AN ELEMENTARY PROOF OF AN EQUIVALENCE THEOREM AND A DUALITY CONSEQUENCE

O. L. MANGASARIAN

(*Received Feb. 14, 1962*)

SUMMARY

The purpose of this note is to state and give a simple proof of an equivalence theorem associated with maximizing (minimizing) a certain nonlinear function subject to linear equalities. Two duality principles are also derived from this equivalence theorem.

INTRODUCTION

An equivalence theorem gives the necessary and sufficient conditions for solving certain constrained maximization problems.[1)] There does not seem to be in the literature an *explicit* statement of the equivalence theorem that is associated with maximizing a differentiable concave function subject to linear equalities, despite the fact that such a theorem is nothing but a special case of the more general and powerful Kuhn-Tucker Equivalence Theorem [7, Theorem 3]. In what follows we shall state such a theorem and give an elementary proof that is considerably simpler than that of Kuhn and Tucker.

In certain cases duality principles may be derived from an equivalence theorem [9], [5], [6], and [8]. A duality principle relates a constrained maximization and a constrained minimization problem in such a way that the existence of a solution to one of these problems insures a solution to the other and the extrema of the two problems are equal. One of the two problems is called the *primal* and the other the *dual.*

Once again the literature does not seem to have an *explicit* and *precise* statement of the duality principles which we shall derive in this note from the equivalence theorem mentioned earlier. For instance, Courant and Hilbert [2] have a somewhat heuristic discussion of this duality that lacks the explicitness of the principles given here. Our results

1) A minimization problem may be readily reduced to a maximization problem by multiplying the function to be minimized by -1.

are again special cases of the more general results, [4], [9], [5], [6], [8], but are more explicit and in some cases more precisely stated.

In what follows, matrix notation will be used. With obvious exceptions, lower case Roman letters will denote column vectors, capital letters matrices, and Greek letters scalars. A prime will indicate the transpose of a vector or matrix. Thus $x'y$ will indicate the scalar product of the row vector x' by the column vector y. The operators ∇, ∇_u and ∇_v are colum vectors whose components are respectively

$$\left(\frac{\partial}{\partial x_1}, \cdots\cdots, \frac{\partial}{\partial x_n}\right)', \left(\frac{\partial}{\partial u_1}, \cdots\cdots, \frac{\partial}{\partial u_m}\right)', \left(\frac{\partial}{\partial v_1}, \cdots\cdots, \frac{\partial}{\partial v_n}\right)'.$$

A function $\varphi(x)$ is convex if for $0 \leqq \alpha \leqq 1$

$$(1-\alpha)\varphi(x^1)+\alpha\varphi(x^2) \geqq \varphi[(1-\alpha)x^1+ax^2]$$

for all x^1 and x^2 in the (convex) region of definition of $\varphi(x)$. If $\varphi(x)$ is convex and differentiable, then it follows from the definitions of convexity and differentiability that

$$\varphi(x^1)-\varphi(x^2) \geqq (x^1-x^2)'\nabla\varphi(x^2).$$

A function is strictly convex if the equality sign holds only for $x^1=x^2$ or $\alpha=0$, 1.

A function $\varphi(x)$ is concave if $-\varphi(x)$ is convex, that is the above two interpolation inequalities hold with $\leqq$ instead of $\geqq$.

EQUIVALENCE THEOREM

Theorem. *If $\varphi(x)$ is a differentiable, concave, scalar function of the n-vector x, then the necessary and sufficient condition that x° be a solution of the maximum problem*

$$\text{Maximize } \varphi(x) \tag{1}$$

$$\text{subject to } Ax-b=0, \tag{2}$$

where A is an m by n matrix of rank $m(m \leqq n)$ and b is an m-vector, is that x° and some m-vector u° satisfy

$$\phi(x,\ u^\circ) \leqq \phi(x^\circ,\ u^\circ) = \phi(x^\circ,\ u) \qquad \text{(3a, b)}^{2)}$$

for all x and m-vectors u, where

$$\phi(x,\ u) \equiv \phi(x)+u'(Ax-b) \tag{4}$$

2) The inequality between $\phi(x, u^\circ)$ and $\phi(x^\circ, u^\circ)$ is referred to as (3a) and the equality between $\phi(x^\circ, u^\circ)$ and $\phi(x^\circ, u)$ as (3b).

Proof of Sufficiency

Assume that (x°, u°) satisfies (3a, b). Then by (3b), $(u-u^\circ)'(Ax-b)=0$ for all u. Thus $Ax^\circ-b=0$ and x° satisfies the constraints (2) of the maximum problem. Now

$$\begin{aligned}\varphi(x)+u^{\circ\prime}(Ax-b)&=\phi(x, u^\circ)\\ &\leqq\phi(x^\circ, u^\circ) && \text{(by 3a)}\\ &=\varphi(x^\circ)+u^{\circ\prime}(Ax^\circ-b)\\ &=\varphi(x^\circ) && \text{(by 3a)}\end{aligned}$$

Hence $\varphi(x^\circ)\geqq\varphi(x)$ for $Ax-b=0$. This completes the sufficiency proof.

Proof of Necessity

From the elementary theory of Lagrange multipliers [1], the necessary conditions that x° be a solution of the maximum problem (1), (2) are that there exists an m-vector u° such that the gradients of the Lagrangian function $\phi(x, u)$ with respect to x and u vanish at (x°, u°) that is

$$\nabla\phi(x^\circ, u^\circ)\equiv\nabla\varphi(x^\circ)+A'u^\circ=0 \tag{5}$$

$$\nabla_u\phi(x^\circ, u^\circ)\equiv Ax^\circ-b=0 \tag{6}$$

The validity of these conditions is insured A having rank m. Now

$$\begin{aligned}\phi(x^\circ, u)&=\varphi(x^\circ)+u'(Ax^\circ-b)\\ &=\varphi(x^\circ)+u^{\circ\prime}(Ax^\circ-b) && \text{(by 6)}\\ &=\phi(x^\circ, u^\circ)\\ &=\varphi(x^\circ) && \text{(by 6)}\\ &\geqq\varphi(x)-(x-x^\circ)'\nabla\varphi(x^\circ) && \text{(by the concavity of } \varphi)\\ &=\varphi(x)+(x-x^\circ)'A'u^\circ && \text{(by 5)}\\ &=\varphi(x)+u^{\circ\prime}(Ax-Ax^\circ)\\ &=\varphi(x)+u^{\circ\prime}(Ax-b) && \text{(by 6)}\\ &=\phi(x, u^\circ)\end{aligned}$$

Hence $\phi(x, u^\circ)\leqq\phi(x^\circ, u^\circ)=\phi(x^\circ, u)$, and the necessity proof is complete.

DUALITY THEOREMS

Theorem I *If x° is a solution of the primal (maximum) problem (1), (2) with $\varphi(x)$ and A satisfying the restrictions thereof, then there exists some m-vector u° such that (x°, u°) is a solution of the dual problem*

$$\text{Minimize } \phi(x, u)\equiv\phi(x)+u'(Ax-b) \tag{7}$$

$$\text{subject to } \nabla\phi\ (x,u)\equiv\nabla\varphi\ (x)+A'u=0 \tag{8}$$

Also

$$\varphi\ (x^\circ)=\phi\ (x^\circ, u^\circ) \tag{9}$$

Theorem II *If $\varphi(x)$ and A satisfy the restrictions mentioned under the maximum problem* (1), (2) *and if in addition $\varphi(x)$ is twice continuously differentiable and strictly concave*[3] *in the neighborhood of x°, then the converse of Theorem* I *is true, namely, thatf if (x°, u°) is a solution of the dual problem* (7), (8), *then x° is a solution of the primal problem* (1), (2). *Also equation* (9) *holds.*

Proof of Theorem I:

Suppose that x° solves the primal problem (1), (2), then by the Equivalence Theorem proved earlier x° and some u° must satisfy (3a, b). Hence by (3a)

$$\nabla\psi(x^\circ, u^\circ) \equiv \nabla\psi(x^\circ) + A'u^\circ = 0.$$

and thus (x°, u°) satisfies the dual constraints (8). From (3b) we have the fact that x° satisfies the primal constraints

$$Ax^\circ - b = 0. \tag{10}$$

Now

$$\begin{aligned}
\psi(x,u) - \psi(x^\circ,u^\circ) &= \varphi(x) + u'(Ax-b) - \varphi(x^\circ) && \text{(by 10)}\\
&\geqq (x-x^\circ)'\nabla\varphi(x) + u'(Ax-b) && \text{(by concavity of } \psi)\\
&= x'(\nabla\varphi(x) + A'u) - x^{\circ\prime}\nabla\varphi(x) - u'b\\
&= x'(\nabla\varphi(x) + A'u) - x^{\circ\prime}\nabla\varphi(x) - u'Ax^\circ && \text{(by 10)}\\
&= (x - x_\circ)'(\nabla\varphi(x) + A'u)
\end{aligned}$$

Hence $\psi(x,u) \geqq \psi(x^\circ,u^\circ)$ for $\nabla\dot{\psi}(x) + A'u = 0$, which is precisely the statement that (x°, u°) is the solution of the dual problem (7), (8). Equation (9) holds as a consequence of (10). This completes the proof.

Proof of Theorem II:

This will be proved by showing that the sufficient conditions (3a, b) which guarantee that x° is a solution of the primal problem follow from the dual solution.

Suppose (x°, u°) solves the dual problem (7), (8). The *necessary* conditions for (x°, u°) to be such a solution are that the gradients with respect to x, u and v of the Lagrangian function

$$\theta(x, u, v) \equiv \varphi(x) + u'(Ax-b) + v'(\nabla\varphi(x) + A'u)$$

must vanish for (x°, u°) and some v°, where v is an n-dimensional vector

3) For quadratic functions it is sufficient to require that $\varphi(x)$ be merely concave and twice differentiable. See [3].

of Lagrange multipliers. Thus[4)]

$$\nabla\theta(x^\circ,\ u^\circ,\ v^\circ)=\nabla\varphi(x^\circ)+A'u^\circ+\nabla v^{\circ\prime}\nabla\varphi(x^\circ)=0 \tag{11}$$

$$\nabla_u\theta(x^\circ,\ u^\circ,\ v^\circ)=Ax^\circ-b+Av^\circ=0 \tag{12}$$

$$\nabla_v\theta(x^\circ,\ u^\circ,\ v^\circ)=\nabla\varphi(x^\circ)+A'u^\circ=0 \tag{13}$$

If an n by n matrix R is defined whose $i\ j$ th element is $\dfrac{\partial^2\varphi(x^\circ)}{\partial x_i\,\partial x_j}$ then equation (11) becomes after sustitution from (13)

$$Rv^\circ=0 \tag{14}$$

Because $\varphi(x)$ is assumed to be twice continuouslydiffer entiable and strictly concave in the neighborhood of x°, R is a symmetric, negative definite matrix and thus nonsingular. It follows from (14) that $v^\circ=0$. Equation (12) becomes

$$Ax^\circ-b=0 \tag{15}$$

and hence $\psi(x^\circ,\ u^\circ)=\psi(x^\circ,\ u)$.

Now since $(x^\circ,\ u^\circ)$ satisfies (8)

$$\nabla\psi(x^\circ,\ u^\circ)\equiv\nabla\varphi(x^\circ)+A'u^\circ=0,$$

and since $\psi(x,\ u)$ is a concave function of x

$$\psi(x,\ u^\circ)-\psi(x^\circ,\ u^\circ)\leqq(x-x^\circ)'\nabla\psi(x^\circ,\ u^\circ)=0$$

Hence
$$\psi(x,\ u^\circ)\leqq\psi(x^\circ,\ u^\circ). \tag{16}$$

Thus conditions (3a, b) follow from the relations (15), (16) and x° is a solution of the primal problem (1), (2). Equation (9) holds because $Ax^\circ-b=0$. This proves Theorem II.

BIBLIOGRAPHY

1. Courant, R.: "Differential and Integral Calculus" Vol. II, Interscience Publishers, New York, 1936, pp. 190—199.
2. Courant, R., and Hilbert, D.: "Methods of Mathematical Physics", Vol. 1, Interscience Publishers, New York, 1953, pp. 231—242.
3. Dorn, W. S.: "Duality in Quadratic Programming", Quarterly of Applied Math., Vol. 18, No. 2, 1960, pp. 155—162.
4. Dorn, W. S.: "A Duality Theorem for Convex Programs" IBM Journal of Research and Development, Vol. 4, No. 4, 1960, pp. 407—413.

4) To insure the validity of the necessary conditions (11), (12), (13), the Jacobian of $\nabla\psi(x,\ u)$ with respect to x must be nonzero at $(x^\circ,\ u^\circ)$. That this is indeed the case follows from the assumption of strict concavity of $\varphi(x)$ in the neighborhood of x° which implies that the matrix R of equation (14) is positive definite and hence its determinant (which is the Jacobian in question) is nonzero.

5. Hanson, M. A.: "A Duality Theorem in Nonlinear Programming with Nonlinear Constraints", Australian Journal of Statistics, Vol. 3, 1961, pp. 64—72.
6. Huard, P.: "Dual Programs", IBM Journal of Research and Development, Vol. 6, No. 1, January 1962, pp. 137—139.
7. Kuhn, H. W. and Tucker, A. W.: "Nonlinear Programming", Proceedings of the second Berkeley symposium in mathematical statistics and probability, August 1950, University of California Press, 1951, pp. 481—492.
8. Mangasarian, O. L.: "Duality in Nolinear Programming", Shell Development Company Emeryville, California, Paper 1054, 1961, to appear in Quarterly of Applied Mathematics.
9. Wolfe, Philip: "A Duality Theorem for Nonlinear Programming", Quarterly of Applied Math., Vol. 19, No. 3, 1961, pp. 239—244.