—NOTES—

A SYMMETRIC DUAL THEOREM FOR NON-LINEAR PROGRAMS*

By BERTRAM MOND (*Aerospace Research Laboratories, Wright-Patterson Air Force Base, Ohio*)

1. Introduction. Following Dorn [3], a pair of dual programs will be called symmetric if 'the dual of the dual is the original program,' i.e. if, when the dual program is recast in the form of the primal, its dual is the primal. A linear program and its dual [6] are symmetric in this sense. Non-linear programs and their duals, as given in [4] for quadratic programs, in [5] for convex programs with linear constraints, and in [7, 8, 10 & 11] for convex programs with more general constraints are not symmetric.

In what follows, a pair of symmetric non-linear dual programs will be exhibited. Symmetric non-linear programs are given in [1 & 3] for quadratic programs and in [2] for more general programs. The relationship of this development to that of Dantzig, Eisenberg and Cottle [2] will be considered in Section 4.

2. Notation. Small letters will generally denote vectors and capital letters matrices. $x \geq y$ means that every component of x is greater than or equal to the corresponding component of y. Prime will denote transpose.

$f(x, y)$ will be a real-valued differentiable function of x and y, where x and y have dimension n and m respectively. $\nabla_x f(x, y)$ will denote the vector that is the gradient of f with respect to the x variable at the point (x, y) and $\nabla_y f(x, y)$ the gradient of f with respect to y at the point (x, y), i.e.

$$\nabla_x f(x, y) = \left[\frac{\partial f(x, y)}{\partial x_1}, \frac{\partial f(x, y)}{\partial x_2}, \cdots, \frac{\partial f(x, y)}{\partial x_n}\right]',$$

$$\nabla_y f(x, y) = \left[\frac{\partial f(x, y)}{\partial y_1}, \frac{\partial f(x, y)}{\partial y_2}, \cdots, \frac{\partial f(x, y)}{\partial y_m}\right]'.$$

3. Symmetric dual programs. Consider the following two programs:

Primal Program (P):

Minimize
$$H(x, y) \equiv f(x, y) - y'\nabla_y f(x, y) \tag{1}$$
subject to
$$-\nabla_y f(x, y) \geq 0, \tag{2}$$
$$x \geq 0. \tag{3}$$

Dual Program (P*):

Maximize
$$G(x, y) \equiv f(x, y) - x'\nabla_x f(x, y) \tag{4}$$
subject to
$$-\nabla_x f(x, y) \leq 0, \tag{5}$$
$$y \geq 0. \tag{6}$$

*Received August 3, 1964; revised manuscript received November 2, 1964.

If $f(x, y) = b'y - y'Ax + c'x$, (P) and (P^*) reduce to a pair of dual linear programs. If $f(x, y) = b'y - y'Ax + \frac{1}{2}x'Cx + p'x$, (P) and (P^*) become the quadratic dual programs of [4]. If $f(x, y) = b'y - y'Ax - \frac{1}{2}y'Dy + \frac{1}{2}x'Cx + p'x$, one obtains the symmetric dual programs of [1]. Letting $f(x, y) = g(x) - y'Ax + b'y$, yields the non-linear dual programs with linear constraints of [5]. Letting $f(x, y) = g(x) - y'F(x)$ (here $F(x)$ is an m-dimensional vector valued function of x) yields the non-linear dual programs with more general constraints of [7].

Some additional *assumptions* about $f(x, y)$ are needed in order to establish a dual relationship between (P) and (P^*).

1. $f(x, y)$ has continuous first partial derivatives.
2. For each fixed y, in the region defined by the constraints (2), (3), (5) and (6), $f(x, y)$ is convex in x.
3. For each fixed x, in the region defined by (2), (3), (5) and (6), $f(x, y)$ is concave in y.
4. Let X_1 be the set of all x for which a vector y exists such that constraints (5) and (6) hold. Let Y_1 be the set of all y for which an x exists such that (5) and (6) hold. Define

$$z(x, y) \equiv \nabla_x f(x, y) \tag{7}$$

and let Z be the set of all vectors $z(x, y)$ with $x \in X_1$ and $y \in Y_1$. It will be assumed that there exists a differentiable function that determines, for a given $z_1 \in Z$ and $y_1 \in Y_1$, a unique $x_1 \in X_1$ such that $z_1 = \nabla_x f(x_1, y_1)$.

5. Let Y_{11} be the set of all y for which there exists an x such that constraints (2) and (3) hold. Let X_{11} be the set of all x for which there exists a y such that (2) and (3) hold. Define

$$w(x, y) \equiv \nabla_y f(x, y) \tag{8}$$

and let W be the set of all $w(x, y)$ with $x \in X_{11}$ and $y \in Y_{11}$. It will be assumed that there exists a differentiable function that determines, for a given $w_{11} \in W$ and $x_{11} \in X_{11}$, a unique $y_{11} \in Y_{11}$ such that $w_{11} = \nabla_y f(x_{11}, y_{11})$.

THEOREM. *If (x_0, y_0) is an optimal solution for (P) or (P^*),*

a) *(x_0, y_0) is optimal for both (P) and (P^*); and*
b) *the minimum of (P) equals the maximum of (P^*).*

PROOF: Assume that (x_0, y_0) is a minimizing solution of (P). Then (x_0, y_0, w_0) is a minimizing solution of (P) recast as

Minimize
$$h(x, y, w) \equiv f(x, y) - y'w \tag{9}$$

subject to
$$-w \geq 0, \tag{10}$$

$$w = \nabla_y f(x, y), \tag{11}$$

$$x \geq 0. \tag{12}$$

Thus $h(x_0, y_0, w_0) \leq h(x, y, w)$ for all (x, y, w) satisfying (10), (11) and (12). Let us now consider the linear programming problem (L):

Minimize
$$\psi(x, w) \equiv f(x_0, y_0) - y_0'w + x'\nabla_x f(x_0, y_0) \tag{13}$$

subject to
$$- w \geq 0, \tag{14}$$
$$x \geq 0. \tag{15}$$

It is obvious that (x_0, w_0) is a feasible solution of L. We will now show that it is also an optimal solution.

Suppose there exists an (x^*, w^*) satisfying (14) and (15) such that $\psi(x^*, w^*) < \psi(x_0, w_0)$, i.e.,
$$- y_0'(w^* - w^0) + (x^* - x_0)'\nabla_x f(x_0, y_0) < 0. \tag{16}$$
Define
$$x_1 = x_0 + r(x^* - x_0), \tag{17}$$
$$w_1 = w_0 + r(w^* - w_0), \qquad (0 \leq r \leq 1). \tag{18}$$
(x_1, w_1) is also feasible for (L). Since w_1 is in W, there exists a $y_1 \geq 0$ such that
$$w_1 = \nabla_y f(x_1, y_1), \tag{19}$$
and thus (x_1, y_1, w_1) is a feasible solution of (P).

Consider $h(x_1, y_1, w_1) - h(x_0, y_0, w_0) = f(x_1, y_1) - y_1'w_1 - f(x_0, y_0) + y_0'w_0$. From the mean value theorem we have, with $0 \leq \theta_1 \leq 1$ and $\theta_2 = 1 - \theta_1$,
$$\begin{aligned} h(x_1, y_1, w_1) - h(x_0, y_0, w_0) &= (y_1 - y_0)'\nabla_y f(x_1 - \theta_1(x_1 - x_0), y_1 - \theta_1(y_1 - y_0)) \\ &\quad + (x_1 - x_0)'\nabla_x f(x_1 - \theta_1(x_1 - x_0), y_1 - \theta_1(y_1 - y_0)) - y_1'w_1 + y_0'w_0 \\ &= (y_1 - y_0)'\nabla_y f(x_1 - \theta_1(x_1 - x_0), y_1 - \theta_1(y_1 - y_0)) - y_1'w_1 + y_0'w_0 \\ &\quad + (x_1 - x_0)'\nabla_x f(x_0 + \theta_2(x_1 - x_0), y_0 + \theta_2(y_1 - y_0)) \\ &= (y_1 - y_0)'\{\nabla_y f(x_1 - \theta_1(x_1 - x_0), y_1 - \theta_1(y_1 - y_0)) - \nabla_y f(x_1, y_1)\} \\ &\quad - y_0'(w_1 - w_0) + (x_1 - x_0)'\{\nabla_x f(x_0 + \theta_2(x_1 - x_0), y_0 \\ &\quad + \theta_2(y_1 - y_0)) - \nabla_x f(x_0, y_0)\} + (x_1 - x_0)'\nabla_x f(x_0, y_0). \end{aligned}$$

By virtue of assumption 5 and the mean value theorem,
$$\begin{aligned} (y_1 - y_0)' &= (w_1 - w_0)'M[w_1 - \eta(w_1 - w_0), x_1 - \eta(x_1 - x_0)] \\ &\quad + (x_1 - x_0)'N[w_1 - \eta(w_1 - w_0), x_1 - \eta(x_1 - x_0)], \qquad 0 \leq \eta \leq 1, \end{aligned} \tag{20}$$
where $M[w, x]$ is a matrix whose i, j element is the partial derivative of the jth component of y with respect to the ith component of w and $N[w, x]$ is a matrix whose i, j element is the partial derivative of the jth component of y with respect to the ith component of x. Hence, using (17), (18) and (20),
$$\begin{aligned} h(x_1, y_1, w_1) - h(x_0, y_0, w_0) &= r\{[(w^* - w_0)'M[w_1 \\ &\quad - \eta r(w^* - w_0), x_1 - \eta r(x^* - x_0)] + (x^* - x_0)'N[w_1 - \eta r(w^* - w_0), x_1 \\ &\quad - \eta r(x^* - x_0)]][\nabla_y f(x_1 - \theta_1 r(x^* - x_0), y_1 - \theta_1 r(y^* - y_0)) - \nabla_y f(x_1, y_1)] \\ &\quad - y_0'(w^* - w_0) + (x^* - x_0)'[\nabla_x f(x_0 + \theta_2 r(x^* - x_0), y_0 + r\theta_2(y^* - y_0)) \\ &\quad - \nabla_x f(x_0, y_0)] + (x^* - x_0)'\nabla_x f(x_0, y_0)\}. \end{aligned}$$

Because of the continuity of $\nabla_x f(x, y)$ and $\nabla_y f(x, y)$, it follows that, for a sufficiently small r, $h(x_1, y_1, w_1) - h(x_0, y_0, w_0)$ will have the same sign as

$$-y_0'(w^* - w_0) + (x^* - x_0)'\nabla_x f(x_0, y_0)$$

which by (16) is negative. This contradicts the hypothesis that (x_0, y_0, w_0) is optimal. Hence (x_0, w_0) is an optimal solution of (L).

The dual of (L), is the problem (L^*):

Maximize $$f(x_0, y_0) \tag{21}$$

subject to $$-\nabla_x f(x_0, y_0) \leq 0, \tag{22}$$

$$u = y_0, \tag{23}$$

$$u \geq 0. \tag{24}$$

By the duality theorem for linear programming, (L^*) has an optimal solution. From (23), $u = y_0$ is the optimal solution. Therefore, by the duality theorem for linear programming,

$$-y_0'w_0 + x_0'\nabla_x f(x_0, y_0) = 0. \tag{25}$$

Now, (x_0, y_0), since it satisfies (22) and (24), is a feasible solution of (P^*). Let (u, v) be any other feasible solution of (P^*).

It follows [see 5 or 9] from assumption 2 that $f(x_0, v) - f(u, v) \geq (x_0 - u)'\nabla_x f(u, v)$ and from assumption 3 that $f(x_0, y_0) - f(x_0, v) \geq (y_0 - v)'\nabla_y f(x_0, y_0)$. Hence $f(x_0, y_0) - f(u, v) = f(x_0, y_0) - f(x_0, v) + f(x_0, v) - f(u, v) \geq (y_0 - v)'\nabla_y f(x_0, y_0) + (x_0 - u)'\nabla_x f(u, v)$.

Therefore $G(x_0, y_0) - G(u, v) = f(x_0, y_0) - x_0'\nabla_x f(x_0, y_0) - f(u, v) + u'\nabla_x f(u, v) \geq (y_0 - v)'\nabla_y f(x_0, y_0) + (x_0 - u)'\nabla_x f(u, v) - x_0'\nabla_x f(x_0, y_0) + u'\nabla_x f(u, v)$. From (25), $y_0'\nabla_y f(x_0, y_0) - x_0'\nabla_x f(x_0, y_0) = 0$, from (6) and (2) $-v'\nabla_y f(x_0, y_0) \geq 0$, from (3) and (5) $x_0'\nabla_x f(u, v) \geq 0$. Thus $G(x_0, y_0) - G(u, v) \geq 0$. Hence (x_0, y_0) is an optimal solution of P^*.

The minimum of $(P)$is $f(x_0, y_0) - y_0'\nabla_y f(x_0, y_0)$. The maximum of (P^*) is $f(x_0, y_0) - x_0'\nabla_x f(x_0, y_0)$. It therefore follows from (25) that the minimum of (P) equals the maximum of (P^*).

If (x_0, y_0) is an optimal solution of (P^*), the corresponding portion of the theorem can be proved, either by an argument analogous to the preceding one, or, by recasting (P^*) in the form of (P) and utilizing the results already established.

4. Remarks. It was pointed out earlier that our results are similar to those of Dantzig, Eisenberg and Cottle [2]. There are, however, two essential differences. These are 1) the method of proof and 2) the assumptions needed to establish the theorem.

The proof of the symmetric duality theorem in [2] is based on the Kuhn-Tucker theorem [9]. The proof given here is an extension of Dorn's method [5] that makes use of the duality theorem of linear programming. It is noteworthy that this method is applicable even though neither (P) nor (P^*) will, in general, have linear constraints.

The assumptions made in regard to both the functionals to be minimized or maximized and the constraints are more stringent in [2]; and, in this sense, our results represent a generalization of those of Dantzig, Eisenberg and Cottle. In [2], $f(x, y)$ is assumed to have continuous second partial derivatives whereas only the existence and continuity of the first partial derivatives are needed here. In [2], $f(x, y)$ is required to be strictly convex in x for fixed y and strictly concave in y for fixed x. As was pointed out in [5], assumptions 2 and 4 are less restrictive than strict convexity with respect to x. Similarly, assumptions 3 and 5 are less restrictive than strict concavity of $f(x, y)$ with respect to y.

In [2], the constraints of both (P) and (P^*) include $x \geq 0$ and $y \geq 0$. Only $x \geq 0$ for the primal and $y \geq 0$ for the dual are required here. Finally, Dantzig, Eisenberg and Cottle assume that the constraints satisfy the Kuhn-Tucker constraint qualification [9]. This assumption is not needed for the proof given here.

References

1. R. W. Cottle, *Symmetric dual quadratic programs*, Quarterly of Applied Mathematics **21**, (1963) 237–243
2. G. B. Dantzig, E. Eisenberg and R. W. Cottle, *Symmetric dual nonlinear programs*, University of California, Operations Research Center Report 30 (1962)
3. W. S. Dorn, *A symmetric dual theorem for quadratic programs*, J. Operations Research Society of Japan **2** (1960) 93–97
4. W. S. Dorn, *Duality in quadratic programming*, Q. Appl. Math. **18** (1960) 155–162
5. W. S. Dorn, *A duality theorem for convex programs*, IBM J. Res. Dev. **4** (1960) 407–413
6. A. J. Goldman and A. W. Tucker, *Theory of linear programming*, in Linear Inequalities and Related Systems (Annals of Mathematical Studies No. 38) Princeton University Press, 1956
7. M. A. Hanson, *A duality theorem in non-linear programming with non-linear constraints*, Austral. J. Statistics **3** (1961) 64–72
8. P. Huard, *Dual programs*, IBM J. Res. Dev. **6** (1962) 137–139
9. H. W. Kuhn and A. W. Tucker, *Nonlinear programming*, Proc. 2nd Berkeley Symposium on Mathematical Statistics and Probability, Univ. of California Press 1951, 481–492
10. O. L. Mangasarian, *Duality in nonlinear programming*, Q. Appl. Math. **20** (1962) 300–302
11. P. Wolfe, *A duality theorem for nonlinear programming*, Q. Appl. Math. **19** (1961) 239–244