In particular, if $\mathscr{X}$ is an infinite-dimensional normed space, the strong topology is finer than the weak* topology. (Cf. Bourbaki [8], p. 111, where it is shown that the set of elements of norm one in the conjugate space is not closed in the weak* topology, although it is closed in the strong topology.)

II.1.4. *Separation by hyperplanes in linear topological spaces.*

II.1.4.1. Let $\mathscr{X}$ be a linear system. A subset of $\mathscr{X}$ is called *linear* if it is closed under the operations of addition and scalar multiplication. A translate of a linear set M, i.e., a set of the form $\{x_0\} + M$ where M is a linear set, is called a (linear) *variety*.[18] If M is a linear set such that M is a proper subset of $\mathscr{X}$ and there is no linear proper subset of $\mathscr{X}$ in which M is contained, M is called a *maximal* linear set. A translate of a maximal linear set is called a *maximal variety.*

With each maximal variety V one may associate a non-null (i.e., $\neq 0^*_x$) linear functional x^* on $\mathscr{X}$ and a real number α such that $V = \{x \in \mathscr{X} : x^*(x) = \alpha\}$. On the other hand, every pair (x^*, α) where x^* is a linear functional and α a real number defines a maximal variety.

If $\mathscr{X}$ is a linear topological space, a maximal variety may or may not be a closed set. We shall call a closed maximal variety a *hyperplane.* (Terminologies of various writers differ. In Bourbaki, hyperplane is synonymous with a maximal variety.) In a linear topological space $\mathscr{X}$, a maximal variety $V = \{x \in \mathscr{X} : x^*(x) = \alpha\}$, where x^* is a linear functional and α a real number, is closed if and only if x^* is continuous. I.e., a maximal variety is a hyperplane if and only if the functional defining the variety is continuous.

We may now state a theorem underlying a great many results concerning convex sets in linear topological spaces. The theorem is variously called the Hahn-Banach Theorem (geometric form) (cf. Bourbaki [7], p. 69) and the Bounding Plane Theorem.

THEOREM II.1. *Let $\mathscr{X}$ be a linear topological space, A an open convex (non-empty) subset of $\mathscr{X}$, and M a linear variety disjoint from A (i.e., $A \cap M = \phi$). Then there exists a hyperplane H containing M and disjoint from A (i.e., $M \subseteq H$ and $H \cap A = \phi$).*

Hence, under the hypotheses of the Theorem there exists a *continuous* linear functional x^* and a real α such that $x^*(x) = \alpha$ for all x in M and $x^*(x) < \alpha$ for all x in A.

In what follows we shall need the following

COROLLARY II.1. *Let $\mathscr{X}$ be a linear topological space and A a convex subset with non-empty interior. Then, for any point x_0 of $\mathscr{X}$ which is not in the interior of A, there exists a continuous linear functional x^*_0 such that $x^*_0(x) \leqq x^*_0(x_0)$ for all x in A.*

[18] In particular, every point of the space, viewed as a one-element set, is a linear variety.

The geometric interpretation of the preceding Corollary is that through every point not in the interior of A there is a hyperplane "bounding" the set A, provided A is convex and has a non-empty interior.

In certain contexts, however, we want a somewhat stronger separation property. Given a set A and a point x_0 outside the set, we are interested in the existence of a hyperplane such that A is wholly on one "side" of it (possibly touching H) while x_0 is on the other "side" (not touching H). I.e., we are looking for a continuous functional x_0^* such that

$$\sup_{x \in A} x_0^*(x) < x_0^*(x_0) \text{ .}$$

It is intuitively clear that we shall have to require that A be a closed convex set. But it turns out that restrictions must also be imposed on the nature of the linear topological space. The desired result follows from Prop. 4 in Bourbaki [7], p. 73. It was established by Mazur for Banach spaces and by Bourgin for Hausdorff locally convex spaces; we shall refer to it as the Mazur-Bourgin Theorem.

THEOREM II.2. (Mazur-Bourgin.) *Let $\mathscr{X}$ be a locally convex linear topological space, A a (non-empty) convex closed subset of $\mathscr{X}$, and x_0 a point outside A, i.e., $x_0 \notin A$. Then there exists a hyperplane "strictly separating" x_0 from A, i.e., there exists a continuous linear functional x_0^* such that the inequality*

$$\sup_{x \in A} x_0^*(x) < x_0^*(x_0) \tag{1}$$

holds.

Following Bourgin,[19] we shall refer to a set that can be "strictly separated" from points not in it as *regularly° convex*. Hence the preceding theorem states that in a locally convex space closed convex sets are regularly° convex. (Also, it is the case that a regularly° convex set is closed and convex.) It may be noted, however, that the class of spaces in which a closed convex set is regularly° convex is wider than that of locally convex spaces, as shown by Klee ([25], (10.1), p. 459). This is of interest since the regular° convexity of certain sets is a crucial property in several results of this chapter. If spaces in which closed convex sets are regularly° convex are called c-regular (as suggested by E. Michael, see Klee [26], p. 106), we may note here that many of the results of this chapter which presuppose local convexity of the space are valid for all c-regular spaces. However, this additional generality does not seem of serious applied interest in our problems.

On the other hand, we may in some cases wish to ensure the regular° convexity of certain sets without restricting ourselves to locally convex spaces. This can be accomplished by imposing an additional requirement

[19] In a slightly modified fashion: what we call regularly° convex (regularly circle-convex) he calls regularly $\mathscr{X}$ convex (where $\mathscr{X}$ is the underlying space).

on the nature of the set A, viz., that it possess a non-empty interior (see, for instance, Klee [25], Theorem 9.7, p. 456). However, the assumption of a non-empty interior rules out certain worth-while applications. Specifically, the sets in whose regular° convexity we are interested are those consisting of the vectors with non-negative coordinates (the non-negative cones); in a Euclidean space of finite dimension such a set (the non-negative orthant) does have an interior, but in infinite-dimensional spaces this is not always the case. In particular, for the l_p spaces $(p \geq 1)$, the non-negative cone has no interior points (cf. Klee [24], p. 771); in other spaces, such as the space (m) of infinite sequences, the non-negative cone does have interior points.

Let $\mathscr{X}$ be a linear topological space, $\mathscr{X}^*$ its conjugate space. Given any element x_0 of the space $\mathscr{X}$, we can define a functional f_{x_0} on the conjugate space $\mathscr{X}^*$ by the relation

$$f_{x_0}(x^*) = x^*(x_0) \qquad \text{for all } x^* \in \mathscr{X}^* .$$

It may be verified that f_{x_0} is additive and homogeneous, hence linear. Now it may be noted that f_{x_0} is a continuous functional on $\mathscr{X}^*$ if $\mathscr{X}^*$ is given its weak* topology; in fact, the weak* topology is the coarsest topology for which all functionals f_x are continuous. Since the strong topology of the conjugate space is finer than (or at least as fine as) the weak* topology, it follows that the functionals f_x are also continuous when $\mathscr{X}^*$ is given its strong topology. Hence the set of all functionals f_x obtained by letting x vary over the whole space $\mathscr{X}$ is a subset of the conjugate of $\mathscr{X}^*$, whether the latter has the weak* or the strong topology. When the set of all f_x (as x varies over $\mathscr{X}$) equals the conjugate of $\mathscr{X}^*$, we call $\mathscr{X}$ *reflexive.* (For instance, the Euclidean spaces are reflexive and so is l_2.) Let $\mathscr{X}$ be a linear topological space and $\mathscr{X}^*$ its conjugate. A subset X^* of $\mathscr{X}^*$ is said to be *regularly convex* (this is not to be confused with the notion of regular° convexity defined earlier) if, given an element x_0^* not in X^*, there exists an element x_0 of the underlying space $\mathscr{X}$ such that

$$\sup_{x^* \in X^*} x^*(x_0) < x_0^*(x_0) . \tag{2}$$

The relation (2) can be understood more easily if we rewrite it as

$$\sup_{x^* \in X^*} f_{x_0}(x^*) < f_{x_0}(x_0^*) \tag{2'}$$

where f_{x_0} is defined as above. Now f_{x_0} is a continuous functional on $\mathscr{X}^*$, as just shown, whether the topology of the conjugate space is weak* or strong. Hence (2′) demands that it be possible to strictly separate X^* from a point x_0^* outside of it by a hyperplane (in either topology) and, furthermore, that the hyperplane be of the type defined by an f_x functional. Now we know that in either topology the conjugate space is locally convex; hence, provided X^* is convex and closed, there always exists *some* hyperplane strictly separating the point and the set (by the Mazur-Bourgin Theorem). However, it does not follow that the separating hyperplane will be *of the f_x type*, i.e., determined by an element of the underlying space. It is therefore noteworthy that, as shown by Bourgin ([9], Theorem 18, p. 655), if $\mathscr{X}$ is a Hausdorff[20] linear space, X^* is regularly convex, if and only if it is convex and closed in the weak* topology. Of course, X^* is closed in the strong topology if it is closed in the weak* topology.

II.1.4.2. *Regularly convex envelope.*

LEMMA II.1. *Let $\mathscr{A}$ be a collection of regularly convex subsets A of $\mathscr{W}^*$ and assume the intersection*

$$I_{\mathscr{A}} = \bigcap_{A \in \mathscr{A}} A$$

of these sets to be non-empty. Then $I_{\mathscr{A}}$ is also regularly convex.[21]

PROOF. Let $w_0^* \notin I_{\mathscr{A}}$. Then $w_0^* \notin A_0$ for some $A_0 \in \mathscr{A}$. Since A_0 is regularly convex, there exists $w_0 \in \mathscr{W}$ such that

$$\sup_{w^* \in A_0} w^*(w_0) < w_0^*(w_0) .$$

Now $I_{\mathscr{A}} \subseteq A_0$, so that

$$\sup_{w^* \in I_{\mathscr{A}}} w^*(w_0) \leqq \sup_{w^* \in A_0} w^*(w_0) ,$$

hence,

$$\sup_{w^* \in I_{\mathscr{A}}} w^*(w_0) < w_0^*(w_0)$$

and the conclusion of the Lemma follows.

If $B \subseteq \mathscr{W}^*$, we denote by $\tilde{B}$ the intersection of all regularly convex sets in $\mathscr{W}^*$ containing B. By the preceding Lemma, $\tilde{B}$ is regularly convex; it is called the *regularly convex envelope* of B.

Clearly, $B = \tilde{B}$ if and only if B is regularly convex.

II.1.5. If T is a linear continuous transformation on $\mathscr{X}$ into $\mathscr{Y}$ (where both $\mathscr{X}$ and $\mathscr{Y}$ are topological linear spaces), we may define a functional φ on $\mathscr{X}$ by the relation

$$\varphi(x) = y_0^*[T(x)] \qquad \text{for all } x \in \mathscr{X} ,$$

where y_0^* is a fixed element of $\mathscr{Y}^*$, i.e., a linear continuous functional on $\mathscr{Y}$. We have

$$\varphi(\alpha x) = y_0^*[T(\alpha x)] = y_0^*[\alpha T(x)] = \alpha y_0^*[T(x)] = \alpha \varphi(x)$$

and

$$\begin{aligned} \varphi(x' + x'') &= y_0^*[T(x' + x'')] = y_0^*[T(x') + T(x'')] \\ &= y_0^*[T(x')] + y_0^*[T(x'')] = \varphi(x') + \varphi(x'') . \end{aligned}$$

[20] It may be shown that the restriction to Hausdorff spaces may be removed.

[21] This is stated for Banach spaces in Krein and Šmulian [30], p. 556.

Hence φ is linear; φ is also continuous,[22] hence it is an element of $\mathscr{X}^*$ and may be denoted by $x^*_{y^*_0}$.

Consider now the functional relation associating with each $y^* \in \mathscr{Y}^*$ the corresponding $x^*_{y^*} \in \mathscr{X}^*$, as just defined. This relation is denoted by T^* and is called the *adjoint* of T. We write

$$x^*_{y^*} = T^*(y^*) \qquad (x^*_{y^*} \in \mathscr{X}^*, y^* \in \mathscr{Y}^*)\ ,$$

where

$$x^*_{y^*}(x) = y^*[T(x)] \qquad \text{for all } x \in \mathscr{X}\ .$$

We note that, for all $x \in \mathscr{X}$,

$$x^*_{\alpha y^*}(x) = \alpha y^*[T(x)] = \alpha x^*_{y^*}(x)$$

and

$$x^*_{y^*_1+y^*_2}(x) = (y^*_1 + y^*_2)[T(x)] = y^*_1[T(x)] + y^*_2[T(x)]$$
$$= x^*_{y^*_1}(x) + x^*_{y^*_2}(x)\ .$$

I.e.,

$$T^*(\alpha y^*) = \alpha T^*(y^*)$$

and

$$T^*(y^*_1 + y^*_2) = T^*(y^*_1) + T^*(y^*_2)\ ,$$

so that T^* is a linear transformation.

When $\mathscr{X}$ and $\mathscr{Y}$ are Banach spaces, T^* is also continuous. (Hille [20], Def. 2.13.1 and Theorem 2.13.3, p. 27. Note that here continuity is equivalent to boundedness.)

When $\mathscr{X}$ and $\mathscr{Y}$ are finite-dimensional Euclidean spaces, let A denote the matrix such that

$$T(x) = Ax\ .$$

Here linear functionals belong to their respective spaces ($\mathscr{X} = \mathscr{X}^*$, $\mathscr{Y} = \mathscr{Y}^*$) and $x^*(x) = x^{*\prime}x$, etc., where the prime denotes transposition. Hence the relation $x^*(x) = y^*[T(x)]$ may be written as $x^{*\prime}x = y^{*\prime}Ax$, i.e., $x^{*\prime} = y^{*\prime}A$, so that $x^* = A'y^*$. I.e., the adjoint transformation T^* corresponds to a premultiplication by the transpose A' of the matrix A representing the given transformation T.

II.2.1. Let A be a set and ρ a transitive binary relation in A. When the relation holds for the ordered pair $a', a'' \in A$, we write $a' \rho a''$. When it does not, we write $a' \bar{\rho} a''$. An element $a_0 \in A_1$, $A_1 \subseteq A$ is said to be *ρ-maximal* in A_1 (or, more briefly, maximal) if, for any $a' \in A$, the relations $a' \in A_1$, $a' \rho a_0$ imply $a_0 \rho a'$.

Let ψ be a real-valued function on A. Then ψ is said to be *isotone* (with respect to ρ) if

$$a' \rho a'' \text{ implies } \psi(a') \geq \psi(a'')\ ;$$

[22] Cf. Kuratowski [32], p. 74, (6).

ψ is said to be *strictly isotone* (with respect to ρ) if, in addition,

$$a' \rho a'' \text{ and } a'' \bar{\rho} a' \text{ imply } \psi(a') > \psi(a'') .$$

In what follows we usually deal with transitive reflexive relations denoted by $\geqq$ or similar symbols. (The denial of $\geqq$ is written $\ngeqq$.) We then write $a' \geq a''$ to mean $a' \geqq a''$ and $a'' \ngeqq a'$.

II.2.2. If $\mathscr{W}$ is a linear system and $K \subseteq \mathscr{W}$, K is said to be a *cone*[23] if

$$w \in K, \lambda \geqq 0 \text{ imply } \lambda w \in K.$$

K is said to be a *convex cone* if K is a cone and a convex set.

A set $K \subseteq \mathscr{W}$ is a *convex cone* if and only if it satisfies

$$w \in K, \lambda \geqq 0 \text{ imply } \lambda w \in K ,$$

and

$$w' \in K, w'' \in K \text{ imply } w' + w'' \in K .$$

It may be noted that both the space $\mathscr{W}$ and the one-element set $\{0_w\}$ are convex cones.

II.2.3. Given a convex cone $K \subseteq \mathscr{W}$, a transitive reflexive relation to be denoted by $\geqq$ (or $\geqq_K$ if we wish to be more explicit) may be defined as follows: for any $w', w'' \in \mathscr{W}$, $w' \geqq w''$ if and only if $w' - w'' \in K$. (In particular, $w \geqq 0_w$ if and only if $w \in K$.)

Example. Let $\mathscr{W}$ be the Euclidean two-space of elements $w = (w^{(1)}, w^{(2)})$ where $w^{(1)}, w^{(2)}$ are real numbers. Then the following convex cones are of interest in defining ordering relations:

$$\begin{aligned} K_1 &= \{w : w^{(1)} \geqq 0, w^{(2)} \geqq 0\} , \\ K_2 &= \{w : w^{(1)} \geqq 0, w^{(2)} = 0\} , \\ K_3 &= \{0_w\} , \\ K_4 &= \mathscr{W} . \end{aligned}$$

We see that

$$\begin{aligned} w \geqq_{K_1} 0_w &\text{ means } w^{(1)} \geqq 0, w^{(2)} \geqq 0 , \\ w \geqq_{K_2} 0_w &\text{ means } w^{(1)} \geqq 0, w^{(2)} = 0 , \\ w \geqq_{K_3} 0_w &\text{ means } w^{(1)} = 0, w^{(2)} = 0 , \end{aligned}$$

and

$w \geqq_{K_4} 0_w$ holds for all $w \in \mathscr{W}$, i.e., it is a vacuous constraint.[24] Other relations could be obtained by replacing $\geqq$ by $>$ in the definitions of K_1 and K_2. Thus we have a great range of possibilities

[23] It would be more precise to speak of a cone with the vertex at origin, but we omit the qualifying phrase since no other cones will be considered. (Our use of the term "cone" may seem unnatural, but it permits us to define the "convex cone" as a cone which is convex.)

[24] This makes it possible to cover simultaneously the cases of unconstrained and (non-vacuously) constrained maximization by orderings based on convex cones.

covering equalities, inequalities ($\geqq$ or $>$), and their various combinations. This makes it possible to obtain results which can be specialized in a variety of ways.

Let $\mathscr{W}$ be a linear topological space and K a convex cone in $\mathscr{W}$. In the applications, we are interested in the topological, as well as the algebraic properties of the cone K. In some theorems, we assume that the cone K is *closed.* This is obviously true of the cones K_1, K_2, K_3, K_4 in the natural (Euclidean) topology of the plane. On the other hand, the cone

$$K_5 = \{(w^{(1)}, w^{(2)}) : w^{(1)} > 0, w^{(2)} > 0\}$$

is not closed in the natural topology of the plane. We see that lack of closedness may result from using cones corresponding to strict, rather than weak, inequalities. In the economic applications the inequalities are usually of the weak type, hence the closedness of the corresponding cones is not a serious restriction.

Another topological property assumed for certain convex cones is that they have *non-empty interiors.* Of the preceding examples, using again the Euclidean topology of the plane, K_1, K_4, and K_5 have interior points, while K_2 and K_3 do not. The requirement of a non-empty interior can be troublesome in infinite-dimensional spaces. Thus consider a space $l_p (p \geqq 1)$ whose elements are infinite sequences $x = (x_1, x_2, \cdots)$ such that

$$\sum_{i=1}^{\infty} |x_i|^p < \infty .$$

This space is normable, the norm being of x defined as

$$\left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p} .$$

Now consider the convex cone K consisting of all the elements x of l_p whose every coordinate is non-negative, this being the natural counterpart of the non-negative orthant in a finite-dimensional space. It may be seen that K has no interior, i.e., every element of K is a boundary point. To see this, take an arbitrary element x' of K. Given a positive number ε, however small, one can find an element x'' of l_p whose distance from the element x' is less than ε and such that x'' has at least one negative coordinate; this can be accomplished by taking x'' such that all but one of the coordinates of x'' are the same as the corresponding coordinates of x', while one coordinate of x'' (with a sufficiently high subscript) is the negative of the corresponding coordinate of x'.

On the other hand, let (m) denote the space of infinite bounded sequences $x = (x_1, x_2, \cdots)$, normed by

$$||x|| = \sup_{1 \leqq i} (|x_i|) ,$$

and define K, as in the preceding example, as the set of all x with non-negative coordinates. Here any point x whose coordinates are all positive is an interior point of the cone.

II.2.4. Let $K \subseteq \mathscr{W}$ be a convex cone. Then the conjugate $K^{\oplus}$ of K is defined by

$$K^{\oplus} = \{w^* \in \mathscr{W}^* : w^*(w) \geqq 0 \text{ for all } w \in K\} .$$

Since $K^{\oplus}$ is a cone, it is called the *conjugate cone* of K. Clearly, $K^{\oplus}$ is the set of linear continuous functionals isotone with respect to $\geqq_K$. We note that $K^{\oplus}$ is never empty, since $0^* \in K^{\oplus}$.

In accord with the notational principles of II.2.3, we write $w^* \geqq_{K^{\oplus}} 0^*_w$ (or, more simply, $w^* \geqq 0$), and call w^* *non-negative on* K if $w^* \in K^{\oplus}$. Furthermore, we write $w^* >_{K^{\oplus}} 0$ (or : $w^* > 0^*_w$) and call w^* *strictly positive on*[25] K if $w^* \geqq_{K^{\oplus}} 0$, and $w \geq_K 0$ implies $w^*(w) > 0$. It is seen that $w^* > 0$ if and only if w^* is a linear continuous functional strictly isotone with respect to $\geqq_K$.

II.2.5. LEMMA II.2.[26] *Let K be a closed convex cone in a locally convex linear space $\mathscr{W}$ and let $w_0 \in \mathscr{W}$ be such that*

$$w^*(w_0) \geqq 0 \textit{ for all } w^* \geqq 0 .$$

Then $w_0 \in K$.

PROOF. Suppose $w_0 \notin K$. By virtue of the Mazur-Bourgin Theorem[27] K is regularly° convex, since it is closed and convex and $\mathscr{W}$ is locally convex, so that there exists a $w_0^* \in \mathscr{W}^*$ such that

$$\sup_{w \in K} w_0^*(w) < w_0^*(w_0) .$$

Now, since $w_0^*(w_0)$ is a fixed number and K a cone, we must have

$$\sup_{w \in K} w_0^*(w) = 0 .$$

Let $w_1^* = - w_0^*$. Then

$$w_1^*(w) \geqq 0 \text{ for all } w \in K \text{; i.e., } w_1^* \in K^{\oplus} ,$$

and

$$w_1^*(w_0) < 0 ,$$

which contradict the hypothesis of the Lemma, hence the proof is completed.

II.2.6.1. LEMMA II.3. *If $K \subseteq \mathscr{W}$ is a convex cone, then the conjugate cone $K^{\oplus}$ is regularly convex.*

PROOF. If $K^{\oplus} = \mathscr{W}^*$, no $w_0^* \notin K^{\oplus}$ exists and the condition of regularity is (vacuously) satisfied. Now let $K^{\oplus} \neq \mathscr{W}^*$ and take $w_0^* \notin K^{\oplus}$. Then there exists a $w_1 \in K$ such that $w_0^*(w_1) < 0$. Write $w_0 = - w_1$. Since

[25] The reader should be warned that this term has a somewhat unusual meaning. In particular, if K is the origin, *every* linear functional is strictly positive on K.

[26] For the case of linear normed spaces, cf. Krein and Rutman [29], p. 16.

[27] Cf. Theorem II.2. in II.1.4.1.

$w^*(w_1) \geqq 0$ for all $w^* \in K^\oplus$ (because $w_1 \in K$ and by definition of $K^\oplus$), we have

$$w^*(w_0) = -w^*(w_1) \leqq 0 \qquad \text{for all } w^* \in K^\oplus ,$$

while

$$w_0^*(w_0) = -w_0^*(w_1) > 0 ,$$

which shows that $K^\oplus$ is regularly convex. This completes the proof.

II.2.6.2. For the special case of $\mathscr{W}$ linear normed, the preceding result follows from Krein and Rutman [29], p. 38, where it is proved that $K^\oplus$ is weak* closed; since the convexity of $K^\oplus$ is evident, this implies the regular convexity of $K^\oplus$; cf. II.1.4.1.

II.3.1. A very abstract version of a (partial ordering) maximization problem of the type considered in the present chapter in connection with the Lagrangian saddle-points can be formulated as follows.

Let $\mathscr{X}$ be an arbitrary space, X a subset of $\mathscr{X}$, $\mathscr{Y}$ an arbitrary space with the transitive relation ρ, and $\mathscr{Z}$ a linear system with $z' \geqq z''$ defined to mean $z' - z'' \in P_z$ where P_z is a convex cone.

Furthermore let f be a function on $\mathscr{X}$ into $\mathscr{Y}$ and g a function on $\mathscr{X}$ into $\mathscr{Z}$.

Let the *constraints* be $x \in X$ and $g(x) \geqq 0_z$.

Let X_0 denote the *permissible x-set*, i.e.,

$$X_0 = X \cap g^{-1}(P_z) = \{x \in \mathscr{X} : x \in X, g(x) \geqq 0_z\} ,$$

while

$$Y_0 = f(X_0) = \{y : y = f(x), x \in X, g(x) \geqq 0_z\}$$

is the *permissible y-set*.

Denote by $\hat{Y}_0$ the ρ-maximal subset of Y_0; i.e.,

$$\hat{Y}_0 = \{y_0 \in y_0 : y' \in Y_0, y' \rho y_0 \text{ imply } y_0 \rho y'\}$$

and call $\hat{Y}_0$ the *maximal y-set*, while $\hat{X}_0 = f^{-1}(\hat{Y}_0)$ is called the *maximal x-set*. An element of a (y- or x-) maximal set is called *maximal*.

The objective is typically to characterize $\hat{X}_0$. Hence a maximization problem is uniquely determined by the selection of

$$\pi \equiv (\mathscr{X}, X ; \mathscr{Y}, \rho ; \mathscr{Z}, P_z ; f, g)$$

and we may refer to π as the (*partial ordering*) *maximization problem*.

In some contexts we only need $\mathscr{Y}, \rho$, and Y_0, without reference to how Y_0 is defined. In others specializing assumptions are made with regard to the entities defining π.

II.3.2. For a given maximization problem π, as defined in the preceding section, we define a *generalized Lagrangian expression* Φ_π or (where safe) Φ by

$$\Phi = \Phi_\pi = \Phi_\pi(x, \zeta^* ; \eta^*) = \eta^*[f(x)] + \zeta^*[g(x)] , \quad x \in \mathscr{X} ,$$

where ζ^* and η^* are real-valued functions on $\mathscr{Z}$ and $\mathscr{Y}$ respectively.

That is, Φ_π is a real-valued function in the Cartesian product space $\mathscr{X} \times [\zeta^*] \times [\eta^*]$ where $[\zeta^*]$ and $[\eta^*]$ are the spaces of real-valued functions on $\mathscr{Z}$ and $\mathscr{Y}$ respectively.

II.3.3. Let $\mathscr{Z}$ be a topological linear space, $\mathscr{Z}^*$ its conjugate space. Symbols such as z^*, z_0^* denote elements of $\mathscr{Z}^*$. We say that Φ_{π_1} (where π_1 differs from π in that π_1 requires $\mathscr{Z}$ to be a linear topological space) has an *isotone saddle-point* at $(x_0, z_0^* ; \eta_0^*)$ if

(1) $x_0 \in X, z_0^* \geqq 0, \eta_0^* \in [\eta_0^*]$ and η_0^* is strictly isotone with respect to ρ,

(2) $\Phi_{\pi_1}(x, z_0^* ; \eta_0^*) \leqq \Phi_{\pi_1}(x_0, z_0^* ; \eta_0^*) \leqq \Phi_{\pi_1}(x_0, z^* ; \eta_0^*)$
for all $x \in X$ and all $z^* \geqq 0$.

Now specialize the partial ordering maximization problem π_1 to the *vectorial* (ordering) *maximization problem* π_2 as follows. Let $\mathscr{X}$ be a linear system, P_x a convex cone in $\mathscr{X}$, $x \geqq 0$ be defined as $x \in P_x$, and $X = P_x$. Furthermore, let $\mathscr{Y}$ be a linear topological space, P_y be a convex cone in $\mathscr{Y}$, and let ρ be $\geqq_{P_y}$. (Hence $y_0^* \in \mathscr{Y}^*$, and $y_0^* > 0$ means y_0^* is strictly positive on P_y.)

We then say that the Lagrangian expression Φ_{π_2} has a *non-negative saddle-point* at $(x_0, z_0^* ; y_0^*)$ if

(1′) $$x_0 \geqq 0 , \qquad z_0^* \geqq 0 , \qquad y_0^* > 0$$

and

(2′) $$\Phi_{\pi_2}(x, z_0^* ; y_0^*) \leqq \Phi_{\pi_2}(x_0, z_0^* ; y_0^*) \leqq \Phi_{\pi_2}(x_0, z^* ; y_0^*)$$
for all $x \geqq 0$ and all $z^* \geqq 0$.

II.4. Let $\mathscr{X}$ and $\mathscr{Y}$ be linear systems and let f be a (single-valued) function with a convex domain $\mathscr{D} \subseteq \mathscr{X}$ and range $\mathscr{R} \subseteq \mathscr{Y}$. Then the *function* f is said to be *concave* if, given any $x', x'' \in \mathscr{D}$ and any real number $0 < \theta < 1$, we have

$$(1-\theta)f(x') + \theta f(x'') \leqq f[(1-\theta)x' + \theta x''] ,$$

where $y' \geqq y''$ means $y' - y'' \in K$ for a given convex cone K in $\mathscr{Y}$.

II.5.1. Let $\mathscr{W}$ be a Banach space and h a (single-valued) function whose domain is a set A of reals and the range a subset of $\mathscr{W}$, i.e.,

$$w = h(\alpha) , \qquad \alpha \in A, w \in \mathscr{W} .$$

Following Graves[28] we define the *first derivative* $h'(\alpha_0) = \left.\frac{d}{d\alpha}h(\alpha)\right|_{\alpha=\alpha_0}$ of h with regard to α at α_0 as the element of $\mathscr{W}$ such that

$$\lim_{\alpha \to \alpha_0} \left\| \frac{h(\alpha) - h(\alpha_0)}{\alpha - \alpha_0} - h'(\alpha_0) \right\| = 0 .$$

[28] Reference [17], p. 164.

Similarly,

$$\frac{d^2}{d\alpha^2}h(\alpha)\bigg|_{\alpha=\alpha_0} = \frac{d}{d\alpha}\frac{d}{d\alpha}h(\alpha)\bigg|_{\alpha=\alpha_0}, \text{ etc.}$$

Now let $\mathscr{X}$ and $\mathscr{Y}$ be Banach spaces and f a function on $\mathscr{X}$ into $\mathscr{Y}$. Then $f(x_0+\alpha x')$, α real, $x_0, x' \in \mathscr{X}$, may be regarded, for fixed x_0 and x', as a function of the real variable α with values in $\mathscr{Y}$. We define the *first, second, etc., variation of f at x_0 with increment x'* by

$$\delta f(x_0\,;\,x') = \frac{d}{d\alpha}f(x_0+\alpha x')\bigg|_{\alpha=0},$$

$$\delta^2 f(x_0\,;\,x') = \frac{d^2}{d\alpha^2}f(x_0+\alpha x')\bigg|_{\alpha=0}, \text{ etc.}^{29}$$

When the domain of f is open and $\delta f(x_0\,;\,x')$ exists and is continuous in x', $\delta f(x_0\,;\,x')$ is called the *Fréchet differential* of f at x_0 with increment x'. It has been shown[30] that the Fréchet differential $\delta f(x_0\,;\,x')$ so defined is linear (i.e., homogeneous and additive) as well as continuous in x'; also that

$$\lim_{||x'||\to 0}\frac{1}{||x'||}\left\|f(x_0+x') - f(x_0) - \delta f(x_0\,;\,x')\right\| = 0$$

for all x in the domain of f.

II.5.2. The "function of a function rule" is valid for Fréchet differentials[31] and may be stated as follows.

Let $\mathscr{Y}, \mathscr{X}, \mathscr{Z}$ be Banach spaces; f a function on $\mathscr{X}$ into $\mathscr{Y}$, g on $\mathscr{Z}$ into $\mathscr{X}$.

$$y = f(x)\,, \qquad y_0 = f(x_0)\,,$$
$$x = g(z)\,, \qquad x_0 = g(z_0)\,,$$

and assume that f and g possess Fréchet differentials at x_0 and z_0 respectively. Write

$$f(g(z)) = h(z)$$

so that h is a function in $\mathscr{Z}$ into $\mathscr{Y}$. Then, for $\zeta \in \mathscr{Z}$,

$$\delta h(z_0\,;\,\zeta) = \delta f(x_0\,;\,\delta g(z_0\,;\,\zeta))\,.$$

The reader is referred to Fréchet [11], [13], Hildebrandt and Graves

[29] An equivalent definition of $\delta f(x_0\,;\,x')$ is

$$\delta f(x_0\,;\,x') = \lim_{\alpha\to 0}\frac{f(x_0+\alpha x') - f(x_0)}{\alpha}$$

where, for a function $w=h(\alpha)$ of real variable α with values in $\mathscr{W}$, we write

$$\lim_{\alpha\to\alpha_0} h(\alpha) = w_0,\ w_0 \in \mathscr{W} \quad \text{if and only if} \lim_{\alpha\to\alpha_0}||\,h(\alpha)-w_0\,|| = 0\,.$$

Cf. Hildebrandt and Graves [19], p. 136, and Hille [20], pp. 71–72.

[30] Hille [20], p. 73 and p. 72, Def. 4.3.4.

[31] Cf. Hildebrandt and Graves [19], pp. 141–44; Graves [17], p. 649.

[19], Graves [16], [17], and Hille [20] for an account of the properties of Fréchet differentials.[32]

II.6.1. Let $\mathscr{X}$ and $\mathscr{Y}$ be two Banach spaces. Consider the linear system whose elements are the ordered pairs $(x, y), x \in \mathscr{X}, y \in \mathscr{Y}$, with addition and scalar multiplication defined by

$$(1') \qquad \begin{cases} (x', y') + (x'', y'') = (x' + x'', y' + y'') \\ \alpha(x, y) = (\alpha x, \alpha y), \ \alpha \text{ real.} \end{cases}$$

Then the linear system of the ordered pairs (x, y) will become a Banach space if it is normed in such a way that[33]

$$(1'') \qquad \lim_{n\to\infty} x_n = x_0 \text{ and } \lim_{n\to\infty} y_n = y_0 \quad \text{if and only if } \lim_{n\to\infty} \left\| (x_n, y_n) - (x_0, y_0) \right\| = 0 \,.$$

Such a Banach space of the ordered pairs (x, y) is denoted by $\mathscr{X} \times \mathscr{Y}$ and is called the *(Banach) product* of $\mathscr{X}$ and $\mathscr{Y}$. Writing, for $A \subseteq \mathscr{X}$, $B \subseteq \mathscr{Y}$, $A \times B = \{(x, y) : x \in A, y \in B\}$ we have[34] $A \times B$ closed if and only if both A and B are closed.

More generally, let $\mathscr{X}$ and $\mathscr{Y}$ be linear topological spaces and consider the linear system of the ordered pairs (x, y) with the operations defined by (1′) above.

Then the space of pairs (x, y), again to be denoted by $\mathscr{X} \times \mathscr{Y}$ (and called *linear topological product*), may be topologized by choosing as a base[35] the sets

$$\{(x', y') : x' \in U_x, y' \in \mathscr{Y}\} \,,$$
$$\{(x'', y'') : x'' \in \mathscr{Y}, y'' \in U_y\} \,,$$
$$\{(x''', y''') : x''' \in U_x, y''' \in U_y\} \,,$$

where U_x is any open set in $\mathscr{X}$, U_y any open set in $\mathscr{Y}$. It may be noted for later reference that[36] with this topology $A \times B$ is closed if A and B both are.

It is known [37] that if $\mathscr{X}$ and $\mathscr{Y}$ are linear topological spaces, then so is $\mathscr{X} \times \mathscr{Y}$; if $\mathscr{X}$ and $\mathscr{Y}$ are locally convex, then so is $\mathscr{X} \times \mathscr{Y}$.

II.6.2. Let[38] $\mathscr{X} = \mathscr{X}' \times \mathscr{X}''$ be the (Banach) product of the two Banach spaces $\mathscr{X}', \mathscr{X}''$. The symbols x' and ξ' denote elements of $\mathscr{X}'$, x'' and ξ'' those of $\mathscr{X}''$, x and ξ those of $\mathscr{X}$. If f is a function on

[32] See also V.3.3.8 for a discussion of differentials in a class of spaces wider than Banach.

[33] Banach [3], pp. 181–82, especially eq. (33), where examples of norms satisfying (1) are given. Cf. also Hyers [22], pp. 3, 5, and Tychonoff [43], p. 772.

[34] Cf. Kuratowski [32], 24.II.1, p. 219.

[35] Cf. Lefschetz [34], p. 6 (6.1); p. 10, Section 12.

[36] Lefschetz [34], p. 11 (12.6).

[37] Tychonoff [43], p. 772; Bourgin [9], p. 639; Hyers [22], pp. 3, 5. In these sources it is shown how a linear topological product of an arbitrary family of spaces is formed.

[38] We confine ourselves to the product of two spaces. The treatment of $\mathscr{X}^{(1)} \times \mathscr{X}^{(2)} \times \cdots \times \mathscr{X}^{(n)}$ is quite analogous.

$\mathscr{X}$ into the Banach space $\mathscr{Y}$, $\delta f(x\,;\,\xi)$ will denote the Fréchet differential of f at x with increment ξ.

Then the *partial Fréchet differential* of f *with respect to* x' *at* x_0 *with increment* ξ' is written as $\delta_{x'}f(x_0\,;\,\xi')$ and is defined by

$$\text{(1)} \qquad \delta_{x'}f(x_0\,;\,\xi') = \delta f(x_0\,;\,(\xi', 0_{x''}))\,, \qquad x_0 \in \mathscr{X}, \qquad x', \xi' \in \mathscr{X}'\,.$$

We have[39] the additivity law

$$\text{(2)} \qquad \delta f(x_0\,;\,(\xi', \xi'')) = \delta_{x'}f(x_0\,;\,\xi') + \delta_{x''}f(x_0\,;\,\xi'')\,.$$

II.6.3. We shall now state the "function of a function" rule for the case of a function of several variables.

Let

$$\mathscr{X} = \mathscr{X}^{(1)} \times \mathscr{X}^{(2)} \times \cdots \times \mathscr{X}^{(n)}$$
$$\mathscr{Z} = \mathscr{Z}^{(1)} \times \mathscr{Z}^{(2)} \times \cdots \times \mathscr{Z}^{(m)}$$

where all spaces are Banach and so are the products. Also, let f be a function in $\mathscr{X}$ into the Banach space $\mathscr{Y}$, $g^{(i)}$ in $\mathscr{Z}$ into $\mathscr{X}^{(i)}$.

$$y = f(x)\,, \qquad y_0 = f(x_0)\,,$$
$$x^{(i)} = g^{(i)}(z)\,, \qquad x_0{}^{(i)} = g^{(i)}(z_0) \qquad (i = 1, 2, \cdots, n)$$

and assume that f and each of the $g^{(i)}$ possess Fréchet differentials at x_0 and z_0 respectively. Write

$$h(z) = f((g^{(1)}(z), g^{(2)}(z), \cdots, g^{(n)}(z)))$$

so that h is a function in $\mathscr{Z}$ into $\mathscr{Y}$. Then, for $\zeta \in \mathscr{Z}$,

$$\delta h(z_0\,;\,\zeta) = \sum_{i=1}^{n} \delta f x^{(i)}(x_0\,;\,\sum_{j=1}^{m} \delta z^{(j)} g^{(i)}(z_0\,;\,\zeta^{(i)})$$

where $\zeta^{(i)} \in \mathscr{Z}^{(i)}$.[40]

II.6.4. We shall find it convenient to define a "quasi-saddle-point" for Lagrangian expressions. We say that

$$\Phi(x, z^*\,;\,y_0^*) \equiv y_0^*[f(x)] + z^*[g(x)]$$

has a non-negative *quasi-saddle-point* at $(x_0, z_0^*\,;\,y_0^*)$ if and only if $y_0^* > 0$, $x_0 \geqq 0$, $z_0^* \geqq 0$, and the following relations hold:

$$\delta_x\Phi((x_0, z_0^*)\,;\,\xi) \leqq 0 \qquad \text{for all } x \geqq 0, x = x_0 + \xi\,,$$
$$\delta_x\Phi((x_0, z_0^*)\,;\,x_0) = 0\,,$$
$$\delta_{z^*}\Phi((x_0, z_0^*)\,;\,\zeta^*) = \zeta^*[g(x_0)] \geqq 0 \qquad \text{for all } z^* \geqq 0, \zeta^* = z^* - z_0^*\,,$$
$$\delta_{z^*}\Phi((x_0, z_0^*)\,;\,z_0^*) = z_0^*[g(x_0)] = 0\,.$$

It is seen that if Φ has a non-negative saddle-point at $(x_0, z_0^*\,;\,y_0^*)$, then it necessarily has a non-negative quasi-saddle-point there, but the converse is not true.

---

39 Cf. Hildebrandt and Graves [19], p. 138.

40 Cf. Fréchet [11], pp. 318–21. (The reprinted version in [13] is free of the misprints in [11].)

III. The "Minkowski-Farkas Lemma"

III.1. Throughout III, $\mathscr{X}$ is a linear topological space, $\mathscr{Y}$ a locally convex[41] linear space, $y' \geqq y''$ means $y' - y'' \in P_y$, where P_y is a closed convex cone. T is a linear continuous transformation on $\mathscr{X}$ into $\mathscr{Y}$. $\mathscr{X}, \mathscr{Y}, P_y, T$ are fixed throughout. $\mathscr{X}^*, \mathscr{Y}^*$ are the conjugate spaces of $\mathscr{X}, \mathscr{Y}$, and T^* is the adjoint of T.

III.2. If $x^* \in \mathscr{X}^*$ is such that

$$(1) \qquad x^* = T^*(y^*) \text{ for some } y^* \in \mathscr{Y}^* ,$$

we say that eq. (1) is *solvable*. If x^* is such that eq. (1) holds for some $y^* \geqq 0$, we say that eq. (1) is *positively*[42] *solvable*. We then also say that x^* *makes eq.* (1) *positively solvable.*

The set of all $x^* \in \mathscr{X}^*$ which make eq. (1) positively solvable will be denoted by Z_T, i.e.,

$$(2) \qquad Z_T = \{x^* \in \mathscr{X}^* : x^* = T^*(y^*), y^* \geqq 0\} .$$

Note that, since $\{y^* : y^* \geqq 0\} = P_y^{\oplus}$, we have

$$(3) \qquad Z_T = T^*(P_y^{\oplus}) .$$

The point x^* is said to be *positively normal with regard to* T if

$$(4) \qquad \text{for all } x \in \mathscr{X}, T(x) \geqq 0 \text{ implies } x^*(x) \geqq 0 .$$

We shall denote by V_T the set of all x^* positively normal with regard to a given T, i.e.,

$$(5) \qquad V_T = \{x^* \in \mathscr{X}^* : \text{for all } x \in \mathscr{X}, T(x) \geqq 0 \text{ implies } x^*(x) \geqq 0\} .$$

III.3. THEOREM III.1. *If x^* makes eq.* (1) *positively solvable, then x^* is positively normal with regard to T. In set language,*

$$(6) \qquad Z_T \subseteq V_T .$$

PROOF. Let $x^* = T^*(y^*)$ for some $y^* \geqq 0$. Then we have, by the definition of T^*,

$$(7) \qquad x^*(x) = (T^*y^*)(x) = y^*(Tx), \qquad \text{for all } x \in \mathscr{X} .$$

Therefore, since $y^* \geqq 0$, $T(x) \geqq 0$ implies $x^*(x) \geqq 0$.

III.4. THEOREM III.2. *If every x^* positively normal with regard to T makes eq.* (1) *positively solvable, then the set of all x^* which make eq.* (1) *positively solvable is regularly convex. In set language, if $V_T \subseteq Z_T$, then Z_T is regularly convex.* (*We may note that, in view of* (6), *Theorem* III.2 *may be equivalently restated as follows: if $V_T = Z_T$, then Z_T is regularly convex.*)

PROOF. We note that the set

$$(8) \qquad X_T = \{x \in \mathscr{X} : T(x) \geqq 0\}$$

is a convex cone and $V_T = X_T^{\oplus}$, so that, by Lemma II.3 in II.2.6.1, V_T is regularly convex and hence so is $Z_T = V_T$.

[41] Cf. II.1.1.3.

[42] "Non-negatively" would be more accurate but awkward.

III.5. THEOREM III.3. *V_T coincides with the regular convex envelope of Z_T*:

$$V_T = \tilde{Z}_T . \tag{9}$$

PROOF. In view of Theorem III.1, it will suffice to establish

$$V_T \subseteq \tilde{Z}_T , \tag{10}$$

i.e., that $x^* \notin \tilde{Z}_T$ implies $x^* \notin V_T$.

Consider some $x_0^* \notin \tilde{Z}_T$. We shall find x_1 such that $T(x_1) \geqq 0$ while $x_0^*(x_1) < 0$.

Since $\tilde{Z}_T$ is regular convex (cf. II.1.4.2), there must exist $x_0 \in \mathscr{X}$ such that

$$\sup_{x^* \in \tilde{Z}_T} x^*(x_0) < x_0^*(x_0) . \tag{11}$$

Since a cone Z_T is contained in $\tilde{Z}_T$ and

$$\sup_{x^* \in \tilde{Z}_T} x^*(x_0)$$

is finite, (11) implies

$$x^*(x_0) \leqq 0 < x_0^*(x_0) \text{ for any } x^* \in Z_T . \tag{12}$$

Now write

$$x_1 = - x_0 . \tag{13}$$

Then (12) may be written

$$x_0^*(x_1) = - x_0^*(x_0) < 0 , \tag{14}$$

and

$$x^*(x_1) = y^*(Tx_1) = (T^*y^*)(x_1) = - x^*(x_0) \geqq 0, \text{ for any } y^* \geqq 0, \tag{15}$$

since $x^* = T^*(y^*) \in Z_T$.

Now since $P_y = \{y^* : y^* \geqq 0\}$ is assumed closed and $\mathscr{Y}$ locally convex, the Lemma II.2 in II.2.5 applies. It follows that

$$T(x_1) \geqq 0 . \tag{16}$$

But (14) and (16) together imply $x_0^* \notin V_T$.

III.6. THEOREM III.4. *The positive normality of x^* with regard to T is equivalent to x^* making eq.* (1) *positively solvable if and only if the set of all x^* making eq.* (1) *solvable is regularly convex. In set language,*

$$Z_T = V_T \textit{ if and only if } Z_T \textit{ is regularly convex.} \tag{17}$$

PROOF. If $Z_T = V_T$, the regular convexity of Z_T follows from Theorem III.2. On the other hand, if Z_T is regularly convex, we have $Z_T = \tilde{Z}_T$ (cf. Lemma II.1 in II.1.4.2). The equality $Z_T = V_T$ then follows from Theorem III.3.

III.7. *The finite-dimensional Euclidean case.* In a reflexive Banach space, a set is regularly convex if and only if it is convex and (strongly) closed (cf. II.1.4). Since Z_T is always convex, for reflexive Banach spaces one may substitute "(strongly) closed" for "regularly convex" in Theorem III. 1, 2, 3, and 4.

In particular, if $\mathscr{X}$ and $\mathscr{Y}$ are finite-dimensional Euclidean spaces (hence Banach and reflexive in the Euclidean distance topology) and T is represented by a matrix, Z_T is a polyhedral convex cone (cf. Gale in [14], p. 290, Def. 1′) which is closed in the Euclidean distance topology. Hence for this case Z_T is necessarily regularly convex and $Z_T = V_T$ for all T. The Minkowski-Farkas Lemma as usually stated asserts that $V_T \subseteq Z_T$ in the finite-dimensional Euclidean case. This follows from Theorem III.3, since Z_T is known to be regularly convex.

[Let $y = (y_1, y_2, \cdots, y_n)$ and write $I = \{1, 2, \cdots, n\}$.

Partition I into I' and I'' where $I' \cup I'' = I, I' \cap I'' = \phi$, and either I' or I'' may be empty. The relation $y \geqq 0$ is interpreted as meaning

$$y_i \geqq 0 \quad \text{if } i \in I' ,$$
$$y_i = 0 \quad \text{if } i \in I'' .$$

The Minkowski-Farkas Lemma is usually stated for $I = I'$, but it is clear that $P_y = \{y \in \mathscr{Y} : y \geqq 0\}$, where the meaning of $y \geqq 0$ is that just stated, is necessarily closed.]

III. *Appendix*: *Relationship with Hausdorff's results.* [NOTE: This appendix is incorrect in its present form and should be ignored. For technical reasons, however, it was impossible to eliminate it from the present printing.]

IIIa.1. Suppose that x^* is positively normal with regard to T (cf. III.2) and let, for some $x' \in \mathscr{X}$,

(1) $$T(x') = 0 .$$

Then

(2′) $$T(x') \geqq 0$$

and

(2″) $$T(-x') \geqq 0 .$$

Since x^* is positively normal, the preceding inequalities yield, respectively,

(3′) $$x^*(x') \geqq 0$$

and

(3″) $$x^*(-x') \geqq 0 ,$$

i.e.,

(4) $$x^*(x') = 0 .$$

Hence, if x^* is positively normal with regard to T, we have

(5) for all $x \in \mathscr{X}$, $T(x) = 0$ implies $x^*(x) = 0$.

Call x^* satisfying (5) *normal with regard to* T. I.e., we have shown that *if* x^* *is positively normal with regard to* T, *then it is also normal with regard to* T.

IIIa.2. We shall now show that

(6) *if* x^* *is normal with regard to* T, *then either* x^* *or* $-x^*$ *is positively normal with regard to* T.

For suppose it could happen that (5) holds and neither x^* nor $-x^*$ is positively normal with regard to T. Then there must exist $x_1, x_2 \in \mathscr{X}$ such that

$$T(x_1) > 0 \ , \tag{7.1}$$

$$x^*(x_1) > 0 \ , \tag{7.2}$$

$$T(x_2) > 0 \ , \tag{8.1}$$

$$x^*(x_2) < 0 \ . \tag{8.2}$$

[Suppose no such pair x_1, x_2 exists. Then it must be that either $T(x) > 0$ implies $x^*(x) \geqq 0$ or $T(x) > 0$ implies $x^*(x) \leqq 0$. This, in conjunction with (5), would then yield (6).]

Let

$$\lambda = \frac{T(x_1)}{T(x_2)} \ . \tag{9}$$

Then

$$T(x_1 - \lambda x_2) = T(x_1) - \lambda T(x_2) = 0 \ . \tag{10}$$

On the other hand, by (7), (8), and (9) (which imply $\lambda > 0$),

$$x^*(x_1 - \lambda x_2) = x^*(x_1) - \lambda x^*(x_2) > 0 \ . \tag{11}$$

Hence (5) fails to hold for $x_1 - \lambda x_2 \in \mathscr{X}$ which establishes the validity of (6).

IIIa.3. Write

$$F_T = \{x^* : \text{for all } x \in \mathscr{X}, T(x) = 0 \text{ implies } x^*(x) = 0\} \tag{12}$$

(the set of x^* normal with regard to T) and recall that the set of all x^* positively normal with regard to T is denoted by V_T. Hence the results in IIIa.1 and IIIa.2 may be written as

$$F_T = V_T \cup (-V_T) \ . \tag{13}$$

IIIa.4. We shall now show that[43]

$$F_T = V_T - V_T \ . \tag{14}$$

First,

$$F_T = V_T \cup (-V_T) \subseteq V_T - V_T \ , \tag{15}$$

for any element in $V_T \cup (-V_T)$ is either of the form $x_1^* - 0_x^*$ where $x_1^* \in V_T$ or of the form $0_x^* + x_2^*$ where $x_2^* \in (-V_T)$. Note that $0_x^* \in V_T \cap (-V_T)$.

On the other hand, let x^* be an element of $V_T - V_T$, i.e.,

$$x^* = x_1^* + x_2^* \ , \qquad x_1^* \in V_T \ , \qquad x_2^* \in (-V_T) \ . \tag{16}$$

By (13), $x_i^* \in F_T$ $(i = 1, 2)$. But then $x^* \in F_T$, since F_T is a linear set.

[43] $A - B$ is the set of all elements of the form $a - b$, $a \in A$, $b \in B$. $A - A$ is neither empty nor the null element!

For let $x_i^* \in F_T$ $(i = 1, 2)$. Then $T(x) = 0$ implies $x_i^*(x) = 0$. Consider $x^* = \alpha_1 x_1^* + \alpha_2 x_2^*$ and suppose $T(x) = 0$; then $x^*(x) = \alpha_1 x_1^*(x) + \alpha_2 x_2^*(x) = 0$, hence $x^* \in F_T$.

IIIa.5.0. From now on we shall assume that

(17′) all the spaces considered are Banach (which implies that both T and T^* are bounded, since T was assumed continuous);

(17″) for every $y^* \in \mathscr{Y}^*$, there exist $y_1^* \in P_y^\oplus$, $y_2^* \in P_y^\oplus$ such that $y^* = y_1^* - y_2^*$.

(17″) is equivalent to the condition that P_y is a *normal cone*; cf. Krein and Rutman [29], Def. 2.2, p. 22, and p. 24.

IIIa.5.1. *Example.* Let $\mathscr{Y}$ be the space of all continuous real-valued functions $y(t)$ defined on the closed interval $[0, 1]$. (This space is usually denoted by $C[0, 1]$.) Then[44] every bounded linear functional y^* can be defined by

$$y^*(y) = \int_0^1 y(t)dg \qquad (y \in \mathscr{Y}) \tag{a}$$

where g is a function of bounded variation. Now define $y \in P_y$ (i.e., $y \geqq 0_y$) to mean

$$y(t) \geqq 0 \qquad \text{for all } 0 \leqq t \leqq 1\,. \tag{b}$$

Then $y^* \geqq 0_y^*$ (i.e., $y^* \in P^\oplus$) means that the function g in (a) is monotone non-decreasing. But it is well known (e.g., Titchmarsh [42], p. 355, Sec. 11.4) that if g is a function of bounded variation, it can be expressed as

$$g = g_1 - g_2 \tag{c}$$

where g_1, g_2 are monotone non-decreasing. I.e., the cone P_y is normal.

IIIa.5.2. The condition (17″) may be written as

$$\mathscr{Y}^* = P_y^\oplus - P_y^\oplus\,. \tag{18}$$

Now suppose

$$x^* \in T^*(\mathscr{Y}^*)\,, \tag{19}$$

i.e.,

$$x^* = T^*(y^*) \text{ for some } y^* \in \mathscr{Y}^*\,. \tag{20}$$

Then, by using (17″), we have

$$x^* = T^*(y_1^* - y_2^*) \qquad (y_i^* \in P_y^\oplus,\ i = 1, 2)\,, \tag{21}$$

i.e.,

$$x^* = x_1^* - x_2^* \tag{22}$$

where

$$x_i^* = T^*(y_i^*) \qquad (y_i^* \in P_y^\oplus,\ i = 1, 2)\,, \tag{23}$$

so that, by definition of Z_T (cf. III.2 (2)),

$$x_i^* \in Z_T \qquad (i = 1, 2)\,, \tag{24}$$

[44] Banach [3], Section 4.1, pp. 59–61.

i.e.,

$$x^* \in Z_T - Z_T ;$$

hence

(25) $$T^*(\mathscr{Y}^*) \subseteq Z_T - Z_T .$$

On the other hand, let

(26) $$x^* \in Z_T - Z_T .$$

Then the relations (22), (23) hold for some $y_i^* \in P_y^{\oplus}$ $(i = 1, 2)$, and hence (20) holds for $y^* = y_1^* - y_2^*$, so that (19) follows and

(27) $$T^*(\mathscr{Y}^*) \supseteq Z_T - Z_T .$$

(Note that (27) holds even if P_y is not assumed normal.) Equations (25) and (27) together yield

(28) $$T^*(\mathscr{Y}^*) = Z_T - Z_T .$$

IIIa.5.3. Consider now the case when Z_T is regularly convex. We know (Theorem III.4) that in this case

(29) $$Z_T = V_T .$$

But then, from (14) and (28) we have

(30) $$T^*(\mathscr{Y}^*) = F_T .$$

When (30) holds, Hausdorff ([18], p. 307) says that the equation $x^* = T^*(y^*)$ is *normally solvable*; he calls the equation $y = T(x)$ *normally solvable* if and only if

(31) $$T(\mathscr{X}) = F_{T^*} ,$$

where

(32) $$F_{T^*} = \{y : \text{for all } y^* \in \mathscr{Y}^*, T^*(y^*) = 0 \text{ implies } y^*(y) = 0\} .$$

Hausdorff shows (*ibid.*, Theorem X, pp. 308, 310) that in Banach spaces the following four properties are equivalent: the normal solvability of $x^* = T^*(y^*)$, the normal solvability of $y = T(x)$, the closedness of $T(\mathscr{X})$, and the closedness of $T^*(\mathscr{Y}^*)$, i.e.,

(33) $$(30) \Leftrightarrow (31) \Leftrightarrow T(\mathscr{X}) \text{ closed} \Leftrightarrow T^*(\mathscr{Y}^*) \text{ closed.}$$

IIIa.5.4. Now, under the assumption that P_y is normal and Z_T regularly convex, we have obtained (30). It follows from (33) that both $T(\mathscr{X})$ and $T^*(\mathscr{Y}^*)$ are closed.

The example below[45] shows that Z_T need not be regularly convex when P_y is normal. This is of importance, since it shows that the assumption of regular convexity in the theorems in IV is not automatically satisfied.

Let $\mathscr{X} = C[0, 1]$ and

$$y = T(x)$$

where

[45] Closely related to one suggested by Professor B. Gelbaum.

$$y(t) = \int_0^t x(t)dt \ .$$

Then y is absolutely continuous, hence continuous, and we may take $\mathscr{Y} = C[0, 1]$ also, As noted earlier, we may define $y \geqq 0_y$ to mean $y(t) \geqq 0, 0 \leq t \leq 1$ in which case P_y is normal. Now take any function $y_0 \in C[0, 1]$ which is not absolutely continuous (e.g., the one given by Titchmarsh [42], Sec. 11.72, p. 366). Then y_0 is not[46] in the range $T(\mathscr{X})$. But y_0 is a strong (uniform) limit of a sequence of polynomials,[47] hence y_0 is an element of the closure of $T(\mathscr{X})$. Hence $T(\mathscr{X})$ is not closed, hence (by (33)), eq. (30) fails, so that Z_T cannot be regularly convex.

IIIa.5.5. Consider now the special case when

(34) $$P_y = \{0_y\} \ .$$

(P_y is (vacuously) normal, but this fact is of no relevance in what follows.) Then

(35) $$P_y^{\oplus} = \mathscr{Y}^* \ .$$

In this case we have (cf. III, eq. (3))

(36) $$Z_T = T^*(\mathscr{Y}^*) \ .$$

Also, using (13), we get

(37) $$V_T = F_T$$

since

(38) $$V_T = - V_T \ .$$

[Let $x^* \in V_T$. Then $T(x) \geqq 0_y$ implies $x^*(x) \geqq 0$. But, for $P_y = \{0_y\}$, $y \geqq 0_y$ is equivalent to $-y \geqq 0_y$; hence $T(x) \geqq 0_y$ implies $T(-x) \geqq 0_y$ which in turn yields $x^*(-x) \geqq 0$ or $-x^*(x) \geqq 0$. The latter relation means that $-x^* \in V_T$. Hence $V_T \subseteqq (-V_T)$. That $(-V_T) \subseteqq V_T$ is shown in the same fashion.]

Now suppose that

(39) $$Z_T = V_T \ .$$

This is equivalent to

(40) $$F_T = T^*(\mathscr{Y}^*) \ ,$$

i.e., Hausdorff's normal solvability of the equation $x^* = T^*(y^*)$.

By Theorem III.4, (39) implies that $Z_T = F_T = T^*(\mathscr{Y}^*)$ is regularly convex; hence (cf. II.1.4.1) $T^*(\mathscr{Y}^*)$ is closed in the weak* topology, hence it is (strongly) closed. Thus we have obtained Hausdorff's result (part of his Theorem X), viz., that the normal solvability implies the closure of $T^*(\mathscr{Y}^*)$, as a special case of our Theorem III.4. On the other hand, suppose the space $\mathscr{X}$ to be reflexive[48] and let $T^*(\mathscr{Y}^*)$ be

[46] Cf. Titchmarsh [42], Section 11.71, p. 364.

[47] The "Weierstrass Theorem," cf. Rudin [40], Section 7.24, p. 131.

[48] Cf. II. 1.4.

closed. In this case (cf. II.1.4) regular convexity is equivalent to regular° convexity and the latter is always equivalent to closure with convexity. Hence, since $T^*(\mathscr{Y}^*)$ is closed and convex, it is regularly convex and this implies, by Theorem III.4, the equalities (39) and (40).

I.e., we have shown, as a special case of our results in III, when $\mathscr{X}$ is reflexive, the (strong) closure of $T^*(\mathscr{Y}^*)$ is a sufficient condition for the normal solvability of the equation $x^* = T^*(y^*)$ which is also a part of Hausdorff's Theorem X.

IV. Further Theorems on Linear Inequalities

IV.1. In IV all spaces are assumed locally convex linear. Products of topological spaces are understood to be linear topological products, hence the product spaces are also locally convex linear.

IV.2. Let U denote a linear continuous transformation on $\mathscr{X}$ into $\mathscr{Z}$. We introduce the transformation T (which is easily seen to be linear and will also be shown to be continuous) on $\mathscr{X}$ into the product space $\mathscr{Y} = \mathscr{Z} \times \mathscr{X}$ defined by

$$(1') \qquad T(x) = (U(x), x) \qquad \text{for all } x \in \mathscr{X} .$$

In the notation of the type used in matrix calculus we may write

$$(1'') \qquad T = \begin{pmatrix} U \\ \\ I \end{pmatrix},$$

where $I(x) = x$, for all $x \in \mathscr{X}$. (I.e., I is the identity transformation in $\mathscr{X}$.)

If P_x, P_z are convex cones in $\mathscr{X}$ and $\mathscr{Z}$ respectively, and $x' \geqq x''$, $z' \geqq z''$ mean $x' - x'' \in P_x, z' - z'' \in P_z$ respectively, then for $y = (z, x)$, we write $y' \geqq y''$ if and only if $y' - y'' \in P_y$, where

$$(2) \qquad P_y = P_z \times P_x = \{(z, x) : z \geqq 0, x \geqq 0\} .$$

It may be noted that if P_z and P_x are closed, then so is P_y (cf. II.6.1).

IV.3. Theorem IV.1.

A. *Let $\mathscr{X}$ be a linear topological space, $\mathscr{Z}$ locally convex, U a linear continuous transformation on $\mathscr{X}$ to $\mathscr{Z}$, P_x and P_z closed convex cones in $\mathscr{X}$ and $\mathscr{Z}$ respectively, and assume that the set*

$$(3) \qquad X_T^* = \{x^* \in \mathscr{X}^* : x^* = T^*(y^*), y^* \geqq 0\}$$

is regularly convex.

B. *It follows that, for any $x^* \in \mathscr{X}^*$, if*

$$(4) \qquad U(x) \geqq 0, x \geqq 0 \qquad \textit{imply } x^*(x) \geqq 0 \qquad \textit{for all } x \in \mathscr{X},$$

then there exists a $z_0^* \geqq 0$ such that

(5) $$z_0^*[U(x)] \leqq x^*(x) \quad \textit{for } x \geqq 0 ,$$

and

(6) $$x^*(x) = 0,\ U(x) \geqq 0,\ x \geqq 0 \quad \textit{imply } z_0^*[U(x)] = 0 .$$

PROOF. (5) may be rewritten as

(4′) $$T(x) \geqq 0 \quad \text{implies } x^*(x) \geqq 0 \quad \text{for all } x \in \mathscr{X} .$$

Furthermore, T is continuous in x.[49]

Since X_T^* is assumed regularly convex, Theorem III.4 yields a functional $y_0^* \geqq 0$ such that

(7) $$x^*(x) = y_0^*[T(x)] \quad \text{for all } x \in \mathscr{X} .$$

Now, since

(8) $$y = (z, x) = (z, 0) + (0, x) ,$$

we have

(9) $$y_0^*(y) = y_0^*((z, 0)) + y_0^*((0, x)) .$$

We shall write

(10.1) $$y_0^*((z, 0)) = z_0^*(z) \quad \text{for all } z \in \mathscr{Z} ,$$

(10.2) $$y_0^*((0, x)) = x_0^*(x) \quad \text{for all } x \in \mathscr{X} ,$$

where $y_0^*((z, 0))$ is continuous in z and $y_0^*((0, x))$ is continuous in x. Since $z \geqq 0$ implies $(z, 0) \geqq 0$ and $x \geqq 0$ implies $(0, x) \geqq 0$, it follows that, for z_0^*, x_0^* defined by (10), $y_0^* \geqq 0$ yields

(11.1) $$z_0^* \geqq 0 ,$$

(11.2) $$x_0^* \geqq 0 .$$

Thus

(12) $$\begin{aligned} x^*(x) &= y_0^*[T(x)] = y_0^*[(U(x), x)] \\ &= z_0^*[U(x)] + x_0^*(x) \quad \text{for all } x \in \mathscr{X} . \end{aligned}$$

Since $x_0^* \geqq 0$, (5) follows.

Now let x_1 satisfy the hypotheses of (6), i.e.,

(13.1) $$x^*(x_1) = 0$$

and

(13.2) $$T(x_1) \geqq 0 .$$

Equations (13.1) and (5) yield

(14) $$z_0^*[U(x_1)] \leqq 0 .$$

[49] We have $T(x) = (U(x), I(x))$ where $I(x) = x$ for all $x \in \mathscr{X}$. Then (cf. Lefschetz [34], p. 7 (8.2)), T is continuous if every inverse image of a member of a sub-base in $\mathscr{Y} = \mathscr{Z} \times \mathscr{X}$ is open. Such a sub-base is given (cf. Lefschetz [34], p. 10) by the collection of sets

$$Y' = \{y' = (z_1', x') \colon z' \in N_z,\ x' \in \mathscr{X}\} , \qquad Y'' = \{y'' = (z'', x'') \colon z'' \in \mathscr{Z},\ x'' \in N_x\}$$

where N_z, N_x are open sets in $\mathscr{Z}$ and $\mathscr{X}$ respectively. Now the inverse image $T^{-1}(Y') = \{x \colon T(x) \in Y'\} = \{x \colon U(x) \in N_z,\ I(x) \in \mathscr{X}\} = \{x \colon U(x) \in N_z\} = U^{-1}(N_z)$ which is open since N_z is open and U continuous. Similarly $T^{-1}(Y'') = \{x \colon U(x) \in \mathscr{Z},\ I(x) \in N_x\} = N_x$ which is open.

On the other hand, since $U(x_1) \geqq 0$, and $z_0^* \geqq 0$,

(15) $$z_0^*[U(x_1)] \geqq 0 .$$

Equations (14) and (15) yield the conclusion of (6).

IV.4. Let all hypotheses under A in Theorem IV.1 hold, except that X_T^* is not assumed regularly convex while $\mathscr{X}$ and $\mathscr{Z}$ are taken to be normed spaces. Suppose there exists a $z_0^* \geqq 0$ such that (5) holds. Then define

(16) $$\varphi(x) = x^*(x) - z_0^*[U(x)] \qquad \text{for all } x \in \mathscr{X} .$$

Clearly φ is linear, and, because of (5),

(17) $$x \geqq 0 \qquad \text{implies } \varphi(x) \geqq 0 .$$

Also, φ is bounded, since, for any $x \in \mathscr{X}$,

(18) $$\begin{aligned} |\varphi(x)| = |x^*(x) - z_0^*[U(x)] &\leqq |x^*(x)| + |z_0^*[U(x)]| \\ &\leqq \|x^*\| \cdot \|x\| + \|z_0^*\| \cdot \|U\| \cdot \|x\| \\ &= (\|x^*\| + \|z_0^*\| \cdot \|U\|)\|x\| . \end{aligned}$$

Thus

(19) $$\varphi \in \mathscr{X}^*, \qquad \varphi \geqq 0_x^* .$$

Now define

(20) $$\psi(y) = \psi((z, x)) = z_0^*(z) + \varphi(x) \qquad \text{for all } z \in \mathscr{Z} \text{ and all } x \in \mathscr{X} ,$$

which is linear in y, since

(21.1) $$\begin{aligned} \psi(\alpha y) = \psi(\alpha(z, x)) = \psi((\alpha z, \alpha x)) &= z_0^*(\alpha z) + \varphi(\alpha x) \\ &= \alpha z_0^*(z) + \alpha\varphi(x) = \alpha\psi(y) \end{aligned}$$

and

(21.2) $$\begin{aligned} \psi(y' + y'') = \psi((z' + z'', x' + x'')) &= z_0^*(z' + z'') + \varphi(x' + x'') \\ &= z_0^*(z') + z_0^*(z'') + \varphi(x') + \varphi(x'') \\ &= \psi(y') + \psi(y'') . \end{aligned}$$

Also,

(22) $$y \geqq 0 \text{ implies } \psi(y) \geqq 0$$

since if $(z, x) \geqq 0$ then $z \geqq 0$ and $x \geqq 0$ and both z_0^* and φ are non-negative functionals.

Finally, ψ is continuous. For let $(z_n, x_n) = y_n \to y_0 = (z_0, x_0)$, $n = 1, 2, \cdots$. Then, by II.6.1, eq. (1), $z_n \to z_0$ and $x_n \to x_0$. Hence, since z_0^* and φ are continuous, $z_0^*(z_n) \to z_0^*(z_0)$ and $\varphi(x_n) \to \varphi(x_0)$, and therefore, $\varphi(y_n) \to \varphi(y_0)$.

Hence

(23) $$\psi \in \mathscr{Y}^* , \qquad \psi \geqq 0_y^* .$$

Because of (16), we have

(24) $$x^*(x) = z_0^*[U(x)] + \varphi(x) \qquad \text{for all } x \in \mathscr{X} ,$$

i.e., by (20),

$$(25)\qquad \begin{aligned} x^*(x) &= \psi[(U(x)), x)] \\ &= \psi[T(x)] \qquad \text{for all } x \in \mathscr{X}\,, \end{aligned}$$

or

$$(26)\qquad x^* = T^*(\psi)\,, \qquad \psi \geqq 0_y^*\,.$$

Now (26) holds for all $x^* \in \mathscr{X}^*$; it follows from Theorem III.4 that the set X_T^* is regularly convex. Thus we have shown that, at least in normed spaces, given the other hypotheses under A in Theorem IV.1, *the assumption of regular convexity of X_T^* is necessary* (as well as sufficient) for the validity of the conclusions. We may state this as

THEOREM IV.2.

A. *Let $\mathscr{X}$ and $\mathscr{Z}$ be normed spaces, U a linear bounded transformation on $\mathscr{X}$ to $\mathscr{Z}$, P_x and P_z closed convex cones in $\mathscr{X}$ and $\mathscr{Z}$ respectively. Then the condition that the set*

$$(27)\qquad X_T^* = \{x^* \in \mathscr{X}^* : x^* = T^*(y^*), y^* \geqq 0\}$$

be regularly convex is equivalent to the following: for any $x^* \in \mathscr{X}^*$, if

$$(28)\qquad U(x) \geqq 0, x \geqq 0 \quad \textit{imply } x^*(x) \geqq 0 \quad \textit{for all } x \in \mathscr{X}\,,$$

then there exists a $z_0^* \geqq 0$ such that

$$(29)\qquad z_0^*[U(x)] \leqq x^*(x) \quad \textit{for } x \geqq 0$$

and

$$(30)\qquad x^*(x) = 0, U(x) \geqq 0, x \geqq 0 \quad \textit{imply } z_0^*[U(x)] = 0\,.$$

IV.5. The following result generalizes Theorem IV.1 to situations where non-homogeneous inequalities appear.

THEOREM IV.3.

A. *Let all the hypotheses under* A *in Theorem* IV.1 *hold, the transformation T being defined in* (37), (38) *below.*

B. *It follows that if, for some $\bar{x} \in \mathscr{X}$,*

$$(31)\qquad \bar{x} \geqq 0 \textit{ and } U(\bar{x}) - a \geqq 0\,,$$

and if, for some $x^* \in \mathscr{X}^*$,

$$(32)\qquad x \geqq 0 \textit{ and } U(x) - a \geqq 0 \quad \textit{imply } x^*(x) - \beta \geqq 0\,,$$

then there exists a $z_0^* \geqq 0$ such that

$$(33)\qquad z_0^*[U(x) - a] \leqq x^*(x) - \beta \quad \textit{for } x \geqq 0$$

and

$$(34)\qquad x^*(x) = \beta, U(x) - a \geqq 0, x \geqq 0 \quad \textit{imply } z_0^*[U(x) - a] = 0$$

PROOF. Consider the product space

$$(35)\qquad \mathscr{W} = \{w : w = (\rho, x), \rho \text{ real}, x \in \mathscr{X}\}$$

and the linear transformation

$$(36)\qquad P(w) = P((\rho, x)) = -a\rho + U(x)\,.$$

on $\mathscr{W}$ into $\mathscr{Z}$.

Then define

$$T = \begin{pmatrix} P \\ I \end{pmatrix}, \qquad I(w) = w \qquad \text{for all } w \in \mathscr{W}, \tag{37}$$

i.e.,

$$T(w) = (P(w), w) \qquad \text{or} \qquad T((\rho, x)) = (-a\rho + U(x), (\rho, x)) \tag{38}$$

and T is a linear transformation on $\mathscr{W}$ into $\mathscr{Z} \times \mathscr{W}$.

Now suppose we have shown that

$$T(w) \geqq 0 \qquad \text{implies } w^*(w) \geqq 0 \qquad \text{for all } w \in \mathscr{W} \tag{39}$$

where we define w^* by

$$w^*(w) = w^*((\rho, x)) = -\beta\rho + x^*(x) . \tag{40}$$

One can ascertain easily that Theorem IV.1 applies, with w replacing x, w^* replacing x^*, and P replacing U.

Hence there exists a $z_0^* \geqq 0$ such that

$$z_0^*[P(w)] \leqq w^*(w) \qquad \text{for } w \geqq 0 \tag{41}$$

and

$$w^*(w) = 0, T(w) \geqq 0 \qquad \text{imply} \qquad z_0^*[P(w)] = 0 . \tag{42}$$

Equation (41), written out explicitly, yields, by (36) and (40),

$$z_0^*[-a\rho + U(x)] \leqq -\beta\rho + x^*(x) \qquad \text{for } \rho \geqq 0, x \geqq 0 . \tag{43}$$

Letting $\rho = 1$ we obtain (33).

Similarly, using (36), (40), and (38) in (42), and putting $\rho = 1$, we obtain (34).

Therefore, it remains to establish (39) which, written out explicitly, states that

$$\left.\begin{aligned} -a\rho + U(x) &\geqq 0 \\ \rho &\geqq 0 \\ x &\geqq 0 \end{aligned}\right\} \text{ imply } \; -\beta\rho + x^*(x) \geqq 0 . \tag{44}$$

Suppose (44) is false. Then the hypotheses of (44) must hold and the conclusion fail for some $\rho_0 \geqq 0$, $x_0 \geqq 0$. We shall first consider the case $\rho_0 > 0$. I.e., we have

$$\begin{aligned} -a\rho_0 + U(x_0) &\geqq 0 \\ \rho_0 &> 0 \\ x_0 &\geqq 0 \end{aligned} \tag{45.1}$$

and

$$-\beta\rho_0 + x^*(x_0) < 0 , \tag{45.2}$$

so that

$$\begin{aligned} -a + U\left(\frac{x_0}{\rho_0}\right) &\geqq 0 \\ \frac{x_0}{\rho_0} &\geqq 0 \end{aligned} \tag{46.1}$$

and

$$-\beta + x^*\left(\frac{x_0}{\rho_0}\right) < 0 \; . \tag{46.2}$$

This, however, violates (32). Hence the implication in (44) has been established for $\rho > 0$. We shall now take up the case $\rho = 0$. I.e., we must show that

$$\left.\begin{aligned} U(x) &\geqq 0 \\ x &\geqq 0 \end{aligned}\right\} \text{ imply } x^*(x) \geqq 0 \; . \tag{47}$$

Let x_1 satisfy the hypotheses of (47) and take a real $\lambda > 0$. Then, by (31),

$$-[a - U(\bar{x})]\lambda + U(x_1) \geqq 0 \tag{48}$$

and hence

$$-a\lambda + U(x_1 + \lambda\bar{x}) \geqq 0 \; . \tag{49}$$

Note also that

$$x_1 + \lambda\bar{x} \geqq 0 \; . \tag{50}$$

Hence, for $\rho_0 = \lambda$, $x_0 = x_1 + \lambda\bar{x}$, the hypotheses of (45.1) are satisfied, so that

$$-\beta\lambda + x^*(x_1 + \lambda\bar{x}) \geqq 0 \; . \tag{51}$$

We therefore have

$$x^*(x_1) + \lambda[x^*(\bar{x}) - \beta] \geqq 0 \qquad \text{for all } \lambda > 0 \; . \tag{52}$$

Suppose now that

$$x^*(x_1) = -\varepsilon < 0 \; . \tag{53}$$

Then (52) is false for any $\lambda > 0$ if $x^*(\bar{x}) - \beta \leqq 0$.
Hence suppose

$$x^*(\bar{x}) - \beta = \eta > 0 \; . \tag{54}$$

and take $\lambda = \varepsilon/(2\eta)$. Then (52) becomes

$$-\varepsilon + \frac{\varepsilon}{2\eta}\eta > 0 \; , \tag{55}$$

i.e.,

$$-\frac{\varepsilon}{2} > 0 \tag{56}$$

which contradicts (53). Hence

$$x^*(x_1) \geqq 0 \tag{57}$$

which establishes the validity of (47).

IV.6. Consider now the special case of Theorem IV.3, where $P_x = \mathscr{X}$, so that the restriction $x \geqq 0$ is necessarily satisfied for all x. In this case 0_x^* is the only non-negative element of $\mathscr{X}^*$. [For otherwise there would be some $x_0^* \in \mathscr{X}^*$ with $x_0^*(x_0) > 0$ for some $x_0 \in \mathscr{X}$, hence $x_0^*(-x_0) < 0$ even though $-x_0 \in P_x$, which contradicts $x_0^* \geqq 0$.] Now the counterpart of (12) for Theorem IV.3 is

$$-\beta\rho + x^*(x) = z_0^*[-a\rho + U(x)] + \tau_0^*\rho + x_0^*(x) \qquad \text{for all } \rho \text{ and } x \in \mathscr{X} . \tag{58}$$

When $P_x = \mathscr{X}$, it follows that $x_0^*(x) = 0$ and (58) becomes

(59) $\quad -\beta\rho + x^*(x) = z_0^*[-a\rho + U(x)] + \tau_0^*\rho \qquad$ for all ρ and all $x \in \mathscr{X}$.

Putting $\rho = 0$, (59) reduces to

(60) $$x^*(x) = z_0^*[U(x)] \qquad \text{for all } x \in \mathscr{X}.$$

Furthermore, if $U(\hat{x}) - a \geqq 0$ and $x^*(\hat{x}) \leqq x^*(x)$ for all $U(x) - a \geqq 0$, then (34) in Theorem IV.3 yields

(61) $$z_0^*[U(\hat{x}) - a] = 0 .$$

which, by (60), implies

(62) $$x^*(\hat{x}) = z_0^*(a) .$$

Hence, with $\beta \leqq x^*(\hat{x})$ by hypothesis, we have

(63) $$x^*(\hat{x}) = z_0^*(a) \geqq \beta ,$$

as in Dantzig's Corollary ([10], p. 334).

We may state these results as

COROLLARY IV.3.

A. *Let $\mathscr{X}$ and $\mathscr{Z}$ be locally convex linear spaces, U a linear transformation on $\mathscr{X}$ to $\mathscr{Z}$, P_z a closed convex cone in $\mathscr{Z}$, and assume that the set*

(64) $$X_U^* = \{x^* \in \mathscr{X}^* : x^* = U^*(z^*), z^* \geqq 0\}$$

is regularly convex.

B. *It follows that if, for some $\bar{x} \in \mathscr{X}$,*

(65) $$U(\bar{x}) - a \geqq 0 ,$$

and if, for some $x^* \in \mathscr{X}^*$,

(66) $$U(x) - a \geqq 0 \qquad \textit{implies } x^*(x) - \beta \geqq 0 ,$$

then there exists a $z_0^* \geqq 0$ such that

(67) $$z_0^*[U(x)] = x^*(x) \qquad \textit{for all } x \in \mathscr{X} ;$$

furthermore

(68) $$\min_{U(x) \geqq a} x^*(x) = z_0^*(a) \geqq \beta .$$

It will be noted that, in Banach spaces at least, the assumption of regular convexity of X_U^* is necessary as well as sufficient if U is bounded; this follows from Theorem IV.2.

In finite-dimensional Euclidean spaces the requirement of regular convexity of X_U^* is necessarily satisfied (cf. III.9) and if $z \geqq 0$ means, as usual, that each of its coordinates is non-negative, then P_z is closed. Hence the hypotheses of the regular convexity of X_U^* and the closure of P_z may be omitted, and we obtain, as a special case of Corollary IV.3, the Lemma (and its Corollary) stated by Dantzig in [10], p. 334. This, of course, suggests the possibility of generalizing the Dantzig result on the equivalence of linear programming and game problems, since the Lemma plays a crucial role in Dantzig's proof and Corollary IV.3 above provides its generalization.

V. The Lagrangian Saddle-Point Theorem

V.1. *Isotone Lagrangian saddle-point implies maximality.*

V.1.0. Contrary to the customary sequence, we find it more convenient to start with the theorem indicated in the title, rather than with one in which the implication goes in the opposite direction. This is done in order that the reason for requiring that the functional on $\mathscr{Y}$ be strictly isotone and that on $\mathscr{Z}$ isotone may become more readily apparent.

V.1.1. Let π_{11} denote the partial ordering maximization problem obtained if in π_1 of II.3.3 the following two requirements are added:

(1) $\mathscr{Z}$ is a locally convex linear topological space;

(2) P_z is closed.

V.1.2. THEOREM V.0. *Let the generalized Lagrangian expression* $\Phi_{\pi_{11}}(x, z^*\,;\,\eta_0^*)$ *have an isotone saddle-point at* $(x_0, z_0^*\,;\,\eta_0^*)$. *Then* x_0 *is maximal.*

For the sake of convenience, we give a more explicit statement of the preceding theorem.

THEOREM V.1. *Let* $x \in \mathscr{X}$, *while the values of* $f(x)$ *and* $g(x)$ *are in* $\mathscr{Y}$ *and* $\mathscr{Z}$, *respectively, where* $\mathscr{Y}$ *is ordered by a relation written as* $\geqq$ *and* $\mathscr{Z}$ *is a locally convex, linear topological space such that* $z \geqq 0$ *means* $z \in P_z$, P_z *being a closed convex cone.*

Write

(1) $$\Phi(x, z^*) = \eta_0^*[f(x)] + z^*[g(x)]$$

where η_0^* *is a strictly isotone functional on* $\mathscr{Y}$ *and* z^* *is linear continuous functional on* $\mathscr{Z}$. *Here* $z^* \geqq 0$ *means that* $z^*(z) \geqq 0$ *for all* $z \geqq 0$.

Suppose that, for some $x_0 \in X$, $z_0^* \geqq 0$, $(X \subseteq \mathscr{X})$, *we have*

(2) $$\Phi(x, z_0^*) \leqq \Phi(x_0, z_0^*) \leqq \Phi(x_0, z^*), \qquad \textit{for all } x \in X \textit{ and all } z^* \geqq 0\,.$$

Then

(3.1) $$g(x_0) \geqq 0$$

and, for all $x \in X$,

(3.2) $$g(x) \geqq 0, f(x) \geqq f(x_0) \qquad \textit{imply } f(x_0) \geqq f(x).$$

V.1.3. PROOF. The right-hand inequality in (2) implies that

(4) $$z_0^*[g(x_0)] \leqq z^*[g(x_0)] \qquad \text{for all } z^* \geqq 0\,;$$

hence, in particular,

(5) $$z_0^*[g(x_0)] \leqq (z_0^* + z_1^*)[g(x_0)] \qquad \text{for all } z_1^* \geqq 0\,,$$

since $z_0^* + z_1^* \geqq 0$ if $z_1^* \geqq 0$. But (5) gives

(6) $$0 \leqq z_1^*[g(x_0)] \qquad \text{for all } z_1^* \geqq 0\,,$$

and (3.1) follows from Lemma II.2 in II.2.5 above, based on the Mazur-Bourgin Theorem.

We shall now show the validity of (3.2). Equation (2) yields

(7) $\Phi(x, z_0^*) \leqq \Phi(x_0, z^*)$ for all $x \in X$ and all $z^* \geqq 0$.

Using $z^* = 0_z^*$, this gives

(8) $\eta_0^*[f(x)] + z_0^*[g(x)] \leqq \eta_0^*[f(x_0)]$ for all $x \in X$.

Now let $x' \in X$ be such that the hypotheses of (3.2) are satisfied, i.e.,

(9.1) $$g(x') \geqq 0$$

and

(9.2) $$f(x') \geqq f(x_0) .$$

Suppose that the conclusion of (3.2) is false, i.e.,

(9.3) $$f(x_0) \ngeqq f(x') .$$

Since η_0^* is strictly isotone, (9.2) and (9.3) together imply (cf. II.2.1)

(10) $$\eta_0^*[f(x')] > \eta_0^*[f(x_0)] .$$

Also, since $z_0^* \geqq 0$, (9.1) gives

(11) $$z_0^*[g(x')] \geqq 0 .$$

From (10) and (11) it follows that

(12) $$\eta_0^*[f(x')] + z_0^*[g(x')] > \eta_0^*[f(x_0)] ,$$

which contradicts (8). Hence (9.3) is false and (3.2) follows.

It is important to note, that, in this proof, it would not have been sufficient to assume η_0^* isotone rather than strictly isotone (cf. V.2.6).

V.2. *Scalarization.*

V.2.1. Let $\mathscr{W}$ be a topological linear space and K a convex cone in $\mathscr{W}$. Then there always exists a linear continuous functional non-negative on K, since the null functional [$\psi(w) = 0$ for all $w \in \mathscr{W}$] has this property. However, even with additional assumptions on K (viz., that it is closed and pointed[50]), there may not exist any continuous linear functional strictly positive on K, as shown by example in Krein and Rutman ([29], pp. 21—22). On the other hand, it has been shown (*ibid.*, Theorem 2.1, p. 21) that if K is a closed pointed convex cone and $\mathscr{W}$ a separable Banach space, a linear continuous (equals bounded, in this case) functional strictly positive on K does exist. It is shown below that the requirement of pointedness can be removed (Lemma V.2.2 in V.2.5 below). ["Strictly positive" is defined in II.2.4.]

When a strictly positive functional exists, it may be used to "scalarize" the Lagrangian problem. It has been pointed out in V.1.2 that a non-negative functional is not adequate for our purposes (cf. also V.2.6 below).

V.2.2. Let $P_y \subseteq \mathscr{Y}$ be a convex cone in the linear topological space $\mathscr{Y}$. We write $y' \geqq y''$ if and only if $y' - y'' \in P_y$. $\hat{Y}$ denotes the $\geqq$ —maximal subset of the given permissible set Y.

[50] K is said to be *pointed* if $0_w \neq w \in K$ implies $-w \notin K$.

An element[51] $y_0 \in Y$ may have the property that there exists a $y^*_{y_0} = y^*_0 > 0$ such that

(1) $$y \in Y \text{ implies } y^*_0(y) \leqq y^*_0(y_0) .$$

In the light of the remarks in V.2.1, such a y^*_0 will not always exist in infinite-dimensional spaces. But even in the two-dimensional Euclidean space, where every closed convex cone does possess a strictly positive linear continuous functional, and with P_y chosen as the non-negative quadrant, the required y^*_0 may not exist for some y_0. This, of course, is not surprising since so far nothing has been assumed about the set Y. But even if (as is natural in certain problems) one were to assume Y to be convex, closed, and even bounded, $y^*_0 > 0$ may not exist for certain elements of $\hat{Y}$.[52]

Let $\hat{\hat{Y}}$ denote the subset of $\hat{Y}$ such that $y_0 \in \hat{\hat{Y}}$ if and only if there exists a y^*_0 such that (1) holds.

We shall now formulate a necessary condition for membership in $\hat{\hat{Y}}$.

Suppose that there exists a $y^*_0 > 0$ such that (1) holds for a given $y_0 \in \hat{Y}$. Then

(2) $$y^*_0(y - y_0) \leqq 0 \qquad \text{for all } y \in Y ,$$

i. e.,

(3) $$y^*_0(y) \leqq 0 \qquad \text{for all } y \in Y - y_0 .$$

Furthermore, by definition of strict positiveness,

(4) $$y^*_0(y) \geqq 0 \qquad \text{for all } y \geqq 0$$

and

(5) $$y^*_0(y) > 0 \qquad \text{for all } y \geq 0 .$$

Putting

(6) $$y^*_1 = -y^*_0 ,$$

we may rewrite (3), (4), and (5) as

(7.1) $$y^*_1(y) \geqq 0 \qquad \text{for all } y \in Y - y_0 ,$$

(7.2) $$y^*_1(y) \geqq 0 \qquad \text{for all } y \leqq 0,$$

(7.3) $$y^*_1(y) > 0 \qquad \text{for all } y \leq 0 ,$$

respectively.

Now consider the intersection K_0 of all convex cones containing the set

(8) $$(Y - y_0) \cup (- P_y) .$$

Clearly, K_0 is a convex cone, and furthermore

(9) $$y^*_1(y) \geqq 0 \qquad \text{for all } y \in K_0 .$$

[This follows from the fact that the set $\{y : y^*_1(y) \geqq 0\}$ is a convex cone

---

[51] The element y_0 in this context need not be $\geqq$ $-$maximal: cf. Theorem V.2.3.

[52] Arrow's example: $Y = \{y\colon y = (y_1, y_2), y_i \geqq 0, y_1^2 + y_2^2 \leqq 1\}$, $y_0 = (0, 1)$. Cf. also Kuhn and Tucker [31], p. 488, example.

which, by (7.1) and (7.2), contains $(Y - y_0) \cup (-P_y)$, and hence, by definition of K_0, it includes K_0.]

Writing $\overline{K_0}$ to denote the closure of K_0, we also have

(10) $$y_1^*(y) \geqq 0 \quad \text{for all } y \in \overline{K_0}\ ,$$

by continuity of y_1^*.

[For linear normed (hence Banach) spaces, this has been noted in Krein and Rutman ([29], pp. 16–17). When $\mathscr{Y}$ is a linear topological space, (10) is proved as follows: Let $y' \in \overline{K_0}$ and suppose $y_1^*(y) < 0$, say $y_1^*(y) = -\alpha$. The inverse image by y_1^* of the open interval $(-3\alpha/2, -\alpha/2)$ is an open set containing y', hence containing at least one point, say y'' of K_0. Thus $y_1^*(y'') < -\alpha/2 < 0$, which contradicts (9).]

Now suppose that there exists an element y' with

(11.1) $$y' \in \overline{K_0}$$

and

(11.2) $$y' \geq 0\ .$$

Define

(12) $$y'' = -y'\ ,$$

so that

(13) $$y'' \leq 0\ .$$

Then, by (7.3),

(14) $$y_1^*(y'') > 0\ .$$

But because of (11.1) and (10),

(15) $$y_1^*(y') \geqq 0\ ,$$

hence

(16) $$y_1^*(y'') \leqq 0\ ,$$

which contradicts (14). Hence we have

THEOREM V.2.1. *Let $\mathscr{Y}$ be a linear topological space and P_y a convex cone. If, for[53] $y_0 \in Y$, there exists $y_0^* > 0$ such that* (1) *holds, then the set $\overline{K_0}$ [the closure of the intersection of all the convex cones containing the set $(Y - y_0) \cup (-P_y)$] does not contain any y' such that $y' \geq 0$.*

V.2.3. *Definition.* If $\overline{K_0}$ contains no $y' \geq 0$ and y_0 is $\geqq$ −maximal, y_0 is said to be *properly maximal.*

V.2.4. THEOREM V.2.2. *Let $\mathscr{Y}$ be a linear topological space with the property that for every closed convex cone $K \subseteq \mathscr{Y}$, there is a linear continuous functional $y^* \in \mathscr{Y}^*$ strictly positive on K.*

Then, for every y_0 properly maximal, there exists a $y_0^* > 0$ such that (1) *is satisfied.*

PROOF. By hypothesis, there exists y_1^* strictly positive on $\overline{K_0}$. Then

[53] In this Theorem, y_0 need not be $\geqq$ −maximal. But Theorem V.2.3 asserts that y_0 must be $\geqq$ −maximal.

(7.1) and (7.2) are satisfied because $\overline{K}_0$ contains the sets $Y - y_0$ and $-P_y$ and because $y_1^* \in (\overline{K}_0)^{\oplus}$. Now take $y' \leq 0$. Then $y' \in \overline{K}_0$. Suppose $-y' \in \overline{K}_0$ also. Then, since y_0 is properly maximal, $-y' \not\geq 0$, which contradicts $y' \leq 0$. Hence, $-y' \notin \overline{K}_0$. But then, because of $y_1^* > 0$, $y_1^*(y') > 0$; i. e., (7.3) also holds. It is seen that

$$y_0^* = -y_1^* \tag{17}$$

has the required property.

V.2.5. COROLLARY V.2.2. *Let $\mathscr{Y}$ be a separable linear normed space and y_0 properly maximal. Then there exists a $y_0^* > 0$ such that* (1) *holds.*

PROOF. In view of Theorem V. 2. 2, it will suffice to prove the following:

LEMMA V.2.2. *For every closed convex cone K in a linear normed separable space $\mathscr{Y}$, there is a linear bounded functional strictly positive on K.*

To prove the Lemma, we first note that Theorem 2.1, p. 21, in Krein and Rutman [29] is precisely equivalent to our Lemma for the case where K is pointed, i.e., where $0 \neq w \in K$ implies $-w \notin K$. Hence, it is sufficient to show that the Krein-Rutman proof can be extended to cover the case of K not assumed pointed.

Now the pointedness of K is not used in the Krein-Rutman proof in reaching the conclusion that there exists a y_0^* (in their notation f_0) such that

$$y_0^* \in K^{\oplus} \tag{18}$$

and

$$y_0 \in K,\ y_0^*(y_0) = 0 \qquad \text{imply } y^*(y_0) = 0 \qquad \text{for all } y^* \in K^{\oplus}\,. \tag{19}$$

We may now use Theorem 1.4, p. 17, in Krein and Rutman [29] which asserts[54] that for every $w_0 \in C$ where C is a closed convex cone in $\mathscr{W}$ and $-w_0 \notin C$, there exists $w_0^* \in C^{\oplus}$ such that $w_0^*(w_0) > 0$. This Theorem, together with (19), yields the conclusion that

$$y_0 \in K,\ y_0^*(y_0) = 0 \qquad \text{imply } -y_0 \in K \tag{20}$$

which, with (18), makes y_0^* strictly positive on K.

V.2.6. THEOREM V.2.3. *Let $\mathscr{Y}$ be a topological linear space and y_0^* strictly positive on the convex cone P_y, and let $y_0 \in Y$ be such that* (1) *holds, i.e., that $y \in Y$ implies $y_0^*(y) \leqq y_0^*(y_0)$. Then y_0 is maximal in Y.*

PROOF. Suppose not. Then, for some $y' \in Y$ we have

$$y' \geq y_0\,. \tag{21}$$

Also, by (1), since $y' \in Y$,

[54] This Theorem follows from the Mazur-Bourgin Theorem (II.1.4.1) whenever $\mathscr{W}$ is locally convex.

(22) $$y_0^*(y' - y_0) \leqq 0 ,$$

while (21) together with $y_0^* > 0$ yields

(23) $$y_0^*(y' - y_0) > 0 .$$

The contradiction between (22) and (23) completes the proof.

It would not have been enough to assume $y_0^* \in P_y^{\oplus}$ (or even $y_0^* \geq 0_y^*$) instead of $y_0^* > 0$. For in that case (21) would only have yielded

(23′) $$y_0^*(y' - y_0) \geqq 0 ,$$

which does not contradict (22) since equality could hold in both. (E.g., in the case of P_y closed, $y' - y_0$ could be a boundary point of P_y with $y_0^*(y' - y_0) = 0$.)

V.2.7. Noting that the hypotheses of Theorems V.2.1 and V.2.3 are identical, we summarize the results of V.2 in Theorem V.2.4.

THEOREM V.2.4. *Let $\mathscr{Y}$ be a linear topological space ordered by the relation $\geqq$ (where $y' \geqq y'$ means $y' - y'' \in P_y$, P_y being a convex cone).*

A. *If there exists $y_0^* > 0$ such that* (1) *holds for some $y_0 \in Y$, then y_0 is properly $\geqq -$maximal in Y.*[55]

B. *If for every closed convex cone $K \subseteq \mathscr{Y}$ there is a linear continuous functional $y^* \in \mathscr{Y}^*$ strictly positive on K (as is, for instance, the case in a separable linear normed space), then for every y_0 properly $\geqq -$maximal there exists a $y_0^* > 0$ such that* (1) *is satisfied.*

V.3. *Maximality implies existence of a saddle-point.*

V.3.1. *Lagrangian saddle-points without differentiability.*

V.3.1.1. The basic idea of the Theorem presented in this section goes back to Slater's paper entitled "Lagrange Multipliers Revisited" [41]. The chief accomplishment of Slater's paper was to establish the existence of a saddle-point for the Lagrangian expression without using differentiability properties in any way whatever, the reliance being placed on the concavity properties of the relevant functions. (A more detailed comparison is given at the end of Part I of the present chapter.) Since the differentiability approach also used the concavity properties, Slater's result was a significant improvement. The present writer extended Slater's result (except for a slight strengthening of Slater's concavity requirements to conform with the usual ones) in a Cowles Commission Discussion Paper (Economics No. 2110) of September 1954. The present version differs significantly from the 1954 version. A suggestion, due to Hirofumi Uzawa, has made it possible not only to simplify the proof tremendously, but also to weaken the assumptions on the functions used (which are merely concave, but not necessarily continuous) and on the underlying spaces.

V.3.1.2. THEOREM V.3.1. *Let $\mathscr{X}$ be a linear system, $\mathscr{Y}$ and $\mathscr{Z}$ linear topological spaces. P_y, P_z are convex cones in $\mathscr{Y}$ and $\mathscr{Z}$ with non-empty*

[55] The point y_0 is $\geqq -$maximal by V.2.3; proper maximality then follows from V.2.1.

interiors, $P_y \neq \mathscr{Y}$, *X a (fixed) convex subset of* $\mathscr{X}$, *f a concave function on X to* $\mathscr{Y}$, *g a concave function on X to* $\mathscr{Z}$. *Let there be a point x_* in X such that*

(1) $g(x_*) > 0$ *(i.e., $g(x_*)$ is an element of the interior of P_z).*

If x_0 maximizes $f(x)$ subject to $g(x) \geqq 0$ and $x \in X$, then there exist linear continuous functionals

(2) $$y_0^* \geq 0 \quad and \quad z_0^* \geqq 0$$

such that, for the Lagrangian expression

(3) $$\Phi(x, z^*) = y^*[f(x)] + z^*[g(x)] ,$$

the saddle-point inequalities

(4) $$\Phi(x, z_0^*) \leqq \Phi(x_0, z_0^*) \leqq \Phi(x_0, z^*)$$

hold for all $x \in X$ and all $z^* \geqq 0$.

(We may note that in applications X is usually a convex cone—e.g., the non-negative orthant of the system $\mathscr{X}$.)

PROOF. Let $\mathscr{W}$ be the topological product space $\mathscr{Y} \times \mathscr{Z}$ and consider the subset of $\mathscr{W}$ defined by

(5) $$A = \{(y, z) : y \in \mathscr{Y}, y \leqq f(x), z \in \mathscr{Z}, z \leqq g(x) \quad \text{for some } x \in X\}.$$

The set A is convex because of the concavity of the functions f and g and the convexity of the set X. Also, A has interior points because P_z and P_y have non-empty interiors.

Consider the point $(f(x_0), 0_z) = w_0$ of the space $\mathscr{W}$. The point w_0 is an element of A, since, by hypothesis, $0_z \leqq g(x_0)$. On the other hand, w_0 does not belong to the interior of A; for if w_0 were interior to A, there would exist an element x in X such that $f(x_0) < f(x)$ and $0_z \leqq g(x)$, which cannot happen because of the assumed maximality of x_0.

Hence, we may apply the Corollary of the Hahn-Banach (Bounding Plane) Theorem (see Corollary II. 1 of II.1.4 above) and obtain a non-null functional w_0^* such that

$$w_0^*(w) \leqq w_0^*(w_0) \quad \text{for all } w \in A .$$

Writing $w_0^* = (y_0^*, z_0^*)$, this implies

(6) $$y_0^*(y) + z_0^*(z) \leqq y_0^*[f(x)] \quad \text{for all } (y, z) \text{ in } A .$$

Since $(f(x), g(x))$, with x in X, belongs to A, we have in particular

(7) $$y_0^*[f(x)] + z_0^*[g(x)] \leqq y_0^*[f(x_0)] \quad \text{for all } x \text{ in } X .$$

Also, since $w_0 = (f(x_0), 0_z)$ is in A, it follows that all ordered pairs of the form $(f(x_0), z)$ are in A if $z \leqq 0_z$, which implies $z_0^*(z) \geqq 0$ for all $z \geqq 0$, i.e.,

(8) $$z_0^* \geqq 0 .$$

Similarly, because w_0 is in A, so are all pairs of the form $(y, 0_z)$ for $y \leqq f(x_0)$; this implies

(9) $$y_0^* \geqq 0 .$$

Now suppose $y_0^* = 0_y^*$ (the null functional). It follows from (7) that $z_0^*[g(x)] \leqq 0$ for all x in X, hence for x_*. Also, since w_0^* is non-null, $z_0^* \geq 0$. But then $z_0^*[g(x_*)] = 0$ because $g(x_*)$ was assumed positive and z_0^* is non-negative. However, since $g(z_*)$ is an interior point of the cone P_z and z_0^* is non-null, it must be that $z_0^*[g(x_*)] > 0$. (This follows from an extension of Prop. 5, Bourbaki [7], p. 75, to the case where the cone need not be pointed and the space is merely assumed linear topological; that this extension is valid follows directly from Bourbaki [7], Prop. 16, p. 52.) Hence we have established that

(10) $$y_0^* \geq 0 \, .$$

Now let $x = x_0$ in (7). It follows that

(11) $$z_0^*[g(x_0)] \leqq 0 \, .$$

On the other hand, since both $g(x_0)$ and z_0^* are non-negative,

(12) $$z_0^*[g(x_0)] \geqq 0 \, ,$$

hence

(13) $$z_0^*[g(x_0)] = 0 \, .$$

Since $z^*[g(x_0)] \geqq 0$ for all $z^* \geqq 0$, (13) implies the right-hand saddle-point inequality, while (7) and (13) yield the left-hand inequality. This completes the proof.

V.3.1.3. A case of particular interest is that of $\mathscr{Y}$ being the space of reals. Here $y_0^* \geq 0$ is equivalent to $y_0^* > 0$ and we have a strictly isotone functional of the type needed in Theorem V.1.1 (saddle-point implies maximality). When $\mathscr{Y}$ is multi-dimensional, however, y_0^* is not strictly isotone but merely isotone, which is inadequate in the context of Theorem V.1, for example, to establish "efficiency" of a given resource allocation.

V.3.1.4. It was shown by Slater that the condition $g(x_*) > 0$ cannot be dispensed with. [In his counter-example, all three spaces are one-dimensional (reals), $f(x) = x - 1$, $g(x) = -(x-1)^2$.] A slight modification of Slater's counter-example shows that the condition $g(x_*) \geq 0$ is not sufficient: we again take $f(x) = x - 1$, $\mathscr{Y}$ two-dimensional, with $g_1(x) = -(x-1)^2$, $g_2(x) = -x + 2$.

V.3.2. *Non-negative Lagrangian saddle-points: the linear non-homogeneous case.*[56]

V.3.2.1. Although linear non-homogeneous situations may be handled by theorems covering the non-linear situations as well, it seems more helpful and simpler to give the direct proofs based on the assumption of linearity.

V.3.2.2. We consider the problem of maximizing the linear non-homogeneous real-valued function on $\mathscr{X}$ to $\mathscr{Y}$ ($\mathscr{Y}$ reals)

[56] We call a function $\varphi(x) + y_0$ on $\mathscr{X}$ to $\mathscr{Y}$ *linear non-homogeneous* if $\varphi(x)$ is linear [i.e., if $\varphi(x)$ is homogeneous and additive]. The possibility that y_0 vanishes is not excluded. ("Affine" might be a more appropriate term.)

$$(1) \qquad f(x) = -x^*(x) + \nu \qquad (x^* \in \mathscr{X}^*)$$

subject to the linear constraints

$$(2.1) \qquad g(x) = U(x) - a \geqq 0_z ,$$

$$(2.2) \qquad x \geqq 0_x ,$$

where U is a linear transformation on $\mathscr{X}$ to $\mathscr{Z}$, it being assumed that $\mathscr{X}$ is a linear topological space and $\mathscr{Z}$ a locally convex space, and the convex cones $P_z = \{z: z \geqq 0_z\}$, $P_x = \{x: x \geqq 0_x\}$ are closed.

In this case, the Lagrangian expression (cf. II.3.3) can be written as[57]

$$(3) \qquad \Phi(x, z^*) = [-x^*(x) + \nu] + z^*[U(x) - a] .$$

V.3.2.3. THEOREM V.3.2. *Let $\mathscr{X}$ be a linear topological space, and $\mathscr{Z}$ a locally convex linear space, the convex cones P_z, P_x closed, and assume the regular convexity of*

$$(4) \qquad W_T^* = \{w^* \in \mathscr{W}: w^* = T^*(v^*),\ v^* \geqq 0,\ v^* \in \mathscr{V}^*\} ,$$

where T is the linear continuous transformation on the topological linear product space $\mathscr{W}$ of the pairs $w = (\rho, x)$, ρ real, $x \in \mathscr{X}$, into the topological product space $\mathscr{Z} \times \mathscr{W} = \mathscr{V}$ given by

$$(5) \quad T((\rho, x)) = (-a\rho + U(x), (\rho, x)) \qquad \textit{for all } \rho \textit{ real and all } x \in \mathscr{X} .$$

Then, for x_0 to maximize $f(x)$ subject to the constraints (2), *it is necessary and sufficient that $\Phi(x, z^*)$ have a non-negative saddle-point at (x_0, z_0^*); i.e., for Φ defined by* (3),

$$(6.1) \qquad \Phi(x, z_0^*) \leqq \Phi(x_0, z_0^*) \qquad \textit{for all } x \geqq 0 ,$$

$$(6.2) \qquad \Phi(x_0, z_0^*) \leqq \Phi(x_0, z^*) \qquad \textit{for all } z^* \geqq 0 ,$$

if and only if

$$(7.1) \qquad U(x_0) - a \geqq 0_z ,$$

$$(7.2) \qquad x_0 \geqq 0_x ,$$

and, for any $x \in \mathscr{X}$,

(8) *if* (2.1) *and* (2.2) *hold, then* $-x^*(x) + \nu \leqq -x^*(x_0) + \nu$.

PROOF. In view of Theorem V.1, we need only prove the necessity. Inequality (6.1) may be rewritten as

$$(6.1') \quad -x^*(x) + z_0^*[U(x) - a] \leqq -x^*(x_0) + z_0^*[U(x_0) - a] \qquad \text{for all } x \geqq 0$$

or as

$$(6.1'') \quad z_0^*[U(x) - a] \leqq x^*(x) - x^*(x_0) + z_0^*[U(x_0) - a] \qquad \text{for all } x \geqq 0 .$$

Now write

$$(8) \qquad x^*(x_0) = \beta .$$

Then, for any x satisfying (2), (8) yields

[57] In general, the first term of the right member of (3) is $y_0^*[-x^*(x) + \lambda]$. In this case, since we shall always take $y_0^* > 0$, we may put $y_0^* = 1$ without loss of generality. [I.e., $y_0^*(y) = y$ for all $y \in \mathscr{Y}$.]

(9′) $$-x^*(x) \leqq -\beta ,$$
or
(9″) $$x^*(x) - \beta \geqq 0 .$$
Hence the hypotheses of Theorem IV.3 are satisfied[58] and therefore there exists a $z_0^* \geqq 0$ such that
(10.1) $$z_0^*[U(x) - a] \leqq x^*(x) - \beta \qquad \text{for } x \geqq 0$$
and
(10.2) $$z_0^*[U(x_0) - a] = 0$$
since x_0 satisfies the hypotheses of IV (34). Equations (10.1) and (10.2) imply (6.1″), hence (6.1).

Inequality (6.2) may be written as
(6.2′) $$-x^*(x_0) + z_0^*[U(x_0) - a] \leqq -x^*(x_0) + z^*[U(x_0 - a)] \qquad \text{for all } z^* \geqq 0 ,$$
i.e., because of (10.2),
(6.2″) $$z^*[U(x_0) - a] \geqq 0 \qquad \text{for all } z^* \geqq 0 .$$
But (6.2″) must hold because of (7.1) and $z^* \geqq 0$.

V.3.3. *Non-negative Lagrangian saddle-points and quasi-saddle-points: the differentiable case.*

V.3.3.1. In this section all spaces are Banach and the functions f and g are assumed to possess Fréchet differentials. We shall call them *differentiable.* The convex cones P_x, P_y, P_z are assumed closed.

V.3.3.2. *Definitions.* We shall say that the function g on $\mathscr{X}$ into $\mathscr{Z}$ is *regular* at a point $\bar{x} \in \mathscr{X}$ if and only if, for every
(1) $$\xi \in \mathscr{X}, \qquad \xi \neq 0_x$$
such that the equality
(2) $$x = \bar{x} + \xi$$
implies the two inequalities
(3) $$x \geqq 0$$
and
(4) $$\delta g(\bar{x}; \xi) + g(\bar{x}) \geqq 0 ,$$
there exists a function Ψ on the closed (real) interval $[0, 1]$ into $\mathscr{X}$, say $x' = \Psi(t)$ $(0 \leq t \leq 1)$, with the following properties:
(5)
(a) $\delta\Psi(t; \tau)$ exists for all $0 \leq t \leq 1$
(b) $\bar{x} = \Psi(0)$
(c) $\Psi(t) \geqq 0$ for $0 \leq t \leq 1$
(d) $g[\Psi(t)] \geqq 0$ for $0 \leq t \leq 1$
(e) $\xi = \delta\Psi(0; \tau)$ with $\tau > 0$.

[58] Note that $\bar{x}$ required in IV.3 exists, for x_0 has this property by definition of maximality.

It is easily seen that the condition of regularity is closely related to the Kuhn and Tucker "constraint qualification" ([31], p. 483). In fact, the assertion concerning Ψ is identical with the corresponding assertion in Kuhn and Tucker, while the conditions (1), (2), (3), (4) under which Ψ must exist are not weaker[59] than the corresponding conditions (5), [31], *loc. cit.* Therefore $\bar{x}$ is necessarily regular in our sense if the Kuhn-Tucker "constraint qualification" is satisfied.

It should also be noted that our condition of regularity is closely related to Goldstine's hypothesis (a) ([15], p. 145) whose relationship to the condition in Bliss ([4], Lemma 76.1, p. 210) is similar to that of our regularity concept to the Kuhn-Tucker "constraint qualification."

V.3.3.3. THEOREM V.3.3.1. A. *Let f be a real-valued differentiable function on the Banach space $\mathscr{X}$, g a differentiable function on $\mathscr{X}$ into the Banach space $\mathscr{Z}$. The cones $P_x = \{x: x \geqq 0\}$ and $P_z = \{z: z \geqq 0\}$ are assumed closed.*

Let x_0 maximize $f(x)$ subject to the constraints $x \geqq 0$, $g(x) \geqq 0$ and suppose g is regular at x_0.

B. *It then follows that the relations*

$$x \geqq 0 \tag{6.1}$$

$$\delta g(x_0; \xi) + g(x_0) \geqq 0 \qquad (\xi = x - x_0) \tag{6.2}$$

imply

$$-\delta f(x_0; \xi) \geqq 0. \tag{7}$$

PROOF. Consider the real-valued function $h(t)$, $0 \leq t \leq 1$, of the real variable t, defined by

$$h(t) = f[\Psi(t)] \qquad (0 \leq t \leq 1). \tag{8}$$

[The function Ψ exists since, by virtue of (6), the relations (1), (2), (3), (4) are satisfied and g is assumed regular at x_0.]

Because of 5(b), (c), (d) and the maximality of x_0, $h(t)$ must have a maximum at $t = 0$. It follows that[60] for

$$\tau > 0\ , \tag{9}$$

$$\delta h(0; \tau) = \delta f[\Psi(0)\, ; \delta\Psi(0, \tau)] \leqq 0\ , \tag{10}$$

whence by 5(e), (7) follows.[61]

THEOREM V.3.3.2. (This Theorem is a generalization of the Kuhn-Tucker Theorem 1 [31], p. 484.)

A. *Let all assumptions under* A *in Theorem* V.3.3.1 *hold. Assume*

---

[59] Since Kuhn and Tucker impose their conditions only on certain components of x and g, it should be noted that for those components g_i of g on which Kuhn and Tucker impose the constraint (5) ([31], *loc. cit.*), we have $g_i(\bar{x}) = 0$. Hence (4) is not weaker than the first part of Kuhn-Tucker (5).

[60] Using the "function of a function rule" as applied to Fréchet differentials, cf. II.5.2.

[61] Theorem V.3.3.1 is implicit in the Kuhn-Tucker proof of their Theorem 1. The proof is suggested (*mutatis mutandis*) by Goldstine [15]. The writer is indebted to Kenneth J. Arrow for clarification on this point.

further the regular convexity of the set

$$W_T^* = \{w^* \in \mathscr{W}^*: w^* = T^*(v^*),\ v^* \geqq 0,\ v^* \in \mathscr{V}^*\}\ ,$$

where T *is given by* V.3.2(4), *with* U *and* a *as in* (15) *below.*

B. *Then there exists a* $z_0^* \geqq 0$ *such that the Lagrangian expression* $\Phi(x, z^*) = f(x) + z^*[g(x)]$ *has a non-negative quasi-saddle-point at* $(x_0, z_0^*; y_0^*)$, $y_0^* = 1$, *i.e., it satisfies the following relations*:

(11.1) $\delta_x\Phi((x_0, z_0^*); \xi) \leqq 0$ *for all* $x \geqq 0, x = x_0 + \xi$

(11.2) $\delta_x\Phi((x_0, z_0^*); x_0) = 0$

(12.1) $\delta_{z^*}\Phi((x_0, z_0^*); \zeta^*) = \zeta^*[g(x_0)] \geqq 0$ *for all* $z^* \geqq 0, \zeta^* = z^* - z_0^*$,

(12.2) $\delta_{z^*}\Phi((x_0, z_0^*); z_0^*) = z_0^*[g(x_0)] = 0$.

PROOF. Since $\delta g(x; \xi)$ and $\delta f(x; \xi)$ are additive in ξ, x being fixed, the relations (6.1), (6.2), and (7) of Theorem V.3.3.1 may be rewritten respectively as

(13.1) $x \geqq 0$,

(13.2) $-\delta f(x_0; x) - [\delta g(x_0; x_0) - g(x_0)] \geqq 0$,

and

(14) $-\delta f(x_0; x) - [-\delta f(x_0; x_0)] \geqq 0$.

Since x_0 is assumed maximal, Theorem V.3.3.1 states that (13.1), (13.2) together imply (14). This corresponds to the implication (34) in Theorem IV.3, with the following correspondence:

(15.1) $U(x) = \delta g(x_0; x)$ for all $x \in \mathscr{X}$,

(15.2) $a = \delta g(x_0; x_0) - g(x_0)$,

(15.3) $x^*(x) = -\delta f(x_0; x)$ for all $x \in \mathscr{X}$,

(15.4) $\beta = -\delta f(x_0; x_0)$.

Since all the other hypotheses of Theorem IV.3 are satisfied (in particular, $\bar{x}$ of IV (31) exists since x_0 is maximal and hence has the required properties), there exists a $z_0^* \geqq 0$ such that

(16.1) $z_0^*[\delta g(x_0; x) - (\delta g(x_0; x_0) - g(x_0))] \leqq -\delta f(x_0; x) - (-\delta f(x_0; x_0))$

and

(16.2) $z_0^*[\delta g(x_0; x) - (\delta g(x_0; x_0) - g(x_0))] = 0$ for $x = x_0$.

Equation (16.2) immediately yields

(17) $z_0^*[g(x_0)] = 0$

which is (12.2) in Theorem V.3.3.2.

Since x_0 is maximal,

(18) $g(x_0) \geqq 0$;

hence

(19) $z^* \geqq 0$ implies $z^*[g(x_0)] \geqq 0$.

Hence, because of (17), (12.1) holds for any ζ^* such that $\zeta^* = z^* - z_0^*$, $z^* \geqq 0$.

Using (17) and the additivity of $\delta f(x; \xi)$ and $\delta g(x; \xi)$ as functions of ξ, we may rewrite (16.1) as

(20) $$z_0^*[\delta g(x_0; x - x_0)] \leqq -\delta f(x_0; x - x_0) \qquad \text{for all } x \geqq 0 ,$$

i.e.,[62]

(21) $$\delta f(x_0; x - x_0) + \delta z_0^* g(x_0; x - x_0) \leqq 0 \qquad \text{for all } x \geqq 0 ,$$

which is (11.1) in Theorem V.3.3.2. [$z_0^* g(x) = z_0^*[g(x)]$ for all x.]

Now setting $x = 0$ in (21), we get

(22) $$-\delta_x \Phi((x_0, z_0^*); x_0) \leqq 0 .$$

Rewrite (21) as

(21′) $$\delta_x \Phi((x_0, z_0^*); x) - \delta_x \Phi((x_0, z_0^*); x_0) \leqq 0 \qquad \text{for all } x \geqq 0$$

and suppose that

(23) $$\delta_x \Phi((x_0, z_0^*); x_0) > 0 .$$

But using in (21′) $x = 2x_0$, we reach a contradiction since $\delta_x \Phi((x_0, z_0^*); x)$ is homogeneous in x. Hence the equality sign must hold in (22), and (11.2) in Theorem V.3.3.1 follows.

V.3.3.5. THEOREM V.3.3.3. (This Theorem is a generalization of the "only if" part of the Kuhn-Tucker Theorem 3. The converse—the "if" part of the Kuhn-Tucker Theorem 3—follows from V.1.1.) *Let all the assumptions under* A *in Theorem* V.3.3.2 *hold, and assume further that f and g are concave. Then $\Phi(x, z^*)$ has a non-negative saddle-point at (x_0, z_0^*) where x_0 is the maximal point of the hypothesis.*

PROOF. The Kuhn-Tucker proof of the "only if" part of Theorem 3 [31], p. 487, is valid under our assumptions. For the sake of completeness we reproduce its major steps in our notation. First, if $h(x)$, $x \in \mathscr{X}$, $\mathscr{X}$ Banach, is a concave function with values in a Banach space V, and the ordering relation is given by a *closed*[63] convex cone P_v, we have, for $0 < \theta \leq 1$

(24) $$h(x'') - h(x') \leqq \frac{1}{\theta} \{h[x' + \theta(x'' - x')] - h(x')\} .$$

Now

(25′) $$\delta h(x'; x'' - x') = \lim_{\theta \to 0} \frac{1}{\theta} \{h[x' + \theta(x'' - x')] - h(x')\} .$$

Then, because P_v is closed,

(25″) $$h(x'') - h(x') \leqq \delta h(x'; x'' - x')$$

which corresponds to Lemma 3 in Kuhn and Tucker [31], p. 485. Hence, for $x \geqq 0$ and $z_0^* \geqq 0$ and using (25), since f and g are concave,

---

[62] We use the function of a function rule and the fact that, since z^* is linear, $\delta z^*(z_0; \zeta) = z^*(\zeta)$. (Cf. II.5.2 and II.5.1, footnote 29.)

[63] Closedness is not used for (24) or (25′), but only for (25″).

$$(26)\qquad \begin{aligned}\Phi(x, z_0^*) &\leqq f(x_0) + z_0^*[g(x_0)] + \delta f(x_0; x - x_0) + z_0^*[\delta g(x_0; x - x_0)]\\ &= \Phi(x_0, z_0^*) + \delta_x\Phi((x_0, z_0^*); x - x_0)\\ &\leqq \Phi(x_0, z_0^*)\end{aligned}$$

where the last inequality is based on (11.1) in Theorem V.3.3.2.

On the other hand, for $z^* \geqq 0$,

$$(27)\qquad \Phi(x_0, z^*) - \Phi(x_0, z_0^*) = (z^* - z_0^*)[g(x_0)] \geq 0$$

by (12.1) in Theorem V.3.3.2. Relations (26) and (27) together imply that Φ has a non-negative saddle-point at (x_0, z_0^*).

V.3.3.6. THEOREM V.3.3.4. (This Theorem is a generalization of the Kuhn-Tucker Theorem 4.)

A. *Let* $\mathscr{X}$, $\mathscr{Z}$, P_x, P_z, *and* g *be as in Theorem* V.3.3.2 (*including the assumption of regular convexity of* X_T^* *and regularity of* g *but not that of concavity*) *while* $\mathscr{Y}$ *is a Banach space possessing the property stated at the beginning of Theorem* V.2.2 (*e.g., it would suffice to assume* $\mathscr{Y}$ *separable*). *Assume further that* f *is a differentiable function on* $\mathscr{X}$ *into* $\mathscr{Y}$ *and also that* x_0 *is properly maximal.*

B. *Then for some* $y_0^* > 0$,

$$(28)\qquad \Phi_1(x, z^*) = y^*[f(x)] + z^*[g(x)]$$

has a non-negative quasi-saddle-point at $(x_0, z_0^*; y_0^*)$; *i.e., the relations* (11), (12) *hold with* $f(x)$ *replaced by* $y_0^*[f(x)]$.

PROOF. Using Theorem V.2.2 with $y_0 = f(x_0)$ we obtain y_0^* such that

$$(29)\qquad y \in Y \quad \text{implies} \quad y_0^*(y) \leqq y_0^*(y_0)$$

for

$$(30)\qquad Y = f(P_x \cap g^{-1}(P_z))\,;$$

i.e., the function y_0^*f, given by

$$(31)\qquad F(x) = y_0^*[f(x)] \quad \text{for all } x \in \mathscr{X}$$

has a maximum at x_0 subject to

$$(32)\qquad x \geqq 0, \quad g(x) \geqq 0\,.$$

Thus we may use Theorem V.3.3.2 as applied to $F(x)$ and the Theorem follows. ($F(x)$ is differentiable since f is differentiable and so is y_0^*.)

V.3.3.7. THEOREM V.3.3.5. (Generalization of the "only if" part of the Kuhn-Tucker Theorem 6.) *Let all the assumptions under* A *in Theorem* V.3.3.4 *hold, and assume further that* f *and* g *are concave. Then the function* $\Phi_1(x, z^*)$ *as defined by* (28) *has a non-negative saddle-point at* $(x_0, z_0^*; y_0^*)$ *for some* $y_0^* > 0$.

PROOF. Use Theorem V.3.3.4, then Theorem V.3.3.3 as applied to Φ_1. (Note that F, defined in (31), is concave if f is.)

Note. The converse is found in Theorem V.1 (the "if" part of the Kuhn-Tucker Theorem 6).

V.3.3.8. In Section V.3.3 the spaces have so far been assumed Banach and the differentials Fréchet. It appears, however, that by using a more general concept of a differential (to be called here the MF differential) one can validate the results of V.3.3 for that class of linear topological spaces for which the auxiliary results from previous sections are valid (i.e., locally convex linear, or the alternatives mentioned in II.1.4.1).

The MF differential is that called μ (or μ^*) differential in Michal [38], p. 82, and also defined (later but independently) by Fréchet [12], pp. 64–65.[64] We shall denote this differential by $df(x_0; h)$ if evaluated at x_0 with increment $h \in \mathscr{X}$; $f(x)$ is in $\mathscr{Y}$; $\mathscr{X}$ and $\mathscr{Y}$ are Hausdorff linear spaces. The MF differential is additive and continuous (hence linear) in the increment h and is further characterized by the following *property* (c): There exists a fixed neighborhood W of 0_x such that, given any neighborhood V of 0_y, there is a neighborhood U of 0_x (U depends on W) such that if

$$h \in U, nh \in W ,$$

then

$$n[f(x_0 + h) - f(x_0) - df(x_0; h)] \in V$$

for all positive integers n (or all positive real numbers n).[65] As partly stated in [38] and shown in [12], $df(x_0; h)$ has the important properties of the Fréchet differential; in particular, the "function of a function" rule is valid and the partial differentials are defined in the usual way and are additive. Furthermore, from the remarks and theorems in Michal [36], [37] and [38], it follows that when $df(x_0; h)$ exists, then the Gâteaux differential, i.e.,

$$\lim_{\theta \to 0} \frac{1}{\theta}[f(x_0 + \theta h) - f(x_0)] ,$$

exists and the two are equal.[66] Now let $\mathscr{X}$ and $\mathscr{Z}$ be (say) locally

[64] One could probably also use the slightly more general F or M differentials; cf. Hyers [22], pp. 14–15.

[65] In linear normed spaces the MF differential exists if and only if the Fréchet differential exists, and the two are equal. Property (c) is equivalent to that given at the end of II.5.1. This is shown in [12], pp. 62–64.

[66] In [38], p. 82, it is stated that in what we called linear spaces the μ^* differential (equivalent to the MF differential) is equivalent to what in [38] is called the M_I differential. In [37], Theorem V, the existence of the M_I differential (called "the differential") is asserted to imply the existence of the M differential of [36] and the two are equal. Finally, in Theorem 4 of [36] it is stated that the existence of the M differential implies that of the Gâteaux differential and the two are equal. The precise meaning of the limit in the definition of the Gâteaux differential is as follows. Write $g(\theta) = [f(x_0 + \theta h)]/\theta$ and denote by $d_G f(x_0; h)$ the Gâteaux differential at x_0 with increment h. Then for a given neighborhood V of 0_y there exists a real number $\delta > 0$ such that

$$g(\theta) \in d_G f(x_0;\ h) + V \qquad \text{for all } 0 < |\theta| < \delta .$$

The equality $df(x_0;\ h) = d_G f(x_0;\ h)$ follows easily from property (c) above when the "starlike" neighborhood system $\mathcal{U}$ is used (see Bourgin [9], pp. 638–39).

convex Hausdorff linear spaces (cf. II.1.4.1 for possible alternative assumptions) and consider the results of V.3.3 with the MF differential $df(x_0; \xi)$, etc., substituted for the Fréchet differential, $\delta f(x_0; \xi)$, etc., throughout. Theorem V.3.3.1 obviously remains valid, since the MF differential is linear and obeys the "function of a function" rule. Theorem V.3.3.2 also retains its validity since the partial MF differentials have the required properties. In the proof of Theorem V.3.3.3, the two relations (25) remain valid because the MF differential, like the Fréchet differential, equals the Gâteaux differential, and Theorem V.3.3.3 follows. (It can also be shown that relations (11) and (12) in Theorem V.3.3.2 are satisfied at a non-negative saddle-point.) The remaining two theorems of V.3.3 also go through.

It may finally be noted that the MF differential is defined and retains many of its important properties when the domain and the range of the function whose differential is taken are in topological groups (not necessarily linear spaces). Thus a possibility appears of studying a broader class of spaces and their Lagrangian expressions from the viewpoint of differentiability.

REFERENCES

[1] K. J. Arrow, E. W. Barankin, and D. Blackwell, "Admissible Points of Convex Sets," in H. W. Kuhn and A. W. Tucker (eds.), *Contributions to the Theory of Games*. Princeton: Princeton University Press, 1953, pp. 87–91.

[2] K. J. Arrow and L. Hurwicz, "Dynamic Aspects of Achieving Optimal Allocation of Resources" (Abstract), *Econometrica*, XX (1952), 86–87.

[3] S. Banach, *Théorie des Operations Linéaires*. Monografje Matematyczne, Warszawa, Tom I, 1932.

[4] C. A. Bliss, *Lectures on the Calculus of Variations*. Chicago: University of Chicago Press, 1946.

[5] N. Bourbaki, *Topologie Générale* (Eléments de Mathématique, II, Première Partie, Livre III, Chapters I and II). Paris: Hermann et Cie, 1940.

[6] N. Bourbaki, *Topologie Générale* (Eléments de Mathématique, III, Première Partie, Chapters III and IV). Paris: 2d ed., Hermann et Cie, 1951.

[7] N. Bourbaki, *Espaces Vectoriels Topologiques* (Eléments de Mathématique, XV, Première Partie; Livre V, Chapters I and II). Paris: Hermann et Cie, 1953.

[8] N. Bourbaki, *Espaces Vectoriels Topologiques* (Eléments de Mathématique, XVIII, Première Partie; Livre V, Chapters III, IV, V). Paris: Hermann et Cie, 1955.

[9] D. G. Bourgin, "Linear Topological Spaces," *Amer. J. Math.*, LXV (1943), 637–59.

[10] G. B. Dantzig, "A Proof of the Equivalence of the Programming Problem and the Game Problem," in [28], pp. 330–35.

[11] M. Fréchet, "La Notion de Différentielle dans l'Analyse Générale," *Annales Scientifiques de l'École Normale Superieure*, 3me Serie, XLII (1925), 293–323.

[12] M. Fréchet, "La Notion de la Différentielle sur un groupe Abelian," *Portugaliae Mathematica*, VII (1948), 59–72.

[13] M. Fréchet, "La Notion de Différentielle dans l'Analyse Générale," in Maurice Fréchet, *Pages Choisies d'Analyse Générale*. Paris-Louvain, 1953, pp. 180–204.

[14] D. Gale, "Convex Cones and Linear Inequalities," in [28], pp. 287–97.

[15] H. H. Goldstine, "Minimum Problems in the Functional Calculus," *Bull. Amer. Math. Soc.*, XLVI (1940), 142–49.

[16] L. M. Graves, "Riemann Integration and Taylor's Theorem in General Analysis," *Trans. Amer. Math. Soc.*, XXIX (1927), 163–77.
[17] L. M. Graves, "Topics in the Functional Calculus," *Bull. Amer. Math. Soc.*, XLI (1935), 641–62.
[18] F. Hausdorff, "Zur Theorie der linearen metrischen Räume," *J. für die Reine und Angewandte Mathematik*, CLXVII (1932), 294–311.
[19] T. H. Hildebrandt and L. M. Graves, "Implicit Functions and Their Differentials in General Analysis," *Trans. Amer. Math Soc.*, XXIX (1927), 127–53.
[20] E. Hille, *Functional Analysis and Semi-Groups*, Amer. Math. Soc. Colloquium Publications (New York), XXXI (1948).
[21] H. Hotelling, "The General Welfare in Relation to Problems of Taxation and Railway Rates," *Econometrica*, VI (1938), 248–56.
[22] D. H. Hyers, "Linear Topological Spaces," *Bull. Amer. Math. Soc.* LI, 1 (1945), 1–21.
[23] J. L. Kelley, *General Topology*. New York: Van Nostrand, 1955.
[24] V. L. Klee, Jr., "The Support Property of a Convex Set in a Linear Normed Space," *Duke Math. J.*, XV (1948), 767–72.
[25] V. L. Klee, Jr., "Convex Sets in Linear Spaces," *Duke Math. J.*, XVIII (1951), 443–66.
[26] V. L. Klee, Jr., "Convex Sets in Linear Spaces, III," *Duke Math. J.*, XX (1953), pp. 105–11 (esp. p. 106).
[27] T. C. Koopmans, "Analysis of Production as an Efficient Combination of Activities," in [28], pp. 33-97.
[28] T. C. Koopmans (ed.), *Activity Analysis of Production and Allocation*, Cowles Commission Monograph No. 13. New York: Wiley, 1951.
[29] M. G. Krein and M. A. Rutman, "Linear Operators Leaving Invariant a Cone in a Banach Space," *Uspehi Matem. Nauk* (N.S.) 3 (1948), No. 1 (23), pp. 3–95. *Am. Math. Soc. Translation* No. 26 (New York, 1950).
[30] M. Krein, and V. Šmulian, "On Regularly Convex Sets in the Space Conjugate to a Banach Space," *Ann. Math.*, Second Series, XLI (1940), 556-83.
[31] H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951, pp. 481–92.
[32] C. Kuratowski, *Topologie* I, Deuxième Edition, Monografie Matematyczne, Tom XX, Warszawa-Wroclaw, 1948.
[33] O. Lange and F. M. Taylor, *On the Economic Theory of Socialism* (ed. B. E. Lippincott). Minneapolis: University of Minnesota Press, 1938.
[34] S. Lefschetz, *Algebraic Topology*, Amer. Math. Soc. Colloquium Publications (New York), XXVII, (1942).
[35] A. P. Lerner, *The Economics of Control*. New York: Macmillan, 1944.
[36] A. D. Michal, "Differential Calculus in Linear Topological Spaces," *Proc. Nat. Acad. Sc. U.S.A.*, XXIV (1938), 340–42.
[37] A. D. Michal, "Differentials of Functions with Arguments and Values in Topological Abelian Groups," *Proc. Nat. Acad. Sc. U.S.A.*, XXVI (1940), 357–59.
[38] A. D. Michal, "Functional Analysis in Topological Group Spaces," *Math. Magazine*, XXI (1947–48), 80–90.
[39] P. C. Rosenbloom, "Quelques Classes de Problèmes Extrémaux," *Bull. de la Soc. Math. de France*, LXXIX (1951), 1–58; LXXX (1952), 183–215.
[40] W. Rudin, *Principles of Mathematical Analysis*. New York: McGraw-Hill, 1953.
[41] M. Slater, "Lagrange Multipliers Revisited: A Contribution to Non-Linear Programming," Cowles Commission Discussion Paper, Math. 403, November 1950.
[42] E. C. Titchmarsh, *The Theory of Functions*, 2d ed. Oxford: Oxford University, Press, 1939
[43] A. Tychonoff, "Ein Fixpunktsatz," *Math. Annalen*, CXI (1935), 767–76.

Fritz John

Collected Papers
Volume 2

Jürgen Moser
Editor

Birkhäuser
Boston · Basel · Stuttgart
1985

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, 197
DOI 10.1007/978-3-0348-0439-4_9, © Springer Basel 2014

COURANT
Anniversary Volume
1948

EXTREMUM PROBLEMS WITH INEQUALITIES AS SUBSIDIARY CONDITIONS

Fritz John

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. There may however be an infinite number of inequalities prescribed. Lagrange's rule for the situation considered here differs from the ordinary one, in that the multipliers may always be assumed to be positive. This makes it possible to obtain <u>sufficient</u> conditions for the occurence of a minimum in terms of the first derivatives only.

Two geometric applications will be discussed here. From the point of view of applications it would seem desirable to extend the method used here to cases, where the functions involved are not necessarily differentiable, or where they do not depend on a finite number of independent variables.

1. Necessary conditions for a minimum.

Let R be a set of points x in a space E, 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:

$$G(x,y) \geqq 0, \tag{1}$$

where G is a function defined for all x in R and all "values" of the parameter y. There may be a finite or infinite number of these inequalities. To gain sufficient generality we assume that the "values" of the parameter y vary over a set of points S in a space H. Then $G(x,y)$ is defined in the set $R \times S$. We are interested in conditions a point x^o of R' has to satisfy in order that

$$(2) \qquad M = F(x^o) = \underset{x \subset R'}{\text{Minimum}}\, F(x)$$

In what follows we restrict ourselves to the case, where the space E containing the set R is the n-dimensional euclidean space E_n, and where the set S of parameter values y is a compact set in a metric space H. We make the further assumptions that $F(x)$ and $G(x,y)$ have first derivatives F_i and G_i with respect to the coordinates x_i of the point $(x_1,\ldots,x_n) = x$, and that $F(x)$, $G(x,y)$, $F_i(x)$, $G_i(x,y)$ are continuous functions of (x,y) in $R \times S$.[1)]

Given a function $\Phi(x)$ with continuous derivatives $\Phi_i(x)$ in R, we denote by

$$(3) \qquad \Phi'(x,z) = \sum_{i=1}^{n} \Phi_i(x) z_i$$

the differential of the function. $\Phi'(x,z)$ is then defined for all $x \subset R$ and $z = (z_1,\ldots,z_n) \subset E_n$, and is linear in z.

Theorem I.

Let x^o be an interior point of R, and belong to the set R' of all points x of R, which satisfy (1) for all $y \subset S$. Let $F(x^o) = \underset{x \subset R'}{\text{Minimum}}\, F(x)$.

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

$$(4a) \qquad G(x^o,y^r) = o \text{ for } r = 1,\ldots,s$$

1) Here continuity in $R \times S$ is defined so as to agree with the following definition of convergence in $R \times S$: $\lim_{r\to\infty} (x^r,y^r) = (x,y)$, if $\lim_{r\to\infty} x^r = x$ and $\lim_{r\to\infty} y^r = y$.

(4b) $\lambda_0 \geqq 0, \quad \lambda_1 > 0, \ldots, \quad \lambda_s > 0$

(4c) $0 \leqq s \leqq n$

(4d) the function

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

has a critical point at x^o i.e.

$$\Phi_i(x^o) = o \quad \text{for } i = 1, \ldots, n .$$

Proof:

Let S' denote the subset of points y of S, for which

$$G(x^o, y) = 0.$$

We shall first show that the system of inequalities

(5a) $F'(x^o, z) < 0$

(5b) $G'(x^o, z, y) > 0 \qquad$ for all $y \subset S'$

can have no solution $z = (z_1, \ldots, z_n)$.

For let (5a,b) be satisfied for a certain z. Denote by S'_ε the set of all points of S having a distance $\leqq \varepsilon$ from sone point of S', and by X^o_ε the set of all points of R having a distance $\leqq \varepsilon$ from x^o. Then there exist positive numbers δ, ε such that

(6) $F'(x,z) < -\delta, \quad G'(x,z,y) > \delta \quad$ for all $x \subset X^o_\varepsilon$, $y \subset S'_\varepsilon$.

For otherwise there would exist sequences of points x^r in R, y^r in S, η^r in S', such that

$$\lim_{r\to\infty} x^r = x^o, \quad \lim_{r\to\infty} (\text{distance of } y^r \text{ and } \eta^r) = 0$$

and either

$$\liminf_{r\to\infty} F'(x^r, z) \geqq 0$$

or

$$\limsup_{r\to\infty} G'(x^r, z, y^r) \leqq 0.$$

As S is compact and G is continuous, S' is compact as well. We can then form a suitable subsequence of the r, such that y^r and η^r converge towards a point y of S'. As F' and G' are continuous, it would follow that either $F'(x^0,z) \geqq 0$ or $G'(x^0,z,y) \leqq 0$, contrary to (5a,b).

Hence (6) holds for suitable positive ε, δ. On the other hand there exists a positive constant $\mu = \mu(\varepsilon)$ such that

$$G(x^0,y) > \mu \tag{7}$$

for all y of S outside S'. For $G(x^0,y)$ is non-negative in S (as $x^0 \subset R'$), vanishes only on S', and is continuous on the compact set S.

As x^0 is an interior point of R, we have for sufficiently small positive t

$$F(x^0 + tz) = F(x^0) + tF'(x^0 + \theta tz, z)$$

$$G(x^0 + tz,y) = G(x^0,y) + tG'(x^0 + \theta tz,z,y)$$

where θ stands for any quantity between 0 and 1. If here t is chosen so small that

$$t \sqrt{\sum_i z_i^2} < \varepsilon, \qquad t \cdot \underset{\substack{y \subset S \\ x \subset X^0_\varepsilon}}{\text{Maximum}} |G'(x,z,y)| < \mu ,$$

we can apply (6), (7) and find that

$$F(x^0 + tz) \leqq F(x^0) - t\delta < F(x^0)$$

$$G(x^0 + tz,y) \geqq G(x^0,y) + t\delta > 0 \quad \text{for all } y \subset S'_\varepsilon$$

$$G(x^0 + tz,y) \geqq \mu - t|G'(x^0 + \theta tz,z,y)| > 0 \text{ for all y of S outside } S'_\varepsilon$$

This would however contradict the assumed minimum property of x^0. Consequently, there can be no z satisfying (5a,b).

The non-existence of a solution z of the system of linear homogeneous inequalities (5a,b) can be seen to be equivalent to the existence of non-negative solutions of a certain system of equations.[2] For this purpose we introduce the "representative" points corresponding to (5a,b), i.e. the points in n-space given respectively by

$$(8) \quad \begin{aligned} q &= (-F_1(x^o), \ldots, -F_n(x^o)) \\ p_y &= (G_1(x^o,y), \ldots, G_n(x^o,y)) \quad \text{for } y \subset S'. \end{aligned}$$

The non-existence of a solution z of (5a,b) implies that the set Σ consisting of q and all p_y does not lie in an open half-space bounded by a hyper-plane through the origin. Then the origin is a point of the convex hull of Σ. As in addition, as a consequence of our assumptions, Σ is closed and bounded, it follows that the origin belongs to a simplex with vertices in Σ, where the point q may be chosen as one of the vertices.[3] Then the origin is center of mass of $s+1$ non-negative masses ($s \leq n$), located in q and s other points of Σ. Equations (4a,b,c,d) are the analytic expression for this fact.

2. Sufficient conditions for a minimum. Equivalence with finite systems of inequalities.

Theorem II.

Let x^o be an interior point of R and belong to the set R' of all points x of R, which satisfy

$$G(x,y) \geqq 0 \quad \text{for all } y \subset S.$$

Let there exist a function $\phi(x)$ of the form

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

2) See L. L. Dines: "Linear inequalities," Bull. Am. Math. Soc. vol. 42 (1936), pp. 353-365. R. W. Stokes: "A geometric theory of linear inequalities," Trans. Am. Math. Soc. vol. 33 (1931), pp. 782-805.

3) That one of the vertices can be chosen arbitrarily in Σ, is evident from the proof of the fundamental theorem that any point of the convex hull of Σ belongs to a simplex with vertices in Σ. See Bonnesen-Fenchel: "Theorie der konvexen Körper," p. 9.

where $y^r \subset S$, such that (4a,b,d) hold. Let in addition the matrix

$$A = \begin{pmatrix} \lambda_o F_1(x^o) & G_1(x^o,y^1) & G_1(x^o,y^2) \ldots\ldots G_1(x^o,y^s) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \lambda_o F_n(x^o) & G_n(x^o,y^1) & G_n(x^o,y^2) \ldots\ldots G_n(x^o,y^s) \end{pmatrix}$$

have rank n.

Then F(x) has a relative minimum at x^o in the set defined by the finite number of inequalities

$$(9) \qquad G(x,y^r) \geqq 0 \qquad \text{for } r = 1,\ldots,s$$

and has, a fortiori, a relative minimum at x^o for the set R'.

Proof:

If F(x) did not have a relative minimum at x^o for the set (9), we could find a sequence of positive numbers t_r and a set of points $z^r \subset E_n$, such that

$$\lim_{r\to\infty} t_r \quad 0, \qquad \sum_{i=1}^{n} (z^r{}_i)^2 = 1$$

$$F(x^o + t_r z^r) < F(x^o)$$

$$G(x^o + t_r z^r, y^k) \geqq 0 \quad \text{for } k = 1,\ldots,s.$$

Then with suitable θ between 0 and 1

$$F'(x^o + \theta t_r z^r, z^r) \leqq 0$$

$$G'(x + \theta t_r z^r, z^r, y^k) \geqq 0 \quad \text{for } k = 1,\ldots,s.$$

For a suitable subsequence the z^r converge towards a vector $z \neq 0$, for which then

$$F'(x^o,z) \leqq 0$$

$$G'(x^o,z,y^k) \geqq 0 \quad \text{for } k = 1,\ldots,s.$$

But, as $\phi(x)$ is stationary at x^o, we have

$$0 = \phi'(x^o,z) = \lambda_o F'(x^o,z) - \sum_{k=1}^{s} \lambda_k G'(x^o,z,y^k).$$

Hence, making use of (4b), we see that z satisfies the system of linear homogeneous equations

$$\lambda_o F'(x^o, z) = 0, \qquad G'(x^o, z, y^k) = 0 \quad \text{for } k = 1, \ldots, s.$$

The existence of a solution $z \neq 0$ of this system contradicts however the assumption made on the rank of A. This completes the proof of theorem II.

Theorem II shows that under suitable conditions a relative minimum of F(x) for the set R' determined by infinitely many inequalities is at the same time a relative minimum for the set determined by a suitable finite number of these inequalities. An example will show that this is not the case for every minimum problem. Consider the problem of finding the minimum of the function

$$F(x) = -x^2$$

in the set of all x satisfying the inequalities

(10) $\quad G(x,y) = y^2 - yx^2 \geqq 0 \quad$ for all y with $0 \leqq y \leqq 1$.

The set R' of all x satisfying (10) consists of the point $x = 0$. That point then is also a relative minimum point of F for the set. If, on the other hand, x is only subjected to a finite number of inequalities

$$G(x, y^k) \geqq 0 \quad \text{for } k = 1, \ldots, s,$$

where $0 \leqq y^k \leqq 1$, then all points of a neighbourhood of $x = 0$ are admitted, and F(x) has no relative minimum in the resulting set at 0.

3. Application to minimum sphere containing a set.[4)]

Let S be a bounded set in E_m. A sphere in E_m may be described

4) See H. W. E. Jung: "Ueber die kleinste Kugel, die eine räumliche Figur einschliesst," Journal für die reine und angewandte Mathematik, vol. 123 (1901), pp. 241-257. For a historical account of this well knwon problem see the paper by L. M. Blumenthal and G. E. Wahlin: "On the spherical surface of smallest radius enclosing a bounded subset of n-dimensional euclidean space," Bull. Am.

by $x = (x_1, \ldots, x_{m+1})$, where $x_1, \ldots, x_m$ are the coordinates of its center and x_{m+1} the square of its radius. Let x^o denote the sphere of least radius enclosing S. The existence of a sphere of least positive radius enclosing S is evident, if the assumption is made that S contains at least two distinct points.

Then

$$F(x) = x_{m+1} \tag{11a}$$

has a minimum for $x = x^o$ in the set of all x satisfying the inequalities

$$G(x,y) = x_{m+1} - \sum_{i=1}^{m} (x_i - y_i)^2 \geqq 0 \quad \text{for all } y \subset S. \tag{11b}$$

As every sphere containing S also contains the closure S of S, we can replace S by $\bar{S}$ in (11b).

According to theorem I we can find s points ($s \leqq m+1$) $y^1, \ldots, y^s$ of S and numbers $\lambda_o, \ldots, \lambda_s$ such that

$$\lambda_o = \sum_{r=1}^{s} \lambda_r \tag{12a}$$

$$\sum_{r=1}^{s} \lambda_r (x_i^r - y_i^r) = 0 \quad \text{for } i = 1, \ldots, m \tag{12b}$$

$$x_{m+1}^o - \sum_{i=1}^{m} (x_i^o - y_i^r)^2 = 0 \quad \text{for } r = 1, \ldots, s \tag{12c}$$

$$\lambda_o \geqq 0, \quad \lambda_1 > 0, \ldots, \lambda_s > 0. \tag{12d}$$

It follows from (12a,d) that $\lambda_o > 0$. From (12) we get for any $x = (x_1, \ldots, x_{m+1})$

$$\sum_{r=1}^{s} \lambda_r \left(x_{m+1} - \sum_{i=1}^{m} (x_i - y_i^r)^2 \right) = \lambda_o \left(x_{m+1} - x_{m+1}^o - \sum_{i=1}^{m} (x_i - x_i^o)^2 \right)$$

Math. Soc., vol. 47 (1941), pp. 771-777.

This identity shows that any sphere containing the points $y^1,\ldots,y^s$ has a radius $\geqq \sqrt{x^o_{m+1}}$, where the = sign only holds, if its center is also at $(x^o_1,\ldots,x^o_m)$. Hence the smallest sphere containing S is uniquely determined, and is at the same time the smallest sphere containing the points $y^1,\ldots,y^s$ of the closure of S.

If D_{rt} denotes the distance of the points y^r and y^t $(r,t = 1,\ldots,s)$, we have from (12)

$$\sum_{r\neq t} \lambda_r \lambda_t D^2_{rt} = \sum_{r,t,i} \lambda_r\lambda_t\left[(x^o_i-y^r_i)-(x^o_i-y^t_i)\right]^2 = 2\lambda_o^2 x^o_{m+1}$$

On the other hand

$$\sum_{r\neq t} \lambda_r\lambda_t = \frac{s-1}{s}\left(\sum_r \lambda_r\right)^2 - \frac{1}{2s}\sum_{r,t} (\lambda_r - \lambda_t)^2 \leqq \frac{s-1}{s}\lambda_o^2 .$$

Dividing the last two inequalities by each other and observing that the $\lambda_r\ \lambda_t$ are positive, it follows that

$$D = \text{diameter of } S \geqq \underset{r,t}{\text{Maximum}}\ D_{rt} \geqq \sqrt{\frac{2s}{s-1}\, x^o_{m+1}} \tag{13}$$

As $s \leqq m+1$, this leads to "Jung's inequality"

$$D \geqq \sqrt{\frac{2(m+1)}{m}}\ R \tag{14}$$

between the diameter D of a set S in E_m and the radius R of the smallest sphere containing the set.[5)]

This result can be extended in various directions. Following L. A. Santaló,[6)] we can consider a set S, which lies on the surface K of the unit-sphere in E_m and is contained entirely in a closed subset interior to a hemi-sphere of K.

We consider the set $y^1,\ldots,y^s$ belonging to S through (12).

5) See Jung, l.c., note 4.
6) "Convex regions on the n-dimensional spherical surface," Annals of Mathematics, vol. 47 (1946), pp. 448-459.

If $y^1,\ldots,y^s$ do not lie in a hyper-plane of E_m, then K is the smallest sphere containing S, for the y^r lie on the smallest sphere and lie only on one sphere. This however is impossible, as S lies in a closed subset interior to a hemisphere of K, and hence is certainly contained in spheres of radius < 1.

Consequently $y^1,\ldots,y^s$ must lie in an (m-1)-dimensional linear space. Then however the inequality (18) between diameter of the set of the y^r and the radius of the least sphere containing the y^r applies with m replaced by m-1. As the least sphere containing the y^r is identical with the one containing S, we have for S

$$D \geqq \sqrt{\frac{2m}{m-1}}\; R\,. \tag{15}$$

We can introduce the "spherical diameter" Δ of S as the least upper bound of the lengths of the greatcircle arcs on K joining any two points of S. Then obviously

$$D = 2\sin\frac{\Delta}{2}\,.$$

Similarly we can introduce the "spherical radius" ρ of the least "spherical" (m-1)-dimensional sphere on K containing S. Obviously

$$R = \sin\rho .$$

We then obtain form (15), as analogue of (14) in (m-1)-dimensional spherical space of curvature 1, the inequality

$$2\sin\frac{\Delta}{2} \geqq \sqrt{\frac{2m}{m-1}}\;\sin\rho$$

or

$$\cos\Delta \leqq \frac{m\cos^2\rho - 1}{m-1}\,. \tag{16}$$

This inequality is the best possible one between ρ and Δ as is seen from the example of a set S on K consisting of the vertices of an m-dimensional regular simplex.[7)]

In a different direction an obvious extension of (14) to Hilbert space suggests itself for $m \to \infty$:

If S is a set in Hilbert-space with the property that any two points of S have a distance $\leq D$, then there exists a point in Hilbert space, from which all points of S have a distance $\leq \frac{1}{\sqrt{2}} D$.

For a proof of this statement one forms the projection S_n of S on the $x_1 \ldots x_n$-coordinate plane. It is easily seen that the center

$$(x_1^n, x_2^n, \ldots, x_n^n, 0, 0, \ldots)$$

of the smallest sphere containing S_n converges for $n \to \infty$ towards a point of Hilbert space with the desired properties.

The constant $\frac{1}{\sqrt{2}}$ is again the best possible one in this connestion, as is shown by the example of the set consisting of the points $(1,0,0,\ldots), (0,1,0,\ldots), (0,0,1,0,\ldots)$, etc.

4. Application to the ellipsoid of least volume containing a set S in E_m.[8)]

A solid ellipsoid in running coordinates $y_1, \ldots, y_m$ may be described by a relation

---

7) Santaló l.c. obtains an inequality, which in appearance is stronger than (16) for $\Delta > \pi/2$. The explanation of this discrepancy must lie in the fact that he uses a different definition of "spherical diameter" from the one used here. (No definition of that term is given in his paper.) For sets of spherical diameter $> \pi/2$ (as used here) the diameter of the set need not be the same, as that of its "spherical convex hull," whereas they seem to be the same in Santalo's use of the term.

8) Related questions have been considered for $m = 2$ by F. Behrend: "Ueber einige Affininvarianten konvexer Bereiche", Math. Ann., vol. 113 (1937), pp. 713-747; "Ueber die kleinste umbeschriebene und die grösste einbeschriebene Ellipse eines konvexen Bereiches", ibid., vol. 115 (1938), pp. 379-411; F. John: "Moments of inertia of convex regions", Duke Math. J., vol. 2 (1936), pp. 447-452;

for $m = 3$ by O. B. Ader: "An affine invariant of convex regions", Duke Math. J. vol. 4 (1938) pp. 291-299; for general m by F. John: "An inequality for convex bodies", U. of Kentucky Research Club

(17) $$\sum_{i,k=1}^{m} x_{ik}(y_i - x_i)(y_k - x_k) \leqq 1$$

where

(18) $$x_{ik} = x_{ki}$$

and the x_{ik} are coefficients of a positive definite quadratic form. The volume of the ellipsoid is given by

$$V = \frac{\omega_m}{\sqrt{d}} ,$$

where ω_m denotes the volume of the unit-sphere in E_m and

$$d = \det(x_{ik}) .$$

If the assumption is made that S is not contained in any hyper-plane, the existence of an ellipsoid of least volume containing S can be seen as follows. There is a sphere of radius $r > 0$ contained in the convex hull of S, and hence contained in any ellipsoid, which contains S. Thus any ellipsoid containing S contains the sphere of radius r about the center of the ellipsoid. We have then for any x_{ik}, x_i satisfying (17) for all $y \subset S$

$$\underset{\sum_i u_i^2 = r^2}{\text{Maximum}} \sum_{i,k} x_{ik} u_i u_k \leqq 1$$

As the x_{ik} are also coefficients of a definite form, it follows that

$$|x_{ik}| \leqq \frac{1}{r^2} .$$

Thus the x_{ik} satisfying (17) for all $y \subset S$ form a bounded set. Moreover we have for those x_{ik}, x_i

$$\lim V = \infty \qquad \text{for } (x_1, \ldots, x_m) \to \infty ,$$

as the ellipsoid contains the convex hull of S and the point $(x_1,$

Bull. 8 (1942), pp. 8-11.

$\ldots,x_m)$. Consequently there exists a set x^o_{ik}, x^o_i, for which V is a minimum, among all x_{ik}, which satisfy (18) and (17) for all $y \subset S$, and for which the x_{ik} are coefficients of a positive definite form.[9)]

We are here again more interested in deriving significant properties of the minimum ellipsoid than in actually "determining" it in terms of S.

As V and -d take their least value simultaneously, we can conclude from theorem I that there exist non-negative constants λ_o, $\ldots$, λ_s, which do not all vanish, such that the function

$$(18) \qquad \Phi(x) = \lambda_o d + \sum_{r=1}^{s} \lambda_r \Big[1 - \sum_{i,k} x_{ik}(y^r_i - x_i)(y^r_k - x_k)\Big]$$

of the $n = \frac{m(m+3)}{2}$ independent variables

$$x_i \qquad (i = 1,\ldots,m)$$

$$x_{ik} \qquad (i,k = 1,\ldots,m; \quad i \leqq k)$$

has a critical value at x^o. Here $y^1,\ldots,y^s$ are points on the boundary of the convex hull of S, for which

$$\sum_{i,k} x^o_{ik}\,(y^r_i - x^o)(y^r_k - x^o) = 1.$$

As $\Phi(x)$ is symmetric in x_{ik} and x_{ki}, the first derivatives of Φ with respect to the x_i $(i = 1,\ldots,m)$ and all x_{ik} $(i,k = 1,\ldots,m)$ must vanish at the critical point. We may apply an affine transformation to E_m, so that the minimum ellipsoid becomes the unit sphere about the origin:

$$x^o_{ik} = \delta_{ik}, \qquad x^o_i = 0.$$

9) For a minimizing sequence the x_{ik} cannot tend towards the coefficients of a non-definte form, as the determinant d of the x_{ik} has to become a maximum, and hence is bounded away from 0.

As

$$\left(\frac{\partial d}{\partial x_{rt}}\right)_{x_{ik} = \delta_{ik}} = \delta_{rt} ,$$

we obtain the following relations:

$$\text{(19a)} \qquad \lambda_o \; \delta_{ik} = \sum_{r=1}^{s} \lambda_r y_i^r y_k^r \qquad \text{for } i,k = 1,\dots,m$$

$$\text{(19b)} \qquad 0 = \sum_{r=1}^{s} \lambda_r y_i^r \qquad \text{for } i = 1,\dots,m$$

$$\text{(19c)} \qquad \lambda_o \geqq 0, \; \lambda_1 > 0 ,\dots, \lambda_m > 0$$

$$\text{(19d)} \qquad \sum_{i=1}^{m} (y_i^r)^2 = 1 \qquad \text{for } r = 1,\dots,s$$

Summing (19a) over all $i=k$, we obtain from (19d) the relation

$$\text{(19e)} \qquad m\,\lambda_o = \sum_{r=1}^{s} \lambda_r \;,$$

which shows that λ_o is positive. It follows from (19) for any ellipsoid containing the points y^r

$$m\,\lambda_o = \sum_r \lambda_r \geqq \sum_r \lambda_r \left(\sum_{i,k} x_{ik}(y_i^r - x_i)(y_k^r - x_k) \right)$$

$$= \lambda_o \sum_i x_{ii} + m \sum_{i,k} x_{ik} x_i x_k \geqq \lambda_o \sum_i x_{ii}$$

$$\geqq m\,\lambda_o (\det x_{ik})^{1/m} \; . \qquad \text{10)}$$

Consequently the volume of any ellipsoid containing S is at least

10) For a deinite form the expression $\sum_i x_{ii}$ is the sum of the Eigen-values, $d = \det(x_{ik})$ is the product. Hence by the well known inequality between arithmetic and geometric means it follows that

$$\sum_i x_{ii} \geqq m d^{1/m}$$

equal to that of the unit-sphere. This shows that the ellipsoid of least volume containing S is at the same time the ellipsoid of least volume containing the points $y^1, \ldots, y^s$ of the boundary of the convex hull of S, where $s \leqq \frac{m(m+3)}{2}$.

Let $u_1, \ldots, u_m$ be any numbers with $\sum_i u_i^2 = 1$. Introduce

$$P_r = \sum_{i=1}^{m} u_i y_i^r$$

Then, because of (19d)

(20) $$|P_r| \leqq 1 \qquad \text{for } r = 1, \ldots, s.$$

On the other hand we have for any t, using (19a, b e)

$$\sum_{r=1}^{s} \lambda_r (P_r + t)^2 = (t^2 + \frac{1}{m}) \sum_{r=1}^{s} \lambda_r .$$

It follows that

(21) $$\underset{r}{\text{Maximum}}\ (P_r + t)^2 \geqq t^2 + \frac{1}{m} .$$

Hence for any t there exists an r such that

$$P_r^2 + 2tP_r - \frac{1}{m} \geqq 0 .$$

The lefthand side of this inequality is a quadratic function of P_r, whose roots may be α, β. We then see that for any α, β with $\alpha\beta = -\frac{1}{m}$, there is a P_r outside the interval

$$\alpha < x < \beta .$$

If we put

(22) $$M = \underset{r}{\text{Maximum}}\ P_r, \quad -\mu = \underset{r}{\text{Minimum}}\ P_r$$

it follows that

(23) $$M\mu \geqq \frac{1}{m} .$$

Consequently

(24) $$M + \mu \geqq \frac{2}{\sqrt{m}} ,$$

and, because of (20),

$$M \geqq \frac{1}{m}, \qquad \mu \geqq \frac{1}{m}. \tag{25}$$

As M and μ are the distances of the two planes of support of the set formed by the y^r in the direction u, it follows that the convex hull of the y^r contains the sphere of radius $\frac{1}{m}$ about the origin, and that the distance of any two parallel planes of support of that convex hull is $\geqq \frac{2}{\sqrt{m}}$. The same holds then for the convex hull of S. We have then the following theorem in terms of the original space before the affine transformation:

Theorem III.

If K is the ellipsoid of smallest volume containing a set S in E_m, then the ellipsoid K' which is concentric and homothetic to K at the ratio 1/m is contained in the convex hull of S.

The example of a simplex shows that the constant $\frac{1}{m}$ is the best possible one in this connection.

As the boundary of the convex hull of S may be an arbitrary convex surface, we see that any closed convex surface lies between two concentric homothetic ellipsoids of ratio $= \frac{1}{m}$. We also have from (24):

Any convex body can be transformed by an affine transformation into a body, for which the ratio of diameter and breadth is $\leqq \frac{1}{\sqrt{m}}$.[11]

A stronger inequality can be derived in the case, where S is symmetric to a point, say the origin. Let K be the ellipsoid of least volume containing S, which has its center at the origin. In this case we again obtain $\lambda_0, \ldots, \lambda_s, y^1, \ldots, y^s$, such that (19a,c,d) are satisfied. We can conclude that (21) holds for t = 0, i.e. that

11) Here "breadth" is defined as minimum distance of any two parallel planes of support.

$$\underset{r}{\text{Maximum}}\ P_r^2 \geqq \frac{1}{m}\ .$$

Then of any two parallel planes of support of the convex hull of S (after suitable affine transformation) at least one has a distance $\frac{1}{\sqrt{m}}$ from the origin. As however S is symmetric to the origin, the same holds then for the other plane of support. Hence:

If S is a set symmetric to the point O, and K the ellipsoid of least volume containing S and having its center at O, then the ellipsoid, which is concentric and homothetic to K at the ratio $\frac{1}{\sqrt{m}}$ is contained in the convex hull of S.

Again $\frac{1}{\sqrt{m}}$ is the best possible constant in this connection, as is seen from the example of the m-dimensional "cube" or of the m-dimensional analogon to an octahedron.

If the convex hull of S is represented by its "gage function" ("Distanzfunktion"),[12] we have the following theorem:

For any function $f(x) = f(x_1, \ldots, x_m)$, which satisfies the conditions

$$f(\mu x) = |\mu|\, f(x) \qquad \text{for all numbers } \mu$$

$$f(x) > 0 \quad \text{for } x \neq 0$$

$$f(x + y) \leqq f(x) + f(y)$$

there exists a positive definite quadratic form $Q = Q(x)$, such that

$$\sqrt{\frac{1}{m}\, Q(x)} \leqq f(x) \leqq \sqrt{Q(x)} \qquad \text{for all } x.$$

It is to be expected that for a convex body S the ratio of minimum circumscribed ellipsoid to the volume of S reaches its largest value for a simplex (respectively for a cube, in case S is symmetric to a point). However the author has been unable to prove

12) See Bonnesen-Fenchel, loc. cit. p. 21.

this statement for general m.[13] If true, the statement must be a consequence of the relations (19), which are characteristic for the circumscribed ellipsoid of least volume.

13) For m = 2 this was proved by F. Behrend, loc. cit.

THE UNIVERSITY OF CHICAGO

MINIMA OF FUNCTIONS OF SEVERAL VARIABLES WITH INEQUALITIES AS SIDE CONDITIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF MATHEMATICS

BY

WILLIAM KARUSH

CHICAGO, ILLINOIS

DECEMBER, 1939

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*,
DOI 10.1007/978-3-0348-0439-4_10, © Springer Basel 2014

TABLE OF CONTENTS

MINIMA OF FUNCTIONS OF SEVERAL VARIABLES WITH INEQUALITIES AS SIDE CONDITIONS

1. Introduction. The problem of determining necessary conditions and sufficient conditions for a relative minimum of a function $f(x_1, x_2, \ldots, x_n)$ in the class of points $x = (x_1, x_2, \ldots, x_n)$ satisfying the equations $g_\alpha(x) = 0$ ($\alpha = 1, 2, \ldots, m$), where the functions f and g_α have continuous derivatives of at least the second order, has been satisfactorily treated [1]*. This paper proposes to take up the corresponding problem in the class of points x satisfying the inequalities

$$g_\alpha(x) \geqq 0 \qquad (\alpha = 1, 2, \ldots, m), \tag{1}$$

where m may be less than, equal to, or greater than n.

We shall be interested in a minimizing point x^0 at which all the functions g_α vanish. The reason we limit our attention to this case is that if $f(x^0)$ = minimum and, say, $g_1(x^0) > 0$ then by continuity $g_1(x) \geqq 0$ for all x sufficiently close to x^0 and hence the condition $g_1(x) \geqq 0$ puts no restriction on the problem so far as the theory of relative minima is concerned. Henceforth in this paper whenever we state "$f(x^0)$ is a minimum" or "x^0 is a minimizing point" we assume that $g_\alpha(x^0) = 0$ for every α.

*Numbers in brackets refer to the list of references at the end of the paper.

We shall not limit ourselves to the case when f and g_α are of class C''. In Sections 3 and 4 we consider the minimum problem under the assumption that the functions f and g_α are merely of class C' near a point $x = x^o$. However in Sections 5 and 6 we do restrict attention to the case when the functions are of class C''. Section 2 will deal with some properties of linear inequalities.

We shall have occasion to use all the results in the first part of Bliss's paper [1] and for convenience we list them here. They are concerned with the problem of minimizing $f(x)$ in the class of points x satisfying equations

$$h_\alpha(x) = 0 \qquad (\alpha = 1,2,\ldots,m < n),$$

and may be compared with the results obtained in this paper. One needs only continuous first derivatives for Theorem 1:1 and continuous second derivatives for the other theorems.

THEOREM 1:1. _A first necessary condition for_ $f(x^o)$ _to be a minimum is that there exist constants_ χ_o, χ_α _not all zero such that the derivatives_ H_{x_i} _of the function_

$$H = \chi_o f + \chi_\alpha h_\alpha$$

all vanish at x^o.

LEMMA 1:1. _If_ $\| h_{\alpha x_i}(x^o) \|$ _has rank_ m, _then for every set of constants_ η_i $(i = 1,2,\ldots,n)$ _satisfying the equations_

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

there exists a curve $x_i(t)$ _having continuous second derivatives near_ $t = 0$, _satisfying the equations_ $h_\alpha[x(t)] = 0$, _and such that_

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

THEOREM 1:2. *If* $\|h_{\alpha x_i}(x^0)\|$ *has rank* m *and* $f(x^0)$ *is a minimum then the condition*

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

must hold for every set η_i *satisfying* $h_{\alpha x_i}(x^0)\eta_i = 0$, *where* $H = f + \chi_\alpha h_\alpha$ *is the function formed with the unique set of multipliers* $\chi_0 = 1$, χ_α *belonging to* x^0.

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

THEOREM 1:3. *If a point* x^0 *has a set of multipliers* $\chi_0 = 1$, χ_α *for which the function* $H = f + \chi_\alpha h_\alpha$ *satisfies the conditions*

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

for all sets η_i *satisfying the equations*

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

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

2. *Preliminary theorems on linear inequalities*. To introduce the important theorem which is about to follow we consider the system of linear inequalities

$$(2) \qquad \begin{aligned} L_1 &\equiv A_{11}u_1 + A_{12}u_2 + \dots + A_{1n}u_n \geqq 0 \\ L_2 &\equiv A_{21}u_1 + A_{22}u_2 + \dots + A_{2n}u_n \geqq 0 \\ &\quad \dots \qquad \dots \qquad \dots \qquad \dots \\ L_m &\equiv A_{m1}u_1 + A_{m2}u_2 + \dots + A_{mn}u_n \geqq 0, \end{aligned}$$

in which the A's are real constants. If for every solution u of (2) the inequality

$$(2') \qquad \Phi \equiv A_1u_1 + A_2u_2 + \dots + A_nu_n \geqq 0$$

is satisfied, then the inequality (2') is called a <u>consequence</u> of the system of inequalities (2). Farkas, in his paper [4], proved the following theorem. (See also Corollary 1, p. 47 of Dines and McCoy [3]).

THEOREM 2:1. *If (2') is a consequence of (2) then there exist non-negative constants C_α such that*

$$\Phi \equiv C_1L_1 + C_2L_2 + \dots + C_mL_m.$$

The solution $(u_1,u_2,\dots,u_n) = (0,0,\dots,0)$ will be called a trivial solution of (2). We note that the theorem does not assume that there necessarily exists a non-trivial solution of (2).

We make an inductive proof. If $n = 1$ the conclusion is readily verified. We suppose the theorem true for $n-1$ variables $u_1,u_2,\dots,u_{n-1}$ and make the proof for n variables. If $\Phi \equiv 0$ then the conclusion is obvious. Hence we assume some A_i is different from zero and, for convenience, let $A_n \neq 0$. Solving $\Phi = A_1u_1 + A_2u_2 + \dots + A_nu_n$ for u_n we obtain $u_n = \frac{\Phi}{A_n} - \frac{1}{A_n}\left[A_1u_1 + \dots + A_{n-1}u_{n-1}\right]$ which we substitute in L_i. The result is

$$(3)\quad L_i \equiv A_{i1}u_1 + \dots + A_{in-1}u_{n-1} + \frac{A_{in}}{A_n}\Phi - \frac{A_{in}}{A_n}\left[A_1u_1 + \dots + A_{n-1}u_{n-1}\right] \geqq 0.$$

If the coefficient of Φ is different from zero divide both sides of (3) by $\left|\frac{A_{in}}{A_n}\right|$ and obtain an inequality $\bar{L}_i \geqq 0$ in which the coefficient of Φ is ± 1. Since $\bar{L}_i$ is a positive multiple of L_i we can replace the latter by $\bar{L}_i$ in (2). This we do and, to simplify notation, drop the bar over L_i. With this understanding the system (2) can be rewritten as

$$(4)\quad \begin{array}{llll} L_{i_1} \equiv \Phi + P_1 \geqq 0, & L_{i_2} \equiv \Phi + P_2 \geqq 0, & \dots, & L_{i_r} \equiv \Phi + P_r \geqq 0, \\ L_{j_1} \equiv -\Phi + N_1 \geqq 0, & L_{j_2} \equiv -\Phi + N_2 \geqq 0, & \dots, & L_{j_s} \equiv -\Phi + N_s \geqq 0, \\ L_{k_1} \equiv \quad Z_1 \geqq 0, & L_{k_2} \equiv \quad Z_2 \geqq 0, & \dots, & L_{k_t} \equiv \quad Z_t \geqq 0, \end{array}$$

where $r + s + t = m$ and the $P_1, P_2, \ldots, N_1, N_2, \ldots, Z_1, Z_2, \ldots$ are linear forms in $u_1, u_2, \ldots, u_{n-1}$. If we consider (4) as a system of inequalities with independent variables $u_1, \ldots, u_{n-1}, \Phi$ then from the fact that (2') is a consequence of (2) it follows that

$$\Phi \geqq 0 \tag{4'}$$

is a consequence of (4).

There is at least one linear form in $(u_1, u_2, \ldots, u_{n-1}, \Phi)$ of the type displayed in the first line of (4). For, if this were not the case then $(u_1, u_2, \ldots, u_{n-1}, \Phi) = (0, 0, \ldots, 0, -1)$ would be a solution of (4) in contradiction to (4'). We may also assume that no one of the P's is identically zero, since if for example $P_1 \equiv 0$ then $\Phi \equiv L_{i_1}$ and the conclusion would hold. By adding each inequality in the first line of (4) to each inequality in the second line we obtain

$$\begin{array}{llll}
 & L_{i_1} \equiv \Phi + P_1 \geqq 0, & \ldots, & L_{i_r} \equiv \Phi + P_r \geqq 0 \\
 & L_{i_1} + L_{j_1} \equiv P_1 + N_1 \geqq 0, & \ldots, & L_{i_1} + L_{j_s} \equiv P_1 + N_s \geqq 0 \\
(5) & L_{i_2} + L_{j_1} \equiv P_2 + N_1 \geqq 0, & \ldots, & L_{i_2} + L_{j_s} \equiv P_2 + N_s \geqq 0 \\
 & \ldots \quad \ldots & \ldots & \ldots \quad \ldots \\
 & L_{i_r} + L_{j_1} \equiv P_r + N_1 \geqq 0, & \ldots, & L_{i_r} + L_{j_s} \equiv P_r + N_s \geqq 0 \\
 & L_{k_1} \equiv Z_1 \geqq 0, & \ldots, & L_{k_t} \equiv Z_t \geqq 0.
\end{array}$$

For each solution $(u_1, \ldots, u_{n-1}, \Phi)$ of (5) we must have

$$\Phi \geqq 0. \tag{5'}$$

For, let there be a solution with $\Phi < 0$. Then every P is positive and we may suppose, for convenience, that $P_1 > 0$ is the smallest P. Putting $\Phi = -P_1$ we still have a solution of (5). But by substituting $P_1 = -\Phi$ in the second line of (5) we see that the latter solution is also a solution of (4), which is impossible by (4'). Hence (5') is a consequence of (5).

We now consider the system of inequalities

$$
(6)\qquad
\begin{array}{ccc}
P_1 \geqq 0, & \ldots, & P_r \geqq 0 \\
P_1 + N_1 \geqq 0, & \ldots, & P_1 + N_s \geqq 0 \\
\ldots & \ldots & \ldots \\
P_r + N_1 \geqq 0, & \ldots, & P_r + N_s \geqq 0 \\
Z_1 \geqq 0, & \ldots, & Z_t \geqq 0.
\end{array}
$$

From the assumption made above that no P is identically zero we see that the system (6) contains at least one form which is not identically zero. If (6) has a non-trivial solution then some P must vanish for every solution. For, if this were not the case then for every P_i there would be a solution $u_1^{(i)}, \ldots, u_{n-1}^{(i)}$ of (6) for which $P_i > 0$. The solution $(u_1, \ldots, u_{n-1}) = (u_1^{(1)} + u_1^{(2)} + \ldots, \ldots, u_{n-1}^{(1)} + u_{n-1}^{(2)} + \ldots)$ makes $P_i > 0$ for every $i = 1, 2, \ldots, r$. From this we deduce that (5) has a solution with $\Phi < 0$, which is a contradiction of the fact that (5') is a consequence of (5). Hence we may suppose that $P_1 = 0$ for every solution of (6). It follows that

$$
(6') \qquad -P_1 \geqq 0
$$

is a consequence of (6). In the case that (6) has no non-trivial solution then (6') is still a consequence of (6). By our induction assumption there exist non-negative constants a, b, c such that

$$
\begin{aligned}
-P_1 \equiv\ & a_1P_1 + \ldots + a_rP_r + c_1Z_1 + \ldots + c_tZ_t \\
& + b_{11}(P_1 + N_1) + b_{12}(P_1 + N_2) + \ldots + b_{1s}(P_1 + N_s) \\
& \qquad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \\
& + b_{r1}(P_r + N_1) + b_{r2}(P_r + N_2) + \ldots + b_{rs}(P_r + N_s).
\end{aligned}
$$

Employing the identities in (5) we find

$$\begin{aligned}(1+a_1+\dots+a_r)\Phi \equiv\ & (1+a_1)L_{i_1}+\dots+a_rL_r + c_1L_{k_1}+\dots+c_tL_{k_t} \\ & + b_{11}(L_{i_1}+L_{j_1})+b_{12}(L_{i_1}+L_{j_2})+\dots+b_{1s}(L_{i_1}+L_{j_s}) \\ & + b_{r1}(L_{i_r}+L_{j_1})+\dots+b_{rs}(L_{i_r}+L_{j_s}),\end{aligned}$$

which proves the theorem.

We define $u = (u_1, u_2, \dots, u_n)$ as a solution of the system

$$(7) \qquad A_{\alpha i}u_i >' 0$$

in case $A_{\alpha i}u_i \geqq 0$ is satisfied with the strict inequality holding for at least one value of α. A set of numbers will be called <u>positive definite</u> in case every number of the set is positive.

LEMMA 2:1. <u>A necessary and sufficient condition that</u> (7) <u>admit no solution</u> u <u>is that the system of equalities</u>

$$(8) \qquad A_{\alpha i}v_\alpha = 0 \qquad (i = 1,2,\dots,n),$$

<u>admit a positive definite solution</u> $v = (v_1, v_2, \dots, v_m)$.

This is Theorem 12 of Dines and McCoy [3]. We employ this lemma to obtain the following modification of Theorem 2:1.

THEOREM 2:2. <u>If for every non-trivial solution</u> u <u>of</u> (2) <u>it is true that</u> $\Phi \equiv A_i u_i > 0$ <u>then there exist constants</u> $C_\alpha > 0$ <u>such that</u>

$$\Phi \equiv C_1L_1 + C_2L_2 + \dots + C_mL_m.$$

<u>If the matrix</u> $\| A_{\alpha i} \|$ <u>has rank</u> n <u>then the converse is also true.</u>

To prove the first part of the theorem we note that the system of inequalities

$$A_{11}u_1 + A_{12}u_2 + \ldots + A_{1n}u_n > {}'0$$
$$\ldots \qquad \ldots \qquad \ldots$$
$$A_{m1}u_1 + A_{m2}u_2 + \ldots + A_{mn}u_n > {}'0$$
$$- A_1u_1 - A_2u_2 - \ldots - A_nu_n > {}'0$$

has no solution u. We use Lemma 2:1 with (7) replaced by this system and obtain positive constants C_α such that

$$C_\alpha A_{\alpha i} - A_i = 0 \qquad\qquad (i = 1,2,\ldots,n),$$

as desired. If $\| A_{\alpha i} \|$ has rank n then every non-trivial solution u of $A_{\alpha i}u_i \geqq 0$ is also a solution of $A_{\alpha i}u_i > {}'0$. Hence $\phi \equiv C_\alpha L_\alpha > 0$.

For simplicity we use the letter U to denote the class of all non-trivial solutions u of (2).

THEOREM 2:3. <u>The statements:</u>

(i) <u>there exists a</u> $\bar{u}$ <u>satisfying</u> $A_{\alpha i}\bar{u}_i > 0$ <u>for every</u> α,

(ii) U <u>is</u> n-<u>dimensional</u>,

(iii) U <u>is not null and no linear form</u> $A_{\alpha i}u_i$ <u>vanishes for all</u> u <u>belonging to</u> U,

<u>are all equivalent</u>.

The first statement (i) implies (ii) for, by continuity, there is an n-dimensional neighborhood of $\bar{u}$ which belongs to U. The statement (iii) follows from (ii) since if we suppose, for example, that $A_{1i}\, u_i = 0$ for all u belonging to U then obviously U could not contain n linearly independent vectors u and hence could not be n-dimensional. To prove (iii) implies (i) we notice that there are solutions $u^{(1)}, u^{(2)}, \ldots, u^{(m)}$ of (2) such that

$$A_{1i}u_i^{(1)} > 0$$
$$A_{2i}u_i^{(2)} > 0$$
$$\cdots \quad \cdots \quad \cdots$$
$$A_{mi}u_i^{(m)} > 0.$$

Hence we need only set $\bar{u}_i = u_i^{(1)} + u_i^{(2)} + \ldots + u_i^{(m)}$.

For the next theorem we need to introduce the notion of <u>I-rank</u> of a matrix, an integral valued function of a matrix analogous to ordinary rank. But first some preliminary remarks are necessary. Suppose

$$\mathfrak{M} = \left\| \begin{matrix} A_{11} & A_{12} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{matrix} \right\|$$

is an $m \times n$ matrix whose elements A_{ij} are all real. The matrix $\mathfrak{M}$ is said to be <u>I-definite with respect to a given column</u> in case the elements of that column are all positive, or are all negative; $\mathfrak{M}$ will be called <u>I-definite</u> in case it contains at least one column with respect to which it is I-definite.

If $\mathfrak{M}$ is not I-definite with respect to the qth column we divide the elements of that column into 3 classes,

r positive elements : A_{iq} $(i = i_1, i_2, \ldots, i_r)$,

s negative elements : A_{jq} $(j = j_1, j_2, \ldots, j_s)$,

t zero elements : A_{kq} $(k = k_1, k_2, \ldots, k_t)$.

From $\mathfrak{M}$ we derive the matrix $\mathfrak{M}_1^{(q)}$ as follows:

To each pair of elements A_{iq}, A_{jq}, the first positive and the second negative, corresponds one row of $\mathfrak{M}_1^{(q)}$ given by

$$\begin{vmatrix} A_{iq} & A_{i1} \\ A_{jq} & A_{j1} \end{vmatrix}, \ \dots \ \begin{vmatrix} A_{iq} & A_{iq-1} \\ A_{jq} & A_{jq-1} \end{vmatrix}, \ \begin{vmatrix} A_{iq} & A_{iq+1} \\ A_{jq} & A_{jq+1} \end{vmatrix}, \ \dots \ \begin{vmatrix} A_{iq} & A_{in} \\ A_{jq} & A_{jn} \end{vmatrix} .$$

To each zero element A_{kq} corresponds the row

$$A_{k1}, A_{k2}, \dots, A_{kq-1}, A_{kq+1}, \dots, A_{kn}.$$

The matrix $\mathcal{M}_1^{(q)}$ will consist of the rows so formed, the number of rows being $rs + t$. The order of the rows shall be fixed by the rule: (1) each row corresponding to a pair A_{iq}, A_{jq} shall precede every A_{kq} row; (2) of two A_{iq}, A_{jq} rows that one shall precede which has the smaller i or (in case the i's are equal) that one which has the smaller j; (3) of two A_{kq} rows that one shall precede which has the smaller k.

Thus $\mathcal{M}_1^{(q)}$ is well-defined if $\mathcal{M}$ is not I-definite with respect to its qth column. If $\mathcal{M}$ is I-definite with respect to its qth column we define $\mathcal{M}_1^{(q)}$ as the matrix of 1 row and $(n-1)$ columns all of whose elements are $+1$ or -1 according as the elements of the qth column of $\mathcal{M}$ are all positive or all negative. The matrix $\mathcal{M}_1^{(q)}$ will be called the <u>I-complement</u> of the qth column of $\mathcal{M}$, and the set $\mathcal{J}_1$ of matrices $\mathcal{M}_1^{(1)}$, $\mathcal{M}_1^{(2)}$, ..., $\mathcal{M}_1^{(n)}$ will be called the <u>I-minors of</u> $(n-1)$ <u>columns</u> of the matrix $\mathcal{M}$. We notice that if a matrix is I-definite then all its I-complements are likewise I-definite.

Now we form the I-complements for each matrix $\mathcal{M}_1^{(q)}$, and call the set $\mathcal{J}_2$ of all such I-complements the I-minors of $(n-2)$ columns of $\mathcal{M}$. Continuing this process we obtain a finite sequence of sets $\mathcal{J}_1, \mathcal{J}_2, \dots, \mathcal{J}_{n-1}$ where each matrix in $\mathcal{J}_p$ is an I-minor of $n-p$ columns of $\mathcal{M}$. If we define $\mathcal{M}$ as its own I-minor of n columns then $\mathcal{J}_0 \equiv \mathcal{M}$ and the set $\mathcal{J}_0 + \mathcal{J}_1 + \dots + \mathcal{J}_{n-1}$ of matrices constitute all the I-minors of $\mathcal{M}$.

We are ready to make the definition: A matrix will be said to be of <u>I-rank h</u> if it possesses at least one I-minor of h columns which is I-definite, but does not possess any I-minor of h + 1 columns which is I-definite. If none of its I-minors are I-definite then it will be said to be of <u>I-rank 0</u>.

In his paper [2] Dines proves the following theorem, the proof of which we shall omit.

THEOREM 2:4. *A necessary and sufficient condition for the existence of a solution* $\bar{u} = (\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_n)$ of

$$\begin{aligned} A_{11}u_1 + A_{12}u_2 + \ldots + A_{1n}u_n &> 0 \\ \ldots \quad \ldots \quad \ldots \quad &\ldots \\ A_{m1}u_1 + A_{m2}u_2 + \ldots + A_{mn}u_n &> 0 \end{aligned}$$

is that the I-rank of $\| A_{\alpha i} \|$ *be greater than zero.*

3. *Necessary conditions involving only first derivatives.* We make some preliminary definitions. A solution $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ of

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

will be called an <u>admissible direction</u> if λ is not the zero vector. A regular arc $x_i(t)$ $(i = 1,2,\ldots,n;\ 0 \leqq t \leqq t_0)$, will be called <u>admissible</u> in case $g_\alpha[x(t)] \geqq 0$ for every α and t. A point x^0 is a <u>normal point</u> in case the matrix

$$\| g_{\alpha x_i}(x^0) \|$$

has rank m.

THEOREM 3:1. *If* $f(x^0)$ *is a minimum then there exist multipliers* χ_0, χ_α *not all zero such that the derivatives* F_{x_i} *of the function*

$$F(x) = \chi_0 f(x) + \chi_\alpha g_\alpha(x)$$

<u>all vanish at</u> x^o.

In the class of points $(x,z) = (x_1,\ldots,x_n,z_1,\ldots,z_m)$ satisfying $h_\alpha(x,z) = g_\alpha(x) - z_\alpha^2 = 0$ the point $(x,z) = (x^o,0)$ is a minimizing point for f. Hence, by Theorem 1:1, there exist constants χ_0, χ_α not all zero such that the function $H(x,z) = \chi_0 f + \chi_\alpha h_\alpha = \chi_0 f + \chi_\alpha g_\alpha - \chi_\alpha z_\alpha^2$ has $H_{x_i}(x^o,0) = 0$. It follows that $F(x) = \chi_0 f + \chi_\alpha g_\alpha$ has $F_{x_i}(x^o) = 0$.

We note that if $m < n$ the above proof of Theorem 3:1 is unnecessary. For, if x^o is a minimizing point in the class of points satisfying $g_\alpha(x) \geqq 0$ it certainly is a minimizing point in the class satisfying $g_\alpha(x) = 0$, and Theorem 1:1 can be applied directly.

If x^o is a minimizing point for f which is normal then the multiplier χ_0 is not zero and can be given the value one by dividing each number of the set χ_0, χ_α by χ_0 and obtaining a new set $\chi_0 = 1$, χ_α which satisfies the conclusion of the theorem. Assume, then, that x^o is a normal minimizing point and $\chi_0 = 1$. If we suppose, for the moment, that the functions f and g_α have continuous second derivatives then by employing the necessary condition on the second derivatives of $H(x,z)$ at the minimizing point $(x,z) = (x^o,0)$ of f as given in Theorem 1:2 we can easily show that $\chi_\alpha \leqq 0$ $(\alpha = 1,2,\ldots,m)$. For, by the theorem just referred to, the quadratic form

$$\eta_i \eta_k H_{x_i x_k}(x^o,0) + 2\eta_i \zeta_\alpha H_{x_i z_\alpha}(x^o,0) + \zeta_\alpha \zeta_\beta H_{z_\alpha z_\beta}(x^o,0) =$$

$$\eta_i \eta_k F_{x_i x_k}(x^o) - 2\zeta_\alpha^2 \chi_\alpha$$

must be non-negative for all sets $\eta_1, \eta_2, \ldots, \eta_n, \Upsilon_1, \Upsilon_2, \ldots, \Upsilon_m$ for which $\eta_i g_{\alpha x_i}(x^o) = 0$ and $\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_m$ is arbitrary. Setting every η and Υ except Υ_α equal to zero and substituting in the quadratic form we find that $\chi_\alpha \leqq 0$.

However, in this section we shall make proofs of the non-positive character of the multipliers χ_α which do not involve second derivatives, and the case when the minimizing point is normal will appear as a special instance (see the proof of the corollary to Theorem 5:1).

We use Theorem 2:1 to obtain the following necessary condition.

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

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

all vanish at x^o.

By a curve $x_i(t)$ $(0 \leqq t \leqq t_o)$, "issuing from x^o in the direction λ" we mean, of course, that $x_i(0) = x_i^o$ and $x_i'(0) = \lambda_i$. Consider an admissible direction λ and the corresponding admissible curve $x_i(t)$ given in the hypothesis. Let $\mathfrak{f}(t) = f[x(t)]$. Since $\mathfrak{f}(0) \leqq \mathfrak{f}(t)$ for $0 \leqq t \leqq t_o$, it follows that $\mathfrak{f}'(0) \geqq 0$. But $\mathfrak{f}'(0) = f_{x_i}(x^o)\lambda_i$. Hence $f_{x_i}(x^o)\lambda_i \geqq 0$. Then $f_{x_i}(x^o)u_i \geqq 0$ is a consequence of $g_{\alpha x_i}(x^o)u_i \geqq 0$ $(\alpha = 1, 2, \ldots, m)$, and by Theorem 2:1 there exist multipliers $\chi_\alpha \leqq 0$ such that $f_{x_i}(x^o)u_i + \chi_\alpha g_{\alpha x_i}(x^o)u_i \equiv 0$. Thus $f_{x_i} + \chi_\alpha g_{\alpha x_i} = 0$ for every i, and the theorem is proved.

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

One would naturally like to know what the probability is, roughly, that the functions $g_\alpha(x)$ will satisfy property Q, as well as some conditions on the functions g_α which will ensure the satisfaction of Q. In order to partially answer these questions we shall briefly discuss one geometric interpretation of an admissible direction.

The tangent planes to the surfaces $g_\alpha(x) = 0$ at their common point of intersection x^o are given by

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

The straight line issuing from x^o in the admissible direction λ is

$$(9) \qquad S: \; x_i(t) = \lambda_i t + x_i^o \qquad (0 \leqq t \leqq t_o;\; i = 1,2,\ldots,n).$$

Substituting the equations of S in $T_\alpha(x)$ we obtain $T_\alpha[x(t)] \geqq 0$. We conclude that the line S lies in the set of points x near x^o satisfying $T_\alpha(x) \geqq 0$; and since the latter set, in a sense, approximates the set of points x near x^o satisfying $g_\alpha(x) \geqq 0$, if the functions g_α are regular enough, it seems that the satisfaction of property Q is not a great restriction on the functions g_α. In fact, the following corollary states a condition on g_α which makes the line S an admissible arc.

COROLLARY. <u>Suppose that for every admissible direction</u> λ <u>it is true that</u> $g_{\alpha x_i}(x^o)\lambda_i = 0$ <u>implies that</u> $g_{\alpha x_i x_k}(x^o)\lambda_i\lambda_k > 0$.

<u>Then if</u> $f(x^0)$ = <u>minimum the first necessary condition is satisfied</u>.

Consider any admissible direction λ and the corresponding line S given in (9). Define $\bar{g}_\alpha(t) = g_\alpha[x(t)]$ $(\alpha = 1,2,\ldots,m;$ $0 \leqq t \leqq t_0)$. We have $d\bar{g}_\alpha(t)/dt = g_{\alpha x_i}[x(t)]x_i'(t) = g_{\alpha x_i}[x(t)]\,\lambda_i$. Hence

$$\frac{d\bar{g}_\alpha(0)}{dt} = g_{\alpha x_i}(x^0)\ \lambda_i \geqq 0.$$

If $d\bar{g}_\alpha(0)/dt > 0$ then $\bar{g}_\alpha(t)$ is monotonically increasing near $t = 0$ and $\bar{g}_\alpha(t) = g_\alpha[x(t)] \geqq g_\alpha(x^0) = 0$. Hence S lies in the set of points x satisfying $g_\alpha(x) \geqq 0$. If $d\bar{g}_\alpha(0)/dt = 0$ then $d^2\bar{g}_\alpha(t)/dt^2 = g_{\alpha x_i x_k}[x(t)]\ \lambda_i\ \lambda_k$ and by hypothesis

$$\frac{d^2\bar{g}_\alpha(0)}{dt^2} = g_{\alpha x_i x_k}(x^0)\ \lambda_i\ \lambda_k > 0.$$

Therefore $\bar{g}_\alpha(t)$ is monotonically increasing and, as before, satisfies $g_\alpha[x(t)] \geqq 0$. We have shown that with S the hypotheses of Theorem 3:2 are satisfied, and the conclusion follows.

In Theorem 3:3 we obtain the same necessary condition that Theorem 3:2 yielded but under a different hypothesis.

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

First we prove that if λ is such that $g_{\alpha x_i}(x^0)\ \lambda_i > 0$ for every α then $f_{x_i}(x^0)\ \lambda_i \geqq 0$. Let g represent any one of the g_α and define, as before,

$$\bar{g}(t) = g[x(t)], \qquad \bar{f}(t) = f[x(t)],$$

where x(t) represents the equations of the line S in (9). Since $d\bar{g}(0)/dt = g_{x_i}(x^0)\ \lambda_i > 0$, $\bar{g}(t)$ is monotonically increasing, $g[x(t)] \geqq g(x^0) = 0$, and S is an admissible arc. Thus $\bar{f}(0) \leqq \bar{f}(t)$, and consequently

$$\frac{dF(0)}{dt} = f_{x_i}(x^o)\,\lambda_i \geqq 0.$$

Now suppose μ_i is an admissible direction. We define a family of directions

$$\nu_i(s) = \bar{\lambda}_i + s(\mu_i - \bar{\lambda}_i) \qquad (0 \leqq s \leqq 1),$$

where $\bar{\lambda}$ is given in the hypothesis of the theorem. Rewriting $\nu_i(s) = (1 - s)\bar{\lambda}_i + s\mu_i$, it is clear that $g_{x_i}(x^o)\nu_i(s) > 0$ for $0 \leqq s < 1$. From the first part of the proof,

$$f_{x_i}(x^o)\,\nu_i(s) \geqq 0 \qquad (0 \leqq s < 1),$$

so that

$$\lim_{s \to 1} f_{x_i}(x^o)\,\nu_i(s) = f_{x_i}(x^o)\,\mu_i \geqq 0.$$

Hence the inequality $f_{x_i}(x^o)\,u_i \geqq 0$ is a consequence of $g_{\alpha x_i}(x^o)\,u_i \geqq 0$ $(\alpha = 1,2,\ldots,m)$, and the theorem follows from Theorem 2:1.

Suppose $m = n$ and the determinant of $\| g_{\alpha x_i}(x^o) \|$ is different from zero. For this case we can write the first necessary condition in an entirely equivalent form as follows.

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

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

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

The system of equations

$$g_{1x_i}(x^0)u_i = 1$$

$$\dots \qquad \dots$$

$$g_{nx_i}(x^0)u_i = 1$$

has a solution $u = \bar{\lambda}$ since determinant $\| g_{\alpha x_i}(x^0) \| \neq 0$. Thus

$$g_{\alpha x_i}(x^0)\bar{\lambda}_i > 0,$$

and we can apply Theorem 3:3 to obtain the first necessary condition; that is, there exist multipliers $\gamma_\alpha \leqq 0$ such that

$$f_{x_i}(x^0) = -\gamma_\alpha g_{\alpha x_i}(x^0).$$

Multiplying both sides of the last equation by $G_{i\beta}$ and summing with respect to the index i, we obtain

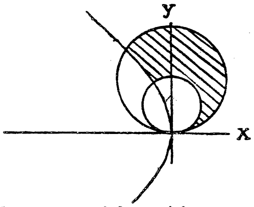
$$f_{x_i}(x^0)G_{i\beta} = -\gamma_\beta \geqq 0,$$

as desired.

The problem of determining necessary and sufficient conditions for the existence of an admissible direction λ satisfying $g_{\alpha x_i}(x^0)\bar{\lambda}_i > 0$ naturally arises in the consideration of Theorem 3:3. The question is answered by Theorems 2:3 and 2:4. In particular, the latter theorem provides a useful method for determining in a finite number of steps whether or not such an admissible vector λ does exist.

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

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



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

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

The only admissible direction is $(a,0)$ with $a > 0$. There is no solution of $g_{\alpha x}(0,0)\lambda_1 + g_{\alpha y}(0,0)\lambda_2 > 0$ for all α. Also $g_{2xx}(0,0)a^2 < 0$ so that the hypothesis of the corollary to Theorem 3:2 is not satisfied. However, it is obvious that there is an admissible arc issuing from $(0,0)$ in the direction $(a,0)$.

4. <u>Sufficient conditions involving only first derivatives.</u> By a proper strengthening of the first necessary condition we can obtain a sufficiency theorem without resorting to second derivatives.

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

By Taylor's expansion formula,

$$f(x) - f(x^0) = f_{x_i}(x')\eta_i$$
$$0 \leqq g_\alpha(x) = g_{\alpha x_i}(x'_\alpha)\eta_i \qquad (\alpha = 1,2,\ldots,m),$$

for x near x^0 and x satisfying $g_\alpha(x) \geqq 0$, where $\eta_i = x_i - x_i^0$, $x'_{\alpha i} = x_i^0 + \theta_\alpha(x_i - x_i^0)$. By hypothesis,

$$
(10)\quad
\begin{aligned}
c_1 g_{1x_1}(x^0) + c_2 g_{2x_1}(x^0) + \dots + c_m g_{mx_1}(x^0) &= f_{x_1}(x^0) \\
\dots \qquad\qquad \dots \qquad\qquad & \dots \\
c_1 g_{1x_n}(x^0) + c_2 g_{2x_n}(x^0) + \dots + c_m g_{mx_n}(x^0) &= f_{x_n}(x^0)
\end{aligned}
$$

where $c_\alpha = -\lambda_\alpha > 0$. For convenience suppose

$$
\begin{vmatrix}
g_{1x_1}(x^0) & \dots & g_{nx_1}(x^0) \\
\dots & \dots & \dots \\
g_{1x_n}(x^0) & \dots & g_{nx_n}(x^0)
\end{vmatrix} \neq 0.
$$

We fix $c_{n+1}, \dots, c_m$ in (10) and solve for $c_1, \dots, c_n$ as continuous functions of the coefficients $g_{\alpha x_i}(x^0)$ and $f_{x_i}(x^0)$. Hence for $A_{\alpha i}$ sufficiently close to $g_{\alpha x_i}(x^0)$ and A_i sufficiently close to $f_{x_i}(x^0)$ there exists a unique solution $\bar{c}_1 > 0$, $\bar{c}_2 > 0$, ..., $\bar{c}_m > 0$ of

$$
\bar{c}_\alpha A_{\alpha i} = A_i \qquad (i = 1, 2, \dots, n).
$$

Hence for x sufficiently close to x^0 there exist constants $\bar{c}_1 > 0$, ..., $\bar{c}_m > 0$ such that

$$
\bar{c}_\alpha g_{\alpha x_i}(x_\alpha') = f_{x_i}(x'),
$$

$$
f(x) - f(x^0) = f_{x_i}(x')\eta_i = \bar{c}_\alpha g_{\alpha x_i}(x_\alpha')\eta_i \geqq 0,
$$

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

We have a sufficiency theorem corresponding to the necessary condition in the corollary to Theorem 3:3.

COROLLARY. _Suppose_ $m = n$ _and determinant_ $\| g_{\alpha x_i}(x^0) \| \neq 0$. _We let_ $\| G_{i\alpha} \|$ _be the inverse matrix of_ $\| g_{\alpha x_i} \|$. _If_ x^0 _is a point satisfying_ $g_\alpha(x^0) = 0$ _such that_

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

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

We define $\chi_\alpha < 0$ by the equation

$$f_{x_i}(x^o)G_{i\alpha} = -\chi_\alpha .$$

Multiplying both sides by $g_{\alpha x_j}(x^o)$ and summing with respect to the index α, we obtain

$$f_{x_j}(x^o) = -\chi_\alpha g_{\alpha x_j}(x^o),$$

and the conclusion follows from Theorem 4:1.

The following sufficiency theorem is entirely equivalent to Theorem 4:1.

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

This result follows at once from Theorem 2:2 and Theorem 4:1.

5. *A necessary condition involving second derivatives.* Suppose $f(x^o)$ is a minimum, $\| g_{\alpha x_i}(x^o) \|$ has rank r, and for convenience the first r row vectors are linearly independent. We also suppose that there exist multipliers χ_α such that $F = f + \chi_\alpha g_\alpha$ has $F_{x_i}(x^o) = 0$, that is,

$$f_{x_i}(x^o) + \chi_\alpha g_{\alpha x_i}(x^o) = 0 \qquad (i = 1,2,\ldots.n).$$

Since all the row vectors are linear combinations of the first r we may suppose $\chi_{r+1} = 0, \ldots, \chi_m = 0$. In this form the multipliers χ_α are unique for, if χ_α' is any other set with $\chi_{r+1}' = 0, \ldots, \chi_m' = 0$ then $\sum_{\alpha=1}^{r} (\chi_\alpha' - \chi_\alpha) g_{\alpha x_i}(x^o) = 0$ and hence $\chi_\alpha' = \chi_\alpha$ $(\alpha = 1,2,\ldots,r)$.

If the hypotheses of Theorem 3:3 are satisfied and $\| g_{\alpha x_i}(x^0) \|$ has rank $r = 1$ or 2 we can show that there are respectively one or two linearly independent rows whose unique multipliers are non-positive. It is obviously sufficient to prove the following proposition: If there exists an admissible direction $\bar{\lambda}$ satisfying $g_{\alpha x_i}(x^0)\bar{\lambda}_i > 0$, every row of $\| g_{\alpha x_i}(x^0) \|$ is a linear combination with non-negative coefficients of some r linearly independent rows ($r = 1$ or 2). If $r = 1$ the proof is obvious. If $r = 2$ we make an inductive proof. The case $m = 2$ is clear. We assume the proposition for $m - 1$ and make the proof for m. By our induction assumption we may suppose that the first two rows of $\| g_{\alpha x_i}(x^0) \|$ are linearly independent and every other row, except possibly the last, is a linear combination with non-negative coefficients of these two. For the last row we have

$$a g_{1x_i} + b g_{2x_i} + c g_{mx_i} = 0 \qquad (i = 1,2,\ldots,n),$$

with $(a,b,c) \neq (0,0,0)$. Hence $a g_{1x_i}\bar{\lambda}_i + b g_{2x_i}\bar{\lambda}_i + c g_{mx_i}\bar{\lambda}_i = 0$. The numbers a, b, c cannot all be of the same sign. For, if they were then the last expression would be different from zero since $g_{\alpha x_i}\bar{\lambda}_i > 0$. Hence one of the three vectors is a linear combination with non-negative coefficients of the other two, and it follows that every row vector is a linear combination with non-negative coefficients of the same two.

That one cannot hope to extend the above proposition to the case when $r \geqq 3$ is shown by the following example. Let

$$\| g_{\alpha x_i}(x^0) \| = \begin{Vmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & -1 \end{Vmatrix}$$

The system of inequalities

$$g_{\alpha x_i}(x^o)\,\lambda_i > 0$$

has a solution $(\bar\lambda_1, \bar\lambda_2, \bar\lambda_3, \bar\lambda_4) = (1,1,1,1)$. If we take the linear combination of the rows with respective coefficients -1, -1, +1, +1 we obtain the zero vector. Since the rank of $\| g_{\alpha x_i}(x^o) \|$ is three any solution v of $g_{\alpha x_i} v_\alpha = 0$ is given by $v = k(-1,-1,+1,+1)$. Hence no row can be a linear combination with positive coefficients of the other three rows.

The next theorem gives a necessary condition involving the second derivatives of the functions f and g_α.

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

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

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

We notice that for any particular η satisfying $g_{\alpha x_i}(x^o)\eta_i = 0$ the selection of the r linearly independent rows that shall satisfy with η the hypotheses of the theorem, depends upon η. In the statement of the theorem we have taken an η and renumbered the functions g_α so that the r linearly independent rows going with η are the first r rows.

We have $g_\alpha[x(t)] \equiv 0$ and hence $g_{\alpha x_i} x_i'(t) \equiv 0$ for $\alpha = 1,2,\ldots,r$. Let $\bar{f}(t) = f[x(t)]$. Then

$$'(t) = f_{x_i}[x(t)]x_i'(t) = (f_{x_i} + \sum_{\alpha=1}^{r} \lambda_\alpha g_{\alpha x_i})x_i'(t) = F_{x_i}[x(t)]x_i'(t),$$

$$\bar{f}'(0) = F_{x_i}[x^o]\eta_i = 0.$$

ut since $f(x^o)$ is a minimum $\bar{f}(0) \leqq \bar{f}(t)$, $\bar{f}''(0) = F_{x_i x_k}[x^o]\eta_i\eta_k \geqq 0$, nd the theorem is proved.

Theorem 5:1 can be applied in the particular case when $x^o)$ is a normal point.

COROLLARY. <u>Suppose x^o is a normal point. Then necessary onditions for $f(x^o)$ to be a minimum are that the first necessary ondition be satisfied and that</u>

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

<u>e satisfied for every admissible direction η satisfying</u>

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

The first necessary condition is easily proved by means of heorem 3:3. For, since the rank of $\| g_{\alpha x_i} \|$ is m there exists solution $\bar{\lambda}$ of $g_{\alpha x_i}(x^o)\bar{\lambda}_i = 1$ $(\alpha = 1,2,\ldots,m)$, and hence a olution of $g_{\alpha x_i}(x^o)\bar{\lambda}_i > 0$.

If the rank of $\| g_{\alpha x_i}(x^o) \|$ is $m = n$ then the second ecessary condition in the corollary is vacuously satisfied since o η exists for which $g_{\alpha x_i}(x^o)\eta_i = 0$. If $m < n$ the second ecessary condition follows if we notice that Lemma 1:1 enables s to satisfy the hypotheses of Theorem 5:1.

6. <u>A sufficiency theorem involving second derivatives.</u> orresponding to Theorem 1:3 we have the following sufficiency heorem.

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

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

or all admissible directions η satisfying

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

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

The proof consists of verifying that the hypotheses of heorem 1:3 are satisfied for the problem of showing that $(x,z) =$ $x^o,0)$ is a minimizing point for f in the class of points $(x,z) =$ $x_1,\ldots,x_n,z_1,\ldots,z_m)$ satisfying

$$h_\alpha(x,z) = g_\alpha(x) - z_\alpha^2 = 0 \qquad (\alpha = 1,2,\ldots,m).$$

et $H(x,z) = f + \chi_\alpha h_\alpha = F(x) - \chi_\alpha z_\alpha^2$. Then

$$H_{x_i}(x^o,0) = F_{x_i}(x^o) = 0, \qquad H_{z_k} = 0.$$

onsider any set $(\eta_i, \zeta_k) \neq (0,0)$ $(i = 1,2,\ldots,n;\ k = 1,2,\ldots,m)$, uch that

$$h_{\alpha x_i}(x^o,0)\eta_i + h_{\alpha z_k}(x^o,0)\zeta_k = 0 \qquad (\alpha = 1,2,\ldots,m),$$

hat is, such that

$$g_{\alpha x_i}(x^o)\eta_i = 0, \qquad \zeta_k \text{ arbitrary.}$$

he quadratic form formed with the second derivatives of H is

$$H_{x_i x_k}(x^o,0)\eta_i\eta_k + 2H_{x_i z_k}(x^o,0)\eta_i\zeta_k + H_{z_k z_\chi}(x^o,0)\zeta_k\zeta_\chi$$

hich reduces to

$$F_{x_i x_k}(x^o)\eta_i\eta_k - 2\lambda_k \zeta_k^2 > 0.$$

ence $(x^o, 0)$ is a minimizing point. It follows that (x^o) is a inimizing point for the original problem.

Under the assumption that the functions f and g_α have ontinuous derivatives of at least the second order, Theorem 4:1 s an immediate corollary of Theorem 6:1. However, as observed efore, Theorem 4:1 also holds for the case when f and g_α have ontinuous derivatives of only the first order.

LIST OF REFERENCES

Bliss, G. A., Normality and Abnormality in the Calculus of Variations, Transactions of the American Mathematical Society, vol. 43 (1938), pp. 365-376.

Dines, L. L., Systems of Linear Inequalities, Annals of Mathematics, vol. 23 (1922), p. 212.

Dines and McCoy, On Linear Inequalities, Transactions of the Royal Society of Canada, vol. 27 (1933), pp. 37-70.

Farkas, J. I., Theorie der einfachen Ungleichungen, Crelle, vol. 124 (1902), p. 1.

VITA

William Karush was born in Chicago, Illinois, on March 1, 1917. He received his early education in the public schools of Chicago, graduating from Murray F. Tuley High School in June, 1934. After attending the Central Y.M.C.A. College in Chicago for two years, he transferred to the University of Chicago; from the latter institution he received the degree of Bachelor of Science in June, 1938. He has been a graduate student in mathematics at the University of Chicago since October, 1938, studying under Professors Albert, Barnard, Bartky, Bliss, Dickson, Graves, Lane, Logsdon, MacLane, and Reid. He wishes to thank these instructors for the helpful part they have played in his mathematical development; particular thanks are due to Professor Graves for his guidance as a teacher and in the writing of this dissertation.

NONLINEAR PROGRAMMING

HAROLD W. KUHN AND ALBERT W. TUCKER
PRINCETON UNIVERSITY AND STANFORD UNIVERSITY

1. Introduction

Linear programming deals with problems such as (see [4], [5]): to maximize a linear function $g(x) \equiv \sum c_i x_i$ of n real variables $x_1, \ldots, x_n$ (forming a vector x) constrained by $m + n$ linear *inequalities*,

$$f_h(x) \equiv b_h - \sum a_{hi} x_i \geqq 0, \; x_i \geqq 0, \qquad h = 1, \ldots, m; \; i = 1, \ldots, n.$$

This problem can be transformed as follows into an equivalent saddle value (minimax) problem by an adaptation of the calculus method customarily applied to constraining *equations* [3, pp. 199–201]. Form the Lagrangian function

$$\phi(x, u) \equiv g(x) + \sum u_h f_h(x).$$

Then, a particular vector x^0 maximizes $g(x)$ subject to the $m + n$ constraints if, and only if, there is some vector u^0 with nonnegative components such that

$$\phi(x, u^0) \leqq \phi(x^0, u^0) \leqq \phi(x^0, u) \quad \text{for all nonnegative } x, u.$$

Such a saddle point (x^0, u^0) provides a solution for a related zero sum two person game [8], [9], [12]. The bilinear symmetry of $\phi(x, u)$ in x and u yields the characteristic duality of linear programming (see section 5, below).

This paper formulates necessary and sufficient conditions for a saddle value of any differentiable function $\phi(x, u)$ of nonnegative arguments (in section 2) and applies them, through a Lagrangian $\phi(x, u)$, to a maximum for a differentiable function $g(x)$ constrained by inequalities involving differentiable functions $f_h(x)$ mildly qualified (in section 3). Then, it is shown (in section 4) that the above equivalence between an inequality constrained maximum for $g(x)$ and a saddle value for the Lagrangian $\phi(x, u)$ holds when $g(x)$ and the $f_h(x)$ are merely required to be concave (differentiable) functions for nonnegative x. (A function is *concave* if linear interpolation between its values at any two points of definition yields a value not greater than its actual value at the point of interpolation; such a function is the negative of a *convex* function—which would appear in a corresponding minimum problem.) For example, $g(x)$ and the $f_h(x)$ can be quadratic polynomials in which the pure quadratic terms are negative semidefinite (as described in section 5).

In terms of *activity analysis* [11], x can be interpreted as an activity vector, $g(x)$ as the resulting output of a desired commodity, and the $f_h(x)$ as unused balances of primary commodities. Then the Lagrange multipliers u can be interpreted as a price vector [13, chap. 8] corresponding to a unit price for the desired commodity, and the Lagrangian function $\phi(x, u)$ as the combined worth of the output of the desired commodity and the unused balances of the primary commodities. These

This work was done under contracts with the Office of Naval Research.

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, 247
DOI 10.1007/978-3-0348-0439-4_11, © Springer Basel 2014

price interpretations seem to relate closely to the price theory in the contemporary paper of K. J. Arrow [1].

A "vector" maximum—of T. C. Koopmans' *efficient point* type [11]—for several concave functions $g_1(x), \ldots, g_p(x)$ can be transformed into a "scalar" maximum for $g(x) \equiv \sum v_k^0 g_k(x)$ by suitable choice of positive constants v_k^0 (as described in section 6). These positive constants can be interpreted as prices to be assigned (for efficient production) to several desired commodities with outputs $g_k(x)$ produced by the activity vector x.

Likewise, a maximum for min $[g_1(x), \ldots, g_p(x)]$ can be transformed into a maximum for $g(x) \equiv \sum v_k^0 g_k(x)$ by suitable choice of nonnegative constants v_k with unit sum (as described in section 7). Such a maximum of a minimum component, example, is the objective of the first player in a zero sum two person game [12].

Modifications resulting from changes in the $m + n$ basic constraints are also considered (in section 8).

Throughout this paper it is assumed that the functions occurring are differentiable. But it seems to be an interesting consequence of the directional derivative properties of general convex (or concave) functions [2, pp. 18–21] that the equivalence between an inequality constrained maximum for $g(x)$ and a saddle value for the Lagrangian $\phi(x, u)$ still holds when the assumption of differentiability is dropped. Then proofs would involve the properties (of linear sum, intersection, and polar) of general closed convex "cones" rather than those of the polyhedral convex "cones" [7], [14] that occur implicitly in this paper through homogeneous linear differential inequalities. However, to assure finite directional derivatives at boundary points of the orthant of nonnegative x, one needs some mild requirement. For this purpose, it is certainly sufficient to assume that the functions are convex (or concave) in some open region containing the orthant of nonnegative x.

NOTATION. Vectors, denoted usually by lower case roman letters, will be treated as one column matrices, unless transposed by an accent $'$ into one row matrices. Vector inequalities or equations stand for systems of such inequalities or equations, one for each component. Thus $x \geqq 0$ means that all the components of the vector x are nonnegative. Rectangular matrices and mapping operators will be denoted by capital letters.

2. Necessary and sufficient conditions for a saddle value

Let $\phi(x, u)$ be a differentiable function of an n-vector x with components $x_i \geqq 0$ and an m-vector u with components $u_h \geqq 0$. Taking partial derivatives, evaluated at a particular point x^0, u^0, let

$$\phi_x^0 = \left[\frac{\partial \phi}{\partial x_i}\right]^0, \qquad \phi_u^0 = \left[\frac{\partial \phi}{\partial u_h}\right]^0.$$

Here ϕ_x^0 is an n-vector and ϕ_u^0 an m-vector.

SADDLE VALUE PROBLEM. *To find nonnegative vectors x^0 and u^0 such that*

$$\phi(x, u^0) \leqq \phi(x^0, u^0) \leqq \phi(x^0, u) \textit{ for all } x \geqq 0,\ u \geqq 0.$$

LEMMA 1. *The conditions*

$$(1) \qquad \phi_x^0 \leqq 0, \qquad \phi_x^{0\prime} x^0 = 0, \qquad x^0 \geqq 0$$

(2) $$\phi_u^0 \geqq 0, \qquad \phi_u^{0\prime} u^0 = 0, \qquad u^0 \geqq 0$$

are necessary that x^0, u^0 *provide a solution for the saddle value problem.*

PROOF. The components of ϕ_x^0 and ϕ_u^0 must vanish except possibly when the corresponding components of x^0 and u^0 vanish, in which case they must be nonpositive and nonnegative, respectively. Hence (1) and (2) must hold.

LEMMA 2. *Conditions* (1), (2) *and*

(3) $$\phi\,(x, u^0) \leqq \phi\,(x^0, u^0) + \phi_x^{0\prime}\,(x - x^0)$$

(4) $$\phi\,(x^0, u) \geqq \phi\,(x^0, u^0) + \phi_u^{0\prime}\,(u - u^0)$$

for all $x \geqq 0$, $u \geqq 0$, *are sufficient that* x^0, u^0 *provide a solution for the saddle value problem.*

PROOF. Applying (3), (1), (2), (4) in turn, one has

$$\begin{aligned} \phi\,(x, u^0) &\leqq \phi\,(x^0, u^0) + \phi_x^{0\prime}\,(x - x^0) \\ &\leqq \phi\,(x^0, u^0) \\ &\leqq \phi\,(x^0, u^0) + \phi_u^{0\prime}\,(u - u^0) \\ &\leqq \phi\,(x^0, u) \end{aligned}$$

for all $x \geqq 0$, $u \geqq 0$.

Conditions (3) and (4) are not as artificial as may appear at first sight. They are satisfied if $\phi(x, u^0)$ is a concave function of x and $\phi(x^0, u)$ is a convex function of u (see section 4).

3. Lagrange multipliers for an inequality constrained maximum

Let $x \longrightarrow u = F(x)$ be a differentiable mapping of nonnegative n-vectors x into m-vectors u. That is, $F(x)$ is an m-vector whose components $f_1(x), \ldots, f_m(x)$ are differentiable functions of x defined for $x \geqq 0$. Let $g(x)$ be a differentiable function of x defined for $x \geqq 0$. Taking partial derivatives, evaluated at x^0, let

$$F^0 = [\partial f_h / \partial x_i]^0\,, \qquad g^0 = [\partial g / \partial x_i]^0\,.$$

Here F^0 is an m by n matrix and g^0 an n-vector.

MAXIMUM PROBLEM. *To find an* x^0 *that maximizes* $g(x)$ *constrained by* $F(x) \geqq 0$, $x \geqq 0$.

CONSTRAINT QUALIFICATION. Let x^0 belong to the boundary of the *constraint set* of points x satisfying $Fx \geqq 0$, $x \geqq 0$. Let the inequalities $F(x^0) \geqq 0$, $Ix^0 \geqq 0$ (where I is the identity matrix of order n) be separated into

$$F_1 x^0 = 0\,, \qquad I_1 x^0 = 0 \quad \text{and} \quad F_2 x^0 > 0\,, \qquad I_2 x^0 > 0\,.$$

It will be assumed for each x^0 of the boundary of the constraint set that any vector differential dx satisfying the homogeneous linear inequalities

(5) $$F_1^0 dx \geqq 0\,, \qquad I_1 dx \geqq 0$$

is tangent to an arc contained in the constraint set; that is, to any dx satisfying (5) there corresponds a differentiable arc $x = a(\theta)$, $0 \leqq \theta \leqq 1$, contained in the constraint set, with $x^0 = a(0)$, and some positive scalar λ such that $[da/d\theta]^0 = \lambda dx$. This assumption is designed to rule out singularities on the boundary of the constraint set, such as an outward pointing "cusp." For example, the constraint set in two dimensions determined by

$$(1 - x_1)^3 - x_2 \geqq 0\,, \qquad x_1 \geqq 0\,, \qquad x_2 \geqq 0$$

does not satisfy the constraint qualification at the boundary point $x_1^0 = 1$, $x_2^0 = 0$, since it does not contain an arc leading from this point in the direction $dx_1 = 1$, $dx_2 = 0$. At such a singular point condition (1) in theorem 1, below, may fail to hold for any u^0—as would be the case for $g(x) \equiv x_1$ subject to the above constraints.

Treating the vector u as a set of m nonnegative Lagrange multipliers [10], form the function

$$\phi(x, u) \equiv g(x) + u'Fx\,.$$

Then

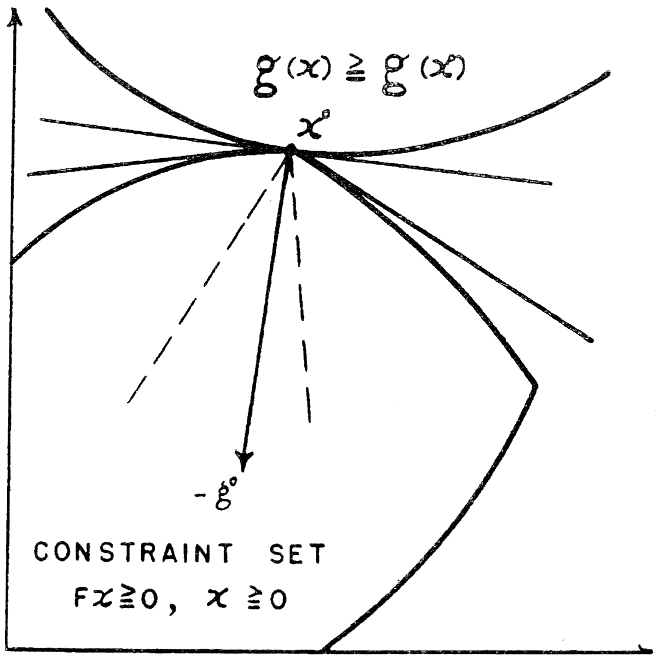
$$\phi_x^0 = g^0 + F^{0\prime}u^0\,, \qquad \phi_u^0 = Fx^0\,.$$

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* $\phi(x, u) \equiv g(x) + u'Fx$.)

PROOF. Let x^0 maximize $g(x)$ constrained by $Fx \geqq 0$, $x \geqq 0$ (subject to the above constraint qualification). Then, the inequality $g^{0\prime}dx \leqq 0$ must hold for all vector differentials dx satisfying (5). But, it is a fundamental property of homogeneous linear inequalities (indicated by H. Minkowski and proved by J. Farkas at the turn of the century) that an inequality $b'x \geqq 0$ holds for all n-vectors x satisfying a system of m inequalities $Ax \geqq 0$ only if $b = A't$ for some m-vector $t \geqq 0$ [6, pp. 5–7], [7, corollary to theorem 2], [9, lemma 1] and [14, theorem 3]. Hence

$$-g^0 = F_1^{0\prime}u_1^0 + I_1'w_1^0 \qquad \text{for some } u_1^0 \geqq 0,\ w_1^0 \geqq 0\,.$$

This equation expresses the intuitively evident geometric fact that at the point x^0 the outward normal $-g^0$ to the set of points x for which $g(x) \geqq g(x^0)$ must belong



to the convex polyhedral "cone" of inward normals to the constraint set. Of course, if x^0 is an interior point of the latter set, then F_1^0 and I_1 are both vacuous. In this case x^0 maximizes $g(x)$ independent of the constraints, so $g^0 = 0$ and conditions (1), (2) hold for $u^0 = 0$.

The above equation may be rewritten as

$$-g^0 = F^{0\prime}u^0 + w^0 \qquad \text{for some } u^0 \geqq 0\,,\ w^0 \geqq 0$$

by adding zeros as components to u_1^0 and w_1^0 to form u^0 and w^0. Consequently,

$$\phi_x^0 = g^0 + F^{0\prime}u^0 \leqq g^0 + F^{0\prime}u^0 + w^0 = 0\,.$$

At the same time, since $w^{0\prime}x^0 = w_1^{0\prime}I_1x^0 = 0$,

$$\phi_x^{0\prime}x^0 = g^{0\prime}x^0 + u^{0\prime}F^0x^0 = 0\,.$$

Moreover,

$$\phi_u^0 = Fx^0 \geqq 0 \quad \text{and} \quad \phi_u^{0\prime}u^0 = u^{0\prime}Fx^0 = u_1^{0\prime}F_1x^0 = 0\,.$$

This completes the proof of theorem 1.

THEOREM 2. *In order that x^0 be a solution of the maximum problem, it is sufficient that x^0 and some u^0 satisfy conditions* (1), (2), *and* (3) *for* $\phi(x, u) \equiv g(x) + u'Fx$.

PROOF. From (3), (1), and (2) one has that

$$\begin{aligned} g(x) + u^{0\prime}Fx = \phi(x, u^0) &\leqq \phi(x^0, u^0) + \phi_x^{0\prime}(x - x^0) \\ &\leqq \phi(x^0, u^0) = g(x^0) + u^{0\prime}Fx^0 = g(x^0) \ \text{ for all } x \geqq 0\,. \end{aligned}$$

But $u^{0\prime}Fx \geqq 0$ for all x satisfying $Fx \geqq 0$. Hence $g(x) \leqq g(x^0)$ for all x satisfying the constraints $Fx \geqq 0$, $x \geqq 0$. This proves theorem 2.

One notes in theorem 2 that (3) need only hold for $Fx \geqq 0$, $x \geqq 0$.

4. Convexity-concavity properties and the equivalence theorem

In this section restrictions are placed on Fx and $g(x)$ which will insure the equivalence of solutions of the maximum problem and the saddle value problem for $\phi(x, u) \equiv g(x) + u'Fx$.

DEFINITIONS. *A function $f(x)$ is* convex *if*

$$(1 - \theta)f(x^0) + \theta f(x) \geqq f\{(1 - \theta)x^0 + \theta x\}$$

for $0 \leqq \theta \leqq 1$ and all x^0 and x in the (convex) region of definition of $f(x)$. A function $f(x)$ is concave *if $-f(x)$ is convex* (that is, if the interpolation inequality holds with $\leqq$ instead of $\geqq$).

LEMMA 3. *If $f(x)$ is convex and differentiable, then*

$$f(x) \geqq f(x^0) + f^{0\prime}(x - x^0) \qquad \left(\textit{where } f^0 = \left[\frac{\partial f}{\partial x_i}\right]^0\right)$$

for all x^0 and x in the region of definition. [With $f(x)$ concave, the inequality is reversed.]

PROOF. From the above definition of convexity one has, for $0 < \theta \leqq 1$,

$$f(x) - f(x^0) \geqq \frac{f\{x^0 + \theta(x - x^0)\} - f(x^0)}{\theta}$$

Hence, in the limit,

$$f(x) - f(x^0) \geqq f^{0\prime}(x - x^0) .$$

THEOREM 3 (Equivalence theorem). *Let the functions* $f_1(x), \ldots, f_m(x)$, $g(x)$ *be concave as well as differentiable for* $x \geqq 0$. *Then,* 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* $\phi(x, u) \equiv g(x) + u'Fx$.

PROOF. By lemma 3 (for concavity)

$$Fx \leqq Fx^0 + F^0(x - x^0)$$
$$g(x) \leqq g(x^0) + g^{0\prime}(x - x^0)$$

for all $x^0 \geqq 0$ and $x \geqq 0$. Hence, for any $u^0 \geqq 0$,

$$\begin{aligned}\phi(x, u^0) &= g(x) + u^{0\prime}Fx \\ &\leqq g(x^0) + u^{0\prime}Fx^0 + (g^{0\prime} + u^{0\prime}F^0)(x - x^0) \\ &= \phi(x^0, u^0) + \phi_x^{0\prime}(x - x^0) .\end{aligned}$$

That is, condition (3) holds for all $x^0 \geqq 0$ and $x \geqq 0$. Under these circumstances, theorems 1 and 2 combine to make conditions (1) and (2) both necessary and sufficient that x^0 provide a solution for the maximum problem.

Condition (4) holds automatically, since the linearity of $\phi(x, u)$ with respect to u implies that

$$\phi(x^0, u) = \phi(x^0, u^0) + \phi_u^{0\prime}(u - u^0)$$

identically. So lemmas 1 and 2 combine to make conditions (1) and (2) both necessary and sufficient that x^0 and u^0 provide a solution for the saddle value problem. This completes the proof of theorem 3.

5. Quadratic and linear problems

LEMMA 4. *A quadratic form*

$$x'Qx = \sum \sum q_{ij}x_i x_j$$

is a convex function for all x, *if* $x'Qx \geqq 0$ *for all* x (*that is, if the form is positive semi-definite*).

PROOF. From the hypothesis, one has

$$\theta(x - x^0)'Q(x - x^0) \geqq \theta^2(x - x^0)'Q(x - x^0)$$

for all $0 \leqq \theta \leqq 1$ and all x, x^0. Hence

$$\begin{aligned}&(1 - \theta)x^{0\prime}Qx^0 + \theta x'Qx \\ &\quad = x^{0\prime}Qx^0 + \theta x^{0\prime}Q(x - x^0) + \theta(x - x^0)'Qx^0 + \theta(x - x^0)'Q(x - x^0) \\ &\quad \geqq x^{0\prime}Qx^0 + \theta x^{0\prime}Q(x - x^0) + \theta(x - x^0)'Qx^0 + \theta^2(x - x^0)'Q(x - x^0) \\ &\quad = \{x^0 + \theta(x - x^0)\}'Q\{x^0 + \theta(x - x^0)\} \\ &\quad = \{(1 - \theta)x^0 + \theta x\}'Q\{(1 - \theta)x^0 + \theta x\}\end{aligned}$$

for all $0 \leqq \theta \leqq 1$ and all x, x^0.

QUADRATIC MAXIMUM PROBLEM. *To find an x^0 that maximizes*

$$g(x) \equiv \sum c_i x_i - \sum \sum c_{ij} x_i x_j$$

constrained by the $m + n$ inequalities

$$f_h(x) \equiv b_h - \sum a_{hi} x_i - \sum \sum a_{hij} x_i x_j \geqq 0 \quad \text{and} \quad x_i \geqq 0 .$$

It is assumed that the quadratic forms in the above double sums (including the preceding sign) are nonpositive for all x (that is, negative semidefinite).

From lemma 4 it follows that these quadratic functions $f_h(x)$ and $g(x)$ are concave for all x, since their linear parts are concave and convex both. Hence, by theorem 3, solution of the quadratic maximum problem is equivalent to solution of the saddle value problem for

$$\phi(x, u) \equiv \sum c_i x_i - \sum \sum c_{ij} x_i x_j + \sum b_h u_h - \sum \sum a_{hi} u_h x_i - \sum \sum \sum a_{hij} u_h x_i x_j .$$

When all of the quadratic terms vanish (an extreme but legitimate special case of semidefiniteness), the quadratic maximum problem reduces to the following problem of *linear programming*.

LINEAR MAXIMUM PROBLEM. *To find an x^0 that maximizes $\sum c_i x_i$ constrained by the $m + n$ linear inequalities*

$$\sum a_{hi} x_i \leqq b_h , \qquad x_i \geqq 0 .$$

Now the equivalent saddle point problem concerns the *bilinear* function

$$\phi(x, u) \equiv \sum c_i x_i + \sum b_h u_h - \sum \sum a_{hi} u_h x_i .$$

The minimum maximum rôles of x and u can be interchanged by replacing $\phi(x, u)$ by $-\phi(x, u)$. Hence, solution of the following *dual problem* of linear programming is equivalent to solution of the saddle point problem for the bilinear function $\phi(x, u)$.

LINEAR MINIMUM PROBLEM. *To find a u^0 that minimizes $\sum b_h u_h$ constrained by the $n + m$ inequalities*

$$\sum a_{hi} u_h \geqq c_i , \qquad u_h \geqq 0 .$$

6. Extension to a vector maximum problem

This section extends the previous results to a maximum problem for a vector function Gx constrained by $Fx \geqq 0$, $x \geqq 0$. Here the concept of maximum—like T. C. Koopmans' *efficient point* [11]—depends on a partial ordering of vectors by the relation $\geq$, where $v \geq v^0$ means that $v \geqq v^0$ but $v \neq v^0$.

Let $x \to v = Gx$ be a differentiable mapping of nonnegative n-vectors x into

p-vectors v. That is, Gx is a p-vector whose components $g_1(x), \ldots, g_p(x)$ are differentiable functions of x defined for $x \geqq 0$. Taking partial derivatives, evaluated at a particular x^0, let

$$G^0 = \left[\frac{\partial g_k}{\partial x_i}\right]^0 .$$

Here G^0 is a p by n matrix. Let g_k^0 denote the n-vector whose components form the k-th row of G^0. Let Fx have the meaning assigned in section 3.

VECTOR MAXIMUM PROBLEM. *To find an x^0 that maximizes the vector function Gx constrained by $Fx \geqq 0$, $x \geqq 0$—that is, to find an x^0 satisfying the constraints and such that $Gx \geq Gx^0$ for no x satisfying the constraints.*

RESTRICTION. Attention will be restricted to solutions x^0 of the vector maximum problem that are ***proper*** in the sense that $G^0dx \geq 0$ for no vector differential dx if x^0 is interior to the constraint set determined by $Fx \geqq 0$, $x \geqq 0$, and for no dx satisfying

$$(5) \qquad F_1^0 dx \geqq 0 , \qquad I_1 dx \geqq 0$$

if x^0 belongs to the boundary of the constraint set (as qualified in section 3).

Example. To maximize $g_1(x) \equiv x$, $g_2(x) \equiv 2x - x^2$, x being a real variable (one dimensional vector) constrained only by $x \geqq 0$. Here, $Gx \geq Gx^0$ for no x if $x^0 \geqq 1$, and $G^0dx \geq 0$ for no dx except at $x^0 = 1$, where $G^0dx \geq 0$ for $dx > 0$. So, any $x > 1$ is a proper solution of this particular vector maximum problem, but $x^0 = 1$ is a solution that is not proper. An argument against admitting $x^0 = 1$ as a "proper" solution is that it would usually be natural to accept a second order loss in $g_2(x) \equiv 2x - x^2$ to achieve a first order gain in $g_1(x) \equiv x$. (The anomaly indicated by $x^0 = 1$ in this example was noticed by C. B. Tompkins. A rather similar anomaly occurs in the paper [1] of K. J. Arrow.)

THEOREM 4. *In order that x^0 be a proper solution of the vector maximum problem, it is necessary that there be some $v^0 > 0$ such that x^0 and some u^0 satisfy conditions* (1) *and* (2) *for $\phi(x, u) \equiv v^{0\prime}Gx + u'Fx$.*

PROOF. Let x^0 be a proper solution of the vector maximum problem. Then, for each $k = 1, \ldots, p$, one must have $g^{0\prime}dx \leqq 0$ for all dx satisfying

$$F_1^0 dx \geqq 0 , \qquad I_1 dx \geqq 0 , \qquad G^0 dx \geqq 0$$

(where F_1^0 and I_1 may be vacuous). Hence, by the fundamental property of homogeneous linear inequalities used in the proof of theorem 1,

$$-g_k^0 = F_1^{0\prime}u_1^k + I_1'w_1^k + G^{0\prime}v^k \quad \text{for some} \quad u_1^k \geqq 0 , \quad w_1^k \geqq 0 , \quad v^k \geqq 0 .$$

Now, summing for $k = 1, \ldots, p$, and transferring the G^0 terms to the left side, one has

$$-G^{0\prime}v^0 = F_1^{0\prime}u_1^0 + I_1'w_1^0 ,$$

where $u_1^0 = \sum u_1^k \geqq 0$, $w_1^0 = \sum w_1^k \geqq 0$, and $v^0 = e + \sum v^k > 0$, e being a p-vector whose components are all 1's.

Let $g(x) \equiv v^{0\prime}Gx$. Then

$$-g^0 = -G^{0\prime}v^0 = F_1^{0\prime}u_1^0 + I_1'w_1^0 .$$

From this point on the proof of theorem 4 is completed by following the remaining steps of theorem 1.

THEOREM 5. *In order that x^0 be a proper solution of the vector maximum problem, it is sufficient that there be some $v^0 > 0$ such that x^0 and some u^0 satisfy conditions* (1), (2), *and* (3) *for* $\phi(x, u) \equiv v^{0\prime}Gx + u'Fx$.

PROOF. From the proof of theorem 2, with $g(x) \equiv v^{0\prime}Gx$, it follows that

$$v^{0\prime}Gx \leqq v^{0\prime}Gx^0$$

for all x satisfying the constraints $Fx \geqq 0$, $x \geqq 0$. But $v^0 > 0$, so $Gx \geq Gx^0$ for no x satisfying the constraints.

If x^0 is interior to the constraint set, then $G^{0\prime}v^0 = 0$ by (1), since $x^0 > 0$, $Fx^0 > 0$, and $u^0 = 0$. So $G^0dx \geq 0$ for no dx. If x^0 belongs to the boundary of the constraint set, then (1) implies that

$$-G^{0\prime}v^0 - F^{0\prime}u^0 = I_1'w_1^0 \quad \text{for some} \quad w_1^0 \geqq 0\,.$$

Through (2) this can be written

$$-G^{0\prime}v^0 = F_1^{0\prime}u_1^0 + I_1'w_1^0 \quad \text{for} \quad u_1^0 \geqq 0\,.$$

Hence $G^0dx \geq 0$ for no dx satisfying

$$\text{(5)} \qquad F_1^0dx \geqq 0\,, \qquad I_1dx \geqq 0\,.$$

This completes the proof of theorem 5.

THEOREM 6 (Equivalence theorem). *Let the functions $f_1(x), \ldots, f_m(x)$, $g_1(x), \ldots, g_p(x)$ be concave as well as differentiable for $x \geqq 0$. Then, x^0 is a proper solution of the vector maximum problem if, and only if, there is some $v^0 > 0$ such that x^0 and some u^0 give a solution of the saddle value problem for $\phi(x, u) \equiv v^{0\prime}Gx + u'Fx$.*

PROOF. Clearly $g(x) \equiv v^{0\prime}Gx$ is concave, since $v^0 > 0$. So the proof of theorem 3 can be duplicated, using theorems 4 and 5 in place of theorems 1 and 2.

7. Another extension

Let Fx and Gx be differentiable mappings, as previously defined (with the constraint qualification on $Fx \geqq 0$, $x \geqq 0$ still in effect). Let min $[Gx]$ denote the (scalar) function whose value for each $x \geqq 0$ is the least among the p values $g_1(x), \ldots, g_p(x)$ of the components of the vector Gx.

MINIMUM COMPONENT MAXIMUM PROBLEM. *To find an x^0 that maximizes min $[Gx]$ constrained by $Fx \geqq 0$, $x \geqq 0$.*

THEOREM 7. *In order that x^0 be a solution of the minimum component maximum problem, it is necessary that there be some nonnegative v^0 with unit component sum satisfying*

$$\text{(6)} \qquad v^{0\prime}Gx^0 = \min\,[Gx^0]$$

and such that x^0 and some u^0 satisfy conditions (1) *and* (2) *for* $\phi(x, u) \equiv v^{0\prime}Gx + u'Fx$.

PROOF. Let F_1x^0 and I_1x^0 have the meanings assigned them in section 3. Further, let Gx^0 be separated into $G_1x^0 = \min\,[Gx^0]$ and $G_2x^0 > \min\,[Gx^0]$ (see note preceding theorem 10, below). Then, since x^0 is assumed to maximize min $[Gx]$ constrained by $Fx \geqq 0$, $x \geqq 0$, one must have that $G_1^0dx > 0$ for no vector differential

dx satisfying

$$F_1^0 dx \geqq 0 , \qquad I_1 dx \geqq 0$$

(or for no dx at all, if F_1^0 and I_1 are vacuous). That is, for each k belonging to a certain nonvacuous subset of the set of indices corresponding to the rows of G^0 that belong to G_1^0 one must have that $g_k^{0\prime} dx \leqq 0$ for all dx satisfying

$$F_1^0 dx \geqq 0 , \qquad I_1 dx \geqq 0 , \qquad G_1^0 dx \geqq 0 .$$

Hence, by the fundamental property of homogeneous linear inequalities used in the proof of theorem 1,

$$-g_k^0 = F_1^{0\prime} u_1^k + I_1' w_1^k + G_1^{0\prime} v_1^k \quad \text{for some} \quad u_1^k \geqq 0 , \quad w_1^k \geqq 0 , \quad v_1^k \geqq 0 .$$

Now, summing for k over the nonvacuous subset and transferring the G_1^0 terms to the left side, one has

$$-G_1^{0\prime} v_1^0 = F_1^{0\prime} u_1^0 + I_1' w_1^0 ,$$

where $u_1^0 = \sum u_1^k \geqq 0$, $w_1^0 = \sum w_1^k \geqq 0$, and $v_1^0 = e_1 + \sum v_1^k \geqq 0$, e_1 being a vector whose components are 0's or 1's—with at least one 1. Here it can be assumed that the sum of the components of v_1^0 is one, since the above vector equation is homogeneous and the sum of the components of v_1^0 is positive. Form v^0 from v_1^0 by adding zeros as components. Then

$$v^{0\prime} G x^0 = v_1^{0\prime} G_1 x^0 = \min [G x^0] .$$

By setting $g(x) \equiv v^{0\prime} G x$, the above vector equation can be rewritten as

$$-g^0 \doteq -G^{0\prime} v_1^0 = -G_1^{0\prime} v_1^0 = F_1^{0\prime} u_1^0 + I_1' w_1^0 .$$

From this point on the proof of theorem 7 is completed by following the remaining steps of theorem 1.

THEOREM 8. *In order that x^0 be a solution of the minimum component maximum problem, it is sufficient that there be some nonnegative v^0 with unit component sum satisfying condition* (6) *and such that x^0 and some u^0 satisfy conditions* (1), (2), *and* (3) *for $\phi(x, u) \equiv v^{0\prime} G x + u' F x$.*

PROOF. From the proof of theorem 2 with $g(x) \equiv v^{0\prime} G x$, it follows that

$$v^{0\prime} G x \leqq v^{0\prime} G x^0$$

for all x satisfying the constraints $Fx \geqq 0$, $x \geqq 0$. But v^0 is nonnegative with unit component sum and satisfies condition (6). Hence

$$\min [Gx] \leqq v^{0\prime} G x \leqq v^{0\prime} G x^0 = \min [G x^0]$$

for all x satisfying the constraints. This proves theorem 8.

THEOREM 9 (Equivalence theorem). *Let the functions $f_1(x), \ldots , f_m(x)$, $g_1(x)$, $\ldots$, $g_p(x)$ be concave as well as differentiable for $x \geqq 0$. Then, x^0 is a solution of the minimum component maximum problem if, and only if, there is some nonnegative v^0 with unit component sum satisfying condition* (6) *and such that x^0 and some u^0 give a solution of the saddle value problem for $\phi(x, u) \equiv v^{0\prime} G x + u' F x$.*

PROOF. Clearly $g(x) \equiv v^{0\prime} G x$ is concave, since v^0 is nonnegative. The proof of theorem 3 can be duplicated, using theorems 7 and 8 in place of theorems 1 and 2.

The fact that the constraints $Fx \geqq 0$ can be written equivalently as $\min [Fx] \geqq 0$

suggests the possibility of interchanging the rôles of Fx and Gx. The following theorem exploits this possibility. As before, constraints are subject to the constraint qualification introduced in section 3. (It is to be noted that a constant, such as min $[Gx^0]$, appearing as a vector in a vector inequality or equation is to be interpreted as a vector all of whose components equal that constant.)

THEOREM 10. *Let the functions $f_1(x), \ldots, f_m(x), g_1(x), \ldots, g_p(x)$ be concave as well as differentiable for $x \geqq 0$. Then, in order that x^0 maximize min $[Gx]$ constrained by $Fx \geqq$ min $[Fx^0]$, $x \geqq 0$, it is sufficient that x^0 maximize min $[Fx]$ constrained by $Gx \geqq$ min $[Gx^0]$, $x \geqq 0$—provided $Fx >$ min $[Fx^0]$ for some $x \geqq 0$.*

PROOF. Let x^0 maximize min $[Fx]$ constrained by $(Gx - \min [Gx^0]) \geqq 0$, $x \geqq 0$, as hypothesized. Then, by theorem 7 applied to this reversed situation, there must be some nonnegative u^0 with unit component sum and some v^0 such that

$$u^{0\prime}Fx^0 = \min [Fx^0] ,$$

$$F^{0\prime}u^0 + G^{0\prime}v^0 \leqq 0 , \quad u^{0\prime}F^0x^0 + v^{0\prime}G^0x^0 = 0 , \quad x^0 \geqq 0$$

$$(Gx^0 - \min [Gx^0]) \geqq 0 , \quad v^{0\prime}(Gx^0 - \min [Gx^0]) = 0 , \quad v^0 \geqq 0 .$$

Assume, if possible, that $v^0 = 0$. Then, using the concavity of the functions forming Fx and the above conditions, one has

$$u^{0\prime}Fx \leqq u^{0\prime}Fx^0 + u^{0\prime}F^0 (x - x^0) \leqq u^{0\prime}Fx^0 \quad \text{for all } x \geqq 0 ,$$

contradicting the proviso that $Fx >$ min $[Fx^0]$ for some $x \geqq 0$. Therefore the vector $v^0 \geq 0$ and one can assume that it has unit component sum by dropping the same assumption concerning u^0. Under these circumstances

$$(Fx^0 - \min [Fx^0]) \geqq 0 , \quad u^{0\prime}(Fx^0 - \min [Fx^0]) = 0 , \quad u^0 \geqq 0 ,$$
$$\text{and } v^{0\prime}Gx^0 = \min [Gx^0] .$$

While, by the concavity of the functions forming Fx and Gx,

$$v^{0\prime}Gx + u^{0\prime}Fx \leqq v^{0\prime}Gx^0 + u^{0\prime}Fx^0 + (v^{0\prime}G^0 + u^{0\prime}F^0)(x - x^0) \quad \text{for all } x \geqq 0 .$$

Consequently, by theorem 8, x^0 is a solution of the minimum component maximum problem for Gx constrained by $(Fx - \min [Fx^0]) \geqq 0$, $x \geqq 0$. This completes the proof of theorem 10.

8. Other types of constraints

The foregoing results admit simple modifications when the constraints $Fx \geqq 0$, $x \geqq 0$ are changed to:

(1) $\quad Fx \geqq 0$, or (2) $Fx = 0$, $x \geqq 0$, or (3) $Fx = 0$.

These modifications are outlined below.

Case 1: $Fx \geqq 0$.

Here, using $\phi(x, u) \equiv g(x) + u'Fx$ defined for all x and constrained only by $u \geqq 0$, one must replace condition (1) by

(1*) $$\phi_x^0 = 0 .$$

Case 2: $Fx = 0$, $x \geqq 0$.

Here, using $\phi(x, u) \equiv g(x) + u'Fx$ defined for all u and constrained only by $x \geqq 0$, one must replace condition (2) by

$$(2^*) \qquad \phi_u^0 = 0 .$$

Case 3: $Fx = 0$.

Here, using $\phi(x, u) \equiv g(x) + u'Fx$ defined for all x and u without constraints, one must replace conditions (1) and (2) by (1*) and (2*). This corresponds to the customary use of the method of Lagrange multipliers for side *equations* [3].

REFERENCES

[1] K. J. Arrow, "A generalization of the basic theorem of classical welfare economics," *Proceedings of the Second Symposium on Mathematical Statistics and Probability*, pp. 507–532.

[2] T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 3, No. 1, Springer, Berlin, 1934; Chelsea, New York, 1948.

[3] R. Courant and D. Hilbert, *Methoden der mathematischen Physik*, 2nd ed., Vol. 1, Springer, Berlin, 1931; Interscience, New York.

[4] G. B. Dantzig, "Programming of interdependent activities: II mathematical model," *Econometrica*, Vol. 17 (1949), pp. 200–211; *Activity Analysis*, see [10].

[5] ———, "Maximization of a linear function of variables subject to linear inequalities," *Activity Analysis*, see [10].

[6] J. Farkas, "Über die Theorie der einfachen Ungleichungen," *Journal für die reine und angewandte Mathematik*, Vol. 124 (1901), pp. 1–27.

[7] D. Gale, "Convex polyhedral cones and linear inequalities," *Activity Analysis*, see [10].

[8] D. Gale, H. W. Kuhn and A. W. Tucker, "Some linear convex problems equivalent to a game problem," unpublished, 1948; "Four equivalent linear convex problems," mimeographed, *Linear Programming Conference Memo* 701, Cowles Commission, 1949.

[9] ———, "Linear programming and the theory of games," *Activity Analysis*, see [10].

[10] F. John, "Extremum problems with inequalities as subsidiary conditions," *Studies and Essays*, Courant Anniversary Volume, Interscience, New York, 1948.

[11] T. C. Koopmans, "Analysis of production as an efficient combination of activities," *Activity Analysis of Production and Allocation*, Cowles Commission Monograph 13, edited by T. C. Koopmans, Wiley, New York, 1951.

[12] J. von Neumann and O. Morgenstern, *The Theory of Games and Economic Behavior*, Princeton University Press, Princeton, 1944; 2nd ed., 1947.

[13] P. A. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, Cambridge, 1948.

[14] H. Weyl, "Elementare Theorie der konvexen Polyeder," *Commentarii Mathematici Helvetici*, Vol. 7 (1935), pp. 290–306. A translation by H. W. Kuhn of this paper appears in "Contributions to the theory of games," *Annals of Math. Study*, No. 24 (1950).

SIAM J. Appl. Math.
Vol. 15, No. 3, May, 1967
Printed in U.S.A.

SECOND ORDER CONDITIONS FOR CONSTRAINED MINIMA*

GARTH P. McCORMICK†

Abstract. This paper establishes two sets of "second order" conditions—one which is necessary, the other which is sufficient—in order that a vector x^* be a local minimum to the constrained optimization problem: minimize $f(x)$ subject to the constraints $g_i(x) \geqq 0$, $i = 1, \cdots, m$, and $h_j(x) = 0$, $j = 1, \cdots, p$, where the problem functions are *twice continuously differentiable*. The necessary conditions extend the well-known results, obtained with Lagrange multipliers, which apply to equality constrained optimization problems, and the Kuhn-Tucker conditions, which apply to mixed inequality and equality problems when the problem functions are required only to have continuous first derivatives. The sufficient conditions extend similar conditions which have been developed only for equality constrained problems. Examples of the applications of these sets of conditions are given.

1. Introduction. Efforts to establish conditions which determine whether or not a point solves an optimization problem have been in progress since the classical work involving Lagrange multipliers. The Lagrange multiplier approach is applied to optimization problems with equality constraints of the form given in Problem L.

Problem L. Find a vector $x^* = (x_1^*, \cdots, x_n^*)^T$ that minimizes $f(x)$ subject to

$$h_j(x) = 0, \qquad j = 1, \cdots, p.$$

Although one is interested in the true or global solution to optimization problems, in general one can only prove theorems about local solutions. A local solution is a point x^* such that in a neighborhood about that point all other points either do not satisfy the constraints of the problem or give values of the objective function greater than or equal to $f(x^*)$.

Using this definition, the basic result of Lagrange multipliers is stated in Theorem 1. (Throughout this paper, the symbol ∇ is the differentiation operator with respect to the vector x, i.e., $\nabla f(x) = (\partial f(x)/\partial x_1, \cdots, \partial f(x)/\partial x_n)^T$; the symbol ∇^2 represents the operator ∇ applied twice, $\nabla^2 f(x)$ is the $n \times n$ matrix whose i, jth element is $\partial^2 f(x)/\partial x_j \partial x_i$. For shorthand, f^* will indicate $f(x^*)$, ∇f^* will indicate $\nabla f(x^*)$; for a parameterized function, a' will represent the derivative with respect to the parameterizing variable, i.e., $f'(\theta) = df[x(\theta)]/d\theta$.) The term "differentiable" will always mean "continuously differentiable."

Theorem 1 (Lagrange multipliers) [3, p. 100]. *If the functions*

* Received by the editors April 29, 1966.

† Advanced Research Department, Research Analysis Corporation, McLean, Virginia 22101.

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_12, © Springer Basel 2014

f, h_1, $\cdots$, h_p are differentiable and if x^* is a point where the vectors ∇h_1^*, $\cdots$, ∇h_p^* are linearly independent, then a necessary condition that x^* be a local minimum to Problem L *is that there exist scalars (called Lagrange multipliers) w_j^*, $j = 1, \cdots, p$, such that*

$$\nabla f^* + \sum_{j=1}^{p} w_j^* \nabla h_j^* = 0. \tag{1}$$

The proof of this theorem will be a by-product of later developments in this paper.

That the necessary conditions expressed by (1) are not always able to distinguish a local minimum from other points is seen in the following example.

Example 1. Minimize $-x_1 - x_2$ subject to

$$x_1^2 + x_2^2 - 2 = 0.$$

Using the necessary condition expressed in (1) we need examine only points $(x_1, x_2)^T$ on the circle $x_1^2 + x_2^2 - 2 = 0$ for which a scalar w_1 exists satisfying the equation

$$(-1, -1)^T + w_1(2x_1, 2x_2)^T = (0, 0)^T. \tag{2}$$

Clearly x_1 must equal x_2, and this leaves the two points $(-1, -1)^T$ and $(1, 1)^T$ as possible local minima. For the former point $w_1 = -\frac{1}{2}$ and for the latter, $w_1 = \frac{1}{2}$ would satisfy (2). Just using (1) then, there is no way to distinguish algebraically between those two points, although geometrically it is clear that $(-1, -1)^T$ is not a local minimum.

The equality constrained optimization problem, Problem L, is a special case of the constrained optimization problem, Problem I.

PROBLEM I. Find a vector x^* that minimizes $f(x)$ subject to

$$g_i(x) \geqq 0, \qquad i = 1, \cdots, m.$$

With respect to Problem I, Fritz John [1] proved a general theorem of which the following is a special case.

THEOREM 2 [1, Theorem 1, p. 188]. *If the functions f, g_1, $\cdots$, g_m are differentiable, then a necessary condition that x^* be a local minimum to Problem* I *is that there exist scalars u_0^*, u_0^*, $\cdots$, u_m^* (not all zero) such that the following inequalities and equalities are satisfied by $(x_1^*, \cdots, x_n^*, u_0^*, u_1^*, \cdots, u_m^*)$:*

$$g_i(x) \geqq 0, \qquad i = 1, \cdots, m, \tag{3}$$

$$u_0 \nabla f(x) - \sum_{i=1}^{m} u_i \nabla g_i(x) = 0, \tag{4}$$

$$(5) \qquad u_i g_i(x) = 0, \qquad i = 1, \cdots, m,$$

$$(6) \qquad u_i \geqq 0, \qquad i = 0, 1, \cdots, m.$$

The proof is omitted here.

With respect to the inequality constraints then, the variables u_i^*, $i = 0, 1, \cdots, m$, are always nonnegative.

In a later paper, Kuhn and Tucker [4] showed that if a condition, called the "first order constraint qualification," holds at x^*, then u_0^* can be taken equal to 1. The statement of the first order constraint qualification and the proof of the Kuhn-Tucker theorem are given in §2.

The failure of conditions (3) through (6) to answer questions which have proper answers is illustrated in the following example.

Example 2. Find the values of the parameter $k > 0$ for which $(0, 0)$ is a local minimum to the problem: minimize $(x_1 - 1)^2 + x_2^2$ subject to

$$-x_1 + x_2^2/k \geqq 0.$$

Using (3) through (6) the following equation must be satisfied:

$$u_0^*(-2, 0)^T - u_1^*(-1, 0)^T = (0, 0)^T.$$

Since $u_0^* = 0$ implies $u_1^* = 0$, and since Theorem 2 says they both cannot be equal to zero, it follows that u_0^* can be taken equal to 1, and $u_1^* = 2$. These values of u_0^* and u_1^* make the necessary conditions (3) through (6) valid for all values of k. But for $k = \frac{1}{4}$, $(0, 0)^T$ is not a local minimum and for $k = 3$, it is.

These two examples provide the motivation for the remainder of the paper. The theorems henceforth will be addressed to the constrained optimization problem, Problem M, containing a mixture of inequality and equality constraints.

PROBLEM M. Minimize $f(x)$ subject to

$$(7) \qquad \begin{aligned} g_i(x) &\geqq 0, \qquad i = 1, \cdots, m, \\ h_j(x) &= 0, \qquad j = 1, \cdots, p. \end{aligned}$$

In §2, using the first order constraint qualification, the Kuhn-Tucker theorem is proved. In §3, by addition of a condition called the "second order constraint qualification", additional necessary conditions are placed on a local minimum to Problem M when the problem functions are assumed twice differentiable. This is a new result although special cases have appeared elsewhere. In [5] similar results are obtained for the problem of maximizing a quadratic indefinite form subject to *linear* constraints. Next, we prove constructively that the first and second order constraint qualifications are satisfied when a regularity condition is placed on x^*. In §4, a sufficiency theorem for a point x^* to be a local minimum to Problem M is given, extending classical results which are valid only for the equality constrained problem Problem L. In §5 it is shown how the "second order" necessary and sufficient conditions solve the two examples for which the first order condition failed.

2. First order necessary conditions. Use will be made of the following lemma which is stated here without proof [2].

FARKAS' LEMMA.[1] *If every vector* z *(*n *components) which satisfies the inequality relations*

$$z^T a_i \geqq 0, \qquad i = 1, \cdots, q,$$

and the equality relations

$$z^T b_j = 0, \qquad j = 1, \cdots, r,$$

also satisfies the inequality

$$z^T c \geqq 0,$$

then there exist nonnegative scalars $t_1, \cdots, t_q$ *and scalars* $s_1, \cdots, s_r$ *(unrestricted in sign) such that*

$$c - \sum_{i=1}^{q} t_i a_i + \sum_{j=1}^{r} s_j b_j = 0.$$

We now state a condition, first introduced by Kuhn and Tucker [4, p. 483], which will be required to hold at any candidate for a local minimum.

First order constraint qualification. Let x^0 be any point satisfying the constraints of Problem M, and assume that the functions $g_1, \cdots, g_m, h_1, \cdots, h_p$ are differentiable at x^0. Then the first order constraint qualification holds at x^0, if for any nonzero vector y such that $y^T \nabla g_i(x^0) \geqq 0$ for all $i \in B^0 = \{i \mid g_i(x^0) = 0\}$, and $y^T \nabla h_j(x^0) = 0$, $j = 1, \cdots, p$, y is tangent to an arc $\alpha(\theta)$ differentiable at x^0 which is contained in the constraint region.

An arc $\alpha(\theta)$ is considered here to be a parameterized curve, differentiable when $0 \leqq \theta \leqq \epsilon$. The first order constraint qualification means that at $\alpha(0) = x^0$, $\alpha'(0) = y$.

THEOREM 3 (Kuhn-Tucker). *If the functions* $f, g_1, \cdots, g_m, h_1, \cdots, h_p$ *are differentiable at a point* x^* *and if the first order constraint qualification holds at* x^*, *the necessary conditions that* x^* *be a local minimum to the constrained optimization problem* M *are that there exist scalars* $u_1^*, \cdots, u_m^*$, $w_1^*, \cdots, w_p^*$ *such that* (x^*, u^*, w^*) *satisfies*

[1] For convenience Farkas' lemma has been restated here in a form different but equivalent to its usual form.

$$g_i(x) \geqq 0, \quad i = 1, \cdots, m, \tag{8}$$

$$h_j(x) = 0, \quad j = 1, \cdots, p, \tag{9}$$

$$u_i g_i(x) = 0, \quad i = 1, \cdots, m, \tag{10}$$

$$u_i \geqq 0, \quad i = 1, \cdots, m, \tag{11}$$

$$\nabla f(x) - \sum_{i=1}^{m} u_i \nabla g_i(x) + \sum_{j=1}^{p} w_j \nabla h_j(x) = 0. \tag{12}$$

Proof. Let $B^* \equiv \{i \mid g_i(x^*) = 0\}$. Consider any nonzero vector y such that $y^T \nabla g_i^* \geqq 0$ for all $i \in B^*$ and $y^T \nabla h_j^* = 0, j = 1, \cdots, p$. By the first order constraint qualification y is the tangent of a differentiable arc, emanating from x^*, which is contained in the constraint region.

Let $\alpha(\theta)$ be that arc. Using the chain rule, the rate of change of f along $\alpha(\theta)$ at $\alpha(0) = x^*$ is

$$f'[\alpha(0)] = \alpha'(0)^T \nabla f(x^*) = y^T \nabla f(x^*).$$

By assumption, x^* is a local minimum and f must increase or remain the same along $\alpha(\theta)$. Thus, $y^T \nabla f(x^*) \geqq 0$.

The hypotheses of Farkas' lemma are satisfied by the vectors $\nabla g_i(x^*)$ for all $i \in B^*$, the vectors $\nabla h_j(x^*), j = 1 \ \cdots \ p$, and the vector $\nabla f(x^*)$. Then there exist values $u_i^* \geqq 0$ for all $i \in B^*$, and $w_j^*, j = 1, \cdots, p$, such that

$$\nabla f(x^*) - \sum_{i \in B^*} u_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} w_j^* \nabla h_j(x^*) = 0.$$

Let $u_i^* = 0$ for all $i \notin B^*$, and the theorem is proved.

3. Second order necessary conditions. The following may be a requirement on some point in the constraint region.

Second order constraint qualification. Let x^0 be some point satisfying the constraints of M, and assume that the functions $g_1, \cdots, g_m, h_1, \cdots, h_p$ are twice differentiable at x^0. The second order constraint qualification holds at x^0 if the following is true. Let y be any nonzero vector such that $y^T \nabla g_i^0 = 0$ for all $i \in B^0 = \{i \mid g_i(x^0) = 0\}$, and such that $y^T \nabla h_j^0 = 0$, $j = 1, \cdots, p$. Then y is the tangent of a twice differentiable arc $\alpha(\theta)$ (where $\theta \geqq 0$) along which $g_i[\alpha(\theta)] = 0$ for all $i \in B^0$, $h_j[\alpha(\theta)] = 0$ for small $\theta, j = 1, \cdots, p$, and $\alpha(0) = x^0$, i.e., $\alpha'(0) = y$.

THEOREM 4 (Second order necessary conditions). *If the functions $f, g_1, \cdots, g_m, h_1, \cdots, h_p$ are twice differentiable at a point x^*, and if the first and second order constraint qualifications hold at x^*, then necessary conditions that x^* be a local minimum to the constrained optimization problem M are that there exist vectors $u^* = (u_1^*, \cdots, u_m^*)^T$ and*

$w^* = (w_1^*, \cdots, w_p^*)^T$ such that (8)–(12) hold, and such that for every vector y, where $y^T \nabla g_i^* = 0$ for all $i \in B^* = \{i \mid g_i(x^*) = 0\}$ and such that $y^T \nabla h_j^* = 0$, $j = 1, \cdots, p$, it follows that

$$y^T\left[\nabla^2 f^* - \sum_{i=1}^m u_i^* \nabla^2 g_i^* + \sum_{j=1}^p w_j^* \nabla^2 h_j^*\right] y \geqq 0. \tag{13}$$

Proof. (i) The first part of the theorem is a repetition of Theorem 3 and follows because the first order constraint qualification is assumed to hold.

(ii) Let y be any nonzero vector such that

$$y^T \nabla g_i^* = 0 \quad \text{for all} \quad i \in B^*, \tag{14}$$

and such that

$$y^T \nabla h_j^* = 0, \qquad j = 1, \cdots, p. \tag{15}$$

(If there are none the theorem is proved.) Let $\alpha(\theta)$ be the twice differentiable arc guaranteed by the second order constraint qualification where $\alpha(0) = x^*$, $\alpha'(0) = y$. Denote $\alpha''(0)$ by z. Then

$$g_i''(0) = y^T(\nabla^2 g_i^*)y + z^T \nabla g_i^* = 0 \quad \text{for all} \quad i \in B^*, \tag{16}$$

$$h_j''(0) = y^T(\nabla^2 h_j^*)\, y + z^T \nabla h_j^* = 0, \qquad j = 1, \cdots, p. \tag{17}$$

Otherwise some g_i, $i \in B^*$, or h_j, $j = 1, \cdots, p$, would not be equal to zero along $\alpha(\theta)$. Using the u^* and w^* given by (i), (14) and (15),

$$f'(0) = y^T \nabla f^* = y^T\left(\sum_{i=1}^m u_i^* \nabla g_i^* - \sum_{j=1}^p w_j^* \nabla h_j^*\right) = 0.$$

Since x^* is a local minimum, and $f'(0) = 0$, $f''(0)$ must be $\geqq 0$. That is,

$$f''(0) = y^T \nabla^2 f^* y + z^T \nabla f^* \geqq 0. \tag{18}$$

From this, (12), (16) and (17), it follows that

$$y^T\left[\nabla^2 f^* - \sum_{i=1}^m u_i^* \nabla^2 g_i^* + \sum_{j=1}^p w_j^* \nabla^2 h_j^*\right] y \geqq 0.$$

The following example illustrates that the first order constraint qualification can be satisfied while the second order constraint qualification fails to hold.

Example 3. Minimize x_2 subject to

$$g_1 = -x_1^9 + x_2^3 \geqq 0,$$

$$g_2 = x_1^9 + x_2^3 \geqq 0,$$

$$g_3 = x_1^2 + (x_2+1)^2 - 1 \geqq 0.$$

The solution is $\begin{pmatrix}0\\0\end{pmatrix}$. Now, $\nabla g_1^* = \nabla g_2^* = \begin{pmatrix}0\\0\end{pmatrix}$, $\nabla g_3^* = \begin{pmatrix}0\\2\end{pmatrix}$. Since all the constraints are equal to zero at $(0, 0)^T$, $B^* = (1, 2, 3)$. Any vector y such that $y \nabla g_i^* \geqq 0$ for all $i \in B^*$ must be of the form $(y_1, y_2^2)^T$. Clearly, any such vector is tangent to an arc pointing into the constraint region. Thus, the first order constraint qualification is satisfied.

Any vector y to be considered for the second order constraint qualification is of the form $\begin{pmatrix}y_1\\0\end{pmatrix}$ (where $y_1 \neq 0$ since y must be "nonzero"). Since there is no arc along which g_1, g_2 and g_3 remain equal to zero, the second order constraint fails to hold.

Note that the first order Lagrange conditions are only satisfied by $(u_1^*, u_2^*, \frac{1}{2})$, where u_1^*, u_2^* are any scalars $\geqq 0$. However,

$$\nabla^2 L(x^*, u^*) = \begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}$$

is negative definite and the second order necessary conditions do not hold. That the second order constraint qualification does not imply the first can be seen in the following example.

$$\begin{aligned} g_1 &= -x_1^2 - (x_2 - 1)^2 + 1 \geqq 0,\\ g_2 &= -x_1^2 - (x_2 + 1)^2 + 1 \geqq 0,\\ g_3 &= x_1 \qquad\qquad\qquad\quad\; \geqq 0.\end{aligned}$$

The point $\begin{pmatrix}0\\0\end{pmatrix}$ is the solution to any problem with these three constraints. Their gradients are $\begin{pmatrix}0\\2\end{pmatrix}$, $\begin{pmatrix}0\\-2\end{pmatrix}$, $\begin{pmatrix}1\\0\end{pmatrix}$.

The second order constraint qualification is satisfied because there are no vectors orthogonal to all three gradients. The first order qualification is not since there are no arcs pointing into the region of feasibility (which is a single point). There are vectors y giving nonnegative inner products, for example, $y = \begin{pmatrix}2\\0\end{pmatrix}$.

In order to use these necessary conditions as criteria for determining if a point can be a local minimum, one must determine if the constraint qualifications are satisfied. One situation which often occurs where this can be done is given in the next theorem.

THEOREM 5 (Condition implying constraint qualifications). *Suppose the functions $g_1, \cdots, g_m, h_1, \cdots, h_p$ are twice differentiable. A sufficient condition that the first order and second order constraint qualifications be satisfied* *at a point* x^* *is that the vectors* ∇g_i^* *for all* $i \in B^*$, $(\nabla h_j^*, j = 1, \cdots, p)$ *be linearly independent.*

Proof. We shall prove that the second order constraint qualification holds by constructing an arc which satisfies the hypotheses. The proof that the first order constraint qualification holds is omitted but is analogous to the one given.

Let y be any nonzero vector satisfying (14) and (15). (If none exist, the second order constraint qualification is trivially satisfied.) Let z be some vector such that

$$y^T(\nabla^2 g_i^*)\, y + z^T \nabla g_i^* = 0 \quad \text{for all} \quad i \in B^*, \tag{19}$$

$$y^T(\nabla^2 h_j^*)\, y + z^T \nabla h_j^* = 0, \qquad j = 1, \cdots, p. \tag{20}$$

Such a z exists because of the independence of the gradients. Assume there are q indices in B^* and that the inequality constraints are reordered so that $g_1, \cdots, g_q$ are those constraints. Let $c = p + q$, which, by the assumption of linear independence and the existence of a nonzero y satisfying (14) and (15), must be less than n. Let

$$M_c(\theta) = \begin{bmatrix} \dfrac{\partial g_1}{\partial x_1} & \cdots & \dfrac{\partial g_1}{\partial x_c} \\ \vdots & & \vdots \\ \dfrac{\partial g_q}{\partial x_1} & \cdots & \dfrac{\partial g_q}{\partial x_c} \\ \dfrac{\partial h_1}{\partial x_1} & \cdots & \dfrac{\partial h_1}{\partial x_c} \\ \vdots & & \vdots \\ \dfrac{\partial h_p}{\partial x_1} & \cdots & \dfrac{\partial h_p}{\partial x_c} \end{bmatrix},$$

let

$$M_{cn}(\theta) = \begin{bmatrix} \dfrac{\partial g_1}{\partial x_{c+1}} & \cdots & \dfrac{\partial g_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial g_q}{\partial x_{c+1}} & \cdots & \dfrac{\partial g_q}{\partial x_n} \\ \dfrac{\partial h_1}{\partial x_{c+1}} & \cdots & \dfrac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \dfrac{\partial h_p}{\partial x_{c+1}} & \cdots & \dfrac{\partial h_p}{\partial x_n} \end{bmatrix},$$

where each partial derivative in $M_c(\theta)$ and $M_{cn}(\theta)$ is evaluated at $\alpha(\theta)$.

We can assume without loss of generality that the vectors in $M_c(\theta)$ at $\theta = 0$ are linearly independent. Let

$$d(\theta) = \begin{bmatrix} \alpha'(\theta)^T \nabla^2 g_1 \, \alpha'(\theta) \\ \vdots \\ \alpha'(\theta)^T \nabla^2 g_q \, \alpha'(\theta) \\ \alpha'(\theta)^T \nabla^2 h_1 \, \alpha'(\theta) \\ \vdots \\ \alpha'(\theta)^T \nabla^2 h_p \, \alpha'(\theta) \end{bmatrix},$$

$$z_{cn} = \begin{bmatrix} z_{c+1} \\ \vdots \\ z_n \end{bmatrix},$$

where each matrix of second partial derivatives in $d(\theta)$ is evaluated at $\alpha(\theta)$. The arc is constructed as follows: let $\alpha(0) = x^*$, $\alpha'(0) = y$, and

$$\alpha''(\theta) = \begin{bmatrix} [M_c(\theta)]^{-1}[-M_{cn}(\theta) z_{cn} - d(\theta)] \\ z_{cn} \end{bmatrix}. \tag{21}$$

The existence of a twice differentiable function (arc) satisfying these conditions is guaranteed since the right-hand side of (21) is defined and continuous in a neighborhood about x^*. (See [6, Theorem 1.2].) It is a nontrivial arc since y was assumed not equal to zero.

That $\alpha''(0) = z$ follows by solving (19) and (20) for $(z_1, \cdots, z_c)^T$ in terms of z_{cn}. This agrees with (21) at $\theta = 0$. Let $e = (g_1, \cdots, g_q, h_1, \cdots, h_p)^T$, and let e_k be any component of e. Then

$$e_k(\theta) = e_k(0) + \theta e_k'(0) + (\theta^2/2)\, e_k''(\phi),$$

where $0 \leqq \phi \leqq \theta$ in a neighborhood about x^*. But $e(0) = 0 = e'(0)$, and

$$\begin{aligned} e''(\phi) &= d^k(\phi) + [M_c^k(\phi) : M_{cn}^k(\phi)] \begin{bmatrix} \{M_c(\phi)\}^{-1}[-M_{cn}(\phi) z_{cn} - d(\phi)] \\ z_{cn} \end{bmatrix} \\ &= 0. \end{aligned}$$

4. Second order sufficient conditions. The following conditions constitute an attempt to add as little as possible to the necessary conditions of Theorem 4 to create ones which are sufficient that a point be a local minimum.

THEOREM 6 (Second order sufficient conditions).[2] *Sufficient conditions that a point x^* be an isolated local minimum to the constrained optimization problem* M, *where f, $g_1, \cdots, g_m$, $h_1, \cdots, h_p$ are twice differentiable functions, are that there exist vectors u^*, w^* such that*

[2] A statement of this theorem when there are no inequality constraints is contained in [3, pp. 115–116].

$$g_i(x^*) \geqq 0, \qquad i = 1, \cdots, m, \tag{22}$$

$$h_j(x^*) = 0, \qquad j = 1, \cdots, p, \tag{23}$$

$$u_i^* g_i(x^*) = 0, \qquad i = 1, \cdots, m, \tag{24}$$

$$u_i^* \geqq 0, \qquad i = 1, \cdots, m, \tag{25}$$

$$\nabla f^* - \sum_{i=1}^{m} u_i^* \nabla g_i^* + \sum_{i=1}^{p} w_j^* \nabla h_j^* = 0, \tag{26}$$

and for every nonzero vector y where $y^T \nabla g_i^* = 0$ for all $i \in D^* = \{i \mid u_i^* > 0\}$ and $y^T \nabla h_j^* = 0, j = 1, \cdots, p$, it follows that

$$y^T \left[\nabla^2 f^* - \sum_{i=1}^{m} u_i^* \nabla^2 g_i^* + \sum_{j=1}^{p} w_j^* \nabla^2 h_j^* \right] > 0. \tag{27}$$

Proof. Assume that x^* is not an isolated local minimum. Then there exists a sequence of points $\{y^k\}$ where $\lim_{k\to\infty} y^k = x^*$ such that (i) each y^k is feasible, and (ii) $f(y^k) \leqq f(x^*)$. We can rewrite each y^k as $x^* + \delta^k s^k$ ($\delta^k > 0$), where s^k is a unit vector. We consider any limit point of the sequence $\{\delta^k, s^k\}$. Clearly any such limit point is of the form $(0, \bar{s})$, where $\bar{s}$ is a unit vector. By (i),

$$g_i(y^k) - g_i(x^*) \geqq 0 \qquad \text{for all} \qquad i \in B^*,$$

$$h_j(y^k) - h_j(x^*) = 0, \qquad j = 1, \cdots, p,$$

and by (ii),

$$f(y^k) - f(x^*) \leqq 0.$$

Dividing each equation above by δ^k, and taking the limit as $k \to \infty$ (using that sequence converging to $\bar{s}$), we have, by the assumed differentiability properties, that

$$\nabla^T g_i^* \bar{s} \geqq 0 \qquad \text{for all} \qquad i \in B^*, \tag{28}$$

$$\nabla^T h_j^* \bar{s} = 0, \qquad j = 1, \cdots, p, \tag{29}$$

$$\nabla^T f^* \bar{s} \leqq 0. \tag{30}$$

We have two cases to consider, and show a contradiction arises from each of them.

(a) For the unit vector $\bar{s}$, $\nabla^T g_i^* \bar{s} > 0$ for at least one $i \in D^*$. This assumption coupled with (28), (29), (30), (24), (25) and (26) means that

$$0 \geqq \nabla^T f^* \bar{s} = \sum_{i \in D^*} u_1^* \nabla^T g_1^* \bar{s} + \sum_{j=1}^{p} w_j^* \nabla^T h_j^* \bar{s} > 0.$$

This is an impossibility.

(b) For the unit vector $\bar{s}$,

$$\nabla^T g_1^* \bar{s} = 0 \quad \text{for all} \quad i \in D^*.$$

Using Taylor's expansion (defining $\mathcal{L}(x, u, w) = f(x) - \sum u_1 g_1(x) + \sum w_j h_j(x)$),

$$\mathcal{L}(y^k, u^*, w^*) = \mathcal{L}(x^*, u^*, w^*) + \delta^k (s^k)^T \nabla \mathcal{L}(x^*, u^*, w^*)$$

$$+ [(\delta^k)^2/2](s^k)^T[\nabla^2 \mathcal{L}(n^k, u^*, w^*)]s^k, \tag{31}$$

where $n^k = \lambda x^* + (1 - \lambda)\delta^k(s^k)$, $0 \leqq \lambda \leqq 1$. Using properties (i) and (ii), (24), (25), (26) to reduce (31) to an inequality, then dividing by $(\delta^k)^2/2$ yields

$$0 \geqq (s^k)^T[\nabla^2 \mathcal{L}(n^k, u^*, w^*)]s^k. \tag{32}$$

Taking the limit as $k \to \infty$ yields, because of assumption (b), a statement contradicting (27).

5. Examples. The application of these theorems to the earlier examples will now be shown. In Example 1, since there is only one equality constraint, the hypotheses of Theorem 4 (by virtue of Theorem 5) are satisfied. If $(-1, -1)^T$ is a local minimum, the matrix $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ must be positive semidefinite for all vectors y such that $(y_1, y_2)(-2, -2)^T = 0$, that is, all vectors of the form $(y_1, -y_1)$. Multiplying yields $-4y_1^2 < 0$ (when $y_1 \neq 0$). Thus $(-1, -1)^T$ cannot be a local minimum. Applying the sufficiency test of Theorem 6 assures us that $(1, 1)^T$ is a local minimum. Since it is the only local minimum, it must also be the global minimum.

In Example 2, $[\nabla^2 f - \sum u_i \nabla^2 g_i]$ at $x^* = (0, 0)^T$ is $\begin{bmatrix} 2 & 0 \\ 0 & 2 - 4/k \end{bmatrix}$. Now since $\nabla g_1^* = (-1, 0)^T$, we need only consider vectors y of the form $(0, y_2)^T$. The number to be investigated is, therefore, $y_2^2[2 - 4/k]$. By Theorem 6, for $k > 2$, $(0, 0)^T$ is a local minimum. By Theorem 4, for $k < 2$, $(0, 0)^T$ is not a local minimum. At $k = 2$, the necessary conditions are satisfied, but not the sufficient ones.

REFERENCES

[1] F. John, *Extremum problems with inequalities as subsidiary conditions*, Studies and Essays, Interscience, New York, 1948, pp. 187–204.

[2] D. Gale, *Convex polyhedral cones and linear inequalities*, Activity Analysis of Production and Allocation, Cowles Commission Monograph 13, T. C. Koopmans, ed., John Wiley, New York, 1951.

[3] H. Hancock, *Theory of Maxima and Minima*, Dover, New York, 1960.

[4] H. W. Kuhn and A. W. Tucker, *Nonlinear programming*, Proceedings of the

Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, Berkeley, California, 1951, pp. 481–492.
[5] K. RITTER, *Stationary points of quadratic maximum problems*, Z. Wahrscheinlichkeitstheorie und Verw. Gebeite, 4 (1965), pp. 149–158.
[6] E. A. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.

AN INDIRECT SUFFICIENCY PROOF FOR THE PROBLEM OF LAGRANGE WITH DIFFERENTIAL INEQUALITIES AS ADDED SIDE CONDITIONS

BY

LOUIS L. PENNISI

1. Introduction. The problem to be considered here consists in finding in a class of arcs C: $y^i(x)$ $(i=1, \cdots, n;\ x^1 \leqq x \leqq x^2)$ joining two fixed points and satisfying a set of differential inequalities and equations of the form

$$\phi^\beta(x, y, \dot{y}) \geqq 0, \qquad \psi^\rho(x, y, \dot{y}) = 0$$

that one which minimizes the integral

$$I(C) = \int_{x^1}^{x^2} f(x, y, \dot{y})dx.$$

Valentine [11](1) has given a brief history of this problem and has derived certain necessary conditions by introducing auxiliary functions $z^\beta(x)$ such that $\phi^\beta(x, y, \dot{y}) = \dot{z}^{\beta^2}$. His sufficiency theorems depended on assumptions of normality and a field theory, and also required all except one of the differential inequalities to be satisfied in the strict sense along the minimizing arc C_0.

Using methods developed by McShane [9] and Hestenes [3; 4; 5; 6] we shall give an indirect proof of a sufficiency theorem. Instead of demanding that all but one of the functions ϕ^β is positive along C_0 we shall impose the more general restriction (2.5) to be described in the next section.

2. Statement of problem and main theorem. Let $\mathfrak{R}$ be a region in $(2n+1)$-dimensional space of points $(x, y, p) = (x, y^1, \cdots, y^n, p^1, \cdots, p^n)$. By an admissible arc C will be meant a set of functions $y^i(x)$ $(i=1, \cdots, n;\ x^1 \leqq x \leqq x^2)$ which are absolutely continuous and have integrable square derivatives $\dot{y}^i(x)$ such that the point $[x, y(x), \dot{y}(x)]$ is in $\mathfrak{R}$ for almost all x on x^1x^2. It will be assumed that the functions $f(x, y, p)$, $\phi^\beta(x, y, p)$, $\psi^\rho(x, y, p)$ are of class C'' on $\mathfrak{R}$ $(\beta=1, \cdots, m;\ \rho=m+1, \cdots, m+t<n)$. The subset on $\mathfrak{R}$ on which $\phi^\beta(x, y, p) \geqq 0$, $\psi^\rho(x, y, p) = 0$ will be denoted by $\mathfrak{D}$. We shall say that C lies in $\mathfrak{D}$ if the point $[x, y(x), \dot{y}(x)]$ lies in $\mathfrak{D}$ for almost all x on x^1x^2.

We shall be concerned with a particular admissible arc C_0: $y^i = y_0^i(x)$ of class C' which lies in $\mathfrak{D}$, satisfies the end conditions

$$(2.1) \qquad y^i(x^s) = y^{is} \qquad (i = 1, \cdots, n;\ x^1 \leqq x \leqq x^2;\ s = 1, 2)$$

and along which the matrix

Received by the editors April 4, 1952.

(1) The numbers in brackets refer to the bibliography at the end of the paper.

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_13, © Springer Basel 2014

$$\left\| \begin{array}{c} \phi^{\beta}_{p^i} \\ \psi^{\rho}_{p^i} \end{array} \right\| \tag{2.2}$$

has rank $m+t$. Suppose that the interval x^1x^2 can be divided into a finite number of open intervals A_τ: $x_\tau < x < x_{\tau+1}$ $(\tau = 0, 1, \cdots, T)$, where $x_0 = x^1$, $x_{T+1} = x^2$, in such a manner that each of the functions $\phi^\beta[x, y_0(x), \dot{y}_0(x)]$ either vanishes identically on A_τ or is positive everywhere on A_τ. Let $\Gamma(x)$ be the set of indices β such that $\phi^\beta[x, y_0(x), \dot{y}_0(x)] = 0$. Then $\Gamma(x)$ is independent of x when x is in A_τ and may be denoted by Γ_τ. Let $A(\beta)$ be the closure of the sum of the intervals A_τ on which $\phi^\beta[x, y_0(x), \dot{y}_0(x)] = 0$.

Consider now a set of continuous functions $\lambda^0 \geqq 0$, $\lambda^\beta(x)$, $\lambda^\rho(x)$ such that if we define

$$F(x, y, p, \lambda) = \lambda^0 f(x, y, p) + \lambda^\beta \phi^\beta(x, y, p) + \lambda^\rho \psi^\rho(x, y, p),$$

then the equations

$$F_{p^i} = \int_{x^1}^{x} F_{y^i} dx + c^i, \tag{2.3}$$

$$\lambda^\beta(x)\phi^\beta(x, y, \dot{y}) = 0 \qquad (\beta \text{ not summed}) \tag{2.4}$$

hold along C_0 with the multipliers $\lambda(x)$ for some set of constants c^i. Let $\Delta(x)$ be the set of indices β such that $\lambda^\beta(x) \neq 0$ and let $B(\beta)$ be the set of points x for which $\lambda^\beta(x) \neq 0$. It is evident from (2.4) that $\overline{B}(\beta)$, the closure of $B(\beta)$, is contained in $A(\beta)$ and that $\Delta(x)$ is contained in $\Gamma(x)$. In fact, $\Delta(x)$ is contained in Γ_τ for all x in $\overline{A}_\tau$. We shall assume that

$$\Gamma(x) - \Delta(x) \text{ contains at most one index.} \tag{2.5}$$

We shall make a further restriction on our choice of multipliers. Let $E_F(x, y, p, q, \lambda)$ be the Weierstrass E-function

$$E_F = F(x, y, q, \lambda) - F(x, y, p, \lambda) - (q^i - p^i)F_{p^i}(x, y, p, \lambda),$$

and define $\bar{\phi}^\beta(x, y, p)$ as

$$\bar{\phi}^\beta(x, y, p) = \phi^\beta(x, y, p)/[1 + \phi^{\beta^2}(x, y, p)]^{1/2}.$$

It will be assumed that there is a neighborhood $\mathcal{D}_1$ of C_0 relative to the set $\mathcal{D}$ and a constant b such that $0 < b < 1$ and the inequality

$$E_F(x, y, p, q, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, q) \geqq bE_L(p, q) - \lambda^\beta(x)\bar{\phi}^\beta(x, y, q) \tag{2.6}$$

holds whenever (x, y, q) is in $\mathcal{D}_1$, (x, y, q) is in $\mathcal{D}$, and $\phi^\beta(x, y, p) = 0$ in case β is in $\Delta(x)$. Here $L(p)$ is the integrand $(1 + p^ip^i)^{1/2}$ of the length integral and $E_L(p, q)$ is the E-function

$$E_L(p, q) = L(q) - (1 + p^iq^i)/L(p) \tag{2.7}$$

for this integrand. We shall call the set of functions $\lambda^0 \geqq 0$, $\lambda^\beta(x)$, $\lambda^\rho(x)$ an *admissible set of multipliers* when the functions are continuous on x^1x^2 and the conditions (2.3), (2.4), (2.5), and (2.6) are satisfied under the conditions described.

Let us now consider a set of functions $\eta^i(x)$ which are absolutely continuous and have integrable square derivatives $\dot\eta^i(x)$ on x^1x^2. If such a set of functions satisfies with the curve C_0 and an admissible set of multipliers the end conditions

$$\eta^i(x^s) = 0, \tag{2.8}$$

the differential equations

$$\phi^\beta_{y^i}\eta^i + \phi^\beta_{p^i}\dot\eta^i = 0 \tag{2.9}$$

for almost all x in $B(\beta)$, the differential inequalities

$$\phi^\beta_{y^i}\eta^i + \phi^\beta_{p^i}\dot\eta^i \geqq 0 \tag{2.10}$$

for almost all x in $A(\beta)-B(\beta)$, and the differential equations

$$\psi^\rho_{y^i}\eta^i + \psi^\rho_{p^i}\dot\eta^i = 0 \tag{2.11}$$

for almost all x on x^1x^2, it will be called an *admissible variation*. For each admissible variation, the second variation

$$\begin{aligned} J_2(\eta) &= \int_{x^1}^{x^2} 2\omega(x, \eta, \dot\eta)dx \\ &= \int_{x^1}^{x^2} (F_{y^iy^k}\eta^i\eta^k + 2F_{y^ip^k}\eta^i\dot\eta^k + F_{p^ip^k}\dot\eta^i\dot\eta^k)dx \end{aligned} \tag{2.12}$$

is well defined.

THEOREM 2.1. *Let C_0 be an arc of class C' which lies in $\mathfrak{D}$, satisfies the end conditions* (2.1), *and along which the matrix* (2.2) *has rank $m+t$. Suppose an admissible set of multipliers can be found for which $J_2(\eta)>0$ for every nonnull admissible variation. Then there is a neighborhood $\mathfrak{J}$ of C_0 in (x, y)-space such that the inequality $I(C)>I(C_0)$ holds for every admissible arc C in $\mathfrak{J}$ which lies in $\mathfrak{D}$, satisfies the end conditions* (2.1), *and is different from C_0.*

3. **The problem with a finite number of variables.** In order to be able to draw conclusions from the inequality (2.6) and also to furnish a model for the calculus of variations problem, we find it convenient to discuss first the problem of minimizing a function $f(x)$ of n variables x^i in the class of points x satisfying m inequalities $\phi^\beta(x) \geqq 0$ and t equations $\psi^\rho(x)=0$, when $m+t<n$. For the normal case, which is the only one we consider, our results are more general than those of Karush [7].

We suppose that the functions $f(x)$, $\phi^\beta(x)$, and $\psi^\rho(x)$ are of class C'' in a neighborhood $\mathcal{R}$ of a point x_0, that $\phi^\beta(x_0)\geqq 0$, $\psi^\rho(x_0)=0$ and that the matrix

$$(3.1) \qquad \left\| \begin{matrix} \phi^{\beta}_{x^i}(x_0) \\ \psi^{\rho}_{x^i}(x_0) \end{matrix} \right\|$$

has rank $m+t$.

THEOREM 3.1. *If x_0 minimizes $f(x)$ in the class of points x near x_0 such that $\phi^\beta(x)\geqq 0$, $\psi^\rho(x)=0$, then there exist unique multipliers μ^β, μ^ρ such that if $F(x, \mu)=f(x)+\mu^\beta\phi^\beta(x)+\mu^\rho\psi^\rho(x)$, then $F_{x^i}(x_0, \mu)=0$. Moreover, $\mu^\beta\leqq 0$ and $\mu^\beta\phi^\beta(x_0)=0$ for each β.*

The first sentence of the theorem follows from well known results [1, p. 210] since the point x_0 is a normal point which minimizes $f(x)$ in the class of points x near x_0 for which $\phi^\beta(x)=\phi^\beta(x_0)$, $\psi^\rho(x)=0$. To prove that $\mu^\beta\leqq 0$ we pick functions $\psi^j(x)$ $(j=m+t+1, \cdots, n)$ of class C'' such that the equations

$$(3.2) \qquad \phi^\beta(x) = v^\beta + \phi^\beta(x_0), \qquad \psi^\rho(x) = 0, \qquad \psi^j(x) = \psi^j(x_0) + v^j$$

have a nonvanishing functional determinant when $x=x_0$, $v^\beta=0$, $v^j=0$. Solutions $x^i(v)$ of class C'' near $v=0$ therefore exist such that $x^i(0)=x_0^i$ and so the function

$$(3.3) \qquad f[x(v)] = F[x(v), \mu] - \mu^\beta v^\beta - \mu^\beta\phi^\beta(x_0)$$

is minimized by $v=0$ in the class of points v near 0 for which each $v^\beta\geqq 0$. Its derivative with respect to v^β, namely $-\mu^\beta$, must therefore be non-negative. If $\phi^\beta(x_0)>0$ for some β, then the function (3.3) has a two-sided minimum when regarded as a function of v^β only and so $\mu^\beta=0$. Hence $\mu^\beta\phi^\beta(x_0)=0$ for each β.

Let us define Γ as the set of all indices β such that $\phi^\beta(x_0)=0$, and Δ as the set of all indices β such that $\mu^\beta<0$. It is clear then that Δ is a subset of Γ.

THEOREM 3.2. *If x_0 minimizes $f(x)$ in the class of points x near x_0 such that $\phi^\beta(x)\geqq 0$, $\psi^\rho(x)=0$, then the function $F(x, \mu)$ defined by Theorem 3.1 is such that*

$$F_{x^i x^k}(x_0, \mu)\pi^i\pi^k \geqq 0$$

for every set π^i such that

$$\psi^{\rho}_{x^i}(x_0)\pi^i = 0, \qquad \phi^{\gamma}_{x^i}(x_0)\pi^i \geqq 0, \qquad \phi^{\delta}_{x^i}(x_0)\pi^i = 0$$

for all ρ, all γ in $\Gamma-\Delta$, and all δ in Δ.

If π^i is a set as described in the theorem, we set $v^\beta(t)=\phi^\beta_{x^i}(x_0)\pi^i t$, $v^j(t)=\psi^j_{x^i}(x_0)\pi^i t$ and observe that the function $f[x\{v(t)\}]$, in which $x(v)$ is the function constructed in the proof of Theorem 3.1, is minimized by $t=0$ in the class of non-negative t near $t=0$. The first derivative of this function obviously vanishes when $t=0$ and so its second derivative must be non-negative when $t=0$, or $F_{x^ix^k}(x_0, \mu)\dot{x}_0^i\dot{x}_0^k \geqq 0$, in which $\dot{x}_0^i = dx^i[v(t)]/dt$ when $t=0$. If we differentiate equations (3.3) after replacing v by $v(t)$, we discover that

$$\phi^\beta_{x^i}(x_0)\dot{x}_0^i = \phi^\beta_{x^i}(x_0)\pi^i, \qquad \psi^\rho_{x^i}(x_0)\dot{x}_0^i = 0 = \psi^\rho_{x^i}(x_0)\pi^i,$$
$$\psi^j_{x^i}(x_0)\dot{x}_0^i = \psi^j_{x^i}(x_0)\pi^i.$$

It follows that $\dot{x}_0^i = \pi^i$ and hence that $F_{x^ix^k}(x_0, \mu)\pi^i\pi^k \geqq 0$, as desired.

COROLLARY. *If x_0 minimizes $f(x)$ in the class of points x near x_0 such that $\phi^\beta(x) \geqq 0$, $\psi^\rho(x)=0$, and if the set $\Gamma-\Delta$ contains at most one index, γ, then the function $F(x, \mu)$ defined by Theorem* 3.1 *is such that $F_{x^ix^k}(x_0, \mu)\pi^i\pi^k \geqq 0$ for every set π^i such that $\psi^\rho_{x^i}(x_0)\pi^i=0$, $\phi^\delta_{x^i}(x_0)\pi^i=0$ for all ρ and all δ in Δ.*

There is nothing new to prove unless π^i is a set satisfying the conditions of the corollary for which $\phi^\gamma_{x^i}(x_0)\pi^i<0$. In this case we define $\bar{\pi}^i = -\pi^i$ and have a set $\bar{\pi}^i$ satisfying the conditions of Theorem 3.2. Since the quadratic form is unaltered when π^i is replaced by $\bar{\pi}^i$, we see that the corollary is true.

THEOREM 3.3. *Suppose there exist multipliers $\mu^0 \geqq 0$, $\mu^\beta \leqq 0$, μ^ρ such that if $F(x, \mu) = \mu^0 f(x) + \mu^\beta\phi^\beta(x) + \mu^\rho\psi^\rho(x)$, then $F_{x^i}(x_0, \mu)=0$, $\mu^\beta\phi^\beta(x_0)=0$ for each β. If the quadratic form $F_{x^ix^k}(x_0, \mu)\pi^i\pi^k>0$ for all nonnull π^i such that $\psi^\rho_{x^i}(x_0)\pi^i=0$, $\phi^\gamma_{x^i}(x_0)\pi^i \geqq 0$, $\phi^\delta_{x^i}(x_0)\pi^i=0$ for all ρ, all γ such that $\mu^\gamma=\phi^\gamma(x_0)=0$, and all δ such that $\mu^\delta<0$, then there is a neighborhood $\mathfrak{J}$ of x_0 such that $f(x)>f(x_0)$ if x is any point in $\mathfrak{J}$ different from x_0 for which $\phi^\beta(x) \geqq 0$, $\psi^\rho(x)=0$.*

If this theorem were false, there would be a sequence of points $x_r \neq x_0$ converging to x_0 such that $\phi^\beta(x_r) \geqq 0$, $\psi^\rho(x_r)=0$, $f(x_r) \leqq f(x_0)$. Define k_r as the positive square root of $|x_r-x_0|^2 - \mu^\beta\phi^\beta(x_r)$. Then k_r converges to zero. If we define $\pi_r^i = (x_r^i - x_0^i)/k_r$, then $|\pi_r| \leqq 1$ since $-\mu^\beta\phi^\beta(x_r) \geqq 0$. By passing to a subsequence we may therefore suppose that π_r^i converges to a limit π_0^i. Since $\mu^0 \geqq 0$, we have that

$$0 \geqq k_r^{-2}[F(x_r, \mu) - F(x_0, \mu) - \mu^\beta\phi^\beta(x_r)].$$

Expanding the right-hand side of this inequality by Taylor's theorem, making use of the fact that $F_{x^i}(x_0, \mu)=0$, and letting r approach ∞, we find that

$$0 \geqq \frac{1}{2} F_{x^ix^j}(x_0, \mu)\pi_0^i\pi_0^j + \limsup\,[-k_r^{-2}\mu^\beta\phi^\beta(x_r)]. \tag{3.4}$$

Moreover, we see from Taylor's theorem and the relations

$$-\frac{k_r}{u^\delta} > \frac{\phi^\delta(x_r)}{k_r} \geqq 0, \qquad \frac{\phi^\gamma(x_r)}{k_r} \geqq 0, \qquad \frac{\psi^\rho(x_r)}{k_r} = 0,$$

that $\phi^{\delta}_{x^i}(x_0)\pi^i_0=0$ if $u^\delta<0$, $\phi^{\gamma}_{x^i}(x_0)\pi^i_0\geqq 0$ if $\phi^\gamma(x_0)=0$, and $\psi^{\rho}_{x^i}(x_0)\pi^i_0=0$. If the set π^i_0 were nonnull, it would then follow from the hypothesis of Theorem 3.3 that $F_{x^ix^j}(x_0)\pi^i_0\pi^j_0>0$ and so (3.4) could not hold since lim sup $[-k_r^{-2}\mu^\beta\phi^\beta(x_r)]\geqq 0$. Hence $\pi^i_0=0$. It then follows from (3.4) that

$$\lim k_r^{-2}\mu^\beta\phi^\beta(x_r) = 0. \tag{3.5}$$

Since we have from the definition of k_r that $1=|\pi_r|^2-k_r^{-2}\mu^\beta\phi^\beta(x_r)$, and since $\pi^i_0=0$, we see that (3.5) cannot hold. We infer the truth of the theorem from this contradiction.

4. **Some consequences of the hypotheses.** For a fixed x on x^1x^2 it follows from our hypotheses (2.4) and (2.6) that the function of q,

$$E_F(x, y_0, \dot y_0, q, \lambda) - bE_L(\dot y_0, q) - \lambda^\beta(x)\phi^\beta(x, y_0, q) + b\lambda^\beta(x)\bar\phi^\beta(x, y_0, q),$$

is minimized by $q^i=\dot y^i_0$ in the class of q^i near $\dot y^i_0$ for which $\phi^\beta(x, y_0, q)\geqq 0$, $\psi^\rho(x, y_0, q)=0$. From Theorem 3.1 we conclude that there exist multipliers μ^β, μ^ρ such that

$$[\mu^\beta - (1-b)\lambda^\beta]\phi^\beta_{p^i} + \mu^\rho\psi^\rho_{p^i} = 0$$

when $y^i=y^i_0(x)$, $p^i=\dot y^i_0(x)$. Since the matrix (2.2) has rank $m+t$, it follows that $\mu^\beta=(1-b)\lambda^\beta$, $\mu^\rho=0$ and since $0<b<1$, we also conclude from Theorem 3.1 that $\lambda^\beta\leqq 0$. From Theorem 3.3 we have that

$$\frac{\partial^2}{\partial q^i\partial q^k}[E_F(x, y_0, \dot y_0, q, \lambda) - bE_L(\dot y_0, q)]\pi^i\pi^k \geqq 0$$

when $q^i=\dot y^i_0(x)$ for all π^i such that $\psi^\rho_{p^i}\pi^i=0$, $\phi^\gamma_{p^i}\pi^i\geqq 0$, $\phi^\delta_{p^i}\pi^i=0$ for all ρ, all γ in $\Gamma(x)-\Delta(x)$ and all δ in $\Delta(x)$. Since (2.5) holds we have the following theorem.

THEOREM 4.1. *If the multipliers satisfy* (2.4) *and* (2.6), *then*

$$\lambda^\beta(x) \leqq 0. \tag{4.1}$$

If they also satisfy (2.5), *then* $F_{p^ip^k}\pi^i\pi^k\geqq bL_{p^ip^k}\pi^i\pi^k$ *holds for every set* π^i *for which* $\psi^\rho_{p^i}\pi^i=\phi^\delta_{p^i}\pi^i=0$ *for all* ρ *and all* δ *in the set* $\Delta(x)$ *for which* $\lambda^\delta(x)<0$.

COROLLARY 1. *If the multipliers satisfy* (2.4), (2.5), *and* (2.6), *then the matrix*

$$\begin{Vmatrix} F_{p^ip^k} & \phi^\gamma_{p^i} & \psi^\rho_{p^i} \\ \phi^\gamma_{p^k} & 0 & 0 \\ \psi^\rho_{p^k} & 0 & 0 \end{Vmatrix}, \tag{4.2}$$

in which γ *ranges over the set* Γ_τ, *is nonsingular when* x *is in* $\bar A_\tau$, $y^i=y^i_0(x)$, $p^i=\dot y^i_0(x)$.

This result is an obvious corollary of Theorem 4.1 when it is remarked that $\Delta(x)$ is a subset of Γ_τ when x is in $\bar{A}_\tau$.

The following differentiability theorem is an immediate consequence of the preceding corollary.

COROLLARY 2. *When the multipliers are admissible, they and the functions* $\dot{y}_0^i(x)$ *are of class* C' *on each interval* $\bar{A}_\tau$.

COROLLARY 3. *There exists a constant* θ *such that if we define*

$$\phi_0^\beta(x, y, p) = \lambda^\beta(x)\phi^\beta(x, y, p),$$

for each β, *and*

$$H(x, y, p) = \theta[\psi^\rho(x, y, p)\psi^\rho(x, y, p) + \phi_0^\beta(x, y, p)\phi_0^\beta(x, y, p)],$$

then the quadratic form $(F_{p^ip^k}+H_{p^ip^k})\pi^i\pi^k>0$ *for all nonnull vectors* π^i *and each* x *on* x^1x^2.

This is a consequence of a known result [10, p. 679] since C_0 satisfies the differential equations $\psi^\rho(x, y, \dot{y})=0, \phi_0^\beta(x, y, \dot{y})=0$ and is such that $F_{p^ip^k}\pi^i\pi^k \geqq 0$ for all nonnull vectors π^i such that $\psi_{p^i}^\rho\pi^i=0$, $\phi_{0p^i}^\beta\pi^i=0$.

LEMMA 4.1. *There exist functions* $\psi^j(x, y, p)$ $(j=m+t+1, \cdots, n)$ *of class* C'' *near* C_0 *such that the determinant*

$$\begin{vmatrix} \phi_{p^i}^\beta \\ \psi_{p^i}^\rho \\ \psi_{p^i}^j \end{vmatrix} \tag{4.3}$$

does not vanish on C_0.

This lemma is a well known consequence of the fact that the matrix (2.2) has rank $m+t$ along C_0 [1, pp. 224–226].

5. **The equivalence of** (2.6) **to** II_N **and condition** III′. We define the class $\mathcal{N}$ as the collection of all sets N of points (x, y, p, ν) such that there exists a neighborhood $\mathcal{R}_1$ of the points $(x, y_0, \dot{y}_0)$ on C_0 and a positive constant a for which (x, y, p, ν) is in N if and only if (x, y, p) is in $\mathcal{R}_1$, $\nu^0=\lambda^0$, $|\nu^\rho-\lambda^\rho(x)| \leqq a$, $|\nu^\beta-\lambda^\beta(x)| \leqq -a\lambda^\beta(x)$ if β is not in $\Gamma(x)-\Delta(x)$, $-a\leqq\nu^\gamma\leqq 0$ if γ is in $\Gamma(x)-\Delta(x)$. We shall say that the arc C_0 with the multipliers λ^0, $\lambda^\beta(x)$, $\lambda^\rho(x)$ satisfies the condition II_N if there is a set N in $\mathcal{N}$ such that

$$E_F(x, y, p, q, \nu) - \nu^\beta\phi^\beta(x, y, q) \geqq 0 \tag{5.1}$$

for all sets (x, y, p, q, ν) such that (x, y, p, ν) is in N, (x, y, p) and (x, y, q) are in $\mathcal{D}$ and $\phi^\beta(x, y, p)=0$ if β is in $\Delta(x)$.

We shall say that the arc C_0 with the multipliers λ^0, $\lambda^\beta(x)$, $\lambda^\rho(x)$ satisfies the condition III′ in case $F_{p^ip^k}[x, y_0(x), \dot{y}_0(x), \lambda(x)]\pi^i\pi^k>0$ for every nonnull vector π^i such that $\psi^\rho_{p^i}[x, y_0(x), \dot{y}_0(x)]\pi^i = \phi^\delta_{p^i}[x, y_0(x), \dot{y}_0(x)]\pi^i = 0$ for all ρ and all δ in $\Delta(x)$.

THEOREM 5.1. *If the arc C_0 with the multipliers λ^0, $\lambda^\beta(x)$, $\lambda^\rho(x)$ satisfies equations* (2.4), *then it satisfies the condition* II_N *and satisfies the condition* III′ *if and only if there is a constant b such that $0 < b < 1$ and a neighborhood $\mathfrak{D}_1$ of C_0 relative to the set $\mathfrak{D}$ such that the inequality* (2.6) *holds whenever (x, y, p) is in $\mathfrak{D}_1$, (x, y, q) is in $\mathfrak{D}$, and $\phi^\beta(x, y, p) = 0$ if β is in $\Delta(x)$.*

The proof of this theorem will follow from a series of lemmas which we now proceed to prove.

LEMMA 5.1. *If the arc C_0 with the multipliers λ^0, $\lambda^\beta(x)$, $\lambda^\rho(x)$ satisfies the condition* II_N *and equations* (2.4), *then $\lambda^\beta(x) \leqq 0$.*

This follows directly from Theorem 3.1 when we observe that

$$E_F(x, y_0, \dot{y}_0, q, \lambda) - \lambda^\beta(x)\phi^\beta(x, y_0, q) \geqq 0$$

for any q^i such that $\phi^\beta(x, y_0, q) \geqq 0$, $\psi^\rho(x, y_0, q) = 0$.

LEMMA 5.2. *If there is a positive constant b such that the inequality* (2.6) *holds for all (x, y, p) in a neighborhood $\mathfrak{D}_1$ of C_0 relative to $\mathfrak{D}$ for which $\phi^\beta(x, y, p) = 0$ if β is in $\Delta(x)$ and all (x, y, q) in $\mathfrak{D}$, and if C_0 with the multipliers λ^0, $\lambda^\beta(x)$, $\lambda^\rho(x)$ satisfies the condition* II_N, *then the constant b may be required to be less than one.*

If b is not less than one, it follows from Lemma 5.1 that if $b' < 1$,

$$b[E_L(p, q) - \lambda^\beta(x)\bar{\phi}^\beta(x, y, q)] \geqq b'[E_L(p, q) - \lambda^\beta(x)\bar{\phi}^\beta(x, y, q)]$$

for all (x, y, q) in $\mathfrak{D}$. Hence the lemma is true.

LEMMA 5.3. *If the arc C_0 with the multipliers $\lambda^0 \geqq 0$, $\lambda^\beta(x) \leqq 0$, $\lambda^\rho(x)$ satisfies the condition* III′ *and satisfies* (2.4), *there exists a positive number b and a neighborhood $\mathfrak{D}_0$ of C_0 relative to $\mathfrak{D}$ such that the inequality* (2.6) *holds for all (x, y, p) in $\mathfrak{D}_0$ such that $\phi^\beta(x, y, p) = 0$ if β is in $\Delta(x)$ and all (x, y, q) in $\mathfrak{D}_0$.*

The lemma may be proved by minor adaptations of the proof of Theorem 3.3.

We can easily infer from the identity

$$\begin{aligned} E_F(x, y, p, q, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, q) &= E_F(x, y, p, q, \nu) \\ &\quad - \nu^\beta\phi^\beta(x, y, q) + (\lambda^\rho - \nu^\rho)E_{\psi^\rho} + (\lambda^\beta - \nu^\beta)(E_{\phi^\beta} - \phi^\beta) \end{aligned} \tag{5.2}$$

and the definition of the class $\mathfrak{N}$ that the following lemma is true.

LEMMA 5.4. *If the arc C_0 with the multipliers λ^0, $\lambda^\beta(x)$, $\lambda^\rho(x)$ satisfies the condition* II_N, *there is a neighborhood $\mathfrak{D}_1$ of C_0 relative to the set $\mathfrak{D}$ and a positive constant a such that*

$$(5.3) \qquad E_F(x, y, p, q, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, q) \geqq a \left| E_{\psi^\rho}(x, y, p, q) \right|$$

whenever (x, y, p) *is in* $\mathcal{D}_1$, $\phi^\beta(x, y, p) = 0$ *if* β *is in* $\Delta(x)$ *and* (x, y, q) *is in* $\mathcal{D}$. *Moreover, the set* $\mathcal{D}_1$ *can be chosen so that for each closed subset* M_γ *of* $B(\gamma)$ *there is a positive constant* a_γ *such that*

$$(5.4) \qquad \begin{aligned} &E_F(x, y, p, q, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, q) \\ &\qquad \geqq a_\gamma \left| E_{\phi^\gamma}(x, y, p, q) - \phi^\gamma(x, y, q) \right| \quad (\gamma \textit{ not summed}) \end{aligned}$$

whenever (x, y, p) *is in* $\mathcal{D}_1$, $\phi^\beta(x, y, p) = 0$ *if* β *is in* $\Delta(x)$, *and* (x, y, q) *is in* $\mathcal{D}$, *and* x *is in* M_γ.

The proof of Theorem 5.1 may now be constructed by modifying the method used by Hestenes to prove a similar theorem [3, Theorem 4.3]. If a constant b and a neighborhood $\mathcal{D}_1$ having the properties described in Theorem 5.1 cannot be found, it follows from Lemmas 5.2 and 5.3 that there exists a sequence (x_k, y_k, p_k, q_k) such that (x_k, y_k, p_k) is in $\mathcal{D}$, (x_k, y_k, q_k) is in $\mathcal{D} - \mathcal{D}_0$, $\phi^\beta(x_k, y_k, p_k) = 0$ if β is in $\Delta(x_k)$,

$$(5.5) \qquad \begin{aligned} &E_F[x_k, y_k, p_k, q_k, \lambda(x_k)] - \lambda^\beta(x_k)\phi^\beta(x_k, y_k, q_k) \\ &\qquad \leqq k^{-1}[E_L(p_k, q_k) - \lambda^\beta(x_k)\bar{\phi}^\beta(x_k, y_k, q_k)], \end{aligned}$$

and for which (x_k, y_k, p_k) converges to a point $(x_0, y_0, \dot{y}_0)$ on C_0. Since $\lambda^\beta(x)$ is continuous, we may suppose that

$$\phi^\beta(x_k, y_k, p_k) = 0$$

if β is in $\Delta(x_0)$.

CASE I. q_k has a finite accumulation point q_0. Then $q_0^i \neq \dot{y}_0^i$ for all i since $(x_0, y_0, \dot{y}_0)$ is on C_0 and (x_0, y_0, q_0) is not in $\mathcal{D}_0$. It follows from (2.7) that

$$(5.6) \qquad E_L(p, q) \leqq 2L(q),$$

and since $\bar{\phi}^\beta(x, y, q) \leqq 1$, it follows from (5.5) and the condition II_N that

$$(5.7) \qquad \lim \left\{ E_F[(x_k, y_k, p_k, q_k, \lambda(x_k)] - \lambda^\beta(x_k)\phi^\beta(x_k, y_k, q_k) \right\} = 0.$$

Suppose for the moment that the set $\Gamma(x_0) - \Delta(x_0)$ is void. By Lemma 4.1 the equations

$$(5.8) \qquad \begin{aligned} &\phi^\delta(x, y, r) = 0 \text{ if } \delta \text{ is in } \Delta(x_0), \quad \phi^\alpha(x, y, r) = \phi^\alpha(x, y, p) + v^\alpha \\ &\text{if } \alpha \text{ is not in } \Gamma(x_0), \ \psi^\rho(x, y, r) = 0, \quad \psi^j(x, y, r) = \psi^j(x, y, p) + v^j \end{aligned}$$

have solutions $r^i(x, y, p, v)$ of class C' near $(x_0, y_0, \dot{y}_0, 0)$ such that $r^i(x_0, y_0, \dot{y}_0, 0) = \dot{y}_0^i$, (x, y, r) is in $\mathcal{D}$, and $\phi^\delta(x, y, r) = 0$ if δ is in $\Delta(x)$. If we set $r_k^i(v) = r^i(x_k, y_k, p_k, v)$, it follows from the condition II_N that for $|v|$ sufficiently small,

$$\liminf \left\{ E_F[x_k, y_k, r_k(v), q_k, \lambda(x_k)] - \lambda^\beta(x_k)\phi^\beta(x_k, y_k, q_k) \right\} \geqq 0,$$

and hence from (5.7)

$$\liminf \{E_F[x_k, y_k, r_k(v), q_k, \lambda(x_k)] - E_F[x_k, y_k, p_k, q_k, \lambda(x_k)]\} \geqq 0,$$

$$\begin{aligned} &F[x_0, y_0, \dot y_0, \lambda(x_0)] - \dot y_0^i F_{p^i}[x_0, y_0, \dot y_0, \lambda)(x_0)] \\ (5.9) \qquad &\quad - F[x_0, y_0, r_0(v), \lambda(x_0)] + r_0^i(v) F_{p^i}[x_0, y_0, r_0(v), \lambda(x_0)] \\ &\quad - q_0^i\{F_{p^i}[x_0, y_0, r_0(v), \lambda(x_0)] - F_{p^i}[x_0, y_0, \dot y_0, \lambda(x_0)]\} = 0. \end{aligned}$$

Define $v^\alpha(e) = \phi^\alpha[x_0, y_0, \dot y_0 + e(q_0 - \dot y_0)] - \phi^\alpha(x_0, y_0, \dot y_0)$ if α is not in $\Gamma(x_0)$, $v^j(e) = \psi^j[x_0, y_0, \dot y_0 + e(q_0 - \dot y_0)] - \psi^j(x_0, y_0, \dot y_0)$, $r^i(e) = r_0^i v(e)$,

$$\begin{aligned} Q(e) = F[x_0, y_0, \dot y_0, \lambda(x_0)] &- \dot y_0^i F_{p^i}[x_0, y_0, \dot y_0, \lambda(x_0)] \\ &- F[x_0, y_0, r(e), \lambda(x_0)] + r^i(e) F_{p^i}[x_0, y_0, r(e), \lambda(x_0)] \\ &- q_0^i\{F_{p^i}[x_0, y_0, r(e), \lambda(x_0)] - F_{p^i}[x_0, y_0, \dot y_0, \lambda(x_0)]\}. \end{aligned}$$

By (5.9) and the fact that $r^i(0) = \dot y_0^i$ we see that $Q(e) \geqq Q(0) = 0$ for all sufficiently small e. Hence $Q'(0) = 0$, or

$$(5.10) \qquad (\dot y_0^i - q_0^i) F_{p^i p^k}[x_0, y_0, \dot y_0, \lambda(x_0)] \dot r^i(0) = 0.$$

Now it follows from (5.3) that $\lim |E_{\psi^\rho}(x_k, y_k, p_k, . q_k)| = 0$. Since $\psi^\rho(x_k, y_k, p_k) = \psi^\rho(x_k, y_k, q_k) = 0$, it follows that

$$(5.11) \qquad (q_0^i - \dot y_0^i)\psi^\rho_{p^i}(x_0, y_0, \dot y_0) = 0.$$

Similarly, it follows from (5.4) and the fact that $\phi^\delta(x_k, y_k, p_k) = 0$ if δ is in $\Delta(x_0)$ that

$$(5.12) \qquad (q_0^i - \dot y_0^i)\phi^\delta_{p^i}(x_0, y_0, \dot y_0) = 0$$

if δ is in $\Delta(x_0)$. If we differentiate the equations satisfied by $r(e)$ and set $e = 0$, we find that $\dot r^i(0) = q_0^i - \dot y_0^i$ and hence

$$F_{p^i p^k}[x_0, y_0, \dot y_0, \lambda(x_0)] \dot r^i(0) \dot r^k(0) = 0$$

although the numbers $\dot r^i(0)$ do not all vanish and satisfy

$$\psi^\rho_{p^i}(x_0, y_0, \dot y_0) \dot r^i(0) = \phi^\delta_{p^i}(x_0, y_0, \dot y_0) \dot r^i(0) = 0$$

for all ρ and all δ in $\Delta(x_0)$, and this is a contradiction of condition III′.

If the set $\Gamma(x_0) - \Delta(x_0)$ is not void, it contains a unique index γ. We show first that $\phi^\gamma_{p^i}(x_0, y_0, \dot y_0)(q_0^i - \dot y_0^i) \geqq 0$. We observe that if we set $\nu^\rho = \lambda^\rho$, $\nu^\beta = \lambda^\beta$ when $\beta \neq \gamma$, it follows from the identity (5.2) and the condition II_N that if $\nu^\gamma = -a$, then the point (x_k, y_k, p_k, ν) is in N for k sufficiently large, and

$$\begin{aligned} &E_F[x_k, y_k, p_k, q_k, \lambda(x_k)] - \lambda^\beta(x_k)\phi^\beta(x_k, y_k, q_k) \\ (5.13) \qquad &\quad \geqq -[a + \lambda^\gamma(x_k)][\phi^\gamma(x_k, y_k, p_k) + (q_k^i - p_k^i)\phi^\gamma_{p^i}(x_k, y_k, p_k)]. \end{aligned}$$

Since (5.7) holds and $\lambda^\gamma(x_0) = \phi^\gamma(x_0, y_0, \dot{y}_0) = 0$, we find that

$$0 \geqq -a(q_0^i - \dot{y}_0^i)\phi^\gamma_{p^i}(x_0, y_0, \dot{y}_0).$$

Since $a > 0$, we see that $\phi^\gamma_{p^i}(x_0, y_0, \dot{y}_0)(q_0^i - \dot{y}_0^i) \geqq 0$.

Suppose $\phi^\gamma_{p^i}(x_0, y_0, \dot{y}_0)(q_0^i - \dot{y}_0^i) = 0$. Then we include the equation $\phi^\gamma(x, y, r) = 0$ with the equations (5.8) and the analysis in the preceding paragraphs is unaltered since $\phi^\gamma_{p^i}(x_0, y_0, \dot{y}_0)\dot{r}^i(0) = 0$. If $\phi^\gamma_{p^i}(x_0, y_0, \dot{y}_0)(q_0^i - \dot{y}_0^i) > 0$, we include with the equations (5.8) the equation $\phi^\gamma(x, y, r) = v^\gamma$. Then the solutions $r^i(x, y, p, v)$ will be such that (x, y, r) will be in $\mathcal{D}$ if $v^\gamma \geqq 0$. We set $v^\gamma = \phi^\gamma[x_0, y_0, \dot{y}_0 + e(q_0 - \dot{y}_0)]$, the v^α and v^j being defined as before. Then $v^\gamma(e) \geqq 0$ if e is positive and sufficiently small and so $Q(e) \geqq Q(0)$ for all nonnegative small e. Hence $Q'(0) \geqq 0$, or $(\dot{y}_0^i - q_0^i)F_{p^ip^k}[x_0, y_0, \dot{y}_0, \lambda(x_0)]\dot{r}^i(0) \geqq 0$. It follows just as above from (5.11) and (5.12) that $\dot{r}^i(0) = q_0^i - \dot{y}_0^i$ and that the last inequality cannot hold when C_0 satisfies condition III′.

This disposes of the case in which q_k has a finite accumulation point. Next consider:

Case II. $L(q_k) \to +\infty$. Then the sequences $q_k^i/L(q_k)$ converge to limits π^i and $\pi^i\pi^i = 1$. It follows from (5.5) and (5.6) that

$$\lim \{E_F[x_k, y_k, p_k, q_k, \lambda(x_k)] - \lambda^\beta(x_k)\phi^\beta(x_k, y_k, q_k)\}/L(q_k) = 0. \tag{5.14}$$

If the set $\Gamma(x_0) - \Delta(x_0)$ is void, we define the functions $r_k^i(v)$ as in the bounded case. Instead of (5.9) we now get

$$-\pi^i\{F_{p^i}[x_0, y_0, r_0(v), \lambda(x_0) - F_{p^i}[x_0, y_0, \dot{y}_0, \lambda(x_0)]\} \geqq 0.$$

We define $v^\alpha(e) = \phi^\alpha(x_0, y_0, \dot{y}_0 + e\pi) - \phi^\alpha(x_0, y_0, \dot{y}_0)$, $v^j(e) = \psi^j(x_0, y_0, \dot{y} + e\pi) - \psi^j(x_0, y_0, \dot{y}_0)$, $r^i(e) = r_0^i[v(e)]$, $Q(e) = -\pi^i\{F_{p^i}[x_0, y_0, r(e), \lambda(x_0)] - F_{p^i}[x_0, y_0, \dot{y}_0, \lambda(x_0)]\}$. Then $Q(e) \geqq Q(0) = 0$ for all sufficiently small e, $Q'(0) = 0$, and

$$-\pi^i F_{p^ip^k}[x_0, y_0, \dot{y}_0, \lambda(x_0)]\dot{r}^k(0) = 0. \tag{5.15}$$

Instead of (5.11) and (5.12) we infer from (5.14), (5.3), and (5.4) that $\psi^\rho_{p^i}(x_0, y_0, \dot{y}_0)\pi^i = \phi^\delta_{p^i}(x_0, y_0, \dot{y}_0)\pi^i = 0$ for all ρ and all δ in $\Delta(x_0)$, and this with (5.15) contradicts the nonsingularity hypothesis, since it is easy to see by differentiating the equations satisfied by $r(e)$ that $\dot{r}^i(0) = \pi^i$. If there is an index γ in $\Gamma(x_0) - \Delta(x_0)$, we see from (5.13) and (5.14) that $\phi^\gamma_{p^i}(x_0, y_0, \dot{y}_0)\pi^i \geqq 0$. The modifications in the analysis for the bounded case can be easily carried out since this inequality holds.

To prove the converse of Theorem 5.1 we utilize the following lemma which is similar to the Corollary of Hestenes [3, Corollary 1, p. 59].

Lemma 5.5. *There is a neighborhood $\mathcal{D}_0$ of C_0 relative to $\mathcal{D}$ and a positive constant b_1 such that*

$$E_{\phi^\beta}(x, y, p, q) - \phi^\beta(x, y, q) \leqq b_1 E_L(p, q) \tag{5.16}$$

whenever (x, y, p) *is in* $\mathcal{D}_0$ *and* (x, y, q) *is in* $\mathcal{D}$. *Moreover, if* $H(x, y, p)$ *is of class* C'' *on* $\mathcal{R}$ *and is such that there exists a positive constant* W *and a neighborhood* $\mathcal{J}_1$ *of* C_0 *in* (x, y)*-space such that* $WL(p) \geqq |H(x, y, p)|$ *whenever* (x, y) *is in* $\mathcal{J}_1$ *and* (x, y, p) *is in* $\mathcal{D}$, *then* $\mathcal{D}_0$ *and* b_1 *may be chosen so that*

$$(5.17) \qquad | E_H(x, y, p, q) | \leqq b_1 E_L(p, q)$$

whenever (x, y, p) *is in* $\mathcal{D}_0$ *and* (x, y, q) *is in* $\mathcal{D}$.

Since $L_{p^i p^k}\pi^i\pi^k > 0$ for all nonnull sets π^i, it follows from reasoning like that used in Lemma 5.3 that there is a neighborhood $\mathcal{R}_1$ of C_0 in (x, y, p)-space and a positive constant b_2 such that (5.17) holds with b_1 replaced by b_2 whenever (x, y, p) and (x, y, q) are in $\mathcal{R}_1$ for both H and ϕ^β. Since $\phi^\beta \geqq 0$ on $\mathcal{D}$, (5.16) also holds with b_1 replaced by b_2 whenever (x, y, p) and (x, y, q) are in $\mathcal{R}_1\mathcal{D}$. Choose a neighborhood $\mathcal{D}_0$ of C_0 relative to $\mathcal{D}$ such that the closure of $\mathcal{D}_0$ is in $\mathcal{R}_1$. Then there exists a positive constant b_3 such that $L(q) \leqq b_3 E_L(p, q)$ whenever (x, y, p) is in $\mathcal{D}_0$ and (x, y, q) is in $\mathcal{D} - \mathcal{R}_1\mathcal{D}$. Let b_4 be an upper bound for $|H(x, y, p) - p^i H_{p^i}(x, y, p)|$, $|H_{p^i}(x, y, p)|$, $|\phi^\beta(x, y, p) - p^i\phi^\beta_{p^i}(x, y, p)|$, $|\phi^\beta_{p^i}(x, y, p)|$ whenever (x, y, p) is in $\mathcal{D}_0$. If (x, y, p) is in $\mathcal{D}_0$ and (x, y, q) is in $\mathcal{D} - \mathcal{R}_1\mathcal{D}$, then

$$| E_H(x, y, p, q) - H(x, y, q) | \leqq b_4(n + 1)L(q) \leqq b_3 b_4(n + 1)E_L(p, q),$$

and the same inequality holds if H is replaced by ϕ^β. It follows that (5.16) is true if $b_1 = b_2 + b_3 b_4(n+1)$ whenever (x, y, p) is in $\mathcal{D}_0$ and (x, y, q) is in $\mathcal{D}$. Reduce $\mathcal{D}_0$ if necessary so that (x, y) is in $\mathcal{J}_1$ if (x, y, p) is in $\mathcal{D}_0$. Then it is clear that (5.17) holds if $b_1 = b_2 + Wb_3 + b_3 b_4(n+1)$ whenever (x, y, p) is in $\mathcal{D}_0$ and (x, y, q) is in $\mathcal{D}$.

We may now complete the proof of Theorem 5.1. Suppose a positive constant b and a neighborhood $\mathcal{R}_1$ of C_0 in (x, y, p) space exist such that the inequality (2.6) holds whenever (x, y, p) is in $\mathcal{R}_1\mathcal{D}$, (x, y, q) is in $\mathcal{D}$, and $\phi^\beta(x, y, p) = 0$ if β is in $\Delta(x)$. By Lemma 5.5 we may find a positive constant b_1 and reduce $\mathcal{D}_1$ if necessary so that $|E_{\psi^\rho}| \leqq b_1 E_L$, $|E_{\bar\phi^\beta}| \leqq b_1 E_L$, $E_{\phi^\beta} - \phi^\beta \leqq b_1 E_L$ if (x, y, p) is in $\mathcal{R}_1\mathcal{D}$ and (x, y, q) is in $\mathcal{D}$. Define N as the member of the class $\mathcal{N}$ determined by the neighborhood $\mathcal{R}_1$ and a positive number a such that $a < b$, $ab_1[t+1+m \max |\lambda^\beta(x)|] < b$. If (x, y, p, q, ν) is such that (x, y, p, ν) is in N, (x, y, p) and (x, y, q) are in $\mathcal{D}$, and $\phi^\beta(x, y, p) = 0$ if β is in $\Delta(x)$, then $\nu^\gamma(E_{\phi^\gamma} - \phi^\gamma) \geqq -ab_1 E_L$ if γ is in $\Gamma(x) - \Delta(x)$,

$$(\nu^\alpha - \lambda^\alpha)E_{\bar\phi^\alpha} \geqq -ab_1 \max |\lambda^\alpha(x)| E_L, \qquad (\nu^\alpha - \lambda^\alpha)\bar\phi^\alpha \geqq a\lambda^\alpha(x)\bar\phi^\alpha,$$

$(\nu^\rho - \lambda^\rho)E_{\psi^\rho} \geqq -ab_1 E_L$. Now if α is in $\Delta(x)$, then $\phi^\alpha(x, y, p) = \bar\phi^\alpha(x, y, p) = 0$, $\phi^\alpha_{p^i}(x, y, p) = \bar\phi^\alpha_{p^i}(x, y, p)$, while if α is not in $\Gamma(x)$, then $\nu^\alpha = \lambda^\alpha = 0$. Hence $(\nu^\alpha - \lambda^\alpha)(E_{\phi^\alpha} - \phi^\alpha) = (\nu^\alpha - \lambda^\alpha)(E_{\bar\phi^\alpha} - \bar\phi^\alpha)$ if α is restricted to the set complementary to $\Gamma(x) - \Delta(x)$. It follows from these relations and the identity (5.2), written in the form

$$E_F(x, y, p, q, \nu) - \nu^\beta\phi^\beta(x, y, q) = E_F(x, y, p, q, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, q) \\ + (\nu^\rho - \lambda^\rho)E_{\psi^\rho} + \nu^\gamma(E_{\phi^\gamma} - \phi^\gamma) + (\nu^\alpha - \lambda^\alpha)E_{\bar\phi^\alpha} - (\nu^\alpha - \lambda^\alpha)\bar\phi^\alpha,$$

that the left-hand side of (5.1) is not less than

$$b[E_L(p, q) - \lambda^\beta(x)\bar\phi^\beta(x, y, q)] - atb_1E_L(p, q) - ab_1E_L(p, q) \\ - a \max |\lambda^\alpha(x)| mb_1E_L(p, q) + a\lambda^\alpha(x)\bar\phi^\alpha(x, y, q) \geqq 0$$

by virtue of our choice of a. This completes the proof of Theorem 5.1 since III′ follows from Theorem 4.1.

6. **Convergent sequences of admissible arcs.** In the proof of Theorem 2.1 we shall need to be able to draw conclusions on the convergence of the derivatives $\dot y_r^i$ of a sequence of admissible arcs C_r which converge to C_0 uniformly in (x, y)-space. The particular results needed can be deduced from the following theorem.

THEOREM 6.1. *Let C_0 satisfy the hypotheses of Theorem 2.1. If C_r is a sequence of admissible arcs in $\mathfrak{D}$ which converge uniformly to C_0 in (x, y)-space such that* $\lim\sup J(C_r) \leqq J(C_0)$, *then it is true that* $\lim K(C_r, C_0) = 0$, *and that there is a subsequence C_{r_k} of the sequence C_r such that* $\lim \dot y_{r_k}^i = \dot y_0^i$ *almost uniformly on x^1x^2.*

Here $J(C)$ and $K(C, C_0)$ are defined as

$$J(C) = \int_{x^1}^{x^2} [F(x, y, \dot y, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, \dot y)]dx,$$

$$K(C, C_0) = \int_{x^1}^{x^2} [L(\dot y - \dot y_0) - 1]dx.$$

Consider the equations

$$\begin{gathered} \phi^\beta(x, y, P) = \phi^\beta(x, y_0, \dot y_0), \qquad \psi^\rho(x, y, P) = 0, \\ \psi^j(x, y, P) = \psi^j(x, y_0, \dot y_0) \end{gathered} \tag{6.1}$$

in which the functions ψ^j are those of Lemma 4.1. By Corollary 2 to Theorem 4.1 the functions involved in equations (6.1) are of class C' for x on each $\bar A_r$ and for (x, y, P) near C_0. Hence there exist solutions $P^i(x, y)$ of class C' if x is in $\bar A_r$ and y is near $y_0(x)$ such that $P^i[x, y_0(x)] = \dot y_0^i(x)$. Since $\dot y_0^i(x)$ is continuous, it follows that equations (6.1) are satisfied when x is an end point x_r of an interval A_r by $P^i(x_r-, y)$ as well as by $P^i(x_r+, y)$. Since the equations (6.1) admit only one solution near $\dot y_0^i(x)$ when y^i is sufficiently near $y_0^i(x)$, it follows that $P^i(x_r-, y) = P^i(x_r+, y)$, and hence that $P^i(x, y)$ is continuous on the whole interval x^1x^2. Differentiating this last equation we see also that $P_{y^k}^i(x, y)$ is continuous on x^1x^2.

With the help of the functions $P^i(x, y)$, $J(C)$ can be written as the sum

$$J(C) = J^*(C) + E^*(C), \tag{6.2}$$

in which

$$J^*(C) = \int_{x^1}^{x^2} [F(x, y, P, \lambda) + (\dot{y}^i - P^i)F_{p^i}(x, y, P, \lambda)]dx,$$

$$E^*(C) = \int_{x^1}^{x^2} [E_F(x, y, P, \dot{y}, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, \dot{y})]dx.$$

For any function $H(x, y, p)$ of class C'' on $\mathfrak{D}$ we define

$$H(C) = \int_{x^1}^{x^2} H(x, y, \dot{y})dx. \tag{6.3}$$

We shall be interested only in the case in which there is a positive constant c and a neighborhood $\mathfrak{D}_0$ of C_0 relative to $\mathfrak{D}$ such that

$$E_F(x, y, p, q, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, q) \geqq c \left| E_H(x, y, p, q) \right| \tag{6.4}$$

whenever (x, y, p) is in $\mathfrak{D}_0$, (x, y, q) is in $\mathfrak{D}$, and $\phi^\beta(x, y, p) = 0$ if β is in $\Delta(x)$. We shall say that H is E^*-dominated by F near C_0 on $\mathfrak{D}$ when this is true.

We shall also restrict the constant c and the neighborhood $\mathfrak{D}_0$ so that

$$E_F(x, y, p, q, \lambda) - \lambda^\beta(x)\phi^\beta(x, y, q) = cE_L(p - P, q - P) \tag{6.5}$$

whenever (x, y, p) is in $\mathfrak{D}_0$, (x, y, q) is in $\mathfrak{D}$, and $\phi^\beta(x, y, p) = 0$ if β is in $\Delta(x)$. The possibility of doing this follows from the E^*-dominance of L by F near C_0 on $\mathfrak{D}$ and from Lemma 5.5 when we observe that $L(p-P)$ satisfies the hypotheses imposed on $H(x, y, p)$ in that Lemma.

Theorem 6.1 will be a consequence of the following lemma, whose proof is identical with that of a similar lemma of Hestenes [4, Theorem 5.1].

Lemma 6.1. *Let C_0 satisfy the hypotheses of Theorem* 2.1. *Given a constant $\epsilon > 0$, there exists a constant $\eta > 0$ and a neighborhood $\mathfrak{J}$ of C_0 in (x, y)-space such that the inequality*

$$H(C) - H(C_0) < \epsilon$$

holds for every admissible arc C in $\mathfrak{J}$ which lies in $\mathfrak{D}$, satisfies the end conditions (2.1), *and is such that*

$$J(C) \leqq J(C_0) + \eta.$$

The first part of Theorem 6.1 will follow from Lemma 6.1 provided that $H(C) = K(C, C_0)$ has an integrand $H(x, y, p) = L(p - \dot{y}_0) - 1$ which is E^*-dominated by F near C_0 on $\mathfrak{D}$. Since this can be shown just as (6.5) was shown, we conclude that

$$\lim K(C_r, C_0) = 0.$$

To prove the second part of the theorem, we observe that Schwarz' inequality implies that

$$\left\{\int_{x^1}^{x^2} |\dot{y}_r - \dot{y}_0|\, dx\right\}^2 = \left\{\int_{x^1}^{x^2} [L(\dot{y}_r - \dot{y}_0) - 1]^{1/2}[L(\dot{y}_r - \dot{y}_0) + 1]^{1/2}dx\right\}^2$$

$$\leqq K(C_r, C_0)\int_{x^1}^{x^2} [L(\dot{y}_r - \dot{y}_0) + 1]dx$$

$$= K(C_r, C_0)[2(x^2 - x^1) + K(C_r, C_0)].$$

It follows that $\dot{y}_r^i$ converges in mean of order one to $\dot{y}_0^i$. The existence of a subsequence which converges almost everywhere to $\dot{y}_0^i$ (and hence almost uniformly to $\dot{y}_0^i$) is a well known consequence of convergence in mean [2, Theorem 23, p. 242 and Theorem 19, p. 239].

7. **The variation η_0^i.** Our proof of Theorem 2.1 is indirect. Suppose the hypotheses of Theorem 2.1 are fulfilled but that the conclusion is not. Then there is a sequence C_r: $y^i = y_r^i(x)$ of curves in $\mathfrak{D}$ which satisfy the end conditions (2.1), are different from C_0, and such that

$$\lim y_r^i(x) = y_0^i(x)$$

uniformly on x^1x^2, and yet $I(C_r) \leqq I(C_0)$. Since C_r and C_0 are in $\mathfrak{D}$ and $\lambda^0 \geqq 0$, $J(C_r) = \lambda^0 I(C_r) \leqq \lambda^0 I(C_0) = J(C_0)$. By virtue of Theorem 6.1 we may replace C_r by a subsequence for which $\dot{y}_r^i$ converges almost uniformly to $\dot{y}_0^i$ on x^1x^2 and hence for which $\lim k_r = 0$, in which

$$k_r \geqq 0, \qquad k_r^2 = K(C_r, C_0) - \int_{x^1}^{x^2} \lambda^\beta(x)\bar{\phi}^\beta(x, y_r, \dot{y}_r)dx.$$

This follows since $\lambda^\beta(x)\bar{\phi}^\beta(x, y_r, \dot{y}_r)$ converges boundedly to $\lambda^\beta(x)\bar{\phi}^\beta(x, y_0, \dot{y}_0) = 0$ almost everywhere on x^1x^2. Let us define

$$\eta_r^i(x) = (y_r^i - y_0^i)/k_r.$$

Then it is clear that

$$\int_{x^1}^{x^2} \{ |\dot{\eta}_r|^2/h_r(x)\} dx \leqq 1, \tag{7.1}$$

in which

$$h_r(x) = 1 + L(\dot{y}_r - \dot{y}_0) = k_r^2 |\dot{\eta}_r|^2/[L(\dot{y}_r - \dot{y}_0) - 1].$$

LEMMA 7.1. *The integrals of the functions $h_r(x)$ are absolutely continuous uniformly with respect to r.*

This follows since $h_r(x)$ differs by 2 from the integrand of $K(C_r, C_0)$ and $K(C_r, C_0)$ tends to zero.

LEMMA 7.2. *The functions* $\eta_r^i(x)$ *are absolutely continuous uniformly with respect to* r.

By Schwartz's inequality and (7.1),

$$\left|\int_M \dot{\eta}_r^i dx\right|^2 \leqq \int_M (|\dot{\eta}_r|^2/h_r)dx \int_M h_r dx \leqq \int_M h_r dx. \tag{7.2}$$

Thus the result follows from Lemma 7.1.

LEMMA 7.3. *The sequence of arcs* C_r *may be chosen so that there exists a function* $\eta_0^i(x)$ *satisfying the end conditions* (2.8) *and such that* $\lim \eta_r^i(x) = \eta_0^i(x)$ *uniformly on* x^1x^2. *Moreover,* $\eta_0^i(x)$ *is absolutely continuous and*

$$\int_{x^1}^{x^2} |\dot{\eta}_0|^2 dx \leqq 2.$$

By Lemma 7.2 the functions $\eta_r^i(x)$ are absolutely continuous uniformly in r. Since $\eta_r^i(x^1)=0$, the functions $\eta_r^i(x)$ are also uniformly bounded. By Ascoli's theorem [2, p. 122] subsequences can be found which converge uniformly to limits $\eta_0^i(x)$ and these limits are obviously absolutely continuous and such that $\eta_0^i(x^s)=0$.

The proof of the integrability of $|\dot{\eta}_0|^2$ to an integral bounded by two is identical with that of the corresponding assertion when there are no differential inequalities [6, pp. 528–529].

LEMMA 7.4. *If* $g(x)$ *is bounded and measurable, and if* $N_{ir}(x)$ *are continuous functions which converge uniformly to* $N_{i0}(x)$ *on* x^1x^2, *then*

$$\lim \int_M g(x)(\eta_r^i - \eta_0^i)dx = 0, \tag{7.3}$$

$$\lim \int_M N_{ir}(x)\dot{\eta}_r^i dx = \int_M N_{i0}(x)\dot{\eta}_0^i dx \tag{7.4}$$

for every measurable subset M *of* x^1x^2. *If* $|g(x)|^2$ *is integrable, then*

$$\lim \int_M g(x)(\dot{\eta}_r^i - \dot{\eta}_0^i)dx = 0 \tag{7.5}$$

for every measurable subset M *of* x^1x^2 *on which* $\dot{y}_r^i$ *converges uniformly.*

The relation (7.3) is obvious since $\eta_r^i(x)$ converges uniformly to $\eta_0^i(x)$. If $|g(x)|^2$ is integrable, there exists for each $\epsilon>0$ a bounded function $G(x)$ such that [8, p. 229]

$$\int_{x^1}^{x^2} | g(x) - G(x) |^2 dx < \epsilon.$$

If M is a set on which $\dot{y}_r^i$ converges uniformly to $\dot{y}_0^i$ on M, then $h_r(x)$ converges uniformly to 2 on M and so it follows from (7.1) and Lemma 7.3 that

$$\int_M |\dot{\eta}_r|^2 dx \leqq 3, \qquad \int_M |\dot{\eta}_0|^2 dx \leqq 2$$

if r is sufficiently large. From Schwarz' inequality we then have

$$\left| \int_M [g(x) - G(x)](\dot{\eta}_r^i - \dot{\eta}_0^i) dx \right|^2$$
$$\leqq 2 \int_M [g(x) - G(x)]^2 dx \int_M (|\dot{\eta}_r|^2 + |\dot{\eta}_0|^2) dx \leqq 10\epsilon.$$

To prove (7.5) it is therefore sufficient to prove it when $g(x)$ is bounded. The relation (7.5) when $g(x)$ is bounded and M is any measurable subset of x^1x^2 and the relation (7.4) follow from known results on Lebesgue-Stieltjes integrals [see 2, Theorem 28, p. 285; Theorem 21, p. 280; Corollary to Theorem 20, p. 280].

8. **Some auxiliary functions.** Let us recall from §6 the definition of the functions $P^i(x, y)$ such that $[x, y, P(x, y)]$ lies in $\mathcal{D}$ whenever (x, y) is in a neighborhood of C_0 and such that $P^i(x, y_0) = \dot{y}_0^i$. Define $p_r^i(x) = P^i[x, y_r(x)]$, $\pi_r^i(x) = [p_r^i(x) - p_0^i(x)]/k_r$, $\pi_0^i(x) = P_{y^v}^i[x, y_0(x)]\eta_0^v(x)$.

LEMMA 8.1. *The following relations hold*:

$$\phi^\beta(x, y_r, p_r) = \phi^\beta(x, y_0, \dot{y}_0), \qquad \psi^\rho(x, y_r, p_r) = 0, \tag{8.1}$$

$$\lim p_r^i(x) = p_0^i(x) = \dot{y}_0^i(x) \textit{ uniformly on } \overset{1}{x}\overset{2}{x}, \tag{8.2}$$

$$\lim \pi_r^i(x) = \pi_0^i(x) \textit{ uniformly on } \overset{1}{x}\overset{2}{x}. \tag{8.3}$$

Equations (8.1) are immediate consequences of the definitions. Equations (8.2) follow from the uniform convergence of $y_r^i(x)$ to $y_0^i(x)$ and the continuity of $P^i(x, y)$ for x on x^1x^2 and y near $y_0(x)$. To prove equations (8.3) observe that Taylor's formula yields

$$\pi_r^i(x) = B_{vr}^i(x)\eta_r^v(x),$$

in which

$$B_{vr}^i(x) = \int_0^1 P_{y^v}^i[x, y_0 + \theta(y_r - y_0)]d\theta.$$

Since y_r^i converges uniformly to y_0^i and since $P_{pv}^i(x, y)$ is continuous for x on x^1x^2 and y near $y_0(x)$, $B_{vr}^i(x)$ converges uniformly to $P_{y^v}^i[x, y_0(x)]$. By Lemma 7.3, $\eta_r^v(x)$ converges uniformly to $\eta_0^v(x)$. Hence equations (8.3) are true.

As a corollary of Lemmas 8.1 and 7.4 we have the following lemma.

LEMMA 8.2. *If $N_{ir}(x)$ are continuous functions which converge uniformly to $N_{i0}(x)$ on x^1x^2, and if $|g(x)|^2$ is integrable, then*

$$\lim \int_M N_{ir}(\dot{\eta}_r^i - \pi_r^i)dx = \int_M N_{i0}(\dot{\eta}_0^i - \pi_0^i)dx$$

for every measurable subset M of x^1x^2, and

$$\lim \int_M g(x)(\dot{\eta}_r^i - \pi_r^i)dx = \int_M g(x)(\dot{\eta}_0^i - \pi_0^i)dx$$

for every measurable subset M of x^1x^2 on which $\dot{y}_r^i$ converges uniformly.

LEMMA 8.3. *If $\phi(x, y, p)$ is any function of class C' near C_0, then*

$$\lim k_r^{-1}[\phi(x, y_r, p_r) - \phi(x, y_0, \dot{y}_0)] = \phi_{y^i}\eta_0^i + \phi_{p^i}\pi_0^i$$

uniformly on x^1x^2.

This follows directly from Taylor's theorem and Lemma 8.1.

If we replace, in Lemma 8.3, ϕ by ψ^ρ and then by ϕ^β and use equations (8.1) we immediately deduce the following lemma.

LEMMA 8.4. *The functions $\eta_0^i(x)$ satisfy with the auxiliary functions $\pi_0^i(x)$ the following equations:*

$$\psi^\rho_{y^i}\eta_0^i + \psi^\rho_{p^i}\pi_0^i = 0, \qquad \phi^\beta_{y^i}\eta_0^i + \phi^\beta_{p^i}\pi_0^i = 0.$$

9. **First order terms.** Let $H(x, y, p)$ be a function of class C' near C_0, and define

$$H(C, M) = \int_M H(x, y, \dot{y})dx,$$

$$H^*(C, M) = \int_M [H(x, y, P) + (\dot{y}^i - P^i)H_{p^i}(x, y, P)]dx,$$

$$E_H^*(C, M) = \int_M E_H(x, y, P, \dot{y})dx,$$

$$H_1(\eta, M) = \int_M (H_{y^i}\eta^i + H_{p^i}\dot{\eta}^i)dx.$$

It is clear that

$$H(C, M) = H^*(C, M) + E_H^*(C, M). \tag{9.1}$$

LEMMA 9.1. *If $H(x, y, p)$ is of class C' near C_0, then*

$$\lim k_r^{-1}[H^*(C_r, M) - H^*(C_0, M)] = H_1(\eta_0, M).$$

This follows at once from Lemmas 8.3 and 8.2.

Let us define

$$J_1(\eta) = \int_{x^1}^{x^2} (F_{y^i}\eta^i + F_{p^i}\dot{\eta}^i)dx.$$

By virtue of equation (2.3) and the fact that $\eta_r^i(x^s) = 0$, we have the following lemma.

LEMMA 9.2. *For each* $r = 0, 1, \cdots, J_1(\eta_r) = F_{p^i}\eta_r^i\big|_{x^1}^{x^2} = 0$.

LEMMA 9.3. *We have that*

$$\lim k_r^{-1}[J(C_r) - J(C_0)] = \lim k_r^{-1}[J^*(C_r) - J^*(C_0)] = \lim k_r^{-1}E^*(C_r) = 0.$$

By Lemmas 9.1 and 9.2 we have that

$$\lim k_r^{-1}[J^*(C_r) - J^*(C_0)] = J_1(\eta_0) = 0.$$

From equation (6.7) we conclude that

$$0 \geqq \limsup k_r^{-1}[J(C_r) - J(C_0)] = \limsup k_r^{-1}E^*(C_r). \tag{9.2}$$

However, $E^*(C_r) \geqq 0$ if r is so large that (x, y_r, p_r) is near enough to C_0 for (2.6) to hold, since both (x, y_r, p_r) and $(x, y_r, \dot{y}_r)$ are in $\mathfrak{D}$ and $\phi^\beta(x, y_r, p_r) = 0$ if β is in $\Delta(x)$. Hence the right-hand side of (9.2) cannot be negative and so the lemma is true.

LEMMA 9.4. *If* $H(x, y, p)$ *is of class* C' *near* C_0 *and is* E^**-dominated by* F *near* C_0 *on* $\mathfrak{D}$, *then*

$$\lim k_r^{-1}[H(C_r, M) - H(C_0, M)] = H_1(\eta_0, M).$$

If r is large enough, we may integrate the inequality (6.4) to see that $|E_H^*(C_r, M)| \leqq c^{-1}E^*(C_r)$ and so we see from Lemma 9.3 that

$$\lim k_r^{-1}E_H^*(C_r, M) = 0.$$

Lemma 9.4 is now an immediate consequence of Lemma 9.1 and equation (9.1).

10. **Admissibility of the variation η_0^i.** We have seen in Lemma 7.3 that the functions η_0^i are absolutely continuous, have integrable square derivatives, and satisfy the end conditions (2.8). We complete the proof of the admissibility of η_0^i in the following lemma.

LEMMA 10.1. *The variation* η_0^i *satisfies* (2.9) *for almost all* x *in* $B(\beta)$, (2.10) *for almost all* x *in* $A(\beta)$, *and* (2.11) *for almost all* x *on* x^1x^2.

It follows from Lemma 5.5 that the functions ψ^ρ and $\bar{\phi}^\beta$ satisfy the conditions imposed on H in Lemma 9.4. We thus infer that

$$0 = \lim k_r^{-1} \int_M [\psi^\rho(x, y_r, \dot{y}_r) - \psi^\rho(x, y_0, \dot{y}_0)]dx = \int_M (\psi^\rho_{y^i}\eta_0^i + \psi^\rho_{p^i}\dot{\eta}_0^i)dx$$

for every measurable subset M of x^1x^2. Hence (2.11) is satisfied for almost all x on x^1x^2. We also have that

$$(10.1) \qquad \begin{aligned} 0 \leqq \lim k_r^{-1} \int_M \phi^\beta(x, y_r, \dot{y}_r)dx &= \int_M (\bar{\phi}^\beta_{y^i}\eta_0^i + \bar{\phi}^\beta_{p^i}\dot{\eta}_0^i)dx, \\ 0 \leqq \int_M (\phi^\beta_{y^i}\eta_0^i &+ \phi^\beta_{p^i}\dot{\eta}_0^i)dx \end{aligned}$$

for every measurable subset M of $A(\beta)$. Hence (2.10) is satisfied for almost all x in $A(\beta)$. If M is a closed subset of $B(\beta)$, there is a number $\epsilon_\beta(M)$ such that $\lambda^\beta(x) \leqq -\epsilon_\beta(M) < 0$ on M. Hence it follows from the definition of k_r that

$$\begin{aligned} 0 \leqq k_r^{-1} \int_M \bar{\phi}^\beta(x, y_r, \dot{y}_r)dx &\leqq - k_r^{-1}\epsilon_\beta^{-1}(M) \int_{x^1}^{x^2} \lambda^\beta(x)\bar{\phi}^\beta(x, x_r, \dot{y}_r)dx \\ &\leqq k_r\epsilon_\beta^{-1}(M). \end{aligned}$$

Hence the inequality (10.1) is an equality for every closed subset M of $B(\beta)$. It follows that (2.9) is satisfied for almost all x in $B(\beta)$.

11. **Second order terms.** The second variation of $J^*(C)$ along C_0 is

$$J_2^*(\eta) = \int_{x^1}^{x^2} [2\omega(x, \eta, \pi_0) + 2(\dot{\eta}^i - \pi_0^i)\omega_{\pi^i}(x, \eta, \pi_0)]dx$$

in which 2ω is defined in (2.12). It is easy to see that

$$J_2^*(\eta) = J_2(\eta) - \int_{x^1}^{x^2} F_{p^ip^v}(\dot{\eta}^i - \pi_0^i)(\dot{\eta}^v - \pi_0^v)dx.$$

LEMMA 11.1. *We have that* $\lim k_r^{-2}[J^*(C_r) - J^*(C_0)] = (1/2)J_2^*(\eta_0)$.

This follows at once from Taylor's theorem and Lemmas 9.2 and 8.2.

LEMMA 11.2. *If $H(x, y, p)$ is a function of the form*

$$H = \theta[\psi^\rho(x, y, p)\psi^\rho(x, y, p) + \phi_0^\beta(x, y, p)\phi_0^\beta(x, y, p)],$$

in which θ is constant and $\phi_0^\beta = \lambda^\beta(x)\phi^\beta(x, y, p)$ (β not summed), then

$$\lim k_r^{-2} \int_M E_H(x, y_r, p_r, \dot{y}_r)dx = 0$$

for every measurable subset M of x^1x^2 on which $\dot{y}_r^i$ converges uniformly.

Since equations (8.1) hold, the lemma is equivalent to proving that

$$\lim k_r^{-2} \int_M \overset{\beta}{\phi}_0(x, y_r, \dot{y}_r)\overset{\beta}{\phi}_0(x, y_r, \dot{y}_r)dx = 0$$

for each β. Since $\dot{y}_r^i$ converges uniformly to $\dot{y}_0^i$ on M and (2.4) holds, there exists for each $\epsilon>0$ an index $R(\epsilon)$ such that

$$0 \geqq \overset{\beta}{\phi}_0(x, y_r, \dot{y}_r)\left[1 + \overset{\beta^2}{\phi}(x, y_r, \dot{y}_r)\right]^{1/2} \geqq -\epsilon$$

for each β if $r>R(\epsilon)$. By the definition of k_r, we have that if $r>R(\epsilon)$,

$$0 \leqq k_r^{-2} \int_M \overset{\beta}{\phi}_0(x, y_r, \dot{y}_r)\overset{\beta}{\phi}_0(x, y_r, \dot{y}_r)dx \leqq -k_r^{-2}\epsilon \int_M \overset{\beta}{\lambda}(x)\overset{\beta}{\bar{\phi}}(x, y_r, \dot{y}_r)dx \leqq \epsilon$$

for each β. Hence the lemma is true.

LEMMA 11.3. *We have that*

$$\liminf k_r^{-2}E^*(C_r) \geqq \frac{1}{2}\int_{x^1}^{x^2} F_{p^ip^v}(\dot{\eta}_0^i - \overset{i}{\pi}_0)(\dot{\eta}_0^v - \overset{v}{\pi}_0)dx.$$

In order to prove this result let M be a measurable subset of x^1x^2 on which $\dot{y}_r^i$ converges uniformly to $\dot{y}_0^i$. With the help of Lemma 11.2 and Corollary 3 to Theorem 4.1 the proof of the relation

$$\liminf k_r^{-2} \int_M E_F(x, y_r, p_r, \dot{y}_r, \lambda)dx \geqq \frac{1}{2}\int_M F_{p^ip^v}(\dot{\eta}_0^i - \overset{i}{\pi}_0)(\dot{\eta}_0^v - \overset{v}{\pi}_0)dx$$

can be made by the method of Hestenes [5, Lemma 10.1]. Since $\lambda^\beta(x)\phi^\beta(x, y_r, \dot{y}_r) \leqq 0$ and since (2.6) holds, we thus find that

$$\begin{aligned}\liminf k_r^{-2}E^*(C_r) &\geqq \liminf k_r^{-2}\int_M \left[E_F(x, y_r, p_r, \dot{y}_r, \lambda) - \overset{\beta}{\lambda}(x)\overset{\beta}{\phi}(x, y_r, \dot{y}_r)\right]dx \\ &\geqq \liminf k_r^{-2}\int_M E_F(x, y_r, p_r, \dot{y}_r, \lambda)dx \\ &\geqq \frac{1}{2}\int_M F_{p^ip^v}(\dot{\eta}_0^i - \overset{v}{\pi}_0)(\dot{\eta}_0^v - \overset{v}{\pi}_0)dx.\end{aligned}$$

Since $\dot{y}_r^i$ converges almost uniformly on x^1x^2, it follows from our choice of M and the integrability of $|\dot{\eta}_0|^2$ that this last inequality also holds when M is replaced by the whole interval x^1x^2. Hence the lemma is true.

12. **Completion of the proof of Theorem 2.1.** By virtue of the definition of C_r and equation (6.7), we have that

$$(12.1)\qquad 0 \geqq k_r^{-2}[J(C_r) - J(C_0)] = k_r^{-2}[J^*(C_r) - J^*(C_0) + E^*(C_r)].$$

By Lemmas 11.1, 11.3 and equation above Lemma 11.1, we have that

$0 \geqq J_2(\eta_0)$. Since η_0^i is an admissible variation by Lemmas 7.3 and 10.1, it follows from the hypothesis of Theorem 2.1 that $\eta_0^i(x)=0$. It then follows from (12.1) and the non-negativeness of $E^*(C_r)$ for sufficiently large r that $\lim k_r^{-2}E^*(C_r)=0$. By Lemma 5.5 there exists a positive number b^*, which we may assume to be less than one, such that $E_L(p_r, \dot y_r) \geqq b^* E_L(p_r - \dot y_0, \dot y_r - \dot y_0)$ for r sufficiently large. It then follows from (2.6) that for sufficiently large r,

$$
\begin{aligned}
k_r^{-2}E^*(C_r) &\geqq bb^*k_r^{-2}\int_{x^1}^{x^2}\left[E_L(p_r - \dot y_0, \dot y_r - \dot y_0) - \lambda^\beta(x)\Phi^\beta(x, y_r, \dot y_r)\right]dx \\
&\geqq bb^*k_r^{-2}\int_{x^1}^{x^2}\left[L(\dot y_r - \dot y_0) - \frac{1 + k_r^2\dot\eta_r^i\pi_r^i}{L(p_r - \dot y_0)} - \lambda^\beta(x)\Phi^\beta(x, y_r, \dot y_r)\right]dx \\
&\geqq bb^*k_r^{-2}\int_{x^1}^{x^2}\left[L(\dot y_r - \dot y_0) - 1 - \lambda^\beta(x)\Phi^\beta(x, y_r, \dot y_r) - \frac{k_r^2\dot\eta_r^i\pi_r^i}{L(p_r - \dot y_0)}\right]dx.
\end{aligned}
$$

By Lemmas 7.4 and 8.1 and the definition of k_r we find that

$$\liminf k_r^{-2}E^*(C_r) \geqq bb^* > 0$$

since $\eta_0^i \equiv 0$, and this is a contradiction from which we infer the truth of Theorem 2.1.

Bibliography

1. Gilbert A. Bliss, *Lectures in the calculus of variations*, University of Chicago Press, 1946.

2. Lawrence M. Graves, *Theory of functions of real variables*, McGraw-Hill, 1946.

3. M. R. Hestenes, *The Weierstrass E-function in the calculus of variations*, Trans. Amer. Math. Soc. vol. 60 (1946) pp. 51–71.

4. ———, *Theorem of Lindberg in the calculus of variations*, ibid. pp. 72–92.

5. ———, *Sufficient conditions for the isoperimetric problem of Bolza in the calculus of variations*, ibid. pp. 93–118.

6. ———, *An indirect sufficiency proof for the problem of Bolza in nonparametric form*, ibid. vol. 62 (1947) pp. 509–535.

7. William Karush, *Minima of functions of several variables with inequalities as side conditions*, University of Chicago Master's Thesis, 1939.

8. E. J. McShane, *Integration*, Princeton University Press, 1944.

9. ———, *Sufficient conditions for a weak relative minimum in the problem of Bolza*, Trans. Amer. Math. Soc. vol. 52 (1942) pp. 344–379.

10. W. T. Reid, *Isoperimetric problems of Bolza in nonparametric form*, Duke Math. J. vol. 5 (1939) pp. 675–691.

11. Frederick A. Valentine, *The problem of Lagrange with differential inequalities as added side conditions*, Contributions to the Calculus of Variations 1933–37, University of Chicago Press.

University of Illinois,
Chicago, Ill.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 80* (Reissue)

(Cowles Commission Discussion Paper: Mathematics 403)
November 7, 1950

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a paper will be filled by the Cowles Foundation within the limits of supply. References in publications to discussion papers (other than mere acknowledgement by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

LAGRANGE MULTIPLIERS REVISITED

Morton Slater

*CFDP 80, distributed in 1959, was an unchanged reissue of the Cowles Foundation, CCDP: Mathematics 403. It is being reprinted because of its historical interest and continuing demand.

November 1, 1980

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_14, © Springer Basel 2014

LAGRANGE MULTIPLIERS REVISITED

A Contribution to Non-Linear Programming

by Morton Slater

November 7, 1950

1. Introduction

The present paper was inspired by the work of Kuhn and Tucker [1] [1]. These authors transformed a certain class of constrained maximum problems into equivalent saddle value (minimax) problems.

Their work seems to hinge on the consideration of still a third type of problem. A very simple but illustrative form of this problem is the following: let $x \epsilon$ positive orthant of some finite dimensional Euclidean space, and let f and g be real valued functions of x with the property that whenever $f \geq 0$, then also $g \geq 0$; under what conditions can one then conclude that $\exists$ a non-negative constant u such that $uf \leq g$ for <u>all</u> $x \geq 0$?

Kuhn and Tucker showed that if f is concave and differentiable, if g is convex and differentiable, and if the set $\{x: f(x) \geq 0\}$ satisfies certain regularity restrictions, then there does indeed exist such a u.

Two directions for generalization are presented:

First of all, the Kuhn-Tucker argument rests heavily on the

---

1. Numerals in brackets refer to the list of references at the end.

differentiability of the functions, although they express the opinion that their theorems are true without this assumption. Is this the case?

Secondly, the inequality $uf \leqq g$ may be thought of as a relation between f and g. From this point of view, it would appear that a best possible theorem which concludes that $uf \leqq g$ would make assumptions about f only in relation to g, and vice versa.

In this paper it is shown how the second generalization may be partly achieved, and that even with this generalization, differentiability may be dispensed with entirely.

The next section will give a detailed account of the generalized theorem hinted at above. The last section will be devoted to an application of this theorem to transforming a class of constrained minimum problems into equivalent saddle value problems.

2. The Main Theorem

Throughout this section, x, u, and v will be points in the positive orthants of n, k, and h dimensional Euclidean spaces respectively.

Theorem 1. Let $f_1, \ldots, f_k$ and $g_1, \ldots, g_h$, be continuous real valued functions of $x = (x_1, \ldots, x_n)$ for $x \geqq 0$ with the following properties:

1° If at any point x all $f_i(x) \geqq 0$, then for that x some $g_j(x) \geqq 0$

2° $\exists x$ such that all $f_i(x) > 0$

3° If for some x^1 and x^2 $\exists A_i \geqq 0$ $i = 1, \ldots, k$, $B_j \geqq 0$ $j=1, \ldots, h$ and C of arbitrary sign such that

$$\Sigma A_i f_i(x^1) + C = \Sigma B_j g_j(x^1)$$

$$\text{and } \Sigma A_i f_i(x^2) + C = \Sigma B_j g_j(x^2)$$

then for all $x = \theta x^1 + (1-\theta) x^2 \quad 0 \leqq \theta \leqq 1$

$$\Sigma A_i f_i(x) + C \geqq \Sigma B_j g_j(x)$$

<u>Assertion</u>: $\exists u_1, \ldots, u_k \geq 0, \; v_1, \ldots, v_h \geq 0 \quad \Sigma v_j = 1$ such that for all $x \geq 0$

$$\Sigma u_i f_i (x) \leqq \Sigma v_j g_j (x)$$

Before proceeding with the sequence of lemmas necessary to prove theorem 1 some discussion of the hypotheses of that theorem is in order.

1° is clearly essential for the truth of the theorem.

2° corresponds to the condition of regularity of the constraint set in the Kuhn-Tucker treatment. It may possibly be weakened, but certainly not dispensed with altogether as the example

$$f (x) = -(x - 1)^2 \qquad g (x) = 1 - x$$

shows. Here $f (x)$ is concave, $g (x)$ convex, and 1° is satisfied. Nevertheless $uf \leq g$ for some $u \geq 0$ is impossible as is easily verified.

3° is of course the most controversial hypothesis of all. On the debit side is the fact that any non-negative linear combination of the f_i, say f, must, as a consequence of 3°, be quasi-concave (i.e. for all real β, $\{x : f (x) \geq \beta\}$ is convex), while any non-negative linear combination of the g_j, say g, must be quasi-convex (i.e. for all real α, $\{x : g (x) \leq \alpha\}$ is convex). This is on the debit side because it puts conditions on the f_i which are independent of the g_j.

Parenthetically we observe that a weaker version of 3° in which "$A_i \geq 0$" and "$B_j \geq 0$" are replaced by "$A_i > 0$" and "$B_j > 0$" which a priori appears to avoid the above difficulty, actually implies 3°. (This can easily be shown.)

On the credit side is that if all the f_i are concave and all the g_j convex (the Kuhn-Tucker case), then 3° is automatically satisfied. What this amounts to is that if a concave function interpolates a convex function at two points, then the concave function dominates the convex function in between.

Moreover, 3° is satisfied by still other functions. Let f and g be strictly increasing functions of a single real variable x, continuous for $x \geqq 0$ and having continuous first and second derivatives for $x > 0$. If in addition $f'(x) > 0$ for all $x > 0$, then by a theorem of M. M. Peixoto [1 , f and g satisfy 3° if and only if

$$g''(x) \geqq \frac{g'(x)}{f'(x)} f''(x) \quad \text{for all} \quad x > 0.$$

Using this theorem, the following examples were easily constructed:

(1) f convex, g convex

$f(x) = x^2 + x - 2 \qquad g(x) = x^2 - 1 \quad u = \frac{2}{3}$

(2) f concave, g concave

$f(x) = \sqrt{x} - 1 \qquad g(x) = \sqrt{x} + x - 2 \quad u = 3$

(3) Neither f nor g convex or concave

$f(x) = 2x - \cos x - (2\pi + 1)$,

$g(x) = x + \sin x - x \cos x - (\pi^2 + \pi)$

$u = \pi$

The main tool in the proof of theorem 1 is the generalized minimax theorem of von Neumann [1] and Kakutani [1], which we shall take as lemma 1.

Lemma 1: Let $\varphi(\zeta,\eta)$ be a continuous real valued function defined for $\xi \in K$ and $\eta \in L$ where K and L are arbitrary bounded closed convex sets of Euclidean spaces R^p and R^q respectively. If for every $\xi^o \in K$ and every real α, the set of all $\eta \in L$ such that $\varphi(\xi^o, \eta) \leq \alpha$ is convex, and if for every $\eta^o \in L$ and every real β, the set of all $\xi \in K$ such that $\varphi(\xi, \eta^o) \geqq \beta$ is convex, then $\exists(\xi^o, \eta^o)$ such that

$$\max_{\xi \in K} \varphi(\xi, \eta^o) = \min_{\eta \in L} \varphi(\xi^o, \eta)$$

Throughout section 2, the functions f_i and g_j will be the functions given in the hypotheses of theorem 1.

Lemma 2: Let $L = \{x : \text{for all } i \;\; f_i(x) \geq 0\}$.

Assertion: $\exists \bar{v} \geq 0$, $\Sigma \bar{v}_j = 1$ such that

$\Sigma \bar{v}_j \, g_j \, (x) \geq 0$ for all $x \in L$

Proof: Since the $f_i(x)$ are quasi concave, L is closed and convex.

Let $K = \{v : v \geq 0, \;\; \Sigma v_j = 1\}$,

$L_N = \{x : x \in L, \;\; \Sigma x_i \leq N\}$,

and $\varphi(v, x) = \Sigma v_j \, g_j \, (x)$

By hypothesis 3^o of theorem 1 we may apply lemma 1 to $\varphi(v, x)$ for $v \in K$ and $x \in L_N$. Hence $\exists(v^o, x^o)$ such that

$$\max_{v \in K} \varphi(v, x^o) = \min_{x \in L_N} \varphi(v^o, x)$$

Using the left hand side of the equality, we see that to prove $\varphi(v^o, x) \geq 0$ for all $x \in L_N$ we need only show that for any $x \in L_N$ $\exists v \in K$ such that $\varphi(v, x) \geq 0$. By hypothesis 1^o of theorem 1 this is clearly the case. Hence $\varphi(v^o, x) \geq 0$ for all $x \in L_N$.

The v^o thus obtained depends on N. Choose $N_n \uparrow \infty$ with n and an associated sequence $\{v^o(n)\}$ such that $\varphi(v^o(n), x) \geq 0$ for all $x \in L_{N_n}$. Since K is compact, some subsequence $\{v^o(n_i)\}$ converges to $\bar{v}$, say. Since $L_{N_{n_i}} \uparrow L$ with i and φ is a continuous function of v we have

$$\varphi(\bar{v}, x) \geq 0 \text{ for all } x \in L \qquad \text{q.e.d.}$$

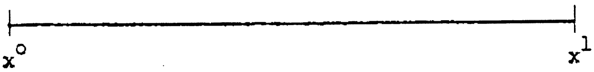
<u>Notation</u>: The function $\varphi(\bar{v}, x) = \Sigma \bar{v}_j g(x)$ defined in lemma 2 will be denoted by 'g(x)" for the remainder of section 2.

<u>Lemma 3</u>: Let $f_i(x^o) > 0$ for $i = 1, \ldots, k$ and $g(x^1) < 0$.

<u>Assertion</u>: $\exists i = i_o$ such that $f_{i_o}(x^1) < 0$ and

$$f_{i_o}(x^o)\, g(x^1) - f_{i_o}(x^1)\, g(x^o) \geq 0$$

<u>Proof</u>: We restrict our attention to the line segment joining x^o to x^1; a fortiori 1^o and 3^o are satisfied on the segment. We suppose (for definiteness) that it is oriented thus:



By 1^o, $g(x^o) \geq 0$. Let $\bar{x}$ be the right-most zero of $g(x)$. $(x^o \leq x < x^1)$. By 1^o and the continuity of the $f_i(x)$, $\exists i = i_o$ such that $f_{i_o}(\bar{x}) \leq 0$.

We show first that $f_{i_o}(x^1) < 0$. Suppose false; then $\exists\, \bar{\bar{x}}$ such that $\bar{x} < \bar{\bar{x}} \leq x^1$, $0 \leq f_{i_o}(\bar{\bar{x}}) < f_{i_o}(x^o)$, and $g(\bar{\bar{x}}) < 0$. Hence $\exists A > 0$ and $B < 0$ such that

$$Af_{i_o}(x^o) + B = g(x^o)$$

$$Af_{i_o}(\bar{\bar{x}}) + B = g(\bar{\bar{x}})$$

so that by 3^o, $$Af_{i_o}(x) + B \geqq g(x) \quad \text{for} \quad x^o \leqq x \leqq \bar{\bar{x}}.$$

In particular $$Af_{i_o}(\bar{x}) + B \geqq 0, \quad \text{so that}$$

$$f_{i_o}(x) \geqq -\frac{B}{A} > 0, \quad \text{a contradiction.}$$

Hence $f_{i_o}(x^1) < 0$, and for

$$C = \frac{g(x^o) - g(x^1)}{f_{i_o}(x^o) - f_{i_o}(x^1)} > 0, \qquad D = \frac{f_{i_o}(x^o)g(x^1) - g(x^o)f_{i_o}(x^1)}{f_{i_o}(x^o) - f_{i_o}(x^1)}$$

we have $$C\,f_{i_o}(x^o) + D = g(x^o)$$

$$C\,f_{i_o}(x^1) + D = g(x^1)$$

so that $$C\,f_{i_o}(x) + D \geqq g(x) \qquad \text{for} \quad x^o \leqq x \leqq x^1.$$

In particular

$$C\,f_{i_o}(\bar{x}) + D \geqq 0$$

so that $$D = -\,C\,f_{i_o}(\bar{x}) \geqq 0, \quad \text{and since}$$

$f_{i_o}(x^o) - f_{i_o}(x^1) > 0$, we have

$$f_{i_o}(x^o)g(x^1) - g(x^o)f_{i_o}(x^1) \geqq 0 \qquad \text{q.e.d.}$$

Lemma 4: Let $\varphi(x, u) = u_1 f_1(x) + \ldots + u_k f_k(x) - g(x)$

Assertion: $\exists$ a finite positive constant M such that for all $x \geq 0$ $\exists\, u \in L_M = \{u : u \geq 0, \ \Sigma u_i \leq M\}$ such that $\varphi(x, u) \leq 0$.

Proof: If $g(x^1) \geq 0$, $\varphi(x^1, 0) \leq 0$

If $g(x^1) < 0$, proceed as follows: let x^o be the point of the hypothesis 2^o at which all $f_i(x) > 0$. Apply lemma 3 to select i_o such that $f_{i_o}(x^1) < 0$ and

$$f_{i_o}(x^o)g(x^1) - f_{i_o}(x^1)g(x^o) \geqq 0$$

Let
$$u_i = 0 \text{ when } i \neq i_o$$
$$u_{i_o} = \frac{g(x^o)}{f_{i_o}(x^o)}$$

Then
$$\varphi(x, u) = \frac{g(x^o)}{f_{i_o}(x^o)} f_{i_o}(x^1) - g(x^1)$$
$$= -\frac{f_{i_o}(x^o)g(x^1) - g(x^o)f_{i_o}(x^1)}{f_{i_o}(x^o)} \leq 0$$

Thus if we take $M = \frac{g(x^o)}{A}$ where $A = \text{glb } f_i(x^o)$ the lemma is proved. The next lemma will not be needed for the proof of theorem 1, but it is convenient to prove it now.

Lemma 5: Let $f_1(x^1) = \ldots = f_\mu(x^1) = 0$, $f_{\mu+1}(x^1), \ldots, f_k(x^1) > 0$ and $g(x^1) = 0$.

Assertion: If $f_1(x^2), \ldots, f_\mu(x^2) \geq 0$, then $g(x^2) \geq 0$

Proof: As in lemma 3 we consider the functions on the segment joining x^1 to x^2:

$$\underset{x^1}{\vdash}\!\!\longrightarrow\!\!\underset{x^2}{\dashv}$$

If $g(x^2) < 0$, it must have a rightmost zero $\bar{x} < x^2$. We will show that this is impossible.

First of all, by quasi-concavity, $f_i(x) \geq 0$ for all x in the segment and $i \leq \mu$. If any $f_i(x^2) \geq 0$ for $i > \mu$, then again by quasi-concavity $f_i(x) \geq 0$ on the whole segment. Finally, suppose some $f_p(x^2) < 0$ for $p > \mu$. Then $\exists A > 0$ such that

$$A\ f_p(x^1) + B = g(x^1)$$

$$A\ f_p(x^2) + B = g(x^2) \quad \text{so that by } 3^\circ$$

$$A\ f_p(x) + B \geq g(x) \quad \text{for } x^1 \leq x \leq x^2.$$

In particular $\quad A\ f_p(\bar{x}) + B \geq 0 \quad$ so that $f_p(\bar{x}) > 0$.

Hence by continuity, $f_p(x) \geq 0$ in some right neighborhood of $\bar{x}$.

Combining all this information we see that $\exists\ \bar{\bar{x}} > \bar{x}$ such that all $f\ (\bar{\bar{x}}) \geq 0$. But by 1°, this implies $g(\bar{\bar{x}}) \geq 0$, a contradiction. q.e.d.

Corollary: If all $f_i(x^1) > 0$ and $g(x^1) = 0$, then $g(x) \geq 0$ for all x.

The proof of theorem 1 is now easy.

Proof: Choose M as in lemma 4 and $N > 0$ arbitrarily. Consider the function

$$\varphi(x, u) = u_1\ f_1(x) + \ldots + u_k\ f_k(x) - g(x)$$

over $\quad x \epsilon \quad K_N = \{x : x \geq 0 \text{ and } \Sigma x_i \leq N\}$

$\quad u \epsilon \quad L_M = \{u : u \geq 0 \text{ and } \Sigma u_i \leq M\}$

By lemma 1 and lemma 4 $\exists x^\circ \epsilon K_N, u^\circ \epsilon L_M$ such that

$$\underset{x \epsilon K_N}{\text{Max}}\ \varphi(x, u^\circ) = \underset{u \epsilon L_M}{\min}\ \varphi(x^\circ, u) \leq 0$$

Thus $\varphi(x, u^o) \leq 0$ for all $x \in K_N$.

The same kind of compactness argument as in lemma 2 is now used to complete the proof:

Choose $N_n \uparrow \infty$ and an associated sequence $\{u^o(n)\} \subset L_M$ (it is essential to observe that M is independent of N) such that

$$\varphi(x, u^o(n)) \leq 0 \quad \text{for all} \quad x \in K_{N_n}$$

By the compactness of L_M $\exists$ a subsequence $\{u^o(n_i)\}$ converging to $\tilde{u}^o$, say. Since $\varphi(x, u)$ is a continuous function of u and $K_{N_{n_i}} \uparrow P = \{x : x \geq 0\}$ we have $\varphi(x, u^o) \leq 0$ for all $x \geq 0$ q.e.d.

3. Applications

Definition: Let $g_1(x), \ldots, g_h(x)$ be any set of functions defined on a set K. A point $x^o \in K$ will be called a minimal point of $g_1, \ldots, g_h$ over K if for all $x \in K$ it is false that for all j $g_j(x) < g_j(x^o)$. A point x^o will be called an essential minimal point of $g_1, \ldots, g_h$ over K if it is minimal and if the deletion of any $g_j(x)$ will cause it to fail to be minimal. A point x^o will be called a strictly minimal point of $g_1, \ldots, g_h$ over K if for all x, if all $g_j(x) \leq g_j(x^o)$ then all $g_j(x) = g_j(x^o)$.

Theorem 2. Let $f_1(x), \ldots, f_k(x)$ and $g_1(x), \ldots, g_h(x)$ be real valued continuous functions which satisfy conditions 2^o and 3^o of theorem 1. Let K be the set $\{x : f_i(x) \geq 0 \quad i = 1, \ldots, k\}$. Let x^o be an essential minimal point of $g_1(x), \ldots, g_h(x)$ over K.

Assertion: x^o is a strictly minimal point of $g_1(x), \ldots, g_h(x)$ over K.

Proof: The functions $f_i(x)$, $i = 1, \ldots, k$ and $g_j(x) - g\ (x^0)$ $j = 1, \ldots, h$ satisfy all the hypotheses of theorem 1. Hence by lemma 2 $\exists\ v_1, \ldots, v_h \geq 0$, $\Sigma v_i = 1$ such that

$$\Sigma\ v_j\ g_j(x) \geq \Sigma\ v_j g_j(x^0) \quad \text{for all} \quad x \in K.$$

If for some j, $v_j = 0$, then that $g_j(x)$ may be deleted and x^0 will, by the above inequality, remain minimal. Hence all $v_j > 0$. But then, again by the above inequality, if for any $x \in K$ all $g_j(x) \leq g_j(x^0)$, then all $g_j(x) = g_j(x^0)$. q.e.d.

We now proceed to the last theorem, the equivalence of a constrained minimum problem and a certain saddle value problem.

Theorem 3: Let $f_1(x), \ldots, f_k(x)$ and $g_1(x), \ldots, g_h(x)$ be continuous real valued functions of x which satisfy conditions 2^0 and 3^0 of theorem 1. Let $K = \{x : f_i(x) \geq 0 \quad i = 1, \ldots, k\}$.

Assertion: x^0 is a minimal point for $g_1(x), \ldots, g_h(x)$ over K if and only if $\exists v_1^0, \ldots, v_h^0 \geq 0$, $\Sigma v_j^0 = 1$, and $u_1^0, \ldots, u_h^0 \geq 0$ such that the function

$$\varphi(x, u) = \Sigma v_j^0 g_j(x) - \Sigma u_i^0 f_i(x) \quad \text{satisfies}$$

$$\varphi(x^0, u) \leq \varphi(x^0, u\ ) \leq \varphi(x, u^0) \tag{1}$$

for all $x \geq 0$. and $u \geq 0$. In other words, $\varphi(x, u)$ has a saddle point at (x^0, u^0).

Proof: Suppose x^0 is a minimal point for the $g_j(x)$ over K. Then the functions $f_i(x)$ $i = 1, \ldots, k$ and $g_j(x) - g_j(x_o)$ $j = 1, \ldots, h$ satisfy 1^0, 2^0, 3^0 of theorem 1. By lemma 2 we choose $v^0 = v_1^0, \ldots, v_h^0)$

so that $\Sigma v_j^o(g_j(x) - g_j(x^o)) \geqq 0$ for all $x \in K$. Write $g(x) = \Sigma v_j^o g_j(x)$.

Now suppose $f_1(x^o) = \ldots = f_\mu(x^o) = 0$ and $f_{\mu+1}(x^o), \ldots, f_k(x^o) > 0$. By lemma 5 $g(x) \geq g(x^o)$ for all x such that $f_1(x), \ldots, f_\mu(x) \geq 0$. (In particular, if all $f_i(x^o) > 0$, then $g(x) \geq g(x^o)$ for all x). Hence by theorem 1 $\exists u_1^o, \ldots, u_\mu^o \geqq 0$ such that

$$u_1^o f_1(x) + \ldots + u_\mu^o f_\mu(x) \leqq g(x) - g(x^o) \quad \text{for all} \quad x \geqq 0.$$

Since $f_i(x^o) = 0$ for $i = 1, \ldots, \mu$ this may be rewritten as

$$g(x^o) - \Sigma_1^\mu u_i f_i(x^o) \leqq g(x^o) - \Sigma_1^\mu u_i^o f_i(x^o) \leqq g(x) - \Sigma_1^\mu u_i^o f_i(x)$$

for all $x \geqq 0$ and all $u_i \geq 0$, $i = 1, \ldots, \mu$. Now take $u_{\mu+1}^o = \ldots = u_k^o = 0$. Then

$$g(x^o) - \Sigma_1^k u_i f_i(x^o) \leq g(x^o) - \Sigma_1^k u_i^o f_i(x^o) \leq g(x) - \Sigma_1^k u_i^o f_i(x)$$

for all $x \geqq 0$ and all $u = (u\ , \ldots, u\ ) \geqq 0$ q.e.d.

Conversely, suppose (1) is satisfied for some v^o and u^o. From the first and last members of the inequality, we find, setting $u = 0$

$$\Sigma_1^h v_j{}^o(g_j(x) - g_j(x^o)) \geq \Sigma_1^k u_i{}^o f_i(x) \quad \text{for all} \quad x \geq 0$$

Hence if $x \in K$, $g_j(x) < g_j(x^o)$ for all j is impossible. This completes the proof.

References

Kakutani, S., 1. "A Generalization of Brouwer's Fixed Point Theorem," Duke Mathematical Journal, Volume 8 (1941), pp 457-459.

Kuhn, H. E., and A. W. Tucker 1. "Non-Linear Programming," RAND Publication LOGS 107 (August 28, 1950).

von Neumann, J., 1. "Uber ein okonomisches Gleichungs-system und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes," Erg. Math. Kolloqu., Volume 8 (1937), pp 73-83.

Peixoto, M. M., 1. "Generalized Convex Functions and Second Order Differential Inequalities," Bulletin of American Mathematical Society, Volume 55 No. 6 (June 1949).

3

THE KUHN-TUCKER THEOREM IN CONCAVE PROGRAMMING

HIROFUMI UZAWA

1. Introduction

In order to solve problems of constrained extrema, it is customary in the calculus to use the method of the Lagrangian multiplier. Let us, for example, consider a problem: maximize $f(x_1, \cdots, x_n)$ subject to the restrictions $g_k(x_1, \cdots, x_n) = 0$ $(k = 1, \cdots, m)$. First, formulate the so-called Lagrangian form

$$\varphi(x, y) = f(x_1, \cdots, x_n) + \sum_{k=1}^{m} y_k g_k(x_1, \cdots, x_n)$$

where unknown $y_1, \cdots, y_m$ are called the Lagrangian multipliers. Then solutions $x_1, \cdots, x_n$ are found among extreme points of $\varphi(x, y)$, with unrestricted x and y, which in turn are characterized as the solutions of

$$\varphi_{x_i}(x, y) = \frac{\partial f}{\partial x_i} + \sum_k y_k \frac{\partial g_k}{\partial x_i} = 0 \qquad (i = 1, \cdots, n) ,$$

$$\varphi_{y_k}(x, y) = g_k(x_1, \cdots, x_n) = 0 \qquad (k = 1, \cdots, m) .$$

This method, although not necessarily true without certain qualifications, has been found to be useful in many particular problems of constrained extrema.

The method is with a suitable modification applied to solve the programming problems also where we are concerned with maximizing a function $f(x_1, \cdots, x_n)$ subject to the restrictions $x_i \geqq 0$ $(i = 1, \cdots, n)$ and $g_k(x_1, \cdots, x_n) \geqq 0$ $(k = 1, \cdots, m)$. Kuhn and Tucker [2] first proved that under some qualifications, concave programming is reduced to finding a saddle-point of the Lagrangian form $\varphi(x, y)$. This Kuhn-Tucker Theorem was further elaborated by Arrow and Hurwicz [1] so that non-concave programming may be handled. In the present chapter we shall, under different qualifications, prove the Kuhn-Tucker Theorem for concave programming.

2. Maximum Problem and Saddle-Point Problem

Let $g(x) = \langle g_1(x), \cdots, g_m(x) \rangle$ (see fn. 1, p. 2) be an m-dimensional vector-valued function and $f(x)$ be a real-valued function, both defined for non-negative vectors $x = \langle x_1, \cdots, x_n \rangle$.

Consider the following

MAXIMUM PROBLEM. *Find a vector $\bar{x}$ that maximizes*

(1) $$f(x)$$

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*,
DOI 10.1007/978-3-0348-0439-4_15, © Springer Basel 2014

subject to the restrictions

$$(2) \qquad x \geqq 0 , \qquad g(x) \geqq 0 .$$

A vector x will be called *feasible* if it satisfies (2), and a feasible vector $\bar{x}$ maximizing $f(x)$ subject to (2) will be called an *optimum* vector of, or a *solution* to, the problem.

Associated with the Maximum Problem, the *Lagrangian form* is defined by

$$(3) \qquad \varphi(x, y) = f(x) + y \cdot g(x) ,$$

where[1]

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \geqq 0 \text{ and } y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \geqq 0 .$$

A pair of vectors $(\bar{x}, \bar{y})$ is called a *saddle-point* of $\varphi(x, y)$ in $x \geqq 0$, $y \geqq 0$, if

$$(4) \qquad \bar{x} \geqq 0, \qquad \bar{y} \geqq 0,$$

$$(5) \qquad \varphi(x, \bar{y}) \leqq \varphi(\bar{x}, \bar{y}) \leqq \varphi(\bar{x}, y) \text{ for all } x \geqq 0 \text{ and } y \geqq 0 ,$$

which may be written as follows:

$$(6) \qquad \varphi(\bar{x}, \bar{y}) = \min_{y \geqq 0} \max_{x \geqq 0} \varphi(x, y) = \max_{x \geqq 0} \min_{y \geqq 0} \varphi(x, y) .$$

SADDLE-POINT PROBLEM. *Find a saddle-point* $(\bar{x}, \bar{y})$ *of* $\varphi(x, y) = f(x) + y \cdot g(x)$.

3. Saddle-Point Implies the Optimality

We are interested in the reduction of a maximum problem to the saddle-point problem of the associated Lagrangian form. First, a proposition will be noted which is true without any qualification on f and g, whenever there exists a saddle-point.

THEOREM 1. *If* $(\bar{x}, \bar{y})$ *is a saddle-point of* $\varphi(x, y)$ *in* $x \geqq 0$, $y \geqq 0$, *then* $\bar{x}$ *is an optimum vector of the maximum problem.*

PROOF. Substituting (3) into (5), we have

$$(7) \qquad f(x) + \bar{y} \cdot g(x) \leqq f(\bar{x}) + \bar{y} \cdot g(\bar{x}) \leqq f(\bar{x}) + y \cdot g(\bar{x}) \quad \text{for all } x \geqq 0,\ y \geqq 0 .$$

---

[1] For two vectors $x = \langle x_1, \cdots, x_n \rangle$ and $u = \langle u_1, \cdots, u_n \rangle$, we shall, as usual, define

$$x \geqq u \text{ if } x_i \geqq u_i , \qquad (i = 1, \cdots, n),$$
$$x \geq u \text{ if } x \geqq u \text{ and } x \neq u ,$$
$$x > u \text{ if } x_i > u_i , \qquad (i = 1, \cdots, n),$$

and $x \cdot u$ stands for the inner product

$$x \cdot u = \sum_{i=1}^{n} x_i u_i .$$

Since the right-hand inequality holds for any $y \geqq 0$, it follows that $g(\bar{x})$ cannot have a negative component, and $\bar{y} \cdot g(\bar{x})$ must be zero:

$$g(\bar{x}) \geqq 0, \qquad \bar{y} \cdot g(\bar{x}) = 0 .$$

Thus the left-hand inequality of (7) may be written as

$$(8) \qquad f(x) + \bar{y} \cdot g(x) \leqq f(\bar{x}) \qquad \text{for all } x \geqq 0 .$$

Since, for any feasible vector x we have $\bar{y} \cdot g(x) \geqq 0$, it follows that $f(x) \leqq f(x) + \bar{y} \cdot g(x) \leqq f(\bar{x})$, which shows that $\bar{x}$ is optimum, q.e.d.

4. The Kuhn-Tucker Theorem

Now a question naturally arises whether, given an optimum vector $\bar{x}$, it is possible to find a vector $\bar{y}$ for which $(\bar{x}, \bar{y})$ is a saddle-point of $\varphi(x, y)$. This, of course, is not true in general, e.g., for convex programming (i.e., where the maximand is a convex function). The following simple example shows that it does not hold even for concave programming:

$$f(x) = x, \qquad g(x) = -x^2 .$$

Regarding concave programming, however, the reduction is shown to be possible provided f and g satisfy certain regularity conditions, e.g., the Kuhn-Tucker Constraint Qualification.[2] We shall give sufficient conditions which make the reduction possible.

THEOREM 2. *Suppose that $f(x)$ and $g(x)$ are concave functions on $x \geqq 0$, and $g(x)$ satisfies the following condition (due to M. Slater* [3]) :[3]

(9) *There exists an $x^0 \geqq 0$ such that $g(x^0) > 0$.*

Then a vector $\bar{x}$ is optimum if, and only if, there is a vector $\bar{y} \geqq 0$ such that $(\bar{x}, \bar{y})$ is a saddle-point of $\varphi(x, y)$.

PROOF. Let $\bar{x}$ be optimum. We shall, in the $(m+1)$-dimensional vector space, define A and B by

$$A = \left\{ \begin{pmatrix} z_0 \\ \\ z \end{pmatrix} ; \begin{pmatrix} z_0 \\ \\ z \end{pmatrix} \leqq \begin{pmatrix} f(x) \\ \\ g(x) \end{pmatrix} \quad \text{for some } x \geqq 0 \right\} ,$$

$$B = \left\{ \begin{pmatrix} z_0 \\ \\ z \end{pmatrix} ; \begin{pmatrix} z_0 \\ \\ z \end{pmatrix} > \begin{pmatrix} f(\bar{x}) \\ \\ 0 \end{pmatrix} \right\} .$$

Since $f(x)$ and $g(x)$ are concave, the set A is convex. Since $\bar{x}$ is opti-

[2] See Kuhn and Tucker [2], p. 483.

[3] A seemingly weaker condition: (9′) For any $u > 0$, there exists a vector $x \geqq 0$ such that $u \cdot g(x) > 0$ is due to S. Karlin. The condition, however, is equivalent to the Slater's condition (9). For the proof, see Chapter 5, pp. 109–10 of the present volume.

mum, A and B have no vector in common. Therefore, by the lemma on the separation of convex sets, there is a non-zero vector $\langle v_0, v\rangle \neq 0$ such that

$$(10) \qquad v_0 z_0 + v \cdot z \leqq v_0 u_0 + v \cdot u \qquad \text{for all } \begin{pmatrix} z_0 \\ \\ z \end{pmatrix} \in A, \begin{pmatrix} u_0 \\ \\ u \end{pmatrix} \in B .$$

By the definition of B, (10) implies $\langle v_0, v\rangle \geqq 0$. Since $\langle f(\bar{x}), 0\rangle$ is on the boundary of B, we also have, by the definition of A,

$$(11) \qquad v_0 f(x) + v \cdot g(x) \leqq v_0 f(\bar{x}) \qquad \text{for all } x \geqq 0 .$$

We have $v_0 > 0$. Otherwise, we have $v \geq 0$ and $v \cdot g(x) \leqq 0$ for all $x \geqq 0$, which contradicts (9).

Let $\bar{y} = v/v_0$. Then

$$(12) \qquad \bar{y} \geqq 0 ,$$

$$(13) \qquad f(x) + \bar{y} \cdot g(x) \leqq f(\bar{x}) \qquad \text{for all } x \geqq 0 .$$

Putting $x = \bar{x}$ in (13), we have $\bar{y} \cdot g(\bar{x}) \leqq 0$. On the other hand, we have

$$(14) \qquad g(\bar{x}) \geqq 0 .$$

Hence

$$(15) \qquad \bar{y} \cdot g(\bar{x}) = 0 .$$

Relations (13), (14), and (15) show that $(\bar{x}, \bar{y})$ is a saddle-point of $\varphi(x, y)$ in $x \geqq 0, y \geqq 0$, q.e.d.

5. A Modification of the Kuhn-Tucker Theorem

Slater's condition (9), however, excludes the case where part of the second half of restriction (2) is

$$h(x) \geqq 0 \qquad \text{and} \qquad -h(x) \geqq 0 ,$$

for linear $h(x)$. In order to make the reduction possible for such cases, we have to modify the Kuhn-Tucker Theorem.

Let sub-sets I and II of $\{1, \cdots, m\}$ be defined by

$$(16) \qquad \mathrm{I} = \{k ;\ g_k(x) = 0 \qquad \text{for all feasible } x\} ,$$

and

$$(17) \qquad \mathrm{II} = \{1, \cdots, m\} - \mathrm{I} .$$

We shall assume that

(18) $\qquad g_k(x)$ *is linear in* x, *for* $k \in \mathrm{I}$.

(19) $\qquad$ *For any* i, *there is a feasible vector* x^i *such that*

$$x_i^i > 0 .$$

Then we have as a modification of Theorem 2 the following:

THEOREM 3. *Suppose that* $f(x)$, $g(x)$ *are concave, and* $g(x)$ *satisfies* (18) *and* (19). *Then a vector* $\bar{x}$ *is optimum if, and only if, there is a vector* $\bar{y}$ *such that* $(\bar{x}, \bar{y})$ *is a saddle-point of* $\varphi(x, y)$ *in* $x \geqq 0$ *and* $y \geqq 0$ (II).[4]

PROOF. It is obvious that, if $(\bar{x}, \bar{y})$ is a saddle-point of $\varphi(x, y)$ in $x \geqq 0$ and $y \geqq 0$ (II), then $\bar{x}$ is optimum.

In order to prove the converse we may assume that

$$\frac{dg_k}{dx}, \; k \in I, \text{ are linear independent.} \tag{20}$$

Let $\bar{x}$ be optimum. We consider two sets A and B defined by

$$A = \left\{ \begin{pmatrix} z_0 \\ z \\ u \end{pmatrix}; \quad z_0 \leqq f(x),\ z_{\mathrm{I}} = g_{\mathrm{I}}(x),\ z_{\mathrm{II}} \leqq g_{\mathrm{II}}(x),\ u \leqq x, \quad \text{for some } x \right\},$$

$$B = \left\{ \begin{pmatrix} z_0 \\ z \\ 0 \end{pmatrix}; \quad z_0 > f(\bar{x}),\ z = 0 \text{ (I)},\ z > 0 \text{ (II)} \right\}.$$

Then A and B are convex, and have no point in common. Therefore, there is a vector $\langle v_0, v, w \rangle \neq 0$ such that

$$v_0 \geqq 0, \qquad v \geqq 0 \text{ (II)}, \qquad w \geqq 0, \tag{21}$$

and

$$v_0 f(x) + v \cdot g(x) + w \cdot x \leqq v_0 f(\bar{x}) \qquad \text{for all } x. \tag{22}$$

It now suffices to prove that $v_0 > 0$. If we had assumed that $v_0 = 0$, then

$$v \cdot g(x) + w \cdot x \leqq 0 \qquad \text{for all } x. \tag{23}$$

For any $k \in \mathrm{II}$, there is a feasible vector x^k such that $g_k(x^k) > 0$. Hence

$$v_k = 0 \qquad \text{for } k \in \mathrm{II}. \tag{24}$$

By (19) and (23),

$$w = 0. \tag{25}$$

Using (24) and (25), the inequality (23) may be written as follows:

$$v_{\mathrm{I}} \cdot g_{\mathrm{I}}(x) \leqq 0 \qquad \text{for all } x, \qquad \text{where } v_{\mathrm{I}} \neq 0. \tag{26}$$

Since we have assumed that $g_{\mathrm{I}}(x)$ is linear, (26) implies

[4] The notation $y \geqq 0$ (II) means $y_k \geqq 0$ for all $k \in \mathrm{II}$. The point $(\bar{x}, \bar{y})$ is said to be a saddle-point of $\varphi(x, y)$ in $x \geqq 0$ and $y \geqq 0$ (II) if $\bar{x} \geqq 0$, $\bar{y} \geqq 0$ (II) and (5) holds for any $x \geqq 0$ and $y \geqq 0$ (II).

$$v_{\mathrm{I}} \cdot \frac{\partial g_{\mathrm{I}}}{\partial x} = 0, \qquad \text{where } v_{\mathrm{I}} \neq 0\,, \tag{27}$$

which contradicts (20), q.e.d.

REFERENCES

[1] K. J. Arrow and L. Hurwicz, "Reduction of Constrained Maxima to Saddle-Point Problems," in J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley and Los Angeles: University of California Press, 1956, Vol. V, pp. 1-20.

[2] H. W. Kuhn and A. W. Tucker, "Nonlinear Programming," in J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley and Los Angeles: University of California Press, 1951, pp. 481-92.

[3] M. Slater, "Lagrange Multipliers Revisited: A Contribution to Non-Linear Programming," Cowles Commission Discussion Paper, Math. 403, November 1950.

CONTRIBUTIONS TO THE CALCULUS OF VARIATIONS

1933–1937

THESES SUBMITTED

TO

THE DEPARTMENT OF MATHEMATICS

OF THE

UNIVERSITY OF CHICAGO

THE UNIVERSITY OF CHICAGO PRESS
CHICAGO · ILLINOIS

Reprinted with the permission of The University of Chicago Press

JOHNSON REPRINT CORPORATION
111 Fifth Avenue, New York, N. Y. 10003

JOHNSON REPRINT COMPANY LIMITED
Berkeley Square House, London, W.1

DOI 10.1007/978-3-0348-0439-4_16, © Springer Basel 2014

First reprinting, 1965, Johnson Reprint Corporation

Printed in the United States of America

THE PROBLEM OF LAGRANGE WITH DIFFERENTIAL INEQUALITIES AS ADDED SIDE CONDITIONS

BY

FREDERICK ALBERT VALENTINE

TABLE OF CONTENTS

THE PROBLEM OF LAGRANGE WITH DIFFERENTIAL INEQUALITIES AS ADDED SIDE CONDITIONS

1. Introduction. The problem of the calculus of variations to be considered here consists in finding in a class of admissible arcs $y_i(x)$ joining two fixed points and satisfying a set of differential equations and inequalities of the form

$$\psi_\alpha(x, y, y') = 0, \qquad \phi_\beta(x, y, y') \geqq 0,$$

that one which minimizes the integral

$$I = \int_{x_1}^{x_2} f(x, y, y')dx.$$

The problem considered is for a space of n + 1 dimensions. A geometric illustration of a three-dimensional problem was suggested by Zermelo.[1] This problem required the finding of the shortest distance between two points on a surface subject to the condition that the direction of the tangent line at any point of the curve make an angle with the perpendicular which is never greater than a given constant. Bolza in a paper[2] issued in 1914 obtained a first necessary condition for a minimum and several corollaries. However he made no sufficiency proofs.

---

[1] E. Zermelo, Jahresberichte der Deutschen Mathematiker-vereinigung, B 11,(1902).

[2] O. Bolza, Über Variationsprobleme mit Ungleichungen als Nebenbedingungen, Mathematische Abhandlungen, H. A. Schwarz, (1914), seite 1.

An equivalent problem is introduced in section 2 of this paper by considering functions $z_\beta(x)$ such that the equations

$$\phi_\beta(x, y, y') = z_\beta'^2$$

hold. This equivalent problem yields a multiplier rule and necessary conditions analogous to those of Weierstrass and Clebsch. These are given in section 3. However as the equivalent problem may become singular, as it does for a composite arc, this method does not provide a complete treatment.

Two sufficiency proofs are made for a composite arc. Such an arc is one without corners composed of two subarcs such that all but one of the functions $\phi_\beta(x, y, y')$ are greater than zero on one subarc, whereas all the functions mentioned are greater than zero on the remaining subarc. An imbedding theorem and a necessary condition analogous to that of Mayer are proved in sections 4, 5 and 6. The first sufficiency proof is made in section 6 and is made with the assumption of normality on subintervals. The second sufficiency proof is made without the above assumption and in part depends upon a necessary condition analogous to that of Hestenes for the problem of Bolza.

It should be noted that although the sufficiency proofs are made for a composite arc, any other subcase which might arise could be handled in a similar manner. It is not due to the fact that other subcases present special difficulties that all of them are not treated, but rather to the fact that each subcase has to be handled separately. The case of the composite arc was treated since it represents a fair sample of the variety of cases which do exist. The treatment applied to the composite arcs will in general apply to all other cases. The singularity of the equivalent problem requires the separate treatment of the various subcases. The case considered affords a fairly complete treatment of the plane and 3-dimensional problems.

The following section describes the analytic setting of the problem and introduces the mechanism by means of which all the necessary conditions, save the analogue of the condition of Mayer, may be obtained.

2. Formulation of the problem. In the following pages the set $(x, y_1, \ldots, y_n, y_1', \ldots, y_n')$ will be denoted by (x, y, y'). The functions $y_i(x)$, $(i = 1, \ldots, n)$, defining the minimizing arc E_{12} and the functions

$$(2:1)\qquad \begin{array}{lll} f(x, y, y'), & \phi_\beta(x, y, y') & (\beta = 1, \ldots, m), \\ \psi_\alpha(x, y, y') & & (\alpha = m+1, \ldots, m+p < n) \end{array}$$

are required to satisfy the following hypotheses:

(1) The functions $y_i(x)$ are continuous on the interval x_1x_2 and have continuous derivatives on this interval except possibly at a finite number of corners.

(2) In a neighborhood N of the set of values (x, y, y') belonging to the arc E_{12} the functions (2:1) have continuous derivatives up to and including those of the third order.

(3) At every element (x, y, y') of the arc E_{12} the $n \times (m + p)$-dimensional matrix

$$\left\| \begin{array}{l} \psi_{\alpha y_i'}(x, y, y') \\ \phi_{\beta y_i'}(x, y, y') \end{array} \right\| \qquad \begin{array}{r} (i = 1, \ldots, n) \\ (\beta = 1, \ldots, m) \\ (\alpha = m+1, \ldots, m+p) \end{array}$$

has rank $m + p$.

Henceforth the subscripts i, β, and α shall have the ranges specified in hypothesis (3). Moreover a repeated index in a term will indicate summation with respect to that index, unless otherwise stated.

An admissible arc is one with the continuity properties (1) and one whose elements (x, y, y') lie in the region N specified in hypothesis (2).

The problem to be treated here consists in finding in the class of admissible arcs $y_i(x)$, joining two fixed points with coordinates (x_1, y_1) and (x_2, y_2), and satisfying the conditions $\phi_\beta \geqq 0$ and $\psi_\alpha = 0$, that one which minimizes the integral

$$J = \int_{x_1}^{x_2} f(x, y, y')dx. \tag{2:2}$$

A problem of Bolza with variable end-points which is equivalent to the problem just formulated may be obtained by setting

$$\phi_\beta = z_\beta'^2(x), \tag{2:3}$$

where the functions $z_\beta(x)$ will obviously have the same continuity properties as the functions $y_i(x)$ in the above problem. The equivalent problem is stated as follows:

To find in the class of admissible arcs

$$y_i = y_i(x), \qquad z_\beta = z_\beta(x)$$

satisfying the differential equations

$$\begin{aligned} \phi_\beta(x, y, y') - z_\beta'^2(x) &= 0, \\ \psi_\alpha(x, y, y') &= 0, \end{aligned}$$

and satisfying the end-conditions

$$\begin{aligned} x_1 &= a_1, & y(x_1) &= y_1, \\ x_2 &= a_2, & y(x_2) &= y_2, \end{aligned}$$

that one which minimizes the integral (2:2).

In view of hypotheses (1) to (3) it follows that the corresponding hypotheses for this equivalent problem are also satisfied. Moreover the above end-conditions are independent. Hence one may apply the theory of the problem of Bolza to this problem so as to obtain a number of necessary conditions. However as the equivalent problem may be singular it does not afford a complete attack. As will be seen later, other methods will be necessary in some cases to complete the theory. The equivalent problem is used primarily in sections 3 and 8.

3. First necessary conditions. From the theory for the problem of Bolza it follows that for every minimizing arc E_{12} there must exist constants c_i, d_β and a function

$$G = \lambda_0 f + \lambda_\alpha(x)\psi_\alpha + \lambda_\beta(x)(\phi_\beta - z_\beta'^2)$$

such that the equations

$$G_{y_i'} = \int_{x_1}^{x} G_{y_i}dx + c_i, \qquad \lambda_\beta z_\beta' = d_\beta$$

are satisfied along E_{12}. In the last m equations the repeated index β does not denote summation. Moreover from the transversality conditions in the problem of Bolza it follows that at the end points of E_{12} the expressions

$$(G - y_i'G_{y_i'} - z_\beta'G_{z_\beta'})dx_s + G_{y_i'}dy_{is} + e_s dx_s$$
$$+ b_s dy_{is} - 2\lambda_\beta z_\beta' dz_{\beta s} \qquad (s = 1, 2)$$

must be identically zero in dx_s, dy_{is} and $dz_{\beta s}$. As a consequence the m conditions

$$\lambda_\beta z_\beta' \Big|^{x_2} = \lambda_\beta z_\beta' \Big|^{x_1} = 0, \qquad (\beta \text{ not summed}),$$

must hold. Hence the functions $\lambda_\beta z_\beta'$ must be identically zero along the arc E_{12}. Therefore one obtains the

FIRST NECESSARY CONDITION I. For every minimizing arc E_{12} joining the fixed points 1 and 2, there must exist constants C_i and a function

$$F = \lambda_0 f + \lambda_\alpha(x)\psi_\alpha + \lambda_\beta(x)\phi_\beta \tag{3:1}$$

such that the equations

$$\begin{gathered} F_{y_i'} = \int_{x_1}^{x} F_{y_i}\,dx + C_i, \\ \phi_\beta(x, y, y') \geqq 0, \qquad \psi_\alpha(x, y, y') = 0 \end{gathered} \tag{3:2}$$

hold at every point of E_{12}. The constant λ_0 and the functions $\lambda_\alpha(x)$ and $\lambda_\beta(x)$ cannot vanish simultaneously at any point of E_{12}, and are continuous except possibly at values of x defining corners of E_{12}. Moreover the m functions

$$\lambda_\beta\phi_\beta \qquad (\beta \text{ not summed})$$

vanish at all points of E_{12}.

The following corollary may be obtained as an immediate consequence of the preceding sentence.

COROLLARY 3:1. If all the functions ϕ_β are greater than zero at every point of E_{12}, the minimizing arc is that one which minimizes the integral (2:2) in the class of admissible arcs satisfying the differential equations

$$\psi_\alpha(x, y, y') = 0.$$

For this case the function F in expression (3:2) reduces to

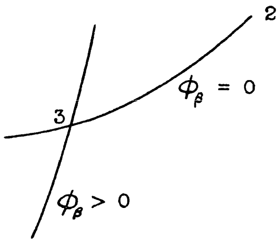
$$F_1 = \lambda_0 f + \lambda_\alpha\psi_\alpha .$$

Since this case is an ordinary problem of Lagrange, a fairly complete treatment of it is known.

The following corollaries and further necessary conditions, with the exception of the necessary condition of Mayer, are obtained for the general problem stated above. In the case of the Mayer condition the problem considered is the one in which all but one of the functions ϕ_β are greater than zero on the interval x_1x_2, whereas the remaining function is zero on certain subintervals of x_1x_2 and greater than zero on the remaining subintervals. It will be no restriction to label this last function by ϕ_1. For this problem the function F occurring in the expression (3:1) has the form

$$F = \lambda_0 f + \lambda_\alpha \psi_\alpha + \lambda_1 \phi_1 .$$

If a minimizing arc E_{12} is composed of two subarcs E_{13} and E_{32}, the functions ϕ_β being greater than zero on E_{13}, and zero on E_{32}, it follows that the functions λ_β are zero on E_{13}. *Hence the arc* E_{12} *is normal if the arc* E_{13} *is normal.* The arc E_{13} is defined by equations (3:2) in which F has been replaced by the function F_1 occurring in corollary (3:1).



The following corollaries are an immediate consequence of the first necessary condition.

COROLLARY 3:2. *On every subarc between corners of a minimizing arc* E_{12} *the differential equations and inequalities*

$$dF_{y_i'}/dx = F_{y_i}, \qquad \phi_\beta(x, y, y') \geqq 0, \qquad \psi_\alpha(x, y, y') = 0$$

must be satisfied, where F *is the function* (3:1).

COROLLARY 3:3. *At every corner of a minimizing arc* E_{12} *the conditions*

$$F_{y_i'}(x, y, y'(x-0), \lambda(x-0)) = F_{y_i'}(x, y, y'(x+0), \lambda(x+0))$$

must be satisfied.

The analogue of the Weierstrass necessary condition for the *equivalent problem* yields the result that at each element $(x, y, z, y', z', \lambda)$ of a minimizing arc which is normal, the inequality

$$\begin{aligned}\mathcal{E} = G(x, y, z, Y', Z', \lambda) - G(x, y, z, y' z', \lambda) \\ - (Y_i' - y_i')G_{y_i'} - (Z_\beta' - z_\beta')G_{z_\beta'} \qquad \geqq 0\end{aligned}$$

must be satisfied for every admissible set $(x, y, z, Y', Z') \neq (x, y, z, y', z')$, satisfying the equations

$$\phi_\beta(x, y, Y') - Z_\beta'^2 = 0 \qquad \psi_\alpha(x, y, Y') = 0.$$

Since the functions $\lambda_\beta z_\beta'$ are identically zero on E_{12} one obtains immediately the

SECOND NECESSARY CONDITION II. *At each element* (x, y, y', λ) *of a minimizing arc* E_{12} *which is normal the inequality*

$$E(x, y, y', Y', \lambda_\alpha, \lambda_\beta) - \lambda_\beta\phi_\beta(x, y, Y') \geqq 0 \tag{3:3}$$

must hold for all sets $(x, y, Y') \neq (x, y, y')$, *and satisfying the differential equations and inequalities*

$$\phi_\beta(x, y, Y') \geqq 0, \qquad \psi_\alpha(x, y, Y') = 0,$$

where $E(x, y, y', Y', \lambda_\alpha, \lambda_\beta)$ *is the function*

$$F(x, y, Y', \lambda_\alpha, \lambda_\beta) - F(x, y, y', \lambda_\alpha, \lambda_\beta) - (Y_i' - y_i')F_{y_i'}.$$

In a similar manner the analogue of the Clebsch condition for the equivalent problem gives the following condition.

THIRD NECESSARY CONDITION III. <u>At every element</u> (x, y, y', λ) <u>of a minimizing arc</u> E_{12} <u>which is normal the inequality</u>

$$(3{:}4) \qquad F_{y_i'y_k'}\pi_i\pi_k - 2\lambda_\beta \varkappa_\beta^2 \geqq 0$$

<u>must be satisfied for every set</u> $[\pi_1, \ldots, \pi_n, \varkappa_1, \ldots, \varkappa_m] \neq [0, \ldots, 0, 0, \ldots, 0]$ <u>and satisfying the equations</u>

$$\psi_{\alpha y_i'}\pi_i = 0, \qquad \phi_{\beta y_i'}\pi_i - 2z_\beta'\varkappa_\beta = 0.$$

At any point of E_{12} where any one of the functions z_β', say z_1', is zero, choose $[\pi_i] = [0]$, and all the $\varkappa_\beta$ except $\varkappa_1$ zero. Hence at such a point of E_{12} the condition $\lambda_1 \leqq 0$ must hold. Where $z_1' \neq 0$, it follows from the first necessary condition that $\lambda_1 = 0$. Hence one obtains the following corollary.

COROLLARY 3:4. <u>At every element</u> (x, y, y') <u>of a minimizing arc</u> E_{12} <u>it is necessary that the inequalities</u>

$$\lambda_\beta \leqq 0$$

<u>be satisfied</u>.

As a consequence of the paragraph preceding the above corollary the condition III yields the following result.

COROLLARY 3:5. <u>At every element of a minimizing arc</u> E_{12} <u>which is normal the inequality</u>

$$(3{:}5) \qquad F_{y_i'y_k'}\pi_i\pi_k \geqq 0$$

<u>must be satisfied for every set</u> $[\pi_1, \ldots, \pi_n] \neq [0, \ldots, 0]$ <u>and satisfying the equations</u>

$$\psi_{\alpha y_i'}\pi_i = 0, \qquad \phi_{\beta y_i'}\pi_i = 0.$$

The equivalent problem was used to obtain the preceding necessary conditions. In the following sections 4 to 7 special methods are used to obtain the necessary condition of Mayer and a sufficiency proof.

4. Imbedding theorem. In the following section an imbedding theorem is established for the case in which all but one of the functions ϕ_β are greater than zero on E_{12}. The remaining function, which will be denoted by ϕ_1, is to be greater than zero on one subarc E_{13} of E_{12} and zero on the remaining subarc E_{32}. Let R_1 and R_2 represent the determinants

$$(4:1)\qquad R_1 = \begin{vmatrix} F_{y_i'y_k'} & \psi_{\alpha y_i'} \\ \psi_{\delta y_k'} & 0 \end{vmatrix}, \qquad R_2 = \begin{vmatrix} F_{y_i'y_k'} & \psi_{\alpha y_i'} & \phi_{1y_i'} \\ \psi_{\delta y_k'} & 0 & 0 \\ \phi_{1y_k'} & 0 & 0 \end{vmatrix}$$

where (α, δ = m + 1, ..., m + p) and (i, k = 1, ..., n). Let F_1 and F_2 denote the functions

$$(4:2)\qquad \begin{aligned} F_1 &= \lambda_0 f + \lambda_\alpha \psi_\alpha, \\ F_2 &= \lambda_0 f + \lambda_\alpha \psi_\alpha + \lambda_1 \phi_1. \end{aligned}$$

The symbol F_1 represents the function F which occurs in the first necessary condition for the problem in which E_{13} is an extremal; similarly F_2 denotes the corresponding function for the problem in which E_{32} is an extremal. The class of arcs defined by equations (3:2) with the function F replaced by F_1 will be denoted by A; whereas the class of arcs which are defined by equations (3:2) with F replaced by F_2, and along which the equation $\phi_1 = 0$ is satisfied, will be represented by B. A composite arc is defined to be one composed of two subarcs, one subarc belonging to A, and the second belonging to B, such that the functions $y_i(x)$ defining

<u>the entire arc and their derivatives</u> $y_i'(x)$ <u>are continuous</u>. From the first necessary condition it follows that the multipliers $\lambda_\alpha(x)$ and $\lambda_1(x)$ are continuous on a composite arc. With this definition in mind, one may prove the following theorem.

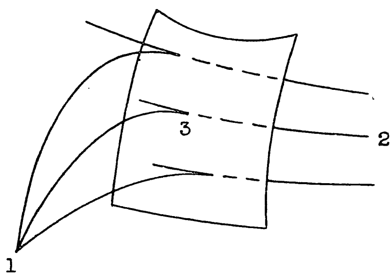
IMBEDDING THEOREM. <u>Consider a composite arc</u> $E_{12} = E_{13} + E_{32}$ <u>satisfying the conditions that</u> R_1 <u>and</u> R_2 <u>be different from zero on</u> E_{13} <u>and</u> E_{32} <u>respectively, and that</u> $\Phi_1' \neq 0$ <u>on</u> E_{13} <u>at</u> 3. <u>Such an arc is a member of an</u> n <u>parameter family of composite arcs defined by the equations</u>

$$\left.\begin{aligned} y_i &= y_i(x, a_1, \ldots, a_n) \\ \lambda_\alpha &= \lambda_\alpha(x, a_1, \ldots, a_n) \\ \lambda_1 &= \lambda_1(x, a_1, \ldots, a_n) \end{aligned}\right. \qquad [x_1 \leqq x \leqq x_3(a)],$$

$$\left.\begin{aligned} y_i &= Y_i(x, a_1, \ldots, a_n) \\ \lambda_\alpha &= \Lambda_\alpha(x, a_1, \ldots, a_n) \\ \lambda_1 &= \Lambda_1(x, a_1, \ldots, a_n) \end{aligned}\right. \qquad [x_3(a) \leqq x \leqq x_2]$$

<u>for the special values</u> a_0 <u>of the parameters</u>.

Proof: Henceforth the letter a will stand for the set $(a_1, \ldots, a_n)$. Consider a composite extremal arc $E_{12} = E_{13} + E_{32}$. Since $R_1 \neq 0$ on the arc E_{13} it follows from the theory of the problem of Lagrange that E_{13} can be imbedded in an n-parameter family of extremals belonging to A, passing through the point 1 or through a point 0 on the extension of E_{13}. Similarly it



is known that if $R_2 \neq 0$ on E_{32}, then E_{32} may be imbedded in a

2n-parameter family of extremals of class[1] B. Denote the n-parameter family of extremals passing through 1 and containing E_{13} by

$$(4:3) \qquad y_i = y_i(x, a), \qquad \lambda_\alpha = \lambda_\alpha(x, a), \qquad \lambda_1 = 0,$$

and the 2n-parameter family containing E_{32} by

$$y_i = Y_i(x, c), \qquad \lambda_\alpha = \Lambda_\alpha(x, c), \qquad \lambda_1 = \Lambda_1(x, c),$$

where $(c = c_1, \ldots, c_{2n})$, and E_{32} is defined for the value of the parameters $c = c_0$. It is also known that at the special values (x_3, c_0) the condition

$$D = \begin{vmatrix} Y_{ic_k} \\ u_{ic_k} \end{vmatrix} \neq 0 \qquad (k = 1, \ldots, 2n)$$

holds, where

$$u_i = F_{y_i'}(x, Y, Y', \Lambda).$$

The necessary conditions

$$(4:4) \qquad \begin{aligned} & y_i(x_4, a) - Y_i(x_4, c) = 0, \\ & F_{1y_i'}[x_4, y(x_4,a), y'(x_4,a), \lambda(x_4,a)] \\ & \quad - F_{2y_i'}[x_4, Y(x_4,c), Y'(x_4,c), \Lambda(x_4,c)] = 0, \\ & \Phi_1[x_4, y(x_4,a), y'(x_4,a)] = 0 \end{aligned}$$

must hold at the point 3, that is for the values $x_4 = x_3$, $a = a_0$, and $c = c_0$. The functional determinant of these equations (4:4) with respect to x_4 and c is

$$(4:5) \qquad \begin{vmatrix} 0 & -Y_{ic_k} \\ F_{1y_i'}' - F_{2y_i'}' & -u_{ic_k} \\ \Phi_1' & 0 \end{vmatrix} = -\Phi_1' D.$$

---

[1] G. A. Bliss, Problem of Lagrange in the calculus of variations, American Journal of Mathematics, vol. 52 (1930), p. 687.

The above determinant will be different from zero at 3 if the function ϕ_1' at $x = x_3$ is different from zero. In the theorem it was assumed that $\phi_1' \neq 0$ holds at $x = x_3$. From the theory of implicit functions it follows that one may solve equations (4:4) for x_4 and c as functions of a. Denote these solutions by

$$x_4 = x_4(a), \qquad c = c(a). \tag{4:6}$$

There remains to show that for values of a sufficiently close to a_0, the subarcs defined by the equations

$$\begin{aligned} y_i &= y_i(x,\ a) && [x_1 \leqq x \leqq x_4(a)], \\ y_i &= Y_i(x,\ c(a)) && [x_4(a) \leqq x \leqq x_2], \end{aligned} \tag{4:7}$$

are tangent along the n-space defined by the first n equations of (4:3) and by (4:6). To show this consider the equations

$$\begin{aligned} u_i &= F_{2y_i'}(x,\ Y,\ Y',\ \Lambda), \\ 0 &= \psi_\alpha(x,\ Y,\ Y'), \\ 0 &= \phi_1(x,\ Y,\ Y'). \end{aligned} \tag{4:8}$$

Since R_2 as defined in expression (4:1) is different from zero, equations (4:8) have a unique solution for Y', Λ, Λ_1. Moreover since $Y = y$ is a solution of (4:8) with

$$u_i = F_{1y_i'}[x_4,\ y(x_4,a),\ y'(x_4,a),\ \lambda(x_4,a)]$$

it is plain that $Y' = y'$, and $\Lambda = \lambda$ at $x = x_4(a)$. Hence the arcs defined by equations (4:7) are composite arcs. Thus there exists an n-parameter family of composite arcs imbedding the composite arc $E_{12} = E_{13} + E_{32}$.

5. The Mayer condition for a composite minimizing arc. In developing this condition a geometric argument will be given first. In section 8 another proof is given by means of the accessory minimum problem associated with the second variation.

Consider an n-parameter family of composite extremals through the point 1 defined by the equations

$$
\begin{aligned}
&y_i = y_i(x, a) && [x_1 \leqq x \leqq x_4(a)], \\
(5{:}1) \qquad &y_i = Y_i(x, a) && [x_4(a) \leqq x \leqq x_2], \\
&y_i(x_3, a) = Y_i(x_3, a). &&
\end{aligned}
$$

Also consider a one-parameter family of these arcs having an envelope D obtained by letting $a = a(t)$. Let the equation of D be

$$x = x(t), \qquad y_i = g_i(t).$$

The fact that D is tangent at each of its points to an extremal

$$y_i = Y_i[x, a(t)]$$

may be expressed by the equations

$$x'(t) = k, \qquad Y_i'x' + Y_{ia_j}a_j' = g_{it} = kY_i'.$$

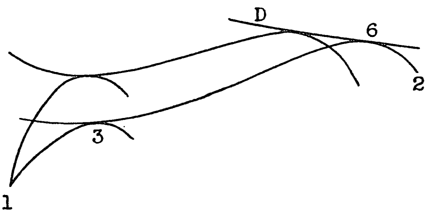
These equations have the solution $(a_j') \neq (0)$ if and only if the determinant

$$\Delta(x, a) \equiv |Y_{ia_j}|$$

is identically zero in t, when x and a are replaced by $x(t)$ and $a(t)$.

DEFINITION. A value x_6 is said to define a point conjugate to the point 1 if it is a root of the determinant $\Delta(x, a)$ belonging to an n-parameter family of composite arcs (5:1).

To prove that if Δ vanishes at x_6, the equations (5:1) do have an envelope to which E_{32} is tangent at $x = x_6$, let E_{12} be contained in an n-parameter family of composite extremals with equations of the form (5:1) for values of the parameters $a = a_0$. All the extremals satisfy the equations



$$y_i'(x_4, a) = Y_i'(x_4, a)$$

where $x_4(a)$ is defined by the first of equations (4:6). Let x_6 define a conjugate point to 3 on E_{32}. We assume for purposes of the proof that $\Delta_x(x_6, a_0) \neq 0$. Hence at least some one n - 1 rowed minor of the determinant $|Y_{ia_j}|$ is different from zero. Suppose for example that the determinant

$$|Y_{ka_t}| \qquad (k, t = 1, \ldots, n - 1)$$

is different from zero. Then the first n differential equations of the set

$$\begin{aligned} \Delta_x dx + \Delta_{a_j} da_j &= 0 \\ Y_{ia_j} da_j &= 0 \end{aligned} \qquad (j = 1, \ldots, n),$$

can be solved for dx/da_n, da_t/da_n. They determine uniquely a solution

$$x = x_5(a_n), \qquad a_t = a_t(a_n) \qquad (t = 1, \ldots, n - 1)$$

through the initial point (x_6, a_0). The determinant $\Delta(x, a)$ is identically zero on this solution since it vanishes at (x_6, a_0) and since its total derivative with respect to a_n is identically zero. Hence the last equation is also satisfied. A similar argument can be made for any other n - 1 rowed minor which may be different from zero. One thus determines

$$x = x_5(t), \qquad a_s = a_s(t) \qquad (s = 1, \ldots, n),$$

t being a properly selected one of the parameters a.

On the one-parameter family of extremals

$$y_i = Y_i(x, a(t)) = Y_i(x, t) \tag{5:2}$$

the curve D is define by the equations

$$x = x_5(t), \qquad y_i = Y_i[x_5(t), a(t)] = y_i(t),$$

and satisfies the equations

$$Y_i'x' + Y_{ia_j}a_j' = ky_i'$$

since

$$Y_{ia_j}a_j' = 0.$$

Hence the family (5:2) is a one-parameter family of composite extremals with an envelope D, touching the extremal E_{32} at the conjugate point 6.

FOURTH NECESSARY CONDITION IV. <u>Let</u> $E_{12} = E_{13} + E_{32}$ <u>be a composite arc which is normal on every subinterval of</u> x_1x_2 <u>and which is imbedded in an n-parameter family of composite arcs. Moreover suppose that</u> R_1 <u>and</u> R_2 <u>are different from zero on</u> E_{13} <u>and</u> E_{32} <u>respectively. Then if</u> E_{12} <u>is a minimizing arc there can exist no conjugate point to</u> 1 <u>on the arc</u> E_{12}.

In the following proof it is assumed that the envelope D of the one-parameter family of arcs (5:1) has a branch projecting backward from 6 to the point 1, as shown in the figure below. It is also assumed that the envelope D is not tangent anywhere to the n-space of tangency defined by the first n equations of (4:3) and by (4:6). That there can exist no conjugate point to 1 on E_{13} between 1 and 3 follows from the theory of the problem of Lagrange which applies to extremals E_{13}. To prove that there can exist no point conjugate to 1 on E_{32} between 3 and 2 consider the integral

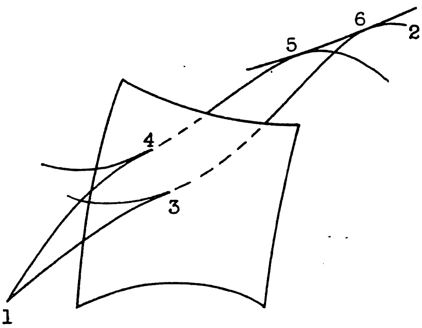
$$I(E_{14} + E_{45} + D_{56}) = \int_{x_1}^{x_4(t)} f[x,\ y(x,t),\ y'(x,t)]dx$$
$$+ \int_{x_4(t)}^{x_5(t)} f[x,\ Y(x,\ t),\ Y'(x,\ t)]dx$$
$$+ \int_{t}^{t_0} f\big[x(u),\ Y[x(u),u],\ Y'[x(u),u]\big]x'(u)\,du.$$

In this expression the equations

$$y_i = y_i(x,\ t) \qquad (x_1 \leqq x \leqq x_4),$$
$$y_i = Y_i(x,\ t) \qquad (x_4 \leqq x \leqq x_2),$$

define the one-parameter family of composite arcs, having an envelope D which has the equations

$$x = x_5(t), \qquad y_i = Y_i[x_5(t),\ t] = g_i(t).$$



Add $\Lambda_1 \phi_1(x,\ Y,\ Y') + \Lambda_\alpha \psi_\alpha(x,\ Y,\ Y')$ to the integrands of the second and third integrals in the above expression for $I(E_{14} + E_{45} + D_{56})$, and add $\lambda_\alpha \psi_\alpha$ to the integrand of the first integral. Then the derivative of I with respect to x_5 is

$$\frac{dI}{dx_5} = \frac{dI}{dt}\frac{dt}{dx_5} = -E(x, Y, Y', g', \lambda_\alpha, \lambda_1) + \lambda_1\phi_1(x, g, g')|^5$$

where g' is the slope of D. But since

$$Y_i'[x, t] = g_i'$$

at every point of D, it follows that

$$dI/dx_5 \equiv 0$$

in t. Consequently one obtains the result

$$I(E_{14} + E_{45} + E_{32}) = I(E_{32}).$$

Hence by the usual argument[1] $I(E_{12})$ cannot be a minimum.

6. Sufficiency proof. The four necessary conditions have been denoted by I, II, III, IV, the order being the same as they occur in this paper. The notation II' will be used to designate the condition II when the equality sign in expression (3:3) is omitted. The condition III' is defined for a composite arc as follows: The normal composite arc will be denoted by $E_{12} = E_{13} + E_{32}$, where E_{13} belongs to A and E_{32} belongs to B. For every element (x, y, y', λ) of E_{13} the inequality

$$F_{1y_i'y_k'}\pi_i\pi_k > 0$$

holds for every set $(\pi_i) \neq (0)$ satisfying the equations

$$\psi_{\alpha y_i'}\pi_i = 0,$$

whereas for every element (x, y, y', λ) of E_{32} the inequality

$$F_{2y_i'y_k'}\pi_i\pi_k > 0$$

is satisfied for every set $(\pi_i) \neq (0)$ satisfying the equations

$$\psi_{\alpha y_i'}\pi_i = 0, \qquad \phi_{1y_i'}\pi_i = 0.$$

---

[1]See Bliss, loc. cit., p. 722.

Condition IV' excludes the point 2 as well as the interior points of E_{12} from being a conjugate point to 1. An arc E_{12} satisfies condition II'_N if the inequality

$$E(x, y, y', Y', \lambda) - \lambda_1 \phi_1(x, y, Y') > 0$$

holds for all sets (x, y, y', Y', λ) for which the sets (x, y, y', λ) are in a neighborhood of similar sets belonging to E_{12}, and $(x, y, Y') \neq (x, y, y')$ satisfies

$$\phi_1(x, y, Y') \geqq 0, \qquad \psi_\alpha(x, y, Y') = 0.$$

We can now state the following theorem.

SUFFICIENCY THEOREM FOR A STRONG RELATIVE MINIMUM. _If an admissible composite arc_ $E_{12} = E_{13} + E_{32}$ _with an extension normal on every subinterval satisfies the conditions_ II'_N, III', IV', _then there exists a neighborhood_ M _of the points_ (x, y) of E_{12} _such that the inequality_ $I(C_{12}) > I(E_{12})$ _holds for every admissible arc_ C_{12} _satisfying_

$$\phi_1 \geqq 0, \qquad \psi_\alpha = 0,$$

which is in M, _and which is not identical with_ E_{12}.

In the first place since E_{12} is normal and satisfies I it is true that there exists a unique set of multipliers $\lambda_0 = 1$, λ_α, λ_1 and constants c_i which with the equations of E_{12} satisfy equations (3:2). In order to complete the proof, the following lemma is established.

LEMMA 6:1. _The condition_ III' _for a composite arc_ $E_{12} = E_{13} + E_{32}$ _implies that the determinants_ R_1 _and_ R_2 _defined by_ (4:1) _are different from zero on_ E_{13} _and_ E_{32} _respectively._

The fact that $R_1 \neq 0$ on E_{13} follows from the theory of the Lagrange problem which applies to E_{13}. However along the extremal E_{32} if the determinant $R_2 = 0$, the equations

$$F_{y_i'y_k'}\pi_k + \lambda_\alpha \psi_{\alpha y_k'} + \lambda_1 \phi_{1y_k'} = 0, \tag{6:1}$$

$$\psi_{\alpha y_k'}\pi_k = 0, \qquad \phi_{1y_k'}\pi_k = 0, \tag{6:2}$$

would have solutions $(\pi_i, \lambda_\alpha, \lambda_1) \neq (0, 0, 0)$ with π_k not all zero since by hypothesis the matrix

$$\begin{Vmatrix} \psi_{\alpha y_k'} \\ \phi_{1y_k'} \end{Vmatrix}$$

must have rank p + 1. By multiplying equations (6:1) by $(\pi_1, \ldots, \pi_n)$ respectively, and adding the result, one obtains

$$F_{y_i'y_k'}\pi_i \pi_k = 0$$

on account of equations (6:2). But this contradicts the latter part of condition III' which states that

$$F_{2y_i'y_k'}\pi_i \pi_k > 0$$

is satisfied for every set $(\pi_i) \neq (0)$ satisfying the equations

$$\psi_{\alpha y_i'}\pi_i = 0, \qquad \phi_{1y_i'}\pi_i = 0.$$

Thus condition III' implies that $R_2 \neq 0$ on E_{32}.

According to the imbedding theorem in section 4 a point O can be chosen on the normal extension of E_{13}, so that E_{12} can be imbedded in an n-parameter family of composite extremals passing through O. From the first n of equations (4:4) it follows that along the n-space of tangency of this composite family we have

$$y_{ia}(x, a) = Y_{ia}(x, a)$$

for values of a close to a_0 which defines E_{12}. Hence $\Delta(x, a)$

defined in section 5 is continuous in x. On account of condition IV' this n-parameter family simply covers a region containing $E_{12} = E_{13} + E_{32}$. For $\Delta(x, a) \neq 0$ on x_1x_2 implies from implicit function theory that there exists a neighborhood M of the points (x, y) on E_{12} in which the equations

$$
\begin{aligned}
&y_i = y_i(x, a) && (x_1 \leqq x \leqq x_3),\\
(6{:}3)\qquad &y_i = Y_i(x, a) && (x_3 \leqq x \leqq x_2),\\
&y_i(x_3, a) = Y_i(x_3, a),
\end{aligned}
$$

have solutions

$$a_i = a_i(x, y).$$

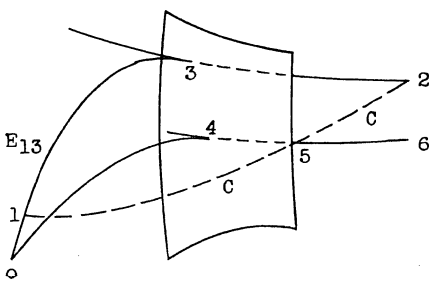
If the region M is taken sufficiently small the values (x, y, p, λ) belonging to M will remain in so small a neighborhood of the sets (x, y, y', λ) of E_{13} and E_{32} that according to II_N' the inequality

$$E(x, y, p, y', \lambda) - \lambda_1 \phi_1(x, y, y') > 0$$

will be satisfied for all sets $(x, y, y') \neq (x, y, p)$ in M, where $p_i(x, y) = y_{ix}$ or Y_{ix} according as the notation refers to an arc of class A or B. Hence one may show that $I(E_{12})$ is a minimum in M as follows.

Let any admissible curve C in M satisfying the conditions $\phi_\beta(x, y, y') \geqq 0$ and $\psi_\alpha(x, y, y') = 0$ be defined by the equations

$$y_i = g_i(x).$$



The integral $I(x_5)$ is de-

$$
\begin{aligned}
I(x_5) = &\int_{x_1}^{x_4} f[x,\ y(x,a),\ y'(x,a)]dx \\
&+ \int_{x_4}^{x_5} f[x,\ Y(x,a),\ Y'(x,a)]dx \\
&+ \int_{x_5}^{x_2} f[x,\ g(x,a),\ g'(x,a)]dx,
\end{aligned}
$$

where $y(x, a)$ and $Y(x, a)$ define the unique composite arc $E_{06} = E_{04} + E_{46}$ joining an arbitrary point 5 on C_{12} to the point 0. If the point 5 lies between the point 0 and the point of tangency 4 on E_{04}, then $I(x_5)$ has the derivative

$$I'(x_5) = -E(x,\ y,\ y',\ g',\ \lambda_\alpha,\ \lambda_1),$$

whereas if 5 lies between 4 and 6 on E_{46}, then $I(x_5)$ has the derivative

$$I'(x_5) = -E(x,\ Y,\ Y',\ g',\ \lambda_\alpha, \lambda_1) + \lambda_1 \phi_1(x,\ g,\ g').$$

In either case $I'(x_5)$ is less than or equal to zero, since $\lambda_1 = 0$ along arcs of class A. Moreover we have the equations

$$
\begin{aligned}
I(x_0) &= I(E_{01}) + I(C_{12}), \\
I(x_2) &= I(E_{01}) + I(E_{12}).
\end{aligned}
$$

The condition II_N' now implies that $I(E_{12})$ is a minimum.

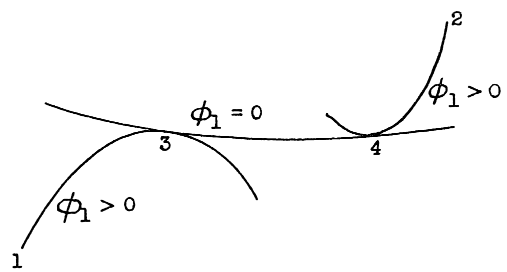
7. Generalizations to more complex arcs. In sections 4, 5, and 6 the composite arc as defined consisted of only two subarcs. The proofs made in these sections for the imbedding theorem, the Mayer condition, and for the sufficiency proof may be extended to apply to arcs without corners composed of n subarcs, ϕ_1 being zero on some of these subarcs, and greater than zero on the remaining subarcs. An imbedding theorem is here established for an arc without corners consisting of three sub-

arcs. It will then be obvious how to construct an imbedding theorem for an arc composed of four or more subarcs.

Consider an extremal $E_{12} = E_{13} + E_{34} + E_{42}$ without corners along which all the functions ϕ_β save one are greater than zero. The function ϕ_1 is to be greater than zero on the subarcs E_{13} and E_{42}, but zero on the arc E_{34}. Suppose that R_1 is different from zero on E_{13} and E_{42}, whereas R_2 is different from zero on E_{34}. From the imbedding theorem for composite arcs, we know that the arc $E_{14} = E_{13} + E_{34}$ may be imbedded in an n-parameter family of composite arcs of the same form. Denote this n-parameter family by

$$\begin{aligned} y_i &= h_i(x, a) & (x_1 \leqq x \leqq x_3), \\ y_i &= Y_i(x, a) & (x_3 \leqq x \leqq x_4), \\ \lambda_\alpha &= \lambda_\alpha(x, a), & \\ \lambda_1 &= \lambda_1(x, a). & \end{aligned}$$

Moreover it is known that under the above hypotheses E_{42} may be imbedded in a 2n-parameter family of arcs of class A. Let these extremals be defined by the equations



$$\begin{aligned} y_i &= y_i(x, b) & \\ \lambda_\alpha &= \lambda_\alpha(x, b) & (x_4 \leqq x \leqq x_2). \\ \lambda_1 &= \lambda_1(x, b) & \end{aligned}$$

The following conditions

$$
(7:1) \qquad \begin{aligned}
&Y_i(x, a) - y_i(x, b) = 0, \\
&F_{2y_i'}[x, Y(x,a), Y'(x,a), \lambda(x,a)] \\
&\quad - F_{1y_i'}[x, y(x,b), y'(x,b), \lambda(x,b)] = 0, \\
&\phi_1[x, y(x,b), y'(x,b)] = 0
\end{aligned}
$$

hold at the point 4 on E_{12}. Moreover it is known that the determinant D satisfies the condition

$$
D = \begin{vmatrix} y_{ib_k} \\ u_{ib_k} \end{vmatrix} \neq 0, \qquad u_i = F_{1y_i'}
$$

at the point 4 on E_{12}. The functional determinant of the above expression with respect to x and b is

$$
C = \begin{vmatrix} Y_i' - y_i' & -y_{ib_k} \\ F_{1y_i'}' - F_{2y_i'}' & -u_{ib_k} \\ \phi_1' & \phi_{1b_k} \end{vmatrix} .
$$

Since at the point 4 we know that

$$
\begin{aligned}
y_i(x_4, b) &= Y_i(x_4, a), \\
y_i'(x_4, b) &= Y_i'(x_4, a),
\end{aligned}
$$

holds, it follows from Corollary 3:2 that the determinant C has the value

$$
C = \phi_1' \begin{vmatrix} y_{ib_k} \\ u_{ib_k} \end{vmatrix} .
$$

For purposes of the proof we assume that $\phi_1' \neq 0$ holds at the point 4. Hence it is true that $C \neq 0$ at the point 4 on E_{12}. Thus one may solve equations (7:1) for x and b as functions of a. Consequently under the above hypotheses the arc $E_{12} = E_{13} + E_{34} + E_{42}$ can be imbedded in an n-parameter family of arcs of the same kind, that is, consisting of three subarcs, two belonging to class A, and one belonging to class B. A proof similar to that given in section 4 shows that the members of this family have no corners. An imbedding theorem for an arc E_{12} composed of n subarcs, ϕ_1 being greater than zero on every other subarc, and zero on the remaining subarcs, can now be made by alternately repeating the processes described in section 6 and in this section.

The proof of the Mayer condition and the sufficiency proof for the arcs considered in this section are so similar to those given for a composite arc that they will need no repetition.

8. The analogue of the Mayer condition and the second variation. The following section establishes the condition IV, formulated geometrically in section 6, by means of the second variation. The equivalent problem stated in section 2 will now be used again.

For a normal extremal E_{12} of the equivalent problem it is known that if $\eta_i(x)$, $\zeta_\beta(x)$ is a set of admissible variations satisfying the equations

$$\begin{aligned} &\Psi_\alpha(x, \eta, \eta') = \psi_{\alpha y_i}\eta_i + \psi_{\alpha y_i'}\eta_i' = 0, \\ (8{:}1)\quad &\Phi_\beta(x, \eta, \eta') - 2z_\beta' \zeta_\beta' = \phi_{\beta y_i}\eta_i + \phi_{\beta y_i'}\eta_i' - 2z_\beta' \zeta_\beta', \\ &\eta_i(x_1) = \eta_i(x_2) = 0, \end{aligned}$$

then there exists a one-parameter family of admissible arcs

$$y_i = y_i(x, b), \qquad z_\beta = z_\beta(x, b)$$

containing E_{12} for $b = b_0$, and having the set $\eta_i(x)$, $\zeta_\beta(x)$ as its variations along E_{12}. In this section the second variation is to be calculated for an admissible arc E_{12} without corners satisfying the equations $\psi_\alpha = 0$ and $\phi_\beta(x, y, y') - z_\beta'^2 = 0$, and also satisfying the multiplier rule with multipliers $\lambda_0 = 1$, $\lambda_\alpha(x)$ and $\lambda_1(x)$.

When the members of the equations

$$\begin{aligned} I(b) &= \int_{x_1}^{x_2} f[x,\ y(x,b),\ y'(x,b)]dx, \\ 0 &= \psi_\alpha[x,\ y(x,b),\ y'(x,b)], \\ 0 &= \phi_\beta[x,\ y(x,b),\ y'(x,b)] - z_\beta'^2(x,\ b), \end{aligned}$$

are differentiated twice with respect to b, one may obtain[1] the equation

$$I''(b_0) = \int_{x_1}^{x_2} [2\omega(x,\ \eta,\ \eta') - 2\ \zeta_\beta'^2 \lambda_\beta]dx, \tag{8:2}$$

where

$$2\omega = F_{y_i\, y_k}\, \eta_i\eta_k + 2F_{y_i y_k'}\eta_i\eta_k' + F_{y_i' y_k'}\, \eta_i'\eta_k'.$$

The accessory minimum problem for this problem consists in finding in the class of arcs $\eta_i(x)$, $\zeta_\beta(x)$ satisfying the equations (8:1) that one which minimizes the second variation (8:2). The case to be considered here is the one in which the minimizing arc is a composite one $E_{12} = E_{13} + E_{32}$. The extremals for the accessory minimum problem for this case must satisfy the differential equations

$$\begin{aligned} &\frac{d}{dx}\Omega_{\eta_i'} = \Omega_{\eta_i}, && \zeta_1'\lambda_1 + \mu_1 z_1' = d_1 \\ &\psi_\alpha(x,\ \eta,\ \eta') = 0, && \Phi_1(x,\ \eta,\ \eta') - 2z_1'\zeta_1' = 0 \end{aligned} \tag{8:3}$$

where

$$\Omega = \mu_0(\omega - \zeta_1'^2\lambda_1) + \mu_\alpha\psi_\alpha + \mu_1(\Phi_1 - 2z_1'\zeta_1')$$

and d_1 is a constant. From the transversality condition one finds that

$$\zeta_1'\lambda_1 + \mu_1 z_1'\,|^1 = \zeta_1'\lambda_1 + \mu_1 z_1'\,|^2 = 0$$

---

[1]See Bliss, loc. cit., p. 723.

holds. Hence it is true that

$$\zeta_1'\lambda_1 + \mu_1 z_1' \equiv 0$$

on x_1x_2. Since λ_1 is zero on E_{13} and z_1' is zero on E_{32}, it follows that

$$\zeta_1'\lambda_1 \equiv \mu_1 z_1' \equiv 0 \qquad (x_1 \leqq x \leqq x_2)$$

holds. The functions $\eta(x)$, $\zeta_1(x)$ which define the minimizing arc for the accessory minimum problem are determined by the equations

$$\Omega_{\eta_i'} = \int_{x_1}^{x} \Omega_{\eta_i} dx + c_i,$$
$$\Psi_\alpha = 0, \qquad \Phi_1 - 2z_1'\zeta_1' = 0.$$

It follows that η_i', μ_α, μ_1 are continuous at x_3 as well as at all other points on x_1x_2 since $\eta_i(x)$ are continuous, and thus all three terms not involving η_i', μ_α, and μ_1 are continuous since the determinant of coefficients of η_i', μ_α, μ_1 is R_2 or R_1 which are different from zero on x_3x_2 and x_1x_3 respectively.

The functions η, ζ_1 are defined for the intervals x_1x_3 and x_3x_2 by the following equations,

(8:4)
$$\frac{d}{dx}\Omega_{2\eta_i'} = \Omega_{2\eta_i}, \qquad (x_3 \leqq x \leqq x_2),$$
$$\Psi_\alpha = 0, \qquad \Phi_1 - 2z_1'\zeta_1' = 0,$$

(8:5)
$$\frac{d}{dx}\Omega_{1\eta_i'} = \Omega_{1\eta_i}, \qquad (x_1 \leqq x \leqq x_3),$$
$$\Psi_\alpha = 0,$$

where

$$\Omega_1 = \mu_0\omega + \mu_\alpha\Psi_\alpha,$$
$$\Omega_2 = \mu_0\omega + \mu_\alpha\Psi_\alpha + \mu_1\Phi_1.$$

On the interval x_1x_3 the function $\zeta_1(x)$ is defined by the equation $\Phi_1 - 2z_1'\zeta_1' = 0$. The function $\zeta_1(x)$ is admissible, since $z_1' \neq 0$ for $(x_1 \leqq x < x_3)$ and the equations

$$\Phi_1[x_3, \eta(x_3-0), \eta'(x_3-0)] = \Phi_1[x_3, \eta(x_3+0), \eta'(x_3+0)] = 0,$$
$$z_1'(x_3) = 0,$$

hold at the point 3. Also the function $\zeta_1'(x)$ is zero at $x = x_3$ since the limit

$$\lim_{x=x_3-0} \Phi_1/z_1' \tag{8:6}$$

exists at $x = x_3$ and is zero. For if the numerator and denominator of the function

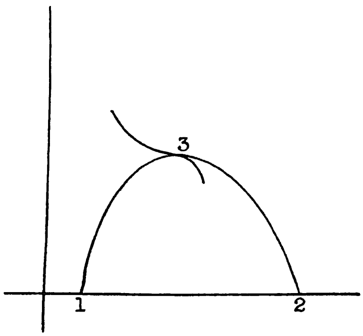
$$\Phi_1{}^2/z_1'{}^2$$

are differentiated separately, one gets

$$2\Phi_1\Phi_1'(x, \eta, \eta')/\Phi_1'(x, y, y').$$

Since it has already been assumed that $\Phi_1'(x_3, y, y') \neq 0$, it is true that the limit (8:6) is zero at 3. Thus the minimizing arcs for the accessory minimum problem are defined by equations (8:5) and (8:4).

In the following argument the determinants R_1 and R_2 defined by expression (4:1) are assumed to be different from zero on E_{13} and E_{32} respectively.



Definition. A value x_4 is said to be conjugate to x_1 on the arc $E_{13} + E_{32}$

if there exists an extremal of the accessory minimum problem of the form

$$\eta_i(x) = u_i(x), \qquad \mu_\alpha = \rho_\alpha(x) \qquad (x_1 \leqq x \leqq x_3),$$

$$\eta_i(x) = v_i(x), \qquad \begin{matrix} \mu_\alpha = \rho_\alpha(x) \\ \mu_1 = \rho_1(x) \end{matrix} \qquad (x_3 \leqq x \leqq x_2),$$

continuous and having continuous derivatives on x_1x_2 and satisfying

$$\eta_i(x_1) = \eta_i(x_4) = 0,$$

but not identically zero on x_1x_2.

ANALOGUE OF THE MAYER CONDITION. <u>Suppose</u> $E_{12} = E_{13} + E_{32}$ <u>is a composite arc which is normal on every subinterval, and which is such that</u> R_1 <u>and</u> R_2 <u>are different from zero on</u> E_{13} <u>and</u> E_{32} <u>respectively. If</u> E_{12} <u>is a minimizing arc there can exist no point conjugate to</u> 1 <u>between the points</u> 1 <u>and</u> 2.

To prove this statement consider the special solution

$$\left.\begin{matrix} \eta_i(x) \equiv u_i(x), & \mu_\alpha(x) \equiv \rho_\alpha(x), & (x_1 \leqq x \leqq x_4) \\ \eta_i(x) \equiv 0, & & (x_4 \leqq x \leqq x_2) \end{matrix}\right. \text{ when } x_4 \leqq x_3,$$

$$\left.\begin{matrix} \eta_i(x) \equiv u_i(x), & \mu_\alpha(x) \equiv \rho_\alpha(x), & (x_1 \leqq x \leqq x_3) \\ \eta_i(x) \equiv v_i(x), & \mu_\alpha(x) \equiv \rho_\alpha(x), & \\ & \mu_1(x) \equiv \rho_1(x), & (x_3 \leqq x \leqq x_4) \\ \eta_i(x) \equiv 0, & & (x_4 \leqq x \leqq x_2) \end{matrix}\right\} \text{ when } x_3 < x_4.$$

For this choice of $\eta_i(x)$ the second variation has the value

$$I''(b_0) = \int_{x_1}^{x_3} 2\omega(x, u, u')dx + \int_{x_3}^{x_4} 2\omega(x, v, v')dx,$$

which has the form

$$I''(b_0) = \int_{x_1}^{x_3}(u_i\Omega_{1u_i} + u_i'\Omega_{1u_i'} + \rho_\alpha\Omega_{1\rho_\alpha} + \rho_1\Omega_{1\rho_1})dx$$
$$+ \int_{x_3}^{x_4}(v_i\Omega_{2v_i} + v_i'\Omega_{2v_i'} + \rho_\alpha\Omega_{2\rho_\alpha} + \rho_1\Omega_{1\rho_1})dx.$$

Upon using equations (8:5) and (8:4) this integral may be evaluated to be

$$I''(b_0) = u_i\Omega_{1u_i'}\Big|_1^3 + v_i\Omega_{2v_i'}\Big|_3^4.$$

But since the relation

$$u_i\Omega_{1u_i'} = v_i\Omega_{2v_i'}$$

holds at the point 3, $I''(b_0)$ has the value

$$I''(b_0) = v_i\Omega_{2v_i'}\Big|^4 - u_i\Omega_{1u_i}\Big|^1,$$

or

$$I''(b_0) = v_i\Omega_{2v_i} = 0.$$

Since for a minimizing arc $\eta_i(x)$ the corner conditions

$$\Omega_{\eta_i'}[x_4, \eta, \eta'(x_4-0), \mu(x_4-0)]$$
$$-\Omega_{\eta_i'}[x_4, \eta, \eta'(x_4+0), \mu(x_4+0)] = 0,$$

hold,[1] and since

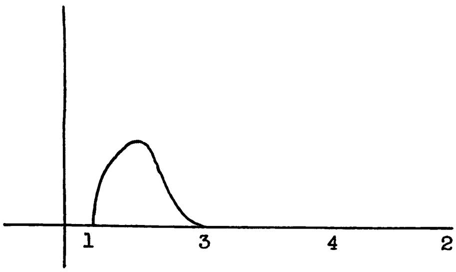
$$\Psi(x, \eta, \eta') = 0, \qquad \Phi_1(x, \eta, \eta') = 0, \qquad R_2 \neq 0,$$

hold, it is true that

$$v_i \equiv v_i' \equiv \rho_\alpha \equiv 0$$
$$(x_3 \leqq x \leqq x_4).$$

Similarly it follows that

$$u_i \equiv u_i' \equiv \rho_\alpha \equiv 0$$



[1]See Bliss, loc. cit., p. 725-6.

for the interval $(x_1 \leqq x \leqq x_3)$. Thus the Mayer condition has been established.

9. Analogue of the necessary condition of Hestenes. In sections 10 and 11 a sufficiency proof is made for a composite arc without the assumption of normality. In order to lead up to this proof another necessary condition, analogous to the necessary condition IV_1, given by Hestenes for the problem of Bolza, is derived.

As shown in section 8 the minimizing arc $\eta_i(x)$ for the accessory minimum problem, when $E_{12} = E_{13} + E_{32}$ is a composite arc, is defined by (8:5) for $(x_1 \leqq x \leqq x_3)$ and by (8:4) for $(x_3 \leqq x \leqq x_2)$. The functional determinant of equations (8:5) with respect to η_i' and μ_α is R_1, whereas the functional determinant of equations (8:4) with respect to η_i', μ_α and μ_1 is R_2, R_1 and R_2 being defined by (4:1). Since we suppose that R_1 and R_2 are different from zero on E_{13} and E_{32} respectively, the equations

$$(9:1) \qquad \begin{aligned} t_i &= \Omega_{1\eta_i'}(x, \eta, \eta', \mu) \\ 0 &= \Psi_\alpha \end{aligned} \qquad (x_1 \leqq x \leqq x_3),$$

have the solutions

$$\begin{aligned} \eta_i' &= \Pi_i(x, \eta, t) \\ \mu_\alpha &= M_\alpha(x, \eta, t) \end{aligned} \qquad (x_1 \leqq x \leqq x_3),$$

and the equations

$$(9:2) \qquad \begin{aligned} t_i &= \Omega_{2\eta_i'}(x, \eta, \eta', \mu) \\ 0 &= \Psi_\alpha \\ 0 &= \Phi_1 \end{aligned} \qquad (x_3 \leqq x \leqq x_2),$$

have the solutions

$$\begin{aligned} \eta_i' &= K_i(x, \eta, t) \\ \mu_\alpha &= N_\alpha(x, \eta, t) \\ \mu_1 &= N_1(x, \eta, t) \end{aligned} \qquad (x_3 \leqq x \leqq x_2).$$

Equations (8:5) and (8:4) may be put into the usual canonical forms by introducing the Hamiltonian functions[1] H_1 and H_2. Equations (8:5) will then be equivalent to

$$(9:3) \qquad \begin{aligned} \eta_i' &= H_{1t_i} \\ t_i' &= -H_{1\eta_i} \end{aligned} \qquad (x_1 \leqq x \leqq x_3),$$

and equations (8:4) will be equivalent to

$$(9:4) \qquad \begin{aligned} \eta_i' &= H_{2t_i} \\ t_i' &= -H_{2\eta_i} \end{aligned} \qquad (x_3 \leqq x \leqq x_2).$$

For an arbitrary pair of solutions of (9:3), (η_i, t_i) and (η_i^*, t_i^*), it is known that

$$(9:5) \qquad \eta_i t_i^* - \eta_i^* t_i = \text{constant} = c.$$

The same relation holds for an arbitrary pair of solutions of (9:4

DEFINITION. <u>The solution</u> (η_i^*, t_i^*) <u>is said to be conjugate to the solution</u> (η_i, t_i) <u>if equation</u> (9:5) <u>holds with</u> $c = 0$.

The sets $[\eta_{ik}, t_{ik}]$ form a conjugate system if any pair of them are conjugate to each other.

A conjugate system of solutions (η_{ik}, t_{ik}) of equations (9:3) and (9:4) may be found such that (η_{ik}, t_{ik}) are continuous. Suppose $\eta_{ik} = \sigma_{ik}$ and $t_{ik} = s_{ik}$ form a conjugate system of solutions of (9:4) on $(x_3 \leqq x \leqq x_2)$ where Ω has been replaced by Ω_2. The solutions $\eta_{ik} = u_{ik}$, $t_{ik} = r_{ik}$ of equations (9:3) with the end conditions

$$u_{ik}(x_3) = \sigma_{ik}(x_3),$$
$$r_{ik}(x_3) = s_{ik}(x_3),$$

---

[1]G. A. Bliss, <u>Problem of Bolza in the Calculus of Variations</u>, Lecture notes at the University of Chicago, Winter 1935, p. 74.

on the interval $(x_1 \leqq x \leqq x_3)$ are well defined. The system of solutions (η_{ik}, t_{ik}) thus obtained is continuous on the entire interval $(x_1 \leqq x \leqq x_2)$ and is a conjugate system.

ANALOGUE OF THE CONDITION OF HESTENES, IV_1. Suppose the arc $E_{12} = E_{13} + E_{32}$ satisfies the hypotheses assumed for the calculation of the second variation. The arc is said to satisfy condition IV_1 if the inequality

$$(\zeta_{ij}u_{ik} - \eta_{ij}v_{ik})a_j b_k \geqq 0 \tag{9:6}$$

is satisfied on $(x_1 \leqq x \leqq x_2)$, where the constants a_j and b_j satisfy the equations

$$\eta_{ij}a_j = u_{ik}b_k, \tag{9:7}$$

and where the set (η_{ij}, ζ_{ij}) is a conjugate system of solutions of equations (9:3), and (u_{ij}, v_{ij}) is a conjugate system of solutions of equations (9:4). The first set (η_{ij}, ζ_{ij}) is defined by the transversality and end-conditions for the point 1, whereas the second set (u_{ij}, v_{ij}) is defined by the corresponding conditions for the point 2. Every normal composite minimizing arc $E_{12} = E_{13} + E_{32}$, for which R_1 and R_2 are different from zero on E_{13} and E_{32} respectively, must satisfy the condition IV_1.

We will first prove the necessity of this condition on E_{32}. Let the set (η_{ij}, ζ_{ij}) be defined as follows

$$\begin{aligned} \eta_{ij} &= \tau_{ij}(x) \\ \zeta_{ij} &= r_{ij}(x) \end{aligned} \qquad (x_1 \leqq x \leqq x_3),$$

and

$$\begin{aligned} \eta_{ij} &= \sigma_{ij}(x) \\ \zeta_{ij} &= s_{ij}(x) \end{aligned} \qquad (x_3 \leqq x \leqq x_2).$$

Moreover let

$$u_{ik} \equiv m_{ik}(x) \qquad\qquad (x_1 \leqq x \leqq x_3),$$
$$v_{ik} \equiv n_{ik}(x)$$

and

$$u_{ik} \equiv p_{ik}(x) \qquad\qquad (x_3 \leqq x \leqq x_2).$$
$$v_{ik} \equiv q_{ik}(x)$$

Consider a solution a_j, b_k of (9:7) for a value x_4 between x_3 and x_2, and let the sets (τ_i, r_i), (σ_i, s_i), (m_i, n_i) and (p_i, q_i) represent

$$\begin{aligned} \tau_i &= \tau_{ij}a_j \\ r_i &= r_{ij}a_j \\ m_i &= m_{ik}b_k \\ n_i &= n_{ik}b_k \end{aligned} \qquad\qquad (x_1 \leqq x \leqq x_3),$$

$$\begin{aligned} \sigma_i &= \sigma_{ij}a_j \\ s_i &= s_{ij}a_j \\ p_i &= p_{ik}b_k \\ q_i &= q_{ik}b_k \end{aligned} \qquad\qquad (x_3 \leqq x \leqq x_2).$$

The arc defined by $\tau_i(x)$ on $(x_1 \leqq x \leqq x_3)$, by $\sigma_i(x)$ on $(x_3 \leqq x \leqq x_4)$, and by $p_i(x)$ on $(x_4 \leqq x \leqq x_2)$ is continuous by (9:7), and satisfies the equations $\Psi_\alpha = 0$ on x_1x_2 and $\Phi_1 = 0$ on x_3x_2. This arc gives to the second variation (8:2) the value

$$I''(b_0) = \int_{x_1}^{x_3} 2\omega(x, \tau, \tau')dx + \int_{x_3}^{x_4} 2\omega(x, \sigma, \sigma')dx + \int_{x_4}^{x_2} 2\omega(x, p, p')dx.$$

If $\mu_\alpha\Psi_\alpha$ is added to the integrand of the first integral, and if $\mu_\alpha\Psi_\alpha + \mu_1\Phi_1$ is added to the integrands of the second and third integrals, $I''(b_0)$ will have the form

$$I''(b_0) = \int_{x_1}^{x_3} 2\Omega_1(x, \tau, \tau')dx + \int_{x_3}^{x_4} 2\Omega_2(x, \sigma, \sigma')dx + \int_{x_4}^{x_2} 2\Omega_2(x, p, p')dx.$$

By the use of the homogeneity property of quadratic forms,[1] one may find the value of $I''(b_0)$ to be

$$I''(b_0) = \tau_i\Omega_{1\tau_i'}\Big|_1^3 + \sigma_i\Omega_{2\sigma_i'}\Big|_3^4 + p_i\Omega_{2p_i'}\Big|_4^2,$$

which reduces to

$$I''(b_0) = \tau_i r_i\Big|_1^3 + \sigma_i s_i\Big|_3^4 + p_i q_i\Big|_4^2.$$

Since the equations

$$\tau_i(x_3) = \sigma_i(x_3), \qquad r_i(x_3) = s_i(x_3),$$
$$\tau_i(x_1) = p_i(x_2) = 0, \qquad \sigma_i(x_4) = p_i(x_4),$$

hold, it follows that

$$I''(0) = s_i(x_4)p_i(x_4) - \sigma_i(x_4)q_i(x_4),$$

and this last expression for $I''(0)$ is

$$(\zeta_{ij}u_{ik} - \eta_{ij}v_{ik})\, a_j b_k \qquad (x_3 \leqq x \leqq x_2).$$

A similar proof can be made when the point 4 lies between the points 1 and 3. In event the point 4 is taken at the point 3, the second variation $I''(b_0)$ will have the form

$$I''(b_0) = \int_{x_1}^{x_3} 2\Omega_1(x, \tau, \tau')dx + \int_{x_3}^{x_2} 2\Omega_2(x, p, p')dx,$$

where $\tau_i(x)$ and $p_i(x)$ are defined above. The completion of the proof for this case is then easily made. Thus the condition IV_1 has been established.

---

[1] Bliss, <u>Problem of Bolza</u>, p. 87.

10. <u>Sufficiency proof without the assumption of normality.</u> One may now prove the following theorem with the aid of the preceding section and some auxiliary lemmas.

THEOREM 10:1. <u>Let</u> $E_{12} = E_{13} + E_{32}$ <u>be an admissible composite arc, satisfying the conditions</u> II_N', III', IV_1', <u>with a set of multipliers</u> $\lambda_0 = 1$, $\lambda_\alpha(x)$, $\lambda_1(x)$. <u>Then there exists a neighborhood</u> F <u>of</u> $E_{12} = E_{13} + E_{32}$ <u>such that</u> $J(C_{12}) > J(E_{12})$ <u>for every admissible arc</u> C_{12} <u>in</u> F <u>joining the points</u> 1 <u>and</u> 2, <u>satisfying</u>

$$\psi_\alpha = 0, \qquad \phi_1 \geqq 0,$$

<u>and distinct from</u> E_{12}.

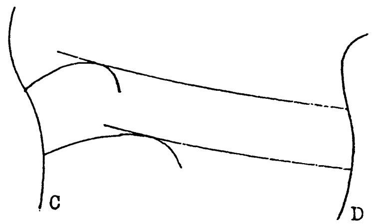
Consider a one-parameter family of composite arcs

$$
\begin{array}{llll}
 & y_i = y_i(x, a), & \lambda_\alpha = \lambda_\alpha(x, a) & (x_1 \leqq x \leqq x_3), \\
(10:1) & y_i = Y_i(x, a), & \lambda_\alpha = \lambda_\alpha(x, a) & (x_3 \leqq x \leqq x_2), \\
 & & \lambda_1 = \lambda_1(x, a), &
\end{array}
$$

and a set of functions $x_1(t)$, $x_2(t)$, $a(t)$ having continuous derivatives, and such that y_i, Y_i, y_{ix}, Y_{ix}, λ_β and λ_1 have continuous first partial derivatives in a neighborhood of the sets (x, a) defined by

$$x_1(t) \leqq x \leqq x_2(t), \qquad a(t) \qquad (t' \leqq t \leqq t'').$$

The end points 1 and 2 of the curves describe two arcs C and D, the equations of C being



$$x = x_1(t),$$
$$y_i = y_i[x_1(t), a(t)],$$

and the curve D being defined by

$$x = x_2(t), \qquad y_i = Y_i[x_2(t), a(t)].$$

The differentials dx, dy_i along the curves C and D are given by the equations

$$(10:2) \qquad \begin{aligned} dx_1 &= x_1'(t)dt, & dy_i &= y_{ix_1}dx_1 + y_{ia}da, \\ dx_2 &= x_2'(t)dt, & dy_i &= Y_{ix_2}dx_2 + Y_{ia}da. \end{aligned}$$

Along the particular composite extremal arc defined by a value t the integral I has the form

$$I(t) = \int_{x_1(t)}^{x_3(t)} F_1[x, y(x,a), y'(x,a), \lambda]dx$$
$$+ \int_{x_3(t)}^{x_2(t)} F_2[x, y(x,a), y'(x,a), \lambda]dx.$$

The derivative of I with respect to t is

$$\frac{dI}{dt} = F_1\frac{dx}{dt}\Big|_1^3 + \int_{x_1}^{x_3}(F_{1y_i}y_{ia}\frac{da}{dt} + F_{1y_i'}y_i'{}_a\frac{da}{dt})dx$$
$$+ F_2\frac{dx}{dt}\Big|_3^2 + \int_{x_3}^{x_2}(F_{2y_i}Y_{ia}\frac{da}{dt} + F_{2y_i'}Y_i'{}_a\frac{da}{dt})dx.$$

Upon integrating by parts and using equations (10:2) one gets

$$(10:3) \qquad dI = [F - y_i'F_{y_i'}]dx + F_{y_i'}dy_i\Big|_1^2.$$

The symbol I^* denotes the integral

$$I^* = \int\Big[[F - y_iF_{y_i'}]dx + F_{y_i'}dy_i\Big].$$

By integrating (10:3) from t' to t'' one obtains the following result.

LEMMA 10:1. *If the composite extremal arcs of the one parameter family (10:1) corresponding to the values t' and t'' of the parameter t are E_{34} and E_{56} respectively, then*

$$I(E_{56}) - I(E_{34}) = I^*(D_{46}) - I^*(C_{35}).$$

Definition of a Field. A field is a region $\mathcal{F}$ of (x, y) space with a set of slope-functions and multipliers

$$p_i(x, y), \quad \ell_0 = 1, \quad \ell_1(x, y), \quad \ell_\beta(x, y),$$

having continuous first partial derivatives in $\mathcal{F}$, and such that the sets (x, y, p) are admissible and satisfy $\psi_\alpha = 0$, $\phi_1 \geqq 0$, and make the I^* integral

$$I^* = \int\{[F - p_i F_{y_i'}]dx + F_{y_i'} dy_i\}$$

independent of the path in $\mathcal{F}$.

LEMMA[1] 10:2. *Let $E_{12} = E_{13} + E_{32}$ be a composite arc such that $R_1 \neq 0$ on E_{13} and $R_2 \neq 0$ on E_{32}, and having a conjugate system of solutions (U_{ik}, V_{ik}) of the accessory equations* (9:3) *and* (9:4). *This solution has the form*

$$\begin{aligned} U_{ik} &= u_{ik} \\ V_{ik} &= v_{ik} \end{aligned} \qquad (x_1 \leqq x \leqq x_3),$$

$$\begin{aligned} U_{ik} &= \sigma_{ik} \\ V_{ik} &= s_{ik} \end{aligned} \qquad (x_3 \leqq x \leqq x_2).$$

Moreover suppose $|U_{ik}| \neq 0$ on x_1x_2. Then E_{12} is an extremal of a field $\mathcal{F}$ consisting of an n-parameter family of composite arcs

$$\begin{aligned} y_i &= y_i(x, \alpha_1, \ldots, \alpha_n), & r_i &= F_{1y_i'} & [x_1 &\leqq x \leqq x_3(\alpha)], \\ y_i &= Y_i(x, \alpha_1, \ldots, \alpha_n), & R_i &= F_{2y_i'} & [x_3(\alpha) &\leqq x \leqq x_2], \end{aligned}$$

and containing E_{12} for values (x, α) satisfying

$$(x_1 \leqq x \leqq x_2), \qquad \alpha_k = 0, \qquad (k = 1, \ldots, n).$$

---

[1] Bliss, *loc. cit.*, Problem of Bolza, p. 103.

The functions y_i, Y_i, y_{ix}, Y_{ix}, r_i, R_i have continuous first partial derivatives in a neighborhood of the values (x, α) belonging to E_{12}, and the variations of that family along E_{12} have the values

$$y_{i\alpha_k}(x, 0) = u_{ik}(x), \quad r_{i\alpha_k}(x, 0) = v_{ik}(x) \quad [x_1 \leqq x \leqq x_3(0)],$$
$$Y_{i\alpha_k}(x, 0) = \sigma_{ik}(x), \quad R_{i\alpha_k}(x, 0) = s_{ik}(x) \quad [x_3(0) \leqq x \leqq x_2].$$

The proof of this lemma can be obtained by an extension of a lemma given by Bliss for the problem of Bolza. By a proof whose details are identical with those given for the imbedding theorem in section 4, it may be proved that the composite arc $E_{12} = E_{13} + E_{32}$ may be imbedded in a 2n-parameter family of composite arcs. As shown by Bliss[1] it is true that this 2n-parameter family may have the form

$$\begin{aligned} y_i &= y_i(x, \alpha_1, \ldots, \alpha_n) \\ r_i &= r_i(x, \alpha_1, \ldots, \alpha_n) \end{aligned} \qquad [x_1 \leqq x \leqq x_3(\alpha)],$$

$$\begin{aligned} y_i &= Y_i(x, \alpha_1, \ldots, \alpha_n) \\ R_i &= R_i(x, \alpha_1, \ldots, \alpha_n) \end{aligned} \qquad [x_3(\alpha) \leqq x \leqq x_2],$$

containing E_{12} for $(\alpha_1, \ldots, \alpha_n) = (0, \ldots, 0)$. It follows from the theory given by Bliss and from the imbedding theorem mentioned that the equations

$$y_{i\alpha_k}(x_3) = Y_{i\alpha_k}(x_3) = \sigma_{ik}(x_3) = u_{ik}(x_3)$$

hold. The remainder of the proof of the above lemma is so similar to that made by Bliss that it will not be repeated.

---

[1] Bliss, loc. cit., Problem of Bolza, p. 105.

THEOREM 10:2. A FUNDAMENTAL SUFFICIENCY THEOREM. *If an arc $E_{12} = E_{13} + E_{32}$ is a composite arc in a field $\mathcal{F}$ and satisfies the condition II_N', then $I(E_{12})$ is a minimum as described in Theorem* 10:1.

In view of the assumption that E_{12} satisfies the condition II_N', the field $\mathcal{F}$ may be restricted to a sufficiently small neighborhood of E_{12} so that all the elements $[x, y, p(x,y), \ell(x,y)]$ belonging to $\mathcal{F}$ lie in the neighborhood N. Then at all points of $\mathcal{F}$ the condition

$$E(x, y, p(x,y), Y', \ell) - \ell_1\phi(x, y, Y') > 0$$

must be satisfied for every set $(x, y, Y') \neq (x, y, y')$ and satisfying $\phi_1(x, y, Y') \geqq 0$, $\psi_\alpha(x, y, Y') = 0$. Since

$$I^*(E_{12}) = I(E_{12})$$

it is true that

$$\begin{aligned} I(C_{12}) - I(E_{12}) &= I(C_{12}) - I^*(E_{12}) \\ &= \int_{x_1}^{x_2} f(x, Y, Y')dx - \int_{x_1}^{x_2} ([F - p_i F_{y_i'}]dx + F_{y_i'}dy_i) \\ &= \int_{x_1}^{x_2} [F - \ell_1\phi_1(x, y, Y') - F(x, y, p_i) - (p_i - Y_i')F_{y_i'}]dx \\ &= \int_{x_1}^{x_2} [E(x, y, p, Y', \lambda) - \ell_1\phi_1(x, y, Y')]dx. \end{aligned}$$

Hence the theorem is established.

LEMMA[1] 10:3. *Let $E_{12} = E_{13} + E_{32}$ be an admissible composite arc satisfying conditions III', IV_1' with a set of multipliers $\lambda_0 = 1$, λ_α, λ_1. Then there exists a conjugate system*

---

[1] Bliss, *loc. cit.*, Problem of Bolza, p. 112.

of solutions $U_{ik}(x)$, $V_{ik}(x)$ *of the canonical accessory equations* (9:3) *and* (9:4) *with* $|U_{ik}| \neq 0$ *on* x_1x_2.

The proof of this lemma is almost identical with that given by Bliss in his notes on the problem of Bolza, hence it will not be repeated.

Now one is in a position to prove the sufficiency theorem 10:1. According to Lemma 6:1 an admissible arc $E_{12} = E_{13} + E_{32}$ satisfying III' must be such that R_1 and R_2 are different from zero on E_{13} and E_{32} respectively. Condition IV_1' and Lemma 10:3 imply the existence of a conjugate system of solutions $U_{ik}(x)$, $V_{ik}(x)$ of the canonical equations (9:3) and (9:4) with determinant $|U_{ik}(x)| \neq 0$ on x_1x_2. Hence by Lemma 10:2 the composite arc E_{12} is in a field $\mathcal{F}$, contained in an n-parameter family of composite arcs. Thus by these conditions and II_N' it follows that the hypotheses of the sufficiency Theorem 10:2 are fulfilled, and therefore the conclusion of Theorem 10:1 is established.

On Linear Inequalities

By LLOYD L. DINES, F.R.S.C. and N.H. MCCOY

1. *Introduction.* Linear inequalities were studied with some degree of generality at least as early as the time of Fourier (1824). However the first significant contribution to their theory was made by Minkowski in his *Geometrie der Zahlen* in 1896. Since that time many papers have appeared in Europe, America, and Japan which have to do more or less directly with the subject. But some of these have been published in places unexpected or not easily accessible, and no general survey has appeared which attempts to take account of all of them. The objects of the present paper are: (1) to give a somewhat comprehensive account of the subject as it has been developed to date, (2) to supplement and extend the theory by some new results, and (3) to supply a more complete bibliography than has yet appeared. This bibliography is to be found at the end of the paper, and references will frequently be made to the sources there listed. To save space in the paper, which is necessarily long, detailed proofs will be omitted when they are satisfactory and easily accessible in the original sources.

2. *Different Types of Algebraic Systems.* Not all of the authors whose works we shall consider have restricted the algebraic systems to pure inequalities. The greater number in fact have treated systems which involve relations expressed by both of the symbols $>$ and $=$. If we consider a system of real linear forms

$$l_i(x) \equiv \sum_{j=1}^{n} a_{ij} x_j, \quad (i=1, 2, \ldots, m),$$

the various types of algebraic systems that have been treated are the following five:

$(a)^1$ $l_i \geqq 0, \quad (i=1, 2, \ldots, m)$

$(b)^2$ $l_i > 0, \quad (i=1, 2, \ldots, m)$

$(c)^3$ $l_i >' 0, \quad (i=1, 2, \ldots, m)$

---

[1]Minkowski 1, and Haar 1.

[2]Dines 1, and Carver 1.

[3]Dines, 3 and 11. The symbol $>'$ has the meaning of $\geqq$ reinforced by the requirement that the inequality shall hold for at least one value of the index i.

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_17, © Springer Basel 2014

$$(d)^4\ l_i \left| \begin{array}{ll} >0, & (i=1, 2, \ldots, p) \\ =0, & (i=p+1, p+2, \ldots, m) \end{array} \right.$$

$$(e)^5\ l_i \left| \begin{array}{ll} \geqq 0, & (i=1, 2, \ldots, p) \\ >0, & (i=p+1, p+2, \ldots, q) \\ =0, & (i=q+1, q+2, \ldots, m)^6. \end{array} \right.$$

Of course these five types are not mutually exclusive. Indeed (e) can be interpreted so as to include all of the others with the possible exception of (c). Nevertheless the less general types are worthy of some special consideration, inasmuch as they give rise to certain theorems which do not appear in connection with the general types, and it is for the most part these theorems which can be carried over to the transcendental cases.

It may be noted also that, from one point of view, only a superficial generality is added by the inclusion of the *equations* in the systems (d) and (e), since these equations can be used immediately to reduce the dimensionality of the problem by elimination of certain of the variables x.

3. *Types of Results Obtained.* The principal results which have been obtained relative to a system of inequalities of any one of the above types may be classified as follows: (1) conditions for the existence of a solution, (2) algorithms for obtaining solutions when such exist, (3) representation of the general solution, (4) relationship between the given system and the adjoint system of equations, and (5) conditions for the mutual dependence or independence of the system.

These five types of results are not mutually exclusive but they indicate the general directions along which the study of systems of linear inequalities has been carried out.

4. *Geometric Interpretations.* For a system of linear inequalities a number of different geometric interpretations are available. Probably the simplest is the one obtained by associating with each linear form $l(x)$ the $(n-1)$-flat or hyperplane defined by the equation $l(x)=0$ in the n-dimensional space of which the general point has coordinates $(x_1, x_2, \ldots, x_n)$. The inequality $l(x)>0$ then restricts

[4]Schlauch 1.

[5]Stokes 1.

[6]Systems of non-homogeneous inequalities of types (a), (b) and (e) have also been studied but each of these systems is essentially equivalent to a properly chosen homogeneous system of one of the above types. Cf. Dines 11, p. 396; Stokes 1, p. 802.

the point $(x_1, x_2, \ldots, x_n)$ to lie in a definite one of the two regions in which the n-space is divided by the hyperplane $l(x)=0$. The weaker condition $l(x) \geqq 0$ defines the same set of points augmented by the points of the bounding hyperplane. The solution of any one of the systems (a), (b), (d), (e) is then represented geometrically by the aggregate of points common to the regions defined by the several conditions of the system. This interpretation, despite its simplicity and suggestiveness, has not proven so directly useful in the theory as certain others which will now be described.

In the second interpretation, the m sets of coefficients

$$(1) \qquad (a_{i1}, a_{i2}, \ldots, a_{in}), \quad (i=1, 2, \ldots, m),$$

of the given linear forms are represented by m fixed points in n-dimensional space. In this space the hyperplanes which contain the origin may be represented by equations of the form

$$L(y) \equiv x_1y_1 + x_2y_2 + \ldots + x_ny_n = 0,$$

in which $x_1, x_2, \ldots, x_n$ are constants. Each such hyperplane divides the n-space into two regions, in one of which the corresponding linear form $L(y)$ is positive while in the other $L(y)$ is negative. A solution of a system of inequalities

$$(b) \qquad \sum_{j=1}^{n} a_{ij}x_j > 0, \quad (i=1, 2, \ldots, m),$$

is then a set of values $(x_1, x_2, \ldots, x_n)$ such that each of the m given points (1) lies on that side of the hyperplane $L(y)=0$ on which $L(y)>0$. For the other types of systems (a), (d), and (e) the necessary modifications are obvious. This geometric interpretation has been used by at least two authors, Haar and Stokes.

A third geometric interpretation[7] results from considering the given coefficients as components of a vector in n-space. Thus we consider m fixed vectors

$$\alpha_i \equiv (a_{i1}, a_{i2}, \ldots, a_{in}), \quad (i=1, 2, \ldots, m),$$

and a variable vector

$$\xi \equiv (x_1, x_2, \ldots, x_n).$$

The given linear forms may then be written

$$\alpha_i \cdot \xi, \quad (i=1, 2, \ldots, m),$$

where the notation $\alpha_i \cdot \xi$ denotes the inner product of the two vectors. The given conditions may be interpreted as restrictions upon the magnitude of the angle which the vector ξ may make with each of the given vectors a_i.

Obviously there is no loss of generality in assuming that the vectors are all normalized, that is that

$$a_{i1}^2 + a_{i2}^2 + \dots + a_{in}^2 = 1, \quad (i = 1, 2, \dots, m)$$
$$x_1^2 + x_2^2 + \dots + x_n^2 = 1.$$

If this be done, the problem may be interpreted in terms of the geometry of the spherical surface

$$x_1^2 + x_2^2 + \dots + x_n^2 = 1.$$

The m given vectors a_i, $(i = 1, 2, \dots, m)$ determine m fixed points on this sphere. The condition $a_i . \xi \geqq 0$ (or $a_i . \xi > 0$) restricts the point determined by the vector ξ to lie on the bounded (or unbounded) hemisphere of which a_i is the pole. And the solution to the problem is represented by the points of the region common to the m hemispheres.

5. *On the Smallest Convex Region Containing a Given Set of Points.* Before leaving the general topic of geometric interpretations it seems best to discuss briefly the above motion which has a very significant relation to linear inequalities, and of which use will be made in later parts of the paper. The notion seems to have been introduced by Minkowski, though he makes no use of it in his treatment of inequalities. In two and three dimensional spaces it has the naive meaning suggested by its name. This meaning has been extended logically to n-dimensional spaces by Minkowski and Carathéodory.[8]

To arrive at an understanding of this extension, let us consider any closed set of points M in n-dimensional Euclidian space. For any point P in this space it is a definite question whether it is possible to pass through P an $(n-1)$-flat which neither contains points of M nor separates points of M. The aggregate of points P for which the answer is negative constitutes the *smallest convex region* containing the set M, and will be denoted by R_M. It can be shown that the point set R_M is perfect, that is it coincides with the set of all limit points of points of the set. A point of R_M is either a *boundary* point or an *inner* point, according as it is or is not a limit point of points not belonging to R_M. A further distinction between boundary points and inner points of R_M is that through each of the former it is possible to pass an $(n-1)$-flat which does not *separate* points of

[7]Dines 2, p. 58.

[8]Carathéodory, 1.

R_M, while every $(n-1)$-flat through an inner point does separate points of R_M.

It is now easy to see the significance of this geometric notion with reference to linear inequalities. If we denote by M the set of m points representing the m sets of coefficients

$$a_i \equiv (a_{i1}, a_{i2}, \ldots, a_{in}), \quad (i=1, 2, \ldots, m)$$

of the m linear forms

$$l_i(x) \equiv \sum_{j=1}^{n} a_{ij}x_j, \quad (1=1. 2, \ldots, m),$$

then the question of the existence of a solution of the corresponding system of linear inequalities depends entirely upon the position of the origin $(0, 0, \ldots, 0)$ relative to the smallest convex region containing M. For the $(n-1)$-flats through this origin have equations of the form

$$L(y) \equiv x_1y_1 + x_2y_2 + \ldots + x_ny_n = 0.$$

If the origin does not belong to R_M, then there is such an $(n-1)$-flat which does not contain nor separate points of M, and so there is a form $L(y)$ such that $L(a_i)$ is positive for every a_i. The coefficients $x=(x_1, x_2, \ldots, x_n)$ of this form satisfy the system of inequalities $l_i(x) > 0$, $(i=1, 2, \ldots, m)$. Conversely, if this system of inequalities admits a solution, the solution furnishes coefficients of a linear form $L(y)$ such that the $(n-1)$-flat $L(y)=0$ does not contain nor separate points of M. Hence the origin does not belong to R_M.

An analogous argument can be made relative to the less restrictive condition that the origin be not an *inner* point of R_M. These considerations lead to the following results[9]

(i) *The system of inequalities*

$$(a) \qquad l_i \geqq 0, \quad (i=1, 2, \ldots, m)$$

admits a non-trivial solution[10] *if and only if the origin is not an inner point of* R_M.

(ii) *The system*

$$(b) \qquad l_i > 0, \quad (i=1, 2, \ldots, m)$$

admits a solution if and only if the origin is not a point of R_M.

[9]Cf. Fujiwara 3, p. 331.

[10]The solution $(0, 0, \ldots, 0)$ of the system (a) is called the trivial solution. Any other solution will be said to be non-trivial.

Carathéodory obtains also the following characterization of the set R_M, which will be recognized as obvious for the cases $n=2$ and $n=3$. The smallest convex region containing a given closed set of points M consists of those points, each of which can be the centroid of a distribution of positive masses (with total mass unity) at properly chosen points of M. He shows further that at most $n+1$ positive masses need to be thus distributed to determine any given point of R_M as a centroid. And indeed if $n+1$ such positive masses are distributed at $n+1$ points of M which do not lie in a common $(n-1)$-flat, then the centroid so determined is an *inner* point of R_M. This can all be expressed analytically as follows:

The smallest convex region R_M containing a given closed set of points M in n-dimensional space consists of the points $\eta \equiv (\eta_1, \eta_2, \ldots, \eta_n)$ whose coordinates are expressible in the form

$$\eta_j = \sum_i m_i a_{ij}, \quad (j=1, 2, \ldots, n) \tag{2}$$

where the points

$$a_i = (a_{i1}, a_{i2}, \ldots a_{in}) \quad (i = i_1, i_2, \ldots, i_n)$$

are a properly chosen set of at most $n+1$ points of M, and the coefficients m are positive constants of sum unity. Furthermore, those points η whose coordinates are expressible by form (2) in terms of $n+1$ points a_i for which the determinant

$$|a_{i1}, a_{i2}, \ldots, a_{in}, 1| \quad (i = i_1, i_2, \ldots, i_{n+1})$$

is different from zero, are inner points of R_M.

6. *Systems of Type (a).* We shall now take up in succession the various types of systems of linear algebraic inequalities and discuss briefly the principal theorems for each type. We consider first a system of the form

$$l_i(x) \geqq 0 \quad (i=1, 2, \ldots, m). \tag{a}$$

We have already noted in the preceding section a necessary and sufficient geometric condition for the existence of a non-trivial solution of this system. An elegant analytic theory was given by Minkowski in 1896. As a matter of fact little has been added to his theory of systems of this type, though his results have been obtained by different methods and in modified form by various authors.[11] This section will be devoted therefore to a presentation of Minkowski's theory. The first two theorems, though of prime importance are stated without

[11]Farkas 1, Haar 1, and Stokes 1.

proof inasmuch as satisfactory proof can be found in the original source. The remaining theorems are proven in some detail, since the introduction of certain new notions seems to afford greater clarity.

First of all, we may assume that among the given linear forms $l_i(x)$ there are n which are linearly independent. For suppose that only h ($h<n$) were linearly independent, and for definiteness that these were the first h forms. Then by a linear transformation

$$\begin{aligned} y_i &= l_i(x), \qquad (i=1, 2, \ldots, h) \\ y_i &= \lambda_i(x), \qquad (i=h+1, h+2, \ldots, n), \end{aligned}$$

where the $n-h$ forms $\lambda_i(x)$ are restricted only by the condition that the transformation be non-singular, the system (a) could be replaced by an equivalent one of the form

$$\begin{aligned} & y_i \geqq 0 \qquad (i=1, 2, \ldots, \mathrm{h}) \\ & \sum_{j=1}^{h} b_{ij} y_j \geqq 0 \quad (1=h+1, h+2, \ldots, m), \end{aligned}$$

which contains only the h variables $y_1, y_2, \ldots, y_h$. It will therefore be assumed in what follows that the number of linearly independent forms in (a) is equal to the number of variables.

We are now ready to develop the notion of fundamental solutions in terms of which Minkowski obtains the general solution of (a).

If $(x_1, x_2, \ldots, x_n)$ is any non-trivial solution of (a) and p is any positive constant, then $(px_1, px_2, \ldots, px_n)$ is also a solution. Two solutions such as these which differ only by a positive factor are not considered as *essentially different* solutions. If $(x'_1, x'_2, \ldots, x'_n)$ is another solution, then $(x_1+x'_1, x_2+x'_2, \ldots, x_n+x'_n)$ is a solution which we shall call the sum of these two solutions. A non-trivial solution of the system (a) which cannot be expressed as the sum of two essentially different non-trivial solutions will be called a *fundamental solution.* A method of obtaining fundamental solutions is indicated by the following important theorem:

THEOREM 1. *A necessary and sufficient condition that a non-trivial solution of the system (a) be a fundamental solution is that for this solution $n-1$ linearly independent $l_i(x)$ vanish.*

That there can be only a finite number of essentially different fundamental solutions is now clear. For if a certain set of $n-1$ linearly independent $l_i(x)$ vanish for a given solution, this solution is thereby determined to within a positive factor and any other solution for which these same forms vanish is not essentially different from the first. Hence there are at most as many essentially different fundamental solutions as there are different sets of $n-1$ linearly independent forms in the given system.

Let

$$X^{(i)}=(X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)}), \quad (i=1, 2, \ldots, N) \tag{3}$$

be a set of essentially different fundamental solutions of the system (a) such that any fundamental solution of the system is a positive multiple of one of these. Such a set of solutions is said to be a *complete set* of fundamental solutions. If we add the conditions,

$$\sum_{k=1}^{n} (X_k^{(i)})^2=1 \quad (i=1, 2, \ldots, N)$$

the set (3) may be said to be the complete set of *normalized* fundamental solutions of the system (a). This set is determined uniquely by the given system. The significance of the concept of complete set of fundamental solutions is brought out in the

THEOREM 2. *If*

$$(X_1^{(i)}, X_2^{(i)}, \ldots X_n^{(i)}), \quad (i=1, 2, \ldots, N),$$

is a complete set of fundamental solutions of the system (a), *then the general solution is*

$$X_j=p_1X_j^{(1)}+p_2X_j^{(2)}+\ldots+p_NX_j^{(N)}, \quad (j=1, 2, \ldots, n)$$

where the p*'s are non-negative parameters. That is, any solution of the system is a linear combination, with non-negative coefficients, of a complete set of fundamental solutions.*

We have seen that there can be only a finite number of fundamental solutions. There may be none, in which case of course the system admits no non-trivial solution. Further information as to the number of fundamental solutions is contained in the following:

THEOREM 3. *If* (i), *the system* (a) *admits at least one fundamental solution and* (ii), *no one of the forms* $l_i(x)$ *vanishes for every fundamental solution, then the system admits at least* n *linearly independent fundamental solutions.*

Before indicating the proof of the theorem, we note that a system which satisfies hypothesis (i) but not hypothesis, (ii) can be reduced by a suitable linear transformation to an equivalent system which explicitly requires that some of the variables be equal to zero.[12] If these variables be put equal to zero, the resulting system in $n'(<n)$

[12]Minkowski 1, p. 44.

variables will satisfy both hypotheses, and the theorem will apply to the new system with n replaced by n'.

Now to prove the theorem, we note first that the hypotheses imply the existence of a solution $a=(a_1, a_2, \ldots, a_n)$ such that each $l_i(a)$ is *positive*. Hence from the continuity of the linear forms it follows that for every $x=(x_1, x_2, \ldots, x_n)$ in a sufficiently restricted neighbourhood of a, the forms $l_i(x)$ will be positive. But in this restricted, but still n-dimensional neighbourhood, there are certainly n linearly independent sets $x^{(j)}=(x_1^{(j)}, x_2^{(j)}, \ldots, x_n^{(j)})$. And these n linearly independent solutions cannot be expressed as linear combinations of less than n linearly independent fundamental solutions. So the theorem follows from Theorem 2.

If the system (a) admits non-trivial solutions, there are two other systems of inequalities uniquely determined by (a), which are intimately related to it and useful in the study of the properties of (a). These we will now define, assuming as we may that hypothesis (ii), as well as (i), of Theorem 3 is satisfied. Using the complete set of *normalized* fundamental solutions

$$(X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)}), \quad (i=1, 2, \ldots, N)$$

of (a) as coefficients, we form the system of inequalities

$$(a') \qquad \sum_{j=1}^{n} X_j^{(i)} y_j \geqq 0, \quad (i=1, 2, \ldots, N)$$

which will be called the *canonical polar system* of (a). It will be useful to note the following obvious relationship between a system and its canonical polar system: *The set of coefficients of each linear form in one system is a solution of the other system.*

Now using the complete set of normalized fundamental solutions of (a') as coefficients, we form the canonical polar system of (a')

$$(a'') \qquad \sum_{j=1}^{n} Y_j^{(i)} x_j \geqq 0, \quad (i=1, 2, \ldots, M)$$

which will be called the *canonical form of system* (a). The name is justified by the following:

THEOREM 4. *If the hypotheses of Theorem 3 are satisfied, the system* (a) *and its canonical form* (a'') *admit precisely the same solutions. Furthermore the linear forms in* (a'') *are identical except for positive numerical factors with a certain subset of the linear forms in* (a).

To prove the theorem we observe first that the system (a'') admits as solutions the sets of coefficients $(X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)})$ of (a')

and hence admits all solutions of (a). Conversely, every solution of (a'') satisfies (a). For suppose $\bar{x}\equiv(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ is a solution of (a''). Then it is a solution of every inequality

$$(4) \qquad a_1x_1+a_2x_2+\ldots+a_nx_n \geqq 0$$

of which the left member is a linear combination of the left members of (a'') with non-negative coefficients. But this is equivalent to saying that $\bar{x}$ is a solution of (4) if the coefficients $a_1, a_2, \ldots, a_n$ are solutions of the system (a'). Since the coefficients of each linear form in (a) constitutes a solution of (a') we conclude that $\bar{x}$ is a solution of (a).

It remains to prove that to each linear form in (a'') there corresponds one in (a) which differs from it only by a positive numerical factor. Let $Y=(Y_1, Y_2, \ldots, Y_n)$ be the set of coefficients of any one of the forms in (a''), and suppose for definiteness that this particular fundamental solution of (a') satisfies the $n-1$ equations

$$(5) \qquad \sum_{j=1}^{n} X_j^{(i)} y_j=0, \quad (i=1, 2, \ldots, n-1).$$

Consider the particular solution $\bar{x}=(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$ of (a) obtained from the linear combination of fundamental solutions

$$\bar{x}\equiv X^{(1)}+X^{(2)}+\ldots+X^{(n-1)}.$$

This solution cannot make all of the linear forms $l_i(x)$ *positive*. For if it did there would exist an n-dimensional neighbourhood of $\bar{x}$ which would consist entirely of solutions of (a), and therefore of solutions of (a''). But on the other hand, since

$$\sum_{j=1}^{n} Y_j\bar{x}_j=0,$$

there are some points in every n-dimensional neighbourhood of $\bar{x}$ for which $\sum_{j=1}^{n} Y_j x_j$ is negative. The contradiction proves that at least one of the forms $l(x)$ vanishes for the solution $\bar{x}$, and so for each of the fundamental solutions $X^{(i)}$, $(i=1, 2, \ldots, n-1)$. The coefficients of this linear form then satisfy the same $n-1$ linear homogeneous equations (5) which determine the set $(Y_1, Y_2, \ldots, Y_n)$. The two sets of coefficients are therefore proportional, and the constant of proportionality must be positive since the solutions of (a) and (a'') are the same. This completes the proof of Theorem 4.

From the relations existing between the given system (a) and its canonical polar system (a') and its canonical form (a'') there follow almost immediately certain results obtained by Minkowski, Farkas, and others.

COROLLARY 1. *If the system* (a) *admits non-trivial solutions, a necessary and sufficient condition that all such solutions shall satisfy a given inequality*

$$\phi(x) \equiv b_1x_1 + b_2x_2 + \ldots + b_nx_n \geqq 0 \tag{6}$$

is that $\phi(x)$ *be expressible in the form*

$$\phi(x) \equiv c_1l_1(x) + c_2l_2(x) + \ldots + c_ml_m(x),$$

where the c_i *are non-negative constants.*

The sufficiency of the condition is obvious. The necessity follows from the fact that if (6) is satisfied by all solutions of (a) then the coefficients of ϕ must satisfy the canonical polar system (a'), and hence ϕ must be a linear combination with non-negative coefficients of the left members of the canonical form (a''). It will be noted that the hypothesis (ii) of Theorem 3 is not essential for Corollary 1. Nor indeed is our general assumption that the system (a) include at least n linearly independent forms. For both of these properties can be secured by appropriate non-singular linear transformations, and the required form of expression for ϕ persists under such transformations.

An inequality $\phi \geqq 0$ which is satisfied by all solutions of a system (a) is said to be a *consequence* of (a). A system in which no one of the inequalities is a consequence of the others may be called an *independent system*.[13] We have then the following condition for independence:

COROLLARY 2. *A system* (a) *is independent if and only if it is identical with its canonical form except for positive numerical factors of the individual linear forms.*

7. *Systems of Type* (b). Let us consider now a system of linear inequalities of the form,

$$l_i(x) > 0, \quad (i = 1, 2, \ldots, m). \tag{b}$$

Fourier studied systems of this type and in 1824 gave an outline of a method of solving such systems by successive elimination of the unknowns, In 1919, Dines[14] considered the problem independently

[13] Cf. Farkas 1.

[14] Dines 1. For a more complete general discussion than that given here see also Dines 11 where extensions to transcendental inequalities are also considered.

from this same point of view and succeeded in systematizing the procedure and formulating the conclusions in a manner which brings out the analogy with the theory of systems of linear equations.

The central feature is the introduction of an integral-valued function of the coefficients of the system which is called the *inequality-rank* or I-*rank* of the matrix $||a_{ij}||$. The I-rank of this matrix may have any one of the values $0, 1, \ldots, n$. The importance of this notion of I-rank is indicated by:

THEOREM 5. *A necessary and sufficient condition for the existence of a solution of the system of inequalities* (*b*) *is that the I-rank of the matrix* $||a_{ij}||$ *be greater than zero. If the I-rank of this matrix is* k (>0), *then the system admits a solution in which* $k-1$ *of the* x_j *are arbitrary.*

We shall not give here a definition of the I-rank but we may remark that the process used in defining the I-rank furnishes a satisfactory algorithm for obtaining the general solution of the system.

The system (*b*) can also be profitably considered in its relationship to the system (*a*), as was remarked by Minkowski. Obviously all solutions of (*b*) are included among the solutions of (*a*). If we assume here, as in the preceding section, that the number of linearly independent forms is equal to the number of variables, the following theorem can be proven.

THEOREM 6.[15] *Necessary and sufficient conditions for the existence of a solution of* (*b*) *are that the system* (*a*) *admit fundamental solutions, and that no one of the linear forms* $l_i(x)$ *vanishes for every fundamental solution of* (*a*). *If these conditions are satisfied, the general solution of* (*b*) *is*

$$x_j = p_1X_j^{(1)} + p_2X_j^{(2)} + \ldots + p_NX_j^{(N)} \quad (j=1, 2, \ldots, n)$$

where

$$(X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)}), \quad (i=1, 2, \ldots, N)$$

is a complete set of fundamental solutions of (*a*), *and the* p'*s are positive parameters.* It will be recalled that in the general solution of (*a*) the parameters are only required to be non-negative.

Still another criterion for the existence of a solution of (*b*) can be obtained by translating into algebraic terms the geometric criterion obtained in § 5. This becomes:

[15]Cf. Stokes 1, p. 793. The firat part of the theorem is obvious, but the second part is not. Miss Stokes obtains it by the geometric method which she uses throughout her paper. An algebraic proof will be found as a corollary to Theorem 12 of the present paper.

THEOREM 7.[16] *A necessary and sufficient condition for the existence of a solution of (b) is that the forms $l_i(x)$ satisfy no identical linear relation*

$$c_1l_1(x)+c_2l_2(x)+\ldots+c_ml_m(x)=0,$$

in which the coefficients c are non-negative constants of unit sum.

This theorem can be thrown into another interesting form which brings into evidence the relationship between a system of linear inequalities

$$(b) \qquad \sum_{j=1}^{n} a_{ij}x_j>0 \quad (i=1, 2, \ldots, m)$$

and what may be called the adjoint or associated system of linear equations

$$(b') \qquad \sum_{i=1}^{m} a_{ij}y_i=0 \quad (j=1, 2, \ldots, n).$$

A few preliminary definitions are essential. A set of real numbers $(x_1, x_2, \ldots, x_n)$ will be said to be *positive* if each x_j is positive, and *negative* if each x_j is negative. In either case the set may be said to be *definite.* The set will be said to be *M-positive* (mildly positive) if at least one element is positive and none are negative; and *M-negative* (mildly negative) if at least one is negative and none are positive. In either of the latter cases the set may be said to be *M-definite.*

THEOREM 8. *A necessary and sufficient condition for the existence of a solution $(x_1, x_2, \ldots, x_n)$ of the system of inequalities (b) is that the associated system of linear equations (b′) admit no M-definite solution $(y_1, y_2, \ldots, y_n)$.*

The following two theorems, due essentially to Carver, are analogues of Corollaries 1 and 2 of the preceding section, and can be proven with little difficulty from the theorems already developed.

THEOREM 9. *If the system (b) admits solutions, a necessary and sufficient condition that all such solutions shall satisfy a given inequality*

$$\phi(x)\equiv b_1x_1+b_2x_2+\ldots+b_nx_n>0$$

is that $\phi(x)$ be expressible in the form

$$\phi(x)=c_1l_1(x)+c_2l_2(x)+\ldots+c_ml_m(x),$$

where the coefficients $(c_1, c_2, \ldots, c_m)$ form an M-positive set.

[16]Obtained in somewhat different form independently by Stiemke and Carver.

C4

THEOREM 10. *If two consistent systems of type* (*b*) *are independent and equivalent, then the two systems contain the same number of inequalities, and the linear forms in one system are identical except for positive numerical factors with those in the other.*

Finally we note the following results, also due to Carver.

THEOREM 11. *If the matrix* $\|a_{ij}\|$ *of coefficients is of rank* m, *the system* (*b*) *certainly admits a solution. If the rank of this matrix is* $r(<m)$, *and the system* (*b*) *admits no solution, then there is a subsystem consisting of at most* $r+1$ *of the given inequalities which admits no solution.*

The truth of the first statement is obvious if the inequalities be thought of as equations with positive right members.

The truth of the second statement is most easily established by consideration of the set of m points

$$M: \qquad (a_{i1}, a_{i2}, \ldots, a_{in}), \quad (i=1, 2, \ldots, m).$$

If the rank of the matrix $\|a_{ij}\|$ is r, these points lie in a common r-flat through the origin and hence the smallest convex region R_M is contained in this r-dimensional manifold. Furthermore, if the system (*b*) admits no solution, the origin is contained in R_M. Hence the coefficients a_{ij} satisfy relations of the form

$$\sum_i m_i a_{ij} = 0, \quad (j=1, 2, \ldots, n)$$

where the summation index i ranges over some subset $i_1, i_2, \ldots, i_{r+1}$ of the numbers $1, 2, \ldots, m$; and the multipliers m_i form an M-definite set. From Theorem 8 it then follows that the subsystem of inequalities $l_i(x) > 0$, $(i=i_1, i_2, \ldots, i_{r+1})$ admits no solution.

8. *Systems of Type* (*c*), *and Positive Solutions of Linear Equations.* The systems considered in this section are of the form

$$(c) \qquad l_i(x) >' 0, \quad (i=1, 2, \ldots, m),$$

and differ from those of type (*a*) only in that the *in*equality is required to hold in at least one instance. If the matrix of coefficients is of rank n the two systems are entirely equivalent except that (*a*) admits the trivial solution $(0, 0, \ldots, 0)$, which does not satisfy the requirements of (*c*). Since, as we have seen in § 6 the given system can always be reduced by a non-singular linear transformation to an equivalent one in which the number of variables is equal to the rank of the matrix, the system (*c*) from one point of view offers no new interest. Of very particular interest however is the case in which it

[17] Cf. Stiemke 1, and Dines 3.

admits *no* solution, inasmuch as that turns out to be precisely the case in which the associated system of linear equations admits a *positive* (and hence of course also a negative) solution. The relationship is expressed in the following theorem obtained[17] independently by Stiemke and Dines, a theorem for which interesting analogues and generalizations are found in the transcendental cases.

THEOREM 12. *A necessary and sufficient condition that the system of inequalities*

$$(c) \qquad \sum_{j=1}^{n} a_{ij}x_j >' 0 \quad (i=1,2,\ldots,m)$$

admit no solution $(x_1, x_2, \ldots, x_n)$ *is that the associated system of equations*

$$(c') \qquad \sum_{i=1}^{m} a_{ij}y_i = 0 \quad (j=1, 2, \ldots, n)$$

admit a definite solution $(y_1, y_2, \ldots, y_m)$.[18]

The system of inequalities (c) is equivalent to the system of equations

$$(7) \qquad \sum_{j=1}^{n} a_{ij}x_j = b_i \quad (i=1, 2, \ldots, m)$$

in which the constants b_i form an M-positive set but are otherwise arbitrary. If the rank of the matrix $\| a_{ij} \|$ is m this system obviously admits a solution, while the associated system of equations (c') admits only the trivial solution $(0, 0, \ldots, 0)$. The theorem is therefore true in this case.

Suppose that the rank of this matrix is r $(<m)$. Then the system of homogeneous equations (c') admits $m-r$ linearly independent solutions which we may denote by

$$(y_{1\kappa}, y_{2\kappa}, \ldots, y_{m\kappa}), \quad (\kappa=1, 2, \ldots, m-r).$$

And the non-homogeneous equations (7) admit a solution if and only if the constants b_i satisfy the $m-r$ conditions

$$(8) \qquad \sum_{i=1}^{m} y_{i\kappa} b_i = 0 \quad (\kappa=1, 2, \ldots, m-r).$$

The system of inequalities (c) therefore admits a solution if and only if the system of equations (8) admits an M-positive solution. But by Theorem 8, the necessary and sufficient condition for this is that the associated system of inequalities (of type (b))

[18]The proof given below is much simpler than the early proofs of Stiemke and Dines. The method is essentially that used in proving an analogous theorem in General Analysis, Dines 7. Cf. also Fujiwara 3.

$$\sum_{\kappa=1}^{m-r} y_{i\kappa} c_\kappa > 0 \quad (i=1, 2, \ldots, m) \tag{9}$$

admit no solution $(c_1, c_2, \ldots, c_{m-r})$.

Now the left members of (9) for arbitrary values of the c_κ's represent the totality of solutions $(y_1, y_2, \ldots, y_m)$ of the system of equations (c'). Hence a necessary and sufficient condition that the system (c) admit a solution is that the system (c') admit no positive solution. This is equivalent to the theorem stated.

This theorem indicates the close relationship between the theory of linear inequalities of type (c) and the theory of *positive* solutions of linear equations. For further developments of this idea the reader may refer to Dines 3, 4 and 6.

We are now in a position to give the deferred proof of Theorem 6, the statement of which we repeat in the following form:

COROLLARY. *If the system (b) admits solutions and the number of linearly independent forms $l_i(x)$ is equal to the number of variables, then the general solution of the system (b) is*

$$x_j = p_1 X_j^{(1)} + p_2 X_j^{(2)} + \ldots + p_N X_j^{(N)}, \quad (j=1, 2, \ldots, n)$$

where

$$(X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)}), \quad (i=1, 2, \ldots, N)$$

is a complete set of fundamental solutions of the system (a), and the p's are positive parameters.

Let

$$\sum_{j=1}^{n} X_j^{(i)} y_j \geqq 0, \quad (i=1, 2, \ldots, N) \tag{a'}$$

denote the canonical polar system of (a), and

$$\sum_{j=1}^{n} Y_j^{(i)} x_j \geqq 0, \quad (i=1, 2, \ldots, M) \tag{a''}$$

the canonical form of system (a). If $(y'_1, y'_2, \ldots, y'_n)$ is any non-trivial solution of (a') we have by Theorem 2,

$$y'_j = q_1 Y_j^{(1)} + q_2 Y_j^{(2)} + \ldots + q_M Y_j^{(M)}, \quad (j=1, 2, \ldots, n)$$

where the q's form an M-positive set. Thus if $(x'_1, x'_2, \ldots, x'_n)$ is any solution of the system (b), and thus by Theorem 4 a solution of (a'') with the *in*equality holding in each instance, we see that

$$\sum_{j=1}^{n} x'_j y'_j > 0.$$

Hence the system of inequalities

$$\sum_{j=1}^{n} X_j^{(i)} y_j \geqq 0, \quad (i=1, 2, \ldots, N)$$

$$-\sum_{j=1}^{n} x'_j y_j \geqq 0$$

admits no non-trivial solution. By the above theorem we may therefore write

$$x'_j = p_1 X_j^{(1)} + p_2 X_j^{(2)} + \ldots + p_N X_j^{(N)} \quad (j=1, 2, \ldots, n)$$

with each $p_i > 0$, which is the desired result.

Theorem 12 and the theorem immediately to follow are analogous respectively to Theorem 8 and 11 of the preceding section.

Theorem 13. *If the matrix* $\| a_{ij} \|$ *is of rank* m, *the system of inequalities* (c) *certainly admits a solution. If the system* (c) *admits no solution, there is a subsystem*

$$\sum_{j=1}^{n} a_{ij} x_j >' 0, \quad (i=i_1, i_2, \ldots, i_s) \tag{10}$$

of s ($\leqq m$) *inequalities which admits no solution, and of which the matrix of coefficients is of rank* $s-1$.

The first statement is obviously true, and the second is also if the matrix of (c) happens to be of rank $m-1$. Let us suppose then that the matrix of (c) is of rank r ($< m-1$); in which case a non-singular linear transformation will reduce the system (c) to an equivalent system

$$\sum_{j=1}^{r} a'_{ij} x'_j >' 0, \quad (i=1, 2, \ldots, m),$$

in which there are only r variables.

Since by hypothesis this system admits no solution, the associated system of equations admits a positive solution (by the preceding theorem). That is the origin is the centroid of suitable positive masses distributed at the m points $(a'_{i1}, a'_{i2}, \ldots, a'_{ir})$ in space of r dimensions. But from the results of Carathéodory (§5), it follows that the origin is then centroid of suitable positive masses distributed at a subset of not more than $r+1$ of these points. Hence there is a system of equations

$$\sum_i a'_{ij} y'_i = 0 \quad (j=1, 2, \ldots, r)$$

in which the subscript i takes only k ($\leqq r+1$) of the values $1, 2, \ldots, m$, which admits a positive solution. The associated system of inequalities

$$\sum_{j=1}^{r} a'_{ij}\, x'_j >' 0, \quad (i = i_1, i_2, \ldots, i_\kappa)$$

therefore admits no solution.

The corresponding inequalities of the system (c) therefore form a subsystem which admits no solution. If the rank of the matrix of this subsystem is $k-1$, we have reached the desired conclusion. If its rank is less than $k-1$, the procedure can by repeated successively until a subsystem satisfying the required condition is obtained.

COROLLARY. *The number s of inequalities in the subsystem* (10) *is not greater than $n+1$.* For the rank $s-1$ cannot possibly be greater than n.

9. *System of Types* (d) *and* (e). A system of inequalities of the form,

$$(e) \qquad l_i(x) \left| \begin{array}{ll} \geqq 0, & (i=1, 2, \ldots, p) \\ > 0, & (i=p+1, p+2, \ldots, q) \\ = 0, & (i=q+1, q+2, \ldots, m) \end{array} \right.$$

includes those of types (a), (b) and (d) as special cases. Systems of type (d) have been studied by Schlauch and those of type (e) by Stokes. We limit ourselves to a statement of the main theorem obtained by Stokes.

We assume that there are n linearly independent $l_i(x)$ in the given system. For convenience denote the set $l_i(x)$, $(i=1, 2, \ldots, p)$ by P; the set $l_i(x)$, $(i=p+1, p+2, \ldots, q)$ by Q and the set $l_i(x)$, $(i=q+1, q+2, \ldots, m)$ by R. Then a solution of the system (e) is a set of real numbers $(x'_1, x'_2, \ldots, x'_n)$ such that for these values of the variables each member of P is non-negative, each member of Q is positive and each member of R vanishes. Necessary and sufficient conditions that the set (e) be consistent are that (i) there exist one or more fundamental solutions

$$(X_1^{(i)}, X_2^{(i)}, \ldots, X_n^{(i)}), \quad (i=1, 2, \ldots, s)$$

of the system (a) for each of which each member of R vanishes, and (ii) no member of Q vanishes for all such fundamental solutions.

Suppose now that the system (e) is consistent and assume for convenience of notation that all the members of Q vanish for the first r ($<s$) of the above mentioned fundamental solutions and only for these. *Then the general solution of the system* (e) *is*

$$x_j = q_1X_j^{(1)} + q_2X_j^{(2)} + \dots + q_rX_j^{(r)} + p_{r+1}X_j^{(r+1)} + \dots + p_sX_j^{(s)}, \quad (j = 1, 2, \dots, n)$$

where the q's are non-negative parameters and the p's are positive parameters.[19]

10. *Generalizations.* Just as the theory of systems of linear algebraic equations has led by extension and analogy to theories of systems of linear equations in infinitely many unknowns and to theories of linear integral equations, so the developments described in the preceding sections suggest analogous theories of transcendental inequalities.

A more comprehensive view of the possibilities is obtained however if we state first an extreme generalization of the algebraic problem which will include the more obvious analogies as special cases. This is the method used by E. H. Moore in his General Analysis.

We note first that the generalization may take either of two quite distinct forms, depending on whether the linear algebraic forms upon which it is based be written

$$(11) \qquad l_i(x) \equiv \sum_{j=1}^{n} a_{ij}x_j \quad (i = 1, 2, \dots, m)$$

or equivalently

$$(12) \qquad l_i(x) \equiv x_i + \sum_{j=1}^{n} \bar{a}_{ij}x_i \quad (i = 1, 2, \dots, m).$$

To obtain the generalization, we replace the variable subscripts i, j by two general variables p, q, whose ranges are classes of elements P and Q respectively, and replace the given matrix a_{ij} (or $\bar{a}_{ij}$) of coefficients by a given real valued function $\alpha(p, q)$ and the set of unknowns x_j by a function $\xi(q)$ to be determined. The summation symbol $\sum_j$ is replaced by a linear operator J_q of such a nature that the result of the operation $J_q\alpha(p, q)\xi(q)$ is a real valued function of the variable p. The generalizations of the sets of forms (11) and (12) may then be written respectively

$$(13) \qquad l[\xi|p] \equiv J_q\alpha(p, q)\xi(q), \quad (p \textit{ on range } P)$$

and

$$(14) \qquad l[\xi|p] \equiv \xi(p) + J_q\alpha(p, q)\xi(q), \quad (p \textit{ on range } P).$$

To each of these forms there corresponds a problem generalizing each of the types (a), (b), (c), (d), and (e) in § 1. By analogy with integral

[19]Stokes 1, p. 792.

equation nomenclature, the inequalities arising from the forms (13) and (14) may properly be called linear inequalities of the *first kind* and *second kind* respectively.

As in the case of integral equations, so with these general inequalities, the *first* kind offers the greater difficulty, and little has been done with this problem in its generality. The only known result[20] for entirely general variables p, q generalizes Theorem 12 of § 8. The given kernel function $\alpha(p, q)$ is assumed to be representable in the form

$$\alpha(p, q) = \sum_{j=1}^{m} \mu_i(p)\nu_j(q), \tag{15}$$

where the functions $\mu_i(p)$ and $\nu_i(q)$ belong respectively to classes of real-valued functions M and N, each of which possess the property of linearity,[21] Two linear functional operators J_p and J_q operate on product functions of the types $\mu'(p)\mu''(p)$ and $\nu'(q)\nu''(q)$ respectively, yielding in each case a definite real number. In addition the operators and classes of functions satisfy the following three postulates:

P. For every function ν of N, $J_q\nu(q)\nu(q) \geqq 0$, the equality holding only if $\nu(q) \equiv 0$.

N. If $\mu(p)$ is an M-definite function (that is if $\mu(p)$ is not identically zero and does not change sign on P) and $\pi(p)$ is a function of M which is everywhere positive, then $J_p\mu(p)\pi(p) \neq 0$.

E. If $(\mu_1, \mu_2, \ldots, \mu_m)$ is a set of functions of M such that no linear combination of them is M-definite, then there exists a positive function $\pi(p)$ in M such that $J_p\mu_i(p)\pi(p) = 0$, $(i = 1, 2, \ldots, n)$ Upon this basis one obtains the following theorem:

THEOREM 14. *For a given kernel function $\alpha(p, q)$ of form (15), the inequality*[22]

$$J_q\alpha(p, q)\zeta(q) >' 0 \quad (p \text{ on range } P)$$

admits a solution $\zeta(q)$, if and only if the associated equation

$$J_q\alpha(p, q)\eta(p) = 0 \quad (q \text{ on range } Q)$$

admits no positive solution $\eta(p)$.

It is certain that the form (15) of the kernel is more restrictive than necessary, but no proof of the theorem has been devised which does not require this form.

---

[20]Dines 10, p. 145.

[21]That is, if $\mu'(p)$ and $\mu''(p)$ belong to M and a and b are real constants then $a\mu'(p) + b\mu''(p)$ belong to M.

[22]The symbol $>'$ has the meaning of $\geqq$ reinforced by the requirement that the inequality shall hold for at least one value of p.

In view of the complementary character of Theorems 8 and 12 one naturally expects to find here a generalization of Theorem 8 analogous to the above generalization of Theorem 12. The fact is however that the suggested generalization is not valid, as may be shown by an example.[23]

Turning now to the inequality of the second kind, we find the situation much simpler. No restriction of form such as (15) is necessary. The inequality of second kind of type (*b*) may be written

$$\xi(p)+J_q\alpha(p,\,q)\xi(q)>0 \quad (p \text{ on range } P), \tag{16}$$

and a satisfactory theory for this type has been obtained[24] upon the postulational basis used by E. H. Moore for his generalized integral equation theory, with the addition of a few new postulates. The reader may refer to the original paper for the complete discussion. It will be sufficient here to indicate the method of treatment, which is to replace the inequality (16) by an equivalent *equation*

$$\xi(p)+J_q\alpha(p,\,q)\,\xi(q)=\pi(p) \quad (p \text{ on range } P)$$

where the function $\pi(p)$ is an arbitrary positive function. An adequate theory of the general equation of this type has been developed upon a postulational basis by Moore and Hildebrandt.

It may be noted here that this same method of treatment will apply, to some extent at least, to the inequalities of the other types (*a*), (*c*), (*d*), and (*e*), the only difference being in the properties to be assigned to the function $\pi(p)$. But due to this difference, it is only for inequalities of type (*b*) that we are able to state the following:

THEOREM 15. *The inequality*

$$\xi(p)+J_q\alpha(p,\,q)\xi(q)>0 \quad (p \text{ on range } P)$$

admits a solution ξ *if and only if the associated equation*

$$\eta(q)+J_p\eta(p)\alpha(p,\,q)=0 \quad (q \text{ on range } Q)$$

admits no M-definite solution η.

This is a generalization of Theorem 8.

In the next section we turn our attention to specializations, particularly of the inequalities of the first kind, for which the general theory is so incomplete.

11. *Specializations; Q a Finite Range.* In this section we consider inequalities arising from the form (13) with the restrictions that Q

[23] Dines 10, p. 141.

[24] Dines 7.

is a finite range and J_q is the finite summation operator. The linear form (13) may in this case be written

$$(17) \qquad \sum_{j=1}^{n} a_j(p)\xi_j, \quad (p \text{ on range } P).$$

The range P is entirely unrestricted, the n given functions $a_j(p)$ are real valued on this range, and the symbols ξ_i, $(j=1, 2, \ldots, n)$ represent real numbers to be determined from the inequalities imposed.

The only case that has been treated[25] without further restriction is that of the inequality

$$(18) \qquad \sum_{j=1}^{n} a_j(p)\xi_j > '0 \quad (p \text{ on range } P)$$

which generalizes an algebraic system of type (c). A certain integral valued function of the set of functions $a_j(p)$ is defined and called the M-rank of the set. This is analogous to the I-rank which we have already mentioned in connection with algebraic inequalities of type (b), and the theorem obtained is analogous to Theorem 5 in §7.

We pass to a consideration of important results which can be obtained upon the additional hypothesis of a certain closure property for the functions $a_j(p)$ in (17).

Let us denote by M the set of points in n-dimensional space with coordinates

$$M \qquad (a_1(p), a_2(p), \ldots, a_n(p)), \quad (p \text{ on range } P).$$

In order that the geometric notions introduced by Carathéodory may be applicable, it is necessary to assume that this set of points be closed, that is that it contains its limit points. When this condition is satisfied, the system of forms (17) will be said to be closed, and any one of the various systems of inequalities to which it gives rise will be said to be closed. For closed systems, it is possible to generalize a number of the results which have been obtained for algebraic systems.

Haar has given for closed systems a theorem generalizing Minkowski's Consequence Theorem (our Corollary 1, § 6). But his proof is open to some criticism, inasmuch as he assumes that the system $l[\xi|p] > 0$ has a solution because the system $l[\xi|p] \geqq 0$ has non-trivial solutions. This same assumption is made and justified by Minkowski in his proof for the algebraic case, but the justification is not quite so simple in the generalization, and the assumption is certainly unwarranted if all of the functions $a_j(p)$ vanish for some one element of the range P.

[25] Dines 6.

But the argument of § 5 gives immediately the following:

THEOREM 16. *If the set of points M is closed,*
(i) *the system of inequalities*

$$\sum_{j=1}^{n} \alpha_j(p)\xi_j \geqq 0, \quad (p \text{ on range } P) \tag{19}$$

admits a non-trivial solution if and only if the origin is not an inner point of R_M, *and*
(ii) *the system*

$$\sum_{j=1}^{n} \alpha_j(p)\xi_j > 0, \quad (p \text{ on range } P) \tag{20}$$

admits a solution if and only if the origin is not a point of R_M.

The analogous theorem for the closed system (18) is not quite so simple, but merits consideration. If the n functions $\alpha_1(p)$, $\alpha_2(p)$, . . . , $\alpha_n(p)$ are linearly independent on P, this system is equivalent (except for the trivial solution) to the system (19). Suppose then that just $r(<n)$ of these functions are linearly independent on P. Then the point set M lies in an r-flat L_r through the origin, and R_M also lies in L_r. The region R_M can therefore have no inner points relative to the fundamental n-dimensional space. But let us in this case define the *inner points of* R_M *relative to* L_r to be those points of R_M which are not limit points of points of L_r not belonging to R_M. Then these inner points of R_M relative to L_r are precisely those through which it is impossible to pass an $(r-1)$-flat lying in L_r which does not separate points of M. On the basis of these notions it is not difficult to complete the proof of the following:

THEOREM 17. *If just* $r(<n)$ *of the functions* $\alpha_1(p)$, $\alpha_2(p)$, . . . , $\alpha_n(p)$ *are linearly independent on* P, *the point set* M *lies in an* r*-flat* L_r *through the origin. A necessary and efficient condition that the closed system*

$$\sum_{j=1}^{n} \alpha_j(p)\xi_j >' 0, \quad (p \text{ on range } P) \tag{21}$$

admit no solution is that the origin be an inner point of R_M *relative to* L_r.

Generalizing the pertinent part of Theorem 11, we have

THEOREM 18. *If* r *is the number of the functions* $\alpha_j(p)$ *which are linearly independent on* P, *and if the closed system (20) admits no solution, then there is a subsystem*

$$\sum_{j=1}^{n} a_j(p_i)\xi_j > 0 \quad (i=1, 2, \ldots, k)$$

consisting of at most $r+1$ inequalities which admits no solution.

The proof is identical with that of Theorem 11.

Similarly generalizing Theorem 13 we have

THEOREM 19. *If r is the number of the functions $a_j(p)$ which are linearly independent on P, and if the closed system* (21) *admits no solution, then there is a subsystem*

$$\sum_{j=1}^{n} a_j(p_i)\xi_j >' 0 \quad (i=1, 2, \ldots, k) \tag{22}$$

consisting of at most $r+1$ conditions which admits no solution.

For from the hypothesis and Theorem 17 it follows that the point set M lies in an r-flat L_r through the origin, and the origin is an inner point of R_M relative to L_r. Hence there is a subset of at most $r+1$ points of M, say $(a_1(p_i), a_2(p_i), \ldots, a_n(p_i))$, $(i=1, 2, \ldots, k)$, to which one can assign *positive* masses so that the origin will be the centroid of the system. That is, the system of linear equations

$$\sum_{i=1}^{k} a_j(p_i)y_i = 0, \quad (j=1, 2, \ldots, n) \tag{23}$$

admits a *positive* solution $(y_1, y_2, \ldots, y_k)$. Hence by Theorem 12 the associated system of inequalities, which is of form (22) admits no solution.[26]

It will be noted that the last two theorems provide *necessary* conditions in finite algebraic terms that certain closed systems of inequalities shall admit no solution. In the case of Theorem 18, it is obvious that the condition is also sufficient. But such is not the case in Theorem 19. For example in the case of the simple algebraic system arising from the three linear forms in two variables

$$l_1(x) \equiv x_1, \; l_2(x) \equiv -x_1, \; l_3(x) \equiv x_2,$$

the subsystem

$$l_i(x) >' 0 \quad (i=1, 2)$$

[26]If the origin belongs to the point set M, the theorem would be very trivially satisfied by taking for the system (22) the single condition corresponding to the element p for which each a (p) vanishes. A study of the proof of Carathéodory's theorem shows however that under the present hypotheses there will always be systems (23) and (22) which are not trivial in the sense just described.

admits no solution, while the system

$$l_i(x) >' 0 \quad (i=1, 2, 3)$$

admits the solution $x_1=0$, $x_2>0$.

The theorem which follows is designed to give a necessary and sufficient finite algebraic criterion for the non-existence of a solution of a closed system (21).

THEOREM 20. *If r (>0) is the number of the functions $a_j(p)$ which are linearly independent on P, and if the system*

$$(21) \qquad \sum_{j=1}^{n} a_j(p)\xi_j >' 0 \quad (p \text{ on range } P)$$

is closed, then a necessary and sufficient condition that this system admit no solution is that there be a finite set $(p_1, p_2, \ldots, p_k)$ consisting of at most $2r$ elements of P, such that the matrix $\|a_j(p_i)\|$ is of rank r and the algebraic system

$$(24) \qquad \sum_{j=1}^{n} a_j(p_i)\xi_j >' 0, \quad (i=1, 2, \ldots, k)$$

admits no solution.

We first prove the theorem for the case $r=n$, and then extend the proof to the general case. The condition is *sufficient.* For if (24) admits no solution, any existing solution of (21) would have to annul all the left members of (24) which would require that $\xi_j=0$, $(j=1, 2, \ldots, n)$. Hence (21) can have no solution.

To prove the *necessity* of the condition, we note that it is obvious for $n=1$, and develop the proof by mathematical induction. If (21) admits no solution, we deduce from Theorems 19 and 13 the existence of an algebraic system

$$(25) \qquad \sum_{j=1}^{n} a_i(p_i)\xi_j >' 0 \quad (i=1, 2, \ldots, s), \ (s \leqq n+1)$$

with matrix of rank $s-1$, which admits no solution. If $s=n+1$, the desired conclusion has been reached. If $s<n+1$, consider the sets $\{\xi_j\}$ which annul the left members of (25). They are all represented by the formulas

$$(26) \qquad \xi_j = \sum_{g=1}^{n-s+1} c_g \xi_j^{(g)} \quad (j=1, 2, \ldots, n),$$

where

$$\xi_1^{(g)}, \xi_2^{(g)}, \ldots, \xi_n^{(g)}) \quad (g=1, 2, \ldots, n-s+1),$$

are $n-s+1$ linearly independent sets.

Substituting (26) in (21) we obtain a system which may be written in the form

$$\sum_{g=1}^{n-s+1} \beta_g(p)c_g >' 0 \quad (p \text{ on range } P), \tag{27}$$

where

$$\beta_g(p) = \sum_{j=1}^{n} \alpha_j(p)\xi_j^{(g)}, \quad (g=1, 2, \ldots, n-s+1). \tag{28}$$

The system (27) in the $n-s+1$ unknowns c_g admits no solution. Furthermore it may easily be verified that the system is closed, and that the functions $\beta_g(p)$ are linearly independent on P. Hence by the assumption of our induction there exists an algebraic subsystem

$$\sum_{g=1}^{n-s+1} \beta_g(p_i)c_g >' 0 \quad (i=s+1, s+2, \ldots, k) \tag{29}$$

consisting of at most $2(n-s+1)$ conditions which admits no solution, and with matrix of rank $n-s+1$.

Consider now the algebraic system of inequalities in the original n unknowns ξ_j, formed by uniting the system (25) and the system obtained from (29) by use of (28) and (26). This may be written in the form (24). The number of its members is at most $s+2(n-s+1)$ which is at most $2n$. And the rank of its matrix is n; for the left members of (29) can simultaneously vanish only if every c_g is zero, and hence in view of (26) the left members of the combined system can vanish only if every ξ_j is zero.

This completes the proof of the theorem for the case $r=n$. Suppose now that $r<n$, and that the first r of the functions $\alpha_i(p)$ are linearly independent on P. Then the remaining functions may be expressed in the form

$$\alpha_j(p) = \sum_{h=1}^{r} c_{jh}\alpha_h(p), \quad (j=r+1, r+2, \ldots, n) \tag{30}$$

where the coefficients c_{jh} are well defined constants.

Upon the substitution of (30) in (21) it becomes apparent that the resulting system can be written in the form

$$\sum_{j=1}^{r} \alpha_j(p)\xi'_j >' 0 \quad (p \text{ on range } P), \tag{31}$$

where the new variables ξ' are related to the old by the transformation

$$\xi'_j = \xi_j + \sum_{g=r+1}^{n} c_{gj}\xi_j, \quad (j=1, 2, \ldots, r),$$

$$\xi'_j = \lambda_j(\xi) \quad (j=r+1, r+2, \ldots, n),$$

the linear forms $\lambda_j(\xi)$ being restricted only by the condition that the transformation be non-singular.

The systems (21) and (31) are related in the following ways: (i) to each soliuton of one corresponds a solution of the other, (ii) to each solution of an algebraic subsystem of one corresponds a solution of the corresponding algebraic subsystem of the other, (iii) the rank of the matrix of any algebraic subsystem of one is equal to the rank of the matrix of the corresponding subsystem of the other. Hence our theorem, which has been proven for the system (31), holds for the system (21).

COROLLARY. *If the function $a_1(p)$, $a_2(p)$, . . . , $a_n(p)$ are linearly independent, Theorem 20 remains valid if the symbol* $>'$ *be replaced by* $\geqq$ *in either one or both of the systems (21) and (24), and trivial solutions be disregarded.*

The number $2r$ occurring in the statement of Theorem 20 cannot be replaced by any smaller number as an example will show. In view of the preceding remarks it will be sufficient to consider the case $r=n$.

Let

$$
\begin{aligned}
l_1 &\equiv x_1+x_2+\ldots+x_n,\\
l_i &\equiv x_1+\ldots+x_{i-1}-x_i+x_{i+1}+\ldots+x_n, \quad (i=2, 3, \ldots, n)\\
l_{i+n} &\equiv -l_i, \quad (i=1, 2, \ldots, n).
\end{aligned}
$$

The algebraic system

$$
l_i >' 0, \quad (i=1, 2, \ldots, 2n) \tag{32}
$$

clearly admits no solution. Furthermore the forms l_i, $(i=1, 2, \ldots, n)$ are linearly independent and we may consider the system (32) as system (21) of Theorem 20, with $r=n$.

The system (32) is based on the two sets of forms l_i, $(i=1, 2, \ldots, n)$ and l_{i+n}, $(i=1, 2, \ldots, n)$. If any one form l_k of either of these sets be omitted, the remaining forms in that set vanish for some non-trivial set $\bar{x}=(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n)$; and either $\pm\bar{x}$ is a solution of the system obtained from (32) by omitting the condition based on l_k. Thus the system corresponding to the system (24) of Theorem 20 must in this case consist of the entire system (32), which contains $2n$ conditions.

We pass now to a consideration of those theorems which have to do with the relationship of systems of inequalities to their associated systems of linear equations. For inequalities based on the specialized type of forms (17) with which we are dealing in this section, the associated equations would have the general character

$$J_p a_j(p)\,\eta\,(p)=0 \quad (j=1, 2, \ldots, n),$$

where $\eta(p)$ is the function to be determined, and J_p is a linear operator which operated on products of functions of the type indicated, producing from each such product a definite constant. Algebraic examples of the kind of theorems to be expected are Theorems 8 and 12. Two other theorems of the same general nature have been obtained by specialization of the range P, the class of functions to which a_j and η shall belong, and the operator J_p. In the first of these, P is taken to be the infinite sequence of positive integers, the functions a_j and η to be infinite sequences of real numbers satisfying certain convergence conditions, and J_p to be infinite summation operator. This case has been treated by Dines,[27] and the interested reader may consult the original paper.

In the second and perhaps more interesting case[28] the range P is a closed linear interval, the functions a_j and η are real continuous functions on this interval and J_p indicates definite integration over the interval. We limit ourselves to a statement of the theorem. The proof given in the paper to which reference is made is somewhat lengthy and difficult, and no adequate simpler proof has been found.[29]

THEOREM 21. *If n real functions $a_1(x), a_2(x), \ldots, a_n(x)$ are continuous on the interval $a \leqq x \leqq b$, a necessary and sufficient condition that the inequality*

$$\sum_{j=1}^{n} a_j(x)\xi_j > 0 \quad (a \leqq x \leqq b)$$

admit no solution $(\xi_1, \xi_2, \ldots, \xi_n)$ is that the associated system of integral equations

$$\int_a^b a_j(x)\eta(x)dx=0 \quad (j=1, 2, \ldots, n)$$

admit a positive continuous solution $\eta(x)$.

If the given functions $a_j(x)$ are assumed to be linearly independent, Theorem 21 can be stated in the following interesting form.

THEOREM 22. *A necessary and sufficient condition that a finite set of real functions, continuous and linearly independent on a closed*

[27]Dines 9.

[28]Dines 8.

[29]A new proof of Theorem 21 made its appearance while the galley proof was being read. See Schoenberg 3.

interval, admit a positive continuous function orthogonal to all of them is that every linear combination of the functions change sign on the interval.

The following somewhat similar theorem for more restricted sets of functions follows immediately from Theorem 16 (i) and a certain very interesting theorem of Kakeya.[30]

THEOREM 23. *A necessary and sufficient condition that a finite set of real functions, continuous and linearly independent on every sub-interval of a closed interval, admit an M-positive*[31] *continuous function orthogonal to all of them is, that every linear combination of the functions change sign on the interval.*

In his statement of the theorem in his first paper Kakeya omits mention of any restriction as to the linear independence of the functions. In his second paper he notes the omission, and states that the functions are assumed to be linearly independent on every sub-interval, but he apparently has in mind the more restrictive condition that no linear combination of the functions is constant on any sub-interval. The theorem which he obtains in attempting to remove this restriction seems to be in error.[32] Apparently he overlooked the fact that when the given functions are simultaneously constant on a sub-interval, the curve represented parametrically by the equations $y_j = a_j(x)$, $(j = 1, 2, \ldots, n)$ has a peculiar point which invalidates his argument relative to the centre of mass.

The statement of Fujiwara,[33] which might be interpreted as implying that our Theorem 22 is a direct consequence of Kakeya's theorem, should not be so interpreted.

In concluding this section, we give two equivalent finite algebraic criteria[34] for the two properties of sets of functions which have been seen to be equivalent in Theorem 22.

THEOREM 24. *A set of real functions $a_1(x), a_2(x), \ldots, a_n(x)$, continuous and linearly independent upon a closed interval has the two properties proven equivalent in Theorem 22 if and only if there exists a point set $(x_1, x_2, \ldots, x_k)$, consisting of at most 2n points of the interval,*

---

[30]Kakeya 1, I.

[31]It will be recalled that *M-positive* means "somewhere positive and nowhere negative".

[32]Kakeya 1, II, p. 90.

[33]Fujiwara 3, p. 332.

[34]For other conditions which are sufficient to assure the properties in question, see Dines 5 and McCoy 1.

C5

such that the given functions are linearly independent upon this point set and (i) the system of inequalities

$$\sum_{j=1}^{n} a_j(x_i)\xi_j >'0 \quad (i=1, 2, \ldots, k)$$

admits no solution $(\xi_1, \xi_2, \ldots, \xi_n)$ *or (ii), what is equivalent, the system of equations*

$$\sum_{i=1}^{k} a_j(x_i)\eta_i = 0 \quad (j=1, 2, \ldots, n)$$

admits a positive solution $(\eta_1, \eta_2, \ldots, \eta_k)$.

This follows immediately from Theorems 21, 20 and 12.

12. *Further Specializations.* In the preceding section we have considered inequalities which arise from the general form (13) by specialization of the range of the variable q to be a finite set. In the present section we shall give attention to inequalities which arise from this general form by specialization of the range of the variable p to be a finite set. In this case the form (13) may be written

$$J_q a_i(q)\xi(q) \quad (i=1, 2, \ldots, n).$$

The range Q of the variable q is entirely unrestricted. The given functions $a_i(q)$ and the unknown function $\xi(q)$ are to belong to a class of real-valued functions F which is assumed to possess the property of linearity; and J_q is a linear operator which operates on products of functions of F, the result of such operation being in every case a definite real unmber. We assume also

P. For every function $f(q)$ of F, $J_q f(q)\, f(q) \geqq 0$, the equality holding if and only if $f(q) \equiv 0$.

The following theorem is a generalization of Theorem 12 of § 8.

THEOREM 25. *If the n functions* $a_1(q), a_2(q), \ldots, a_n(q)$ *belong to the class F, a necessary and sufficient condition that the system of inequalities,*

(33) $$J_q a_i(q)\xi(q) >'0 \quad (i=1, 2, \ldots, n)$$

admit no solution $\xi(q)$ *belonging to F, is that there exist positive constants* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *such that*

$$\lambda_1 a_1(q) + \lambda_2 a_2(q) + \ldots + \lambda_n a_n(q) = 0, \quad (q \text{ on range } Q).$$

The sufficiency of the condition is almost obvious. To establish the necessity of the condition we make use of the

LEMMA.[35] *If $a_i(q)$, $(i=1, 2, \ldots, n)$ are functions of F linearly independent on Q and $a_1, a_2, \ldots, a_n$ are any real constants, then there exists a function $\xi'(q)$ of F such that*

$$J_q a_i(q)\xi'(q) = a_i, \quad (i=1, 2, \ldots, n).$$

If the system (33) admits no solution, then by this Lemma just $r(<n)$ of these functions are lineraly independent on Q. Suppose for convenience that the first r are linearly independent. Then we have

$$a_i(q) = \sum_{j=1}^{r} c_{ij}a_j(q), \quad (i=r+1, r+2, \ldots, n) \tag{34}$$

where the coefficients c_{ij} are well defined constants.

Now the system of algebraic inequalities,

$$\begin{aligned} &x_i \geqq 0, && (i=1, 2, \ldots, r) \\ &\sum_{j=1}^{r} c_{ij}x_j \geqq 0, && (i=r+1, r+2, \ldots, n) \end{aligned} \tag{35}$$

does not admit a solution for which the *in*equality holds in at least one instance. For suppose $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_r)$ were such a solution. By the lemma, there exists a function $\bar{\xi}(q)$ of F such that

$$J_q a_i(q)\bar{\xi}(q) = \bar{x}_i \quad (i=1, 2, \ldots, r),$$

and it is easily verified that $\bar{\xi}(q)$ is then a solution of (33) contrary to hypothesis. Hence by Theorem 12 of § 8 there exist positive constants $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that

$$\sum_{i=r+1}^{n} \lambda_i c_{ij} + \lambda_j = 0, \quad (j=1, 2, \ldots, r). \tag{36}$$

From (34) and (36) we now find that

$$\sum_{i=r+1}^{n} \lambda_i a_i(q) = -\sum_{j=1}^{r} \lambda_j a_j(q), \quad (q \text{ on range } Q)$$

that is

$$\sum_{i=1}^{n} \lambda_i a_i(q) = 0, \quad (q \text{ on range } Q)$$

as required by the theorem.

In a similar manner the complementary theorem generalizing Theorem 8 of § 7 may be established.

The following theorem was obtained by Haar[36] for the case in which F is the class of continuous functions on a closed linear interval, and J_q the definite integral over the interval. We omit the proof as

[35] Dines 10, pp. 143, 146.

[36] Haar 1.

the method used by Haar may be immediately extended to establish this more general result.

THEOREM 26. *If the linearly independent functions* $\alpha_1(q)$, $\alpha_2(q)$, $\ldots$, $\alpha_n(q)$ *belong to* F, *and the system*

$$J_q\, \alpha_i(q)\xi(q) \geqq 0, \quad (i=1, 2, \ldots, n)$$

admits at least one non-zero solution $\xi(q)$ *belonging to* F, *then a necessary and sufficient condition that all such solutions shall satisfy a given inequality*

$$J_q\beta(q)\xi(q) \geqq 0,$$

where $\beta(q)$ *belongs to* F, *is that* $\beta(q)$ *be expressible in the form*

$$\beta(q) \equiv c_1\alpha_1(q) + c_2\alpha_2(q) + \ldots + c_m\alpha_n(q), \quad (q \text{ on range } Q)$$

where the c_i *are non-negative constants.*

This theorem is a generalization of Minkowski's Consequence Theorem (Corollary 1, § 7).

In concluding these remarks we may mention one further type of specialization. If we restrict both P and Q to be the infinite sequence of positive integers and J_q to be the infinite summation operator, the form (13) then gives rise to a system of infinitely many inequalities in an infinite number of unknowns. Schoenberg[37] and Hildebrandt[38] have studied certain restricted systems of this type along lines suggested by Theorem 2 of § 6. The general solution appears in the form of Stieltjes integrals, and this form of result is applied to obtain and generalize results due to Hausdorff, S. Bernstein, and Widder relative to completely monotonic functions.

13. *Linear Differential Inequalities.* We conclude our discussion of linear inequalities with a brief mention of inequalities of a different sort from those referred to previously. Certain linear differential inequalities have been studied by Bohl, Kakeya and Fujiwara.[39] The following problem considered by Kakeya will illustrate sufficiently the connection with the general subject of linear inequalities.

Let the real functions $f_1(x)$, $f_2(x)$, $\ldots$, $f_n(x)$, be continuous and uniform in the interval $\alpha \leqq x \leqq \beta$. To determine under what conditions there exists a continuous function $y(x)$ having continuous n-th derivative and satisfying the differential inequality,

$$P(x\,|y) \equiv y^{(n)} + f_1(x)y^{(n-1)} + \ldots + f_n(x)y \geqq 0$$

in the interval (α, β) and also the boundary conditions,

$$y^{(i)}(\alpha) = a_i, \; y^{(i)}(\beta) = b_i, \quad (i=0, 1, \ldots, n-1).$$

[37]Schoenberg 1 and 2,[38] Hildebrandt and Schoenberg 1.
[39]Bohl 1, Kakeya 2, Fujiwara 2, 3.

Kakeya reduces this problem to the question of the existence of an M-positive solution of a system of integral equations of the type used in establishing Theorem 23. The results are stated in terms of the geometric notion of smallest convex region containing a certain curve.

Bibliography

Bohl, P. 1. *Ueber Differentialungleichungen*, Crelle, vol. 144 (1914), p. 284.

Carathéodory, C. 1. *Ueber den Variabilitätsbereich der Fourier'schen Konstanten . . . etc.*, Rendiconti del Circolo Mat. di Palermo, vol. 32 (1911), p. 193.

Carver, W. B. 1. *Systems of linear inequalities*, Annals of Math., vol. 23 (1922), p. 212.

Dines, L. L. 1. *Systems of linear inequalities*, Annals of Math., vol. 20 (1919), p. 191.

2. *Definite linear dependence*, Annals of Math., vol. 27 (1925), p. 57.

3. *Note on certain associated systems of linear equalities and inequalities*, Annals of Math., vol. 28 (1926), p. 41.

4. *On positive solutions of a system of linear equations*, Annals of Math., vol. 28 (1927), p. 386.

5. *On completely signed sets of functions*, Annals of Math., vol. 28 (1927), p. 393.

6. *On sets of functions of a general variable*, Transactions of the Amer. Math. Soc., vol. 29 (1927), p. 463.

7. *Linear inequalities in general analysis*, Bulletin of the Amer. Math. Soc., vol. 33 (1927), p. 695.

8. *A theorem on orthogonal functions with an application to integral inequalities*, Transactions of the Amer. Math. Soc., vol. 30 (1928), p. 425.

9. *A theory on orthogonal sequences*, Transactions of the Amer. Math. Soc., vol. 30 (1928), p. 439.

10. *On linear integral inequalities of first kind*, Transactions of the Royal Society of Canada, vol. 23 (1929), p. 141.

11. *Linear inequalities and some related properties of functions*, Bulletin of Amer. Math. Soc., vol. 36 (1930), p. 393.

Farkas, J. 1. *Ueber die Theorie der einfachen Ungleichungen*, Crelle, vol. 123, (1902), p. 1.

Fourier, J. 1. *Oeuvres* 2, p. 325, Extract from Histoire de l'Académie, 1824, p. XLVII.

Fujiwara, M. 1. *Ueber den kleinsten eine Kurve enthaltenden konvexen Körper*, Tôhoku Imp. Univ. Science Reports, vol. 4 (1915), p. 339.

2. *Zur Theorie der Differentialungleichungen*, Tôhoku Imp. Univ. Science Reports, vol. 6 (1917), I, p. 101, and II, p. 307.

3. *On the system of linear inequalities, and linear integral inequality*, Proc. of the Imp. Academy, vol. 4 (1928), p. 330.

4. *On the system of linear inequalities*, Proc. of the Imp. Academy, vol. 6 (1930), p. 297.

Gummer, C. F. 1. *Sets of linear equations in positive unknowns*, (Abstract), Amer. Math. Monthly, vol. 33 (1926), p. 488.

Haar, A. 1. *Ueber lineare Ungleichungen*, Szeged Acta, sectio scient-math., vol. 2 (1924), p. 1.

HAUSDORFF, F. 1. *Ueber das Momentenproblem für ein endliches Interval*, Mathematische Zeitschrift, vol. 16 (1923), p. 220.

HILDEBRANDT, T. H., and SCHOENBERG, I. J. 1. *On linear functional operations and the moment problem for a finite interval in one or several dimensions.* To appear in Annals of Mathematics.

HUBER, A. 1. *Eine Erweiterung der Dirichletschen Methode . . . etc.*, Monatshefte für Math. und Physik, vol. 37 (1930), p. 55.

KAKEYA, S. 1. *On some integral equations* (a series of five papers under this title): I, Tôhoku Math. Journal, vol. 4 (1914), p. 186; II, Proc. of Tôkyô Math.-Phys. Soc., vol. 8 (1915), p. 83; III, Tôhoku Math. Journal, vol. 8 (1915), p. 14; IV and V, Proc. of Tôkyô Math.-Phys. Soc., vol. 8, p. 408 and vol. 9, p. 93.

2. *On linear differential inequlaity.* Proc. of Tôkyô Math.-Phys. Soc., vol. 8 (1915), p. 256.

LA MENZA, F., *Los sistemas de inecuaciones lineales y la division del hiperspacio*, Atti del Congresso Internazionale dei Matematici (Bologna), 1928, II, p. 199.

LOVITT, W. V. 1. *Preferential Voting*, Amer. Math. Monthly, vol. 23 (1916), p. 363.

McCOY, N. H. 1. *Note on the existence of a positive function orthogonal to a given set of functions*, Bulletin of Amer. Math. Soc., vol. 36 (1930), p. 878.

MINKOWSKI, H. 1. *Geometrie der Zahlen*, Teubner, First edition 1896, revised edition 1910, p. 39.

SCHLAUCH, H. M. 1. *Mixed systems of linear equations and inequalities*, Amer. Math. Monthly, vo. 39 (1932), p. 218.

SCHOENBERG, I. J. 1. *On finite and infinite completely monotonic sequences*, Bulletin of Amer. Math. Soc., vol. 38 (1932), p. 72.

2. *On finite-rowed systems of linear inequalities in infinitely many variables.* (Two papers under this title, of which the first appeared in the Transactions of the Amer. Math. Soc., vol. 34 (1932), p. 594, and the second has been submitted to the same journal).

3. *Convex domains and linear combinations of continuous functions*, Bulletin of Amer. Math. Soc. (1933) p. 273.

STIEMKE, E. 1. *Ueber positive Lösungen homogener linearer Gleichung en*, Mathematische Annalen, vol. 76 (1915), p. 340.

STOKES, R. W. 1. *A geometric theory of solution of linear inequalities*, Transactions of the Amer. Math. Soc., vol. 33 (1931), p. 782.

SIAM-AMS Proceedings
Volume 9
1976

NONLINEAR PROGRAMMING: A HISTORICAL VIEW

Harold W. Kuhn[1]

ABSTRACT

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

1. INTRODUCTION AND SUMMARY

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

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

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

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

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*, DOI 10.1007/978-3-0348-0439-4_18, © Springer Basel 2014

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

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

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

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

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

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

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

2. WHAT IS A NONLINEAR PROGRAM?

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

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

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

$$\ldots \quad \ldots \quad \ldots$$

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

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

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

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

$$\begin{aligned} \text{Maximize} \quad & f(x_1,\dots,x_n) \quad \text{subject to} \\ & g_1(x_1,\dots,x_n) = 0 \\ & \dots \quad \dots \\ & g_m(x_1,\dots,x_n) = 0. \end{aligned}$$

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

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

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

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

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

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

This is a linear program in standard form.

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

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

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

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

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

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

$$x_1+x_2-1 = -y_1, \quad -x_1-2x_2+1 = -y_2$$
$$x_1 \geqq 0, \quad x_2 \geqq 0, \quad y_1 \geqq 0, \quad y_2 \geqq 0.$$

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

$$(x_1+x_2-1)(x_1+2x_2-1) = -y_1$$
$$x_1 \geqq 0, \quad x_2 \geqq 0, \quad y_1 \geqq 0.$$

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

3. DUALITY IN LINEAR PROGRAMMING AND BEFORE

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

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

$$\text{Maximize } f(x) = c \cdot x \text{ for "feasible" solutions to}$$
$$Ax-b = -y.$$

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

$$\text{Minimize } h(v) = v \cdot b \text{ for "dual feasible" solutions to}$$
$$vA-c = u.$$

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

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

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

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

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

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

while the definition of feasibility for the dual pair implies that

$$h-f \geqq 0$$

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

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

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

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

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

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

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

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

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

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

4. THE KARUSH CONDITIONS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Proof: By the differentiability of f,

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

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

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

for k large enough. Taking limits

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

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

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

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

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

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

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

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

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

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

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

5. THE KUHN-TUCKER PAPER

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

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

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

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

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

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

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

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

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

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

6. THE JOHN CONDITIONS

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

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

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

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

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

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

7. SYLVESTER'S PROBLEM

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

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

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

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

$$\text{(2)} \qquad \begin{aligned} &\text{Minimize } v+x^2/2 \text{ subject to} \\ &\qquad v+x^2/2 \geqq (x-a_j)^2/2 \\ &\text{for } v\epsilon R,\ x\epsilon R^m, \text{ and } j = 1,\dots,n. \end{aligned}$$

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

$$\text{(3)} \qquad \begin{aligned} &\text{Minimize } v+\Sigma_i x_i^2/2 \text{ subject to} \\ &y_j = v+\Sigma_i x_i a_{ij} - \Sigma_i a_{ij}^2/2 \geqq 0 \\ &\text{for } v\epsilon R,\ (x_i)\epsilon R^m, \text{ and } j = 1,\dots,n. \end{aligned}$$

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

$$\begin{array}{c|ccc|c|l}
 & \lambda_1 & \cdots & \lambda_n & -1 & \\
\hline
v & 1 & \cdots & 1 & 1 & = 0 \\
\hline
x_1 & a_{11} & \cdots & a_{1n} & 0 & = z_1 \\
\vdots & \vdots & & \vdots & \vdots & \ \ \vdots \\
x_m & a_{m1} & \cdots & a_{mn} & 0 & = z_m \\
\hline
-1 & a_1^2/2 & \cdots & a_n^2/2 & 0 & = f \\
\hline
 & =y_1 & \cdots & =y_n & =h &
\end{array}$$

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

$$\begin{aligned} &\text{Minimize } H = h+x^2/2 \text{ for} \\ &h = v,\ \ y_j = v+\Sigma_i x_i a_{ij} - a_j^2/2 \geqq 0, \ \text{ all } j. \\ &\text{Maximize } F = f-z^2/2 \text{ for} \end{aligned}$$

$$f = \Sigma_j \lambda_j a_j^2/2, \quad \Sigma_j \lambda_j = 1, \quad z_i = \Sigma_j a_{ij} \lambda_j, \quad \lambda_j \geqq 0, \quad \text{all } i \text{ and } j.$$

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

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

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

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

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

Theorem 7.1:

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

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

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

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

such that

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

and

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

This formulation opens a number of possibilites for computation.

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

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

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

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

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

8. A SERMON

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

APPENDIX

(omitted)

(omitted)

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ŒUVRES

DE LAGRANGE,

PUBLIÉES PAR LES SOINS

DE M. J.-A. SERRET,

SOUS LES AUSPICES

DE SON EXCELLENCE

LE MINISTRE DE L'INSTRUCTION PUBLIQUE.

TOME PREMIER.

PARIS,

GAUTHIER-VILLARS, IMPRIMEUR-LIBRAIRE

DE L'ÉCOLE IMPÉRIALE POLYTECHNIQUE, DU BUREAU DES LONGITUDES,

SUCCESSEUR DE MALLET-BACHELIER,

Quai des Augustins, 55.

—

M DCCC LXVII

G. Giorgi and T.H. Kjeldsen (eds.), *Traces and Emergence of Nonlinear Programming*,
DOI 10.1007/978-3-0348-0439-4_19, © Springer Basel 2014

RECHERCHES

SUR LA

MÉTHODE DE MAXIMIS ET MINIMIS.

RECHERCHES

SUR LA

MÉTHODE DE MAXIMIS ET MINIMIS.

(*Miscellanea Taurinensia,* t. I, 1759.)

1. Les Géomètres savent depuis longtemps que lorsque la première différentielle d'une variable quelconque disparaît sans que la seconde disparaisse en même temps, elle devient toujours un *maximum* ou un *minimum;* et en particulier elle est un *maximum,* si sa différentielle seconde est négative, et un *minimum,* si cette différentielle est positive. Si la différentielle seconde disparaît en même temps que la première, alors la quantité n'est ni un *maximum,* ni un *minimum,* à moins que la troisième différentielle ne disparaisse de même, dans lequel cas la proposée deviendra un *maximum,* si la différentielle quatrième est négative, et un *minimum,* si elle est positive, et ainsi de suite. En général, pour qu'une quantité soit un *maximum* ou un *minimum,* il faut que les ordres successifs des différentielles, qui s'évanouissent ensemble, soient en nombre impair, et alors elle est sûrement un *maximum* ou un *minimum,* selon que la différentielle qui suit la dernière évanouissante se trouve négative ou positive. *Voyez* MACLAURIN, *Traité des Fluxions,* p. 238 et 857.

2. Tout ceci supposé et bien entendu, que Z représente une fonction algébrique des variables t, u, x, y,..., et qu'on se propose de la rendre un *maximum* ou un *minimum.* Soit, selon les règles ordinaires,

$$dZ = p\,dt + q\,du + r\,dx + s\,dy + \ldots,$$

I.

et l'on aura d'abord cette équation

$$p\,dt + q\,du + r\,dx + s\,dy + \ldots = 0.$$

Mais comme la relation entre t, u, x,... est encore indéterminée, de même que celle de leurs différentielles dt, du, dx,..., et que d'ailleurs l'équation donnée doit être vraie quel que soit leur rapport, il est évident que pour les chasser tout à fait de l'équation, il faut égaler séparément à zéro chaque membre $p\,dt$, $q\,du$, $r\,dx$, ..., d'où l'on tire autant d'équations particulières qu'il y a de variables, savoir :

$$p = 0, \quad q = 0, \quad r = 0, \ldots.$$

Par le moyen de toutes ces équations on trouvera les valeurs de chaque inconnue t, u, x, ..., qui, substituées dans la fonction proposée Z, la rendront un *maximum* ou un *minimum*.

3. Passons maintenant à l'examen de la seconde différentielle. En supposant, ce qui est permis, les premières différentielles dt, du, dx,... constantes, on aura

$$d^2Z = dp\,dt + dq\,du + dr\,dx + ds\,dy + \ldots.$$

Soit

$$\begin{aligned}
dp &= \mathrm{A}\,dt + \mathrm{B}\,du + \mathrm{D}\,dx + \mathrm{G}\,dy + \ldots,\\
dq &= \mathrm{B}\,dt + \mathrm{C}\,du + \mathrm{E}\,dx + \mathrm{H}\,dy + \ldots,\\
dr &= \mathrm{D}\,dt + \mathrm{E}\,du + \mathrm{F}\,dx + \mathrm{I}\,dy + \ldots,\\
ds &= \mathrm{G}\,dt + \mathrm{H}\,du + \mathrm{I}\,dx + \mathrm{L}\,dy + \ldots,
\end{aligned}$$

ce qui donnera

$$\begin{aligned}
d^2Z = {} & \mathrm{A}\,dt^2 + 2\mathrm{B}\,dt\,du + \mathrm{C}\,du^2 + 2\mathrm{D}\,dt\,dx + 2\mathrm{E}\,du\,dx\\
& + \mathrm{F}\,dx^2 + 2\mathrm{G}\,dt\,dy + 2\mathrm{H}\,du\,dy + 2\mathrm{I}\,dx\,dy + \mathrm{L}\,dy^2 + \ldots.
\end{aligned}$$

Pour commencer par le cas le plus simple, supposons qu'il n'y ait qu'une seule variable t, de sorte que $d^2Z = \mathrm{A}\,dt^2$; on voit d'abord que, puisque dt^2 est toujours positif, la différentielle d^2Z doit avoir le même signe que la quantité A; donc, si A est positif, Z sera un *minimum*, et si

A est négatif il sera un *maximum;* si $A = 0$ on suivra les règles données (1).

4. Les variables contenues dans Z soient deux, savoir t et u; alors

$$d^2 Z = A dt^2 + 2B dt du + C du^2.$$

Il paraît au premier aspect bien difficile de connaître si cette expression $d^2 Z$ doit être positive ou négative, sans qu'on ait le rapport de dt à du, qui n'est pas donné; car, puisqu'en changeant ce rapport la fonction $d^2 Z$ doit aussi varier, il semble indubitable qu'elle pourra aussi passer du positif au négatif, et du négatif au positif, pendant que les quantités A, B, C restent les mêmes. Qu'on donne cependant à la proposée

$$A dt^2 + 2B dt du + C du^2$$

cette forme

$$A\left(dt + \frac{B\,du}{A}\right)^2 + \left(C - \frac{B^2}{A}\right) du^2;$$

et on verra que, comme les carrés $\left(dt + \frac{B\,du}{A}\right)^2$ et du^2 ont toujours le même signe +, toute la quantité sera nécessairement positive si les deux coefficients A et $C - \frac{B^2}{A}$ sont positifs, et au contraire elle deviendra négative, lorsque ceux-ci seront tous deux négatifs, quel que soit le rapport de dt à du. On aura donc pour le cas du *minimum*

$$A > 0, \quad C - \frac{B^2}{A} > 0,$$

savoir

$$C > \frac{B^2}{A} \quad \text{ou} \quad CA > B^2,$$

ce qui donne de même

$$C > 0;$$

à moins donc que les quantités A, B, C n'aient ces conditions

$$A > 0, \quad C > 0 \quad \text{et} \quad AC > B^2,$$

la proposée Z ne pourra pas être un *minimum*. En second lieu on trouvera pour le *maximum*

$$A < o, \quad C - \frac{B^2}{A} < o,$$

savoir

$$C < \frac{B^2}{A}, \quad CA > B^2,$$

puisque A est négatif, ce qui donne encore

$$C < o;$$

donc les conditions pour le *maximum* seront en partie les mêmes, et en partie précisément contraires à celles du *minimum*.

5. Si A ou C, ou toutes deux sont égales à zéro sans que B le soit aussi, la condition de $AC > B^2$ ne pourra pas subsister, ainsi la quantité proposée ne deviendra jamais un vrai *maximum* ou *minimum;* la même chose arrivera toutes les fois que A et C seront de signe contraire, car puisque B^2 est toujours positif la condition de $AC > B^2$ devient impossible. Si B s'évanouissait encore en même temps que A ou C, d^2Z se trouverait réduite au cas d'une seule variable, et par conséquent pourrait être de nouveau un *maximum* ou un *minimum*, ou ni l'un ni l'autre, selon ce qu'on a dit pour le premier cas. Enfin, si la quantité d^2Z était toute égale à zéro, savoir

$$A = o, \quad B = o, \quad C = o,$$

il faudrait recourir à la différentielle troisième; que si celle-ci se trouve n'être pas égale à zéro, la quantité Z ne peut être ni un *maximum* ni un *minimum;* et au contraire, si elle évanouit en même temps que la seconde, on cherchera tout de suite la quatrième; et si elle n'est pas évanouissante, il sera facile, par la méthode dont nous nous sommes servi ci-devant, de connaître si elle est positive ou négative, ce qui déterminera de nouveau le *maximum* ou le *minimum*.

6. Lorsque les variables sont trois, savoir t, u, x, la différentielle d^2Z prend cette forme

$$d^2Z = A\,dt^2 + 2B\,dt\,du + C\,du^2 + 2D\,dt\,dx + 2E\,du\,dx + F\,dx^2$$

qu'on réduira d'abord à

$$A\left(dt + \frac{B\,du}{A} + \frac{D\,dx}{A}\right)^2 + \left(C - \frac{B^2}{A}\right)du^2 + 2\left(E - \frac{BD}{A}\right)du\,dx + \left(F - \frac{D^2}{A}\right)dx^2.$$

Soit posé

$$C - \frac{B^2}{A} = a, \quad E - \frac{BD}{A} = b, \quad F - \frac{D^2}{A} = c,$$

et on aura

$$d^2Z = A\left(dt + \frac{B\,du}{A} + \frac{D\,dx}{A}\right)^2 + a\,du^2 + 2b\,du\,dx + c\,dx^2;$$

qu'on opère à présent sur ces trois derniers membres, comme on a fait ci-dessus (4), et toute la différentielle proposée d^2Z deviendra

$$A\left(dt + \frac{B\,du}{A} + \frac{D\,dx}{A}\right)^2 + a\left(du + \frac{b\,dx}{a}\right)^2 + \left(c - \frac{b^2}{a}\right)dx^2;$$

or, les carrés $\left(dt + \frac{B\,du}{A} + \frac{D\,dx}{A}\right)^2$, $\left(du + \frac{b\,dx}{a}\right)^2$ et dx^2 étant toujours positifs, toute la différentielle sera de même positive si les coefficients A, a et $c - \frac{b^2}{a}$ ont chacun le signe $+$; on a donc pour le *minimum* les conditions suivantes

$$A > 0, \quad a > 0, \quad ca > b^2,$$

ou, en remettant au lieu de a, b, c leurs valeurs,

$$A > 0, \quad C - \frac{B^2}{A} > 0, \quad \left(C - \frac{B^2}{A}\right)\left(F - \frac{D^2}{A}\right) > \left(E - \frac{BD}{A}\right)^2,$$

savoir

$$A > 0, \quad CA > B^2 \quad \text{et} \quad (CA - B^2)(FA - D^2) > (EA - BD)^2,$$

d'où il résulte encore

$$C > 0, \quad F > 0 \quad \text{et} \quad FA > D^2.$$

On trouvera par les mêmes principes pour le *maximum*

$$A < 0, \quad CA > B^2 \quad \text{et} \quad (CA - B^2)(FA - D^2) > (EA - BD)^2,$$

et par conséquent

$$C < 0, \quad F < 0 \quad \text{et} \quad FA > D^2.$$

7. Si les quantités A et C évanouissent seules, ou toutes deux, ou une simplement, la seconde condition devient impossible; si c'est F qui évanouit, alors la troisième devient impossible; car $(CA - B^2)(- D^2)$, qui est nécessairement négatif à cause de $CA > B^2$, doit toujours se trouver moindre de $(EA - BD)^2$, d'où il suit que Z ne saurait être un *maximum* ou un *minimum,* si A, C, F prises séparément ou ensemble, comme on voudra, sont égales à zéro. Si par l'évanouissement des termes la différentielle d^2Z se réduisait à deux variables, ou à une seulement, elle tomberait dans le second cas ou dans le premier, et on devrait suivre les règles données (3 et suiv.). Enfin, si toute la d^2Z se trouvait égale à zéro, et que la différentielle troisième ne fût pas de même égale à zéro, on serait sûr que la proposée Z ne pourrait jamais devenir ni un *maximum,* ni un *minimum;* et quand cette différentielle troisième évanouirait avec la seconde, par des transformations semblables à celles que nous avons pratiquées, on pourrait dans la quatrième différentielle distinguer le cas du *minimum* et du *maximum* et ceux qui sont inutiles.

8. On peut étendre la même théorie aux fonctions de quatre ou plus variables. Quiconque aura bien saisi l'esprit des réductions que j'ai employées jusqu'ici, pourra sans peine découvrir celles qui conviendront à chaque cas particulier. Au reste, pour ne pas se méprendre dans ces recherches, il faut remarquer que les transformées pourraient bien venir différentes de celles que nous avons données; mais en examinant la chose de plus près, on trouvera infailliblement que, quelles qu'elles

soient, elles pourront toujours se réduire à celles-ci, ou au moins y être comprises.

9. Comme je crois cette théorie entièrement nouvelle, il ne sera peut-être pas inutile d'ajouter les réflexions suivantes. Quel que soit le nombre des variables qui entrent dans la fonction proposée Z, si on les regarde chacune en particulier, et qu'on cherche le *maximum* ou *minimum* qui lui convient pendant que toutes les autres demeurent les mêmes, on trouvera à part les premières différentielles $p\,dt$, $q\,du$, $r\,dx,\ldots$, dont chacune étant égalée à zéro nous donnerait les mêmes équations que ci-dessus (2)

$$p = 0, \quad q = 0, \quad r = 0, \ldots.$$

De la même manière passant aux différentielles secondes, on trouverait celles-ci séparément $A\,dt^2$, $C\,du^2$, $F\,dx^2$, $L\,dy^2,\ldots$, et par conséquent si A, C, F, L,... sont toutes positives ou négatives, on pourrait croire que cela suffit pour que les valeurs de t, u, $x,\ldots$, tirées des équations $p = 0$, $q = 0,\ldots$, rendent nécessairement la proposée Z un *minimum* ou un *maximum*. Il est vrai, en effet, que par rapport à chacune de ces variables considérées à part, la quantité donnée Z devra toujours être la plus grande ou la plus petite; mais est-il certain que ce qui vaut pour chacune prise séparément doive aussi valoir pour toutes ensemble? Examinons la chose plus intimement.

10. Que la proposée Z contienne les seules variables t et u, et on pourra la regarder comme l'ordonnée à une surface, dont t et u sont les deux autres; donc la question dans ce cas se réduit à trouver la plus grande ou la plus petite ordonnée d'une surface dont l'équation est donnée, savoir

$$dZ = p\,dt + q\,du.$$

Si l'on fait u constant, elle se réduit d'abord à

$$dZ = p\,dt,$$

I. 2

et dans ce cas elle exprime toutes les sections de la meme superficie parallèles à l'axe des t, à mesure que la quantité u reçoit des valeurs différentes. Soit donc posé $p = o$, et on aura (2) une valeur de t qui donnera la plus grande ou la plus petite ordonnée Z dans chacune de ces sections parallèles; mais, puisque u est constant, si l'on différentie de nouveau dZ, on a

$$d^2 Z = A dt^2,$$

et par conséquent on jugera du *maximum* ou *minimum* par la seule valeur de A, après y avoir cependant substitué à la place de t la valeur que fournit l'équation $p = o$. Savoir si A se trouve positive ou négative, quelle que soit la valeur de u, ou bien si, en changeant u, elle peut aussi changer de signe, on conclura dans le premier cas que toutes lesdites sections ont un *maximum* ou un *minimum*, et dans le second qu'elles ont entre certaines limites un *maximum*, entre d'autres un *minimum*. Si A est égal à zéro, quelle que soit la valeur de la constante u, alors aucune desdites sections n'aura ni un *maximum* ni un *minimum*. Mais, si A devient seulement égal à zéro, lorsque u a de certaines valeurs données, dans ces cas seulement les sections correspondantes seront destituées du *maximum* ou du *minimum*. Le lieu de toutes ces ordonnées qui sont un *maximum* ou un *minimum*, ou ni l'un ni l'autre, sera contenu dans l'équation $p = o$, en ayant égard à la seule variabilité de u; elles formeront donc dans la même superficie une section qui sera a simple ou à double courbure, et qui sera déterminée par les deux equations conjointes

$$dZ = p\,dt + q\,du \quad \text{et} \quad p = o,$$

ou

$$dZ = q\,du \quad \text{et} \quad p = o.$$

On voit par là que, pour trouver le *maximum* ou le *minimum* de la surface entière, il faudra chercher la plus grande ou la plus petite ordonnée qui convient à cette même section; on aura donc de nouveau

$$q = o,$$

ce qui donnera la valeur de l'autre variable u.

11. Passons maintenant à la différentielle de q; elle a été d'abord supposée (3) égale à $B\,dt + C\,du$; mais puisque dans ce cas t est déterminé par u dans l'équation $p = 0$, ou bien dans sa différentielle $A\,dt + B\,du = 0$, dt est égal à $-\frac{B\,du}{A}$, ce qui rend

$$dq = \left(-\frac{B^2}{A} + C\right) du;$$

il résulte donc que si $-\frac{B^2}{A} + C$ est positif, savoir si $C > \frac{B^2}{A}$, l'ordonnée sera la moindre; si $C < \frac{B^2}{A}$, elle sera la plus grande, et si $C = \frac{B^2}{A}$, elle ne sera ni l'une ni l'autre, à moins que les conditions requises dans les différentielles des genres plus élevés ne soient remplies. Or, en réfléchissant sur ces *maximum* et *minimum*, il sera aisé de comprendre que l'ordonnée Z ne pourra pas être un *maximum* entre toutes les autres, à moins qu'elle ne soit la plus grande de toutes celles qui sont contenues dans la section déterminée par $dZ = q\,du$, et de plus que toutes les ordonnées qui composent cette même section ne soient encore elles-mêmes des *maximum* dans les sections parallèles correspondantes (10). On prouvera de même que la quantité Z ne saurait être absolument un *minimum* sans qu'elle soit de même un *minimum* dans la section qui contient tous les *minimum*. Car dans tous les autres cas l'ordonnée serait ou la plus grande ou la plus petite d'entre celles qui ne sont ni les plus grandes ni les plus petites, ou bien entre les plus grandes ou les plus petites, elle ne serait ni la plus grande ni la plus petite, ou enfin elle serait la plus grande d'entre les plus petites, ou au contraire, ce qui ne donne pas un vrai *maximum* ou *minimum* comme on cherche. De tout ceci je conclus donc qu'après avoir tiré des équations $p = 0$, $q = 0$, les valeurs de t et u, et les avoir substituées dans A et dans $C - \frac{B^2}{A}$, il faut, pour que Z soit un vrai *maximum*, que A soit négatif et

$$C < \frac{B^2}{A}, \quad \text{savoir} \quad CA > B^2;$$

2.

et au contraire, si Z doit être un vrai *minimum,* on doit trouver A positif et

$$C > \frac{B^2}{A}, \quad \text{ou} \quad CA > B^2,$$

conformément à la théorie générale expliquée (4 et suiv.).

12. Si, au lieu de considérer d'abord u constant et t variable, on avait fait u variable et t constant, on serait parvenu aux déterminations suivantes

$$C < o \quad \text{et} \quad AC > B^2$$

pour le *maximum,* et

$$C > o \quad \text{et} \quad AC > B^2$$

pour le *minimum,* ce qui revient au même. Au reste, cette méthode que nous venons d'employer pour découvrir les conditions des *maximum* et *minimum* dans les fonctions à deux seules changeantes, est également applicable à toutes les autres fonctions plus composées, elle a même l'avantage d'être plus analytique et plus directe que la première, c'est pourquoi je tâcherai ici de la développer dans toute sa généralité.

13. Soient les variables contenues dans Z en tel nombre qu'on voudra; je ne considère d'abord qu'une variable seule, et je tire par la différentiation l'équation pour le *maximum* ou *minimum* qui lui convient; puis en passant à la différentielle seconde, je trouve les conditions qui déterminent la proposée à être un *maximum* ou un *minimum,* ou ni l'un ni l'autre. Après cette première opération, je substitue dans Z ou dans ses différentielles simplement la valeur de la première variable trouvée, et je procède sur une autre variable de la même manière; ensuite, mettant de nouveau dans la fonction proposée Z la valeur qu'on aura trouvée pour cette seconde variable, on passera à l'examen d'une troisième variable, et ainsi de suite, etc. Soit t la première variable qu'on veut considérer dans Z, et on aura

$$dZ = p\,dt \quad \text{et} \quad d^2Z = A\,dt^2,$$

d'où $p = 0$, et $A > 0$ pour le *minimum*, $A < 0$ pour le *maximum* (1). Que t et u soient à présent toutes deux variables, il en résultera

$$dZ = p\,dt + q\,du,$$

qui, à cause de $p = 0$, se réduit à

$$dZ = q\,du,$$

d'où l'on tire

$$d^2Z = (B\,dt + C\,du)\,du;$$

mais puisque $p = 0$, dp le sera aussi, et par conséquent

$$A\,dt + B\,du = 0,$$

ce qui donne

$$dt = -\frac{B\,du}{A};$$

cette valeur substituée dans d^2Z la changera en

$$d^2Z = \left(-\frac{B^2}{A} + C\right) du^2,$$

j'aurai donc $q = 0$ et

$$-\frac{B^2}{A} + C > 0$$

pour le *minimum*, et

$$-\frac{B^2}{A} + C < 0$$

pour le *maximum*, savoir, puisque A est positif dans le premier cas et négatif dans le second, en multipliant par A, il résultera toujours la même condition de $AC > B^2$. Si, outre les deux précédentes, il y a encore une troisième variable x à considérer, je cherche la valeur de dZ eu égard à ces trois variables t, u, x, et je trouve

$$dZ = p\,dt + q\,du + r\,dx,$$

ce qui, à cause de $p = 0$, $q = 0$, se change en

$$dZ = r dx;$$

donc la différentielle seconde sera

$$d^2Z = (D dt + E du + F dx) dx.$$

A présent, par le moyen des équations

$$p = 0, \quad q = 0,$$

ou bien de leurs différentielles

$$A dt + B du + D dx = 0 \quad \text{et} \quad B dt + C du + E dx = 0,$$

je cherche des valeurs de dt et du en dx, et je trouve

$$dt = \frac{BE - CD}{AC - B^2} dx, \quad du = \frac{BD - AE}{AC - B^2} dx;$$

je les substitue dans l'expression de d^2Z, ce qui me donne

$$d^2Z = \left(\frac{BE - CD}{AC - B^2} D + \frac{BD - AE}{AC - B^2} E + F\right) dx^2.$$

Il résulte donc en premier lieu pour le *maximum* ou *minimum*

$$r = 0;$$

ensuite

$$\frac{BE - CD}{AC - B^2} D + \frac{BD - AE}{AC - B^2} E + F > 0$$

pour le *minimum*, et < 0 pour le *maximum;* ou bien, en ôtant le dénominateur $AC - B^2$ qui est toujours positif, on a

$$2BDE - CD^2 - AE^2 - FB^2 + ACF > 0$$

pour le *minimum*, et < 0 pour le *maximum*. Soit multipliée cette expression par A, qui est positif dans le premier cas et négatif dans le second, et on aura

$$2ABDE - ACD^2 - A^2E^2 - AB^2F + A^2CF > 0,$$

soit pour le *maximum*, soit pour le *minimum*, savoir

$$(CA - B^2)(FA - D^2) > (EA - BD)^2.$$

On suivra le même procédé pour un plus grand nombre de variables.

14. Cette méthode, étant générale pour quelque nombre de variables que ce soit, ne sera pas bornée aux seules fonctions algébriques, mais pourra encore s'étendre avec succès aux *maximum* et *minimum* qui sont d'un genre plus élevé et qui appartiennent à des formules intégrales indéfinies. Je me réserve de traiter ce sujet, que je crois d'ailleurs entièrement nouveau, dans un ouvrage particulier que je prépare sur cette matière, et dans lequel, après avoir exposé la méthode générale et analytique pour résoudre tous les problèmes touchant ces sortes de *maximum* ou *minimum*, j'en déduirai, par le principe de la moindre quantité d'action, toute la mécanique des corps soit solides, soit fluides.

15. Je finirai ce Mémoire par quelques exemples des plus simples qui éclaircissent la théorie qu'on vient d'établir. Soient tant de corps qu'on voudra parfaitement élastiques et rangés en ligne droite sans se toucher; supposons que le premier vienne choquer le second avec une vitesse donnée c, le second avec la vitesse acquise du premier choque le troisième, et ainsi de suite; les masses du premier et du dernier étant données, on demande celles des corps intermédiaires, afin que le dernier reçoive la plus grande vitesse possible. Soit a la masse du premier, et b celle du dernier; soient ensuite t, u, x, y,... les masses intermédiaires inconnues; par les lois du choc on trouvera la vitesse communiquée par le prêmier corps a au second t égale à $\frac{2ac}{a+t}$, celle que donne celui-ci au

troisième u égale à $\frac{2.2act}{(a+t)(t+u)}$, et ainsi de suite; donc la vitesse que recevra le dernier b sera exprimée par

$$\frac{2\ldots 2catuxy\ldots b}{(a+t)(t+u)(u+x)(x+y)\ldots},$$

expression qui doit devenir un *maximum*. Pour en trouver plus aisément la différentielle, qu'on la suppose égale à Z, et prenant les logarithmes d'une part et de l'autre, on trouvera

$$\left.\begin{array}{l} l2\ldots 2ca + lt + lu + lx + ly + \ldots \\ \quad - l(a+t) - l(t+u) - l(u+x) - l(x+y) - \ldots \end{array}\right\} = l\mathrm{Z},$$

ce qui donne par la différentiation

$$\frac{dt}{t}+\frac{du}{u}+\frac{dx}{x}+\frac{dy}{y}+\ldots-\frac{dt}{a+t}-\frac{dt+du}{t+u}-\frac{du+dx}{u+x}-\frac{dx+dy}{x+y}-\ldots=\frac{d\mathrm{Z}}{\mathrm{Z}};$$

d'où, en mettant ensemble et réduisant au même dénominateur les termes affectés des mêmes différentielles, l'on tire

$$d\mathrm{Z}=\frac{\mathrm{Z}(au-t^2)\,dt}{t(a+t)(t+u)}+\frac{\mathrm{Z}(tx-u^2)\,du}{u(t+u)(u+x)}+\frac{\mathrm{Z}(uy-x^2)\,dx}{x(u+x)(x+y)}+\ldots.$$

On aura donc en premier lieu pour le *maximum* ou *minimum* les équations suivantes

$$au=t^2,\quad tx=u^2,\quad uy=x^2,\ldots,$$

qui donnent les analogies

$$a:t=t:u,\quad t:u=u:x,\quad u:x=x:y,\ldots,$$

savoir

$$\because a:t:u:x:y:\ldots b;$$

d'où l'on voit que toutes les masses doivent constituer une progression géométrique, dont les deux extrêmes sont les données a et b. Pour juger à présent du *maximum* ou *minimum*, soit fait d'abord, pour abréger,

$$\frac{Z}{t(a+t)(t+u)} = \alpha,$$
$$\frac{Z}{u(t+u)(u+x)} = \beta,$$
$$\frac{Z}{x(u+x)(x+y)} = \gamma,$$
$$\dots\dots\dots\dots,$$

on aura

$$p = \alpha(au - t^2),$$
$$q = \beta(tx - u^2),$$
$$r = \gamma(uy - x^2),$$
$$\dots\dots\dots;$$

donc

$$dp = (au - t^2)\, d\alpha + \alpha(a\,du - 2\,t\,dt),$$
$$dq = (tx - u^2)\, d\beta + \beta(x\,dt + t\,dx - 2\,u\,du),$$
$$dr = (uy - x^2)\, d\gamma + \gamma(y\,dx + u\,dy - 2\,x\,dx),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$

Or, comme les termes a, t, u, x, y, ... doivent être en progression continue, si l'on nomme $1 : m$ la raison constante d'un antécédent quelconque à son conséquent, on trouve

$$t = ma, \quad u = m^2 a, \quad x = m^3 a, \quad y = m^4 a, \dots,$$

de plus

$$\beta = \frac{\alpha}{m^3}, \quad \gamma = \frac{\alpha}{m^6}, \dots,$$

lesquelles valeurs substituées dans les expressions précédentes les réduiront à

$$dp = \alpha a(du - 2m\,dt),$$
$$dq = \alpha a\left(dt - \frac{2\,du}{m} + \frac{dx}{m^2}\right),$$
$$dr = \alpha a\left(\frac{du}{m^2} - \frac{2\,dx}{m^3} + \frac{dy}{m^4}\right),$$

3

et ainsi des autres. On aura donc

$$A = -2m\alpha a,\quad B = \alpha a,\quad C = -\frac{2\alpha a}{m},\quad D = 0,\quad E = \frac{\alpha a}{m^2},$$

$$F = -\frac{2\alpha a}{m^3},\quad G = 0,\quad H = 0,\quad I = \frac{\alpha a}{m^4},\ \ldots$$

On voit par là en premier lieu que A est négatif, et que par conséquent la proposée doit être un *maximum* si les autres conditions se trouvent remplies. Or

$$AC = 4\alpha^2 a^2 \quad \text{et} \quad B^2 = \alpha^2 a^2,$$

donc

1° $$AC > B^2;$$

$$AC - B^2 = 3\alpha^2 a^2,\quad FA - D^2 = \frac{4\alpha^2 a^2}{m^2},\quad EA - BD = -\frac{2\alpha^2 a^2}{m},$$

donc

$$(AC - B^2)(FA - D^2) = \frac{12\alpha^4 a^4}{m^2},\quad \text{et} \quad (EA - BD)^2 = \frac{4\alpha^4 a^4}{m^2},$$

et par conséquent

2° $$(AC - B^2)(FA - D^2) > (EA - BD)^2.$$

S'il n'y a que deux masses intermédiaires t et u, il suffit d'avoir égard à la première de ces conditions; s'il y en a trois, il faut encore considérer la seconde; s'il y en avait plusieurs autres, il faudrait avoir recours à autant de conditions qu'il y a de variables. Au reste, dans ce problème, on les trouvera toutes remplies si on veut bien prendre la peine de pousser plus loin le calcul; de sorte qu'on peut franchement assurer que, lorsque les masses intermédiaires, quel que soit leur nombre, sont telles qu'elles forment une progression géométrique entre les deux extrêmes données, la vitesse que reçoit la dernière par leur moyen est toujours la plus grande possible. Ce problème a été traité par M. Huyghens, le premier, et depuis par beaucoup d'autres Géomètres; mais sans avoir aucunement égard aux nouvelles déterminations, que nous avons cependant trouvées nécessaires pour s'assurer de l'existence du *maximum* ou *minimum*.

16. Soit l'équation générale pour les surfaces de second ordre

$$z^2 = ax^2 + 2bxy + cy^2 - ex - fy;$$

qu'on se propose de trouver le point où l'ordonnée z est la plus grande ou la plus petite; on aura, en différentiant,

$$2z\,dz = 2ax\,dx + 2by\,dx + 2bx\,dy + 2cy\,dy - e\,dx - f\,dy,$$

ce qui fournit d'abord les deux équations suivantes

$$ax + by = \frac{e}{2},$$

$$cy + bx = \frac{f}{2},$$

d'où l'on tire

$$x = \frac{ec - fb}{2(ac - b^2)},$$

$$y = \frac{eb - fa}{2(ac - b^2)}.$$

Différentions de nouveau la différentielle trouvée, et on aura, puisque $dz = 0$,

$$2z\,d^2z = 2a\,dx^2 + 4b\,dx\,dy + 2c\,dy^2$$

où les quantités x, y ne se trouvent plus. Or, afin que l'ordonnée z soit un vrai *maximum* ou *minimum*, il faut que a et c soient toutes deux négatives dans le premier cas, et toutes deux positives dans le second; de plus, il faut encore que $ca > b^2$, car sans cela les valeurs trouvées pour les ordonnées x et y ne donneraient jamais ni un *maximum*, ni un *minimum;* en effet, toutes les fois que ca n'est pas plus grand que b^2, le célèbre M. Euler a démontré par une autre voie, dans l'Appendice à l'*Introduction à l'Analyse des infiniment petits*, que la surface proposée s'étend à l'infini et qu'elle a une asymptote conique. Il paraît donc clairement que la méthode pour déterminer les *maximum* et *minimum*,

3.

quand il y a plusieurs variables, en ne les regardant qu'une à la fois, peut souvent être très-fautive. Car, par exemple, dans le cas précédent, en traitant d'abord x comme variable, on trouve la différentielle première $2\left(ax + by - \frac{e}{2}\right) dx$, et la seconde $2adx^2$; de même, en faisant varier y, on a pour la différentielle première $2\left(cy + bx - \frac{f}{2}\right) dy$, et pour la seconde $2cdy^2$. Or les deux différentielles premières posées égales à zéro donnent les mêmes équations qu'on a trouvées, et les deux secondes font voir que si a et c sont toutes deux positives ou toutes deux négatives, l'ordonnée z est un *maximum* ou un *minimum*, si on a simplement égard à la variabilité des x et y considérées séparément; mais on n'est pas en droit de conclure pour cela que z soit un *maximum* ou un *minimum*, par rapport à toutes deux ensemble, comme on vient de le voir.