



Soviet-era science, translated into English

ON THE KUHN– TUCKER THEOREM

1969

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196901.88069>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

UDC 519.3+512.25-26

MATHEMATICS

A. A. Kaplan, G. Sh. Rubinshtein

ON THE KUHN–TUCKER THEOREM

(Presented by Academician L. V. Kantorovich on 5 III 1969)

The equivalence conditions, given in this note, for a convex programming problem and the problem of a saddle point of the corresponding Lagrange function may be regarded as a generalization of the Kuhn–Tucker conditions to the case of nondifferentiable functions.

Problem I (the basic problem). On some open convex set G in the m -dimensional arithmetic space R^m , concave functions g^1, g^2, \dots, g^n, f are given and, in addition, a certain convex set $X^0 \subset G$ is singled out. It is required to maximize the function f in the domain $\Omega = \bigcap_{j=1}^n X^j$, where

$$X^j = \{x \in G : g^j(x) \geq 0\}, \quad j = 1, 2, \dots, n.$$

Problem I^* (the dual problem). Let Y^0 be the set of vectors $y = (y_1, y_2, \dots, y_n)$ from R^n with nonnegative components ($y \geq 0$). It is required to minimize the function $\varphi(y) = \sup_{x \in X^0} G(x, y)$, where

$$G(x, y) = \sum_{j=1}^n y_j g^j(x) + f(x), \quad (1)$$

in the domain $\Omega^* = \{y \in Y^0 : \varphi(y) < +\infty\}$.

Vectors $x \in \Omega$, $y \in \Omega^*$ are called **feasible** in problems I and I^* , and the sought vectors are called **optimal**.

For any feasible vectors x and y , the relation $f(x) \leq G(x, y) \leq \varphi(y)$ is evidently valid. Therefore, for optimality of a vector $x^0 \in X^0$, it is sufficient that it be feasible and that there exist $y^0 \in Y^0$ for which $\varphi(y^0) = f(x^0)$. This condition is equivalent to the existence of $y^0 \in Y^0$ such that $G(x, y^0) \leq G(x^0, y^0) \leq G(x^0, y)$ for all $x \in X^0$, $y \in Y^0$. The corresponding point (x^0, y^0) is called a **saddle point** for the function (1), called the **Lagrange function** of problem I.

A significant place in the theory of convex programming is devoted to clarifying additional assumptions concerning the set X^0 and the functions g^1, g^2, \dots, g^n under which the assertion (!) is valid: *for optimality of a vector $x^0 \in X^0$ in*

problem I, the existence of a vector y^0 such that (x^0, y^0) is a saddle point of the function (1) is not only sufficient but also necessary.

Conditions of this kind for convex programming problems with differentiable functions were first obtained in the work of Kuhn and Tucker ⁽¹⁾. Subsequently, other conditions ensuring the validity of assertion (!) were also found. The theorems given below provide new conditions of the indicated type.

Problem I reduces to finding the maximal t for which the point $te^{n+1} = (0, 0, \dots, 0, t)$ of R^{n+1} belongs to the convex set $A^- = \{a = a' - a'' : a' \in A, a'' \geq 0\}$, where $A = \{a = F(x) : x \in X^0\}$, $F(x) = (g^1(x), \dots, g^n(x), f(x))$. In this case the maximum t coincides with the maximum of f on Ω . Problem I^* reduces to finding the minimal t for which there exists a hyperplane H passing through the point te^{n+1} and strictly separating from the set A^- the ray $\{t'e^{n+1} : t' > t\}$. In this case the minimum t coincides with the minimum of φ on Ω^* , and for a fixed $x^0 \in \Omega$ the existence

for the function (1) a saddle point (x^0, y^0) is equivalent to the existence, for $t = f(x^0)$, of a hyperplane H with the indicated properties (see, for example, (2')).

If x^0 is an optimal vector in problem I, then there always exists a hyperplane H passing through the point $f(x^0)e^{n+1}$ and separating the set A^- and the ray $\{te^{n+1} : t > t(x^0)\}$. If A^- contains an interior point of the form t^*e^{n+1} , this hyperplane strictly separates from A^- the corresponding ray, i.e., assertion (!) is valid. The existence of such an interior point is equivalent to the existence of a point $x^* \in X^0$ at which $g^j(x^*) > 0$, $j = 1, 2, \dots, n$ (Slater's condition (3)). This condition is one of the simplest ones ensuring the validity of assertion (!). When it is violated, not all separating hyperplanes strictly separate the ray of interest to us, and more delicate arguments are required to prove the existence of the needed hyperplane. We give some definitions.

The **limiting cone** of a convex set X at a point $x^0 \in X$ is called

$$X_{x^0} = K_{x^0}(\bar{X}),$$

where

$$K_{x^0} = \bigcup_{x \in X} L(x^0, x), \quad L(x^0, x) = \{z = x^0 + t(x - x^0) : t > 0\},$$

and the bar denotes closure. By $X_{x^0 x^1 \dots x^r}$ is meant the limiting cone of the set $X_{x^0 x^1 \dots x^{r-1}}$ at the point x^r .

A concave function g defined in R^m is called **superlinear relative to the point x^0** if

$$g(x^0 + t(x - x^0)) = g(x^0) + t(g(x) - g(x^0))$$

for all $x \in R^m$, $t > 0$.

For a concave function g defined on an open convex set $G \subset R^m$, by a **support function at the point** $x^0 \in G$ is meant the function defined in R^m

$$g_{x^0}(x) = g(x^0) + g'_{x^0}(x - x^0),$$

where

$$g'_{x^0}(u) = \lim_{t \rightarrow +0} \frac{1}{t} [g(x^0 + tu) - g(x^0)].$$

Here $g_{x^0 x^1 \dots x^r}$ is a function supporting $g_{x^0 x^1 \dots x^{r-1}}$ at the point x^r . It is easy to show that it is superlinear relative to any point from the affine hull of the points x^0, x^1, \dots, x^r , and moreover

$$g_{x^0 x^1 \dots x^r}(x) \geq g_{x^0 x^1 \dots x^{r-1}}(x).$$

Lemma 1. *If a concave function g , defined on an open convex set $G \subset R^m$, is such that $\sup g(x) > 0$, $g(x^0) = 0$, then the limiting cone X_{x^0} of the set*

$$X = \{x \in G : g(x) \geq 0\}$$

is

$$\{x \in R^m : g_{x^0}(x) \geq 0\}.$$

Proof. It is easy to see that $g_{x^0}(x) \geq 0$ for $x \in K_{x^0}(X)$, and hence also for $x \in X_{x^0}$. Conversely, if $g_{x^0}(x) \geq 0$, then for sufficiently large n the points

$$y_n = \frac{n-1}{n}x^0 + \frac{1}{n}x$$

belong to G and

$$g_{x^0}(x) = ng(y_n) + \varepsilon_n,$$

where $\varepsilon_n \geq 0$, $\lim \varepsilon_n = 0$. It can be shown that the points

$$z_n = x + \frac{\varepsilon_n}{g(x^*)}(x^* - y_n),$$

which converge to x , where x^* is a fixed vector for which $g(x^*) > 0$, are contained in $K_{x^0}(X)$, and, consequently, $x \in X_{x^0}$.

Lemma 2. *If x^0 is an optimal vector of problem I , in which*

$$\sup g^j(x) > 0, \quad j = 1, 2, \dots, n, \quad \bigcap_{j=0}^n X_{x^0}^j = \Omega_{x^0},$$

then x^0 is also optimal for the problem of maximizing f_{x^0} in the domain

$$\Omega^0 = \{x \in X_{x^0}^0 : g_{x^0}^j(x) \geq 0, j \in J^0\}, \quad J^0 = \{j : g^j(x^0) = 0\}. \quad (2)$$

Proof. Taking into account that for $j \notin J^0$ the cones $X_{x^0}^j$ coincide with R^m , by the preceding lemma we have

$$\Omega^0 = \bigcap_{j=0}^n X_{x^0}^j = \Omega_{x^0}.$$

Hence, using the continuity of the function f_{x^0} and its expression through f , it is easy to obtain the required inequality

$$f_{x^0}(x) \leq f_{x^0}(x^0)$$

for all $x \in \Omega^0$.

Lemma 3. *Suppose that in problem I $G = R^m$, $X^0 = X_{x^0}^0$, $g^j(x^0) = 0$, $j = 1, 2, \dots, n$, and the functions g^1, g^2, \dots, g^n, f are superlinear relative to x^0 . If, moreover, x^0 is an optimal vector and the set L of optimal vectors of the problem is an affine manifold, then there exists a vector y^0 such that the point (x^0, y^0) is a saddle point for the function (1).*

Proof. It can be shown that, under the assumptions made, the vector function F introduced above maps every affine manifold L' , parallel to L , into a single point. By Q we denote the affine complement of L containing the point x^0 (i.e., $(L - x^0) \oplus (Q - x^0) = R^m$). In view of what has been said, the set $A \subset R^{n+1}$ in the present case coincides with $F(X^0 \cap Q)$, and A and A^- are cones with vertex $F(x^0) = f(x^0)e^{n+1}$. If the vector y^0 of interest to us did not exist, then the point $(f(x^0) + 1)e^{n+1}$ would belong to the closure of the cone A^- , i.e., there would be $t_i > 0$ and vectors $z^i \in X^0 \cap Q$ such that

$$t_i g^j(z^i) \geq -\frac{1}{j} \quad (j = 1, 2, \dots, n), \quad t_i (f(z^i) - f(x^0)) = 1,$$

$$\lim z^i = z^*, \quad z^* \in X^0 \cap Q, \quad z^* \in \bar{L}, \quad t^* = \lim t_i > 0.$$

But this is impossible, since for $t^* < +\infty$, for the vector $z^* \in \bar{Q}$ we would have $f(z^*) > f(x^0)$, while for $t^* = +\infty$ the vector z^* would be optimal and, consequently, would be contained in L .

Theorem 1. If in problem I $\sup g^j(x) > 0$, $j = 1, 2, \dots, n$, then, for the assertion (!) to hold for an arbitrary concave function f , it is necessary and sufficient that

$$\bigcap_{j=0}^n X_{x^0 x^1 \dots x^r}^j = \Omega_{x^0 x^1 \dots x^r}, \quad (3)$$

whatever the vectors $x^0 \in \Omega$, $x^k \in \Omega_{x^0 x^1 \dots x^{k-1}}$, $k = 1, 2, \dots, r$, may be.

Sufficiency. Let x^0 be an optimal vector of problem I. By Lemma 2 it is also optimal in the problem of maximizing the function f_{x^0} in the region Ω^0 , defined according to (2). If the set L of optimal vectors in this problem is not an affine manifold, then we pass to maximizing the function $f_{x^0 x^1}$ in the region $\Omega_{x^0 x^1}$

(x^1 is a fixed interior point of the set L in its affine hull), and so on. Finally, we arrive at the problem of maximizing the function $f_{x^0 x^1 \dots x^r}$ in the cone $\Omega_{x^0 x^1 \dots x^r}$, in which x^0 is an optimal vector and the set of optimal vectors is an affine manifold. The cone $\Omega_{x^0 x^1 \dots x^r}$ admits the representation

$$\Omega^r = \{x \in X_{x^0 x^1 \dots x^r}^0 : g_{x^0 x^1 \dots x^r}^j(x) \geq 0, j \in J^r\}, \quad (4)$$

where $J^r = \{j : g_{x^0 x^1 \dots x^{r-1}}^j(x^r) = 0\}$. On the basis of Lemma 3, the Lagrange function of this problem has a saddle point (x^0, \tilde{y}^0) , from which, by adding components $y_i^0 = 0$ for $j \in J^r$, one obtains a saddle point (x^0, y^0) for the Lagrange function of the original problem I.

Necessity. Suppose that, for x^0, x^1, \dots, x^r , relation (3) is violated, i.e., some ray $\mathcal{L}(x^0, \tilde{x})$ belongs to the difference of the cones $\bigcap_{j=0}^n X_{x^0 x^1 \dots x^r}^j$ and $\Omega_{x^0 x^1 \dots x^r}$. In view of the closedness of the latter, there exists a linear function f for which $f(x) \leq f(x^0)$ if $x \in \Omega_{x^0 x^1 \dots x^r}$, and $f(x) > f(x^0)$ for $x \in \mathcal{L}(x^0, \tilde{x})$. It can be shown that, for the problem of maximizing such a function f in the region $\Omega = \{x \in X^0 : g^j(x) \geq 0, j = 1, \dots, n\}$, assertion (!) is false.

Let us note that the inequalities $\sup g^j(x) > 0$ in the proof of Theorem 1 were used only to justify the coincidence, when

$$g_{x^0 x^1 \dots x^k}^j(x^k) = 0,$$

of the cones

$$\{x \in R^m : g_{x^0 x^1 \dots x^k}^j(x) \geq 0\}$$

and $X_{x^0 x^1 \dots x^k}^j$. This means that the following also holds.

Theorem 2. For the assertion (!) to hold for an arbitrary function f , it is necessary and sufficient that, for any vectors $x^0 \in \Omega$, $x^k \in \Omega_{x^0 x^1 \dots x^{k-1}}$, $k = 1, 2, \dots, r$, the cone (4) coincide with the cone $\Omega_{x^0 x^1 \dots x^r}$.

It is precisely this proposition that should be regarded as a direct generalization of the Kuhn-Tucker theorem to the case of nondifferentiable functions. Further, from Theorem 1 it follows...

Theorem 3*. If in problem I $\sup g^j(x) > 0$, $j = 1, 2, \dots, n$, and the sets X^0, X^1, \dots, X^{n_1} are polyhedra (intersections of a finite number of half-spaces), then, provided there exists a vector $x^* \in \Omega$ such that $g^j(x^*) > 0$, $j = n_1 + 1, \dots, n$, assertion (!) is valid.

Proof. We show that in the present case the nontrivial inclusion of relation (3) is valid. Let $\tilde{x} \in \bigcap_{j=0}^n X_{x^0}$. Then the open segment (x^*, \tilde{x}) is contained in each of the cones $K_{x^0}(X^j)$ (for $j = 0, 1, \dots, n_1$ —by virtue of their closedness, and for $j = n_1 + 1, \dots, n$ —by virtue of the solidity of the corresponding convex sets). That is,

$$(x^*, \tilde{x}) \subset \bigcap_{j=0}^n K_{x^0}(X^j) = K_{x^0}(\Omega),$$

and, consequently, $\tilde{x} \in \Omega_{x^0}$. By means of the same arguments, relation (3) is verified successively for $r = 1, 2, \dots$

Institute of Mathematics
Siberian Branch of the Academy of Sciences of the USSR
Novosibirsk

Received
26 II 1969

REFERENCES

1. H. Kuhn, A. Tucker, Proc. II Symp. on Math. Stat. and Prob., Univ. of California Press, 1951, p. 483–489.
2. G. Sh. Rubinshtein, collection *Mathematical Programming*, “Nauka,” 1966, p. 9.
3. H. Uzawa, collection *Studies in Linear and Nonlinear Programming*, IL, 1962, p. 57.
4. S. Karlin, *Mathematical Methods in the Theory of Games, Programming, and Economics*, Moscow, 1964.

* For the case when X^0 is the nonnegative orthant and the sets X^1, X^2, \dots, X^{n_1} are half-spaces, this theorem was proved in (2) (see also (4), p. 239).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.