P. Huard*

Dual Programs

This letter is motivated by a theorem of W. S. Dorn[1] concerning the dual of a mini-convex program (i.e., a program whose minimal economic function is convex) under linear constraints. We prove this theorem for the more general case of convex nonlinear constraints.

Dorn obtains his theorem by generalizing the result of J. B. Dennis.[2] We give here a proof which is somewhat different: Relying on the conditions of Kuhn and Tucker,[3,4] we obtain, in the particular case of linear constraints, an easier demonstration of the second half of Dorn's theorem.

Notations

$\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{x}, \mathbf{y}, \mathbf{l}, \mathbf{v}, \mathbf{t}$	are column vectors
C	is a matrix
J	the set of indices of the components of $\mathbf{x}$
$K, \bar{K}$	are complementary subsets of J
$f(\mathbf{x}), g(\mathbf{x},\mathbf{v}), h(\mathbf{x}_K)$	are scalar functions of vector argument.

In addition we use the following matrix notation:

$\mathbf{x}_K$ Column vector whose components have indices consisting of the set $K (K \subset J)$

$\mathbf{x}_{\bar{K}}$ Similar definition as above with K substituted by $\bar{K}$

$df/d\mathbf{x}$ A row vector, whose components have the same indexing set as the components of $\mathbf{x}$, and whose j^{th} component is given by $\partial f(\mathbf{x})/\partial \mathbf{x}_j$

$d\mathbf{a}/d\mathbf{x}$ A matrix where the rows have the same indexing set as the components of the column vector $\mathbf{a}$, and where the columns have the same indexing set as the components of the column vector $\mathbf{x}$. The (i, j)th term is given by:

$$\frac{\partial \mathbf{a}_i(\mathbf{x})}{\partial \mathbf{x}_j}$$

$$\frac{d(\mathbf{a}^T)}{d\mathbf{x}} = \left[\frac{d\mathbf{a}}{d\mathbf{x}}\right]^T,$$

the superscript T being used for transpose.

Every sign (e.g., hat, bar) placed on a variable (scalar, vector, matrix) indicates the substitution of the variable by a constant. Also, to simplify notation we set

$$\mathbf{a}(\hat{\mathbf{x}}) = \hat{\mathbf{a}}; \quad f(\hat{\mathbf{x}}) = \hat{\mathbf{f}}; \quad \left[\frac{d^2 f(\mathbf{x})}{d\mathbf{x}^2}\right]_{\mathbf{x}=\bar{\mathbf{x}}} = \overline{\left[\frac{d^2 f}{d\mathbf{x}^2}\right]}$$

and so forth.

Dual programs

Let $f(\mathbf{x})$ and $\mathbf{a}(\mathbf{x})$ be respectively a scalar and a column vector, both being convex functions of the column x and having continuous derivatives.

We call *Problem I* the following program:

$$\left.\begin{array}{ll}\text{To minimize } f(\mathbf{x}) \text{ under} & \\ \text{the condition } \mathbf{a}(\mathbf{x}) \leqslant 0. & (1)\end{array}\right\} \text{Problem I}$$

The necessary and sufficient conditions for $\hat{\mathbf{x}}$ to be an optimal solution of Problem I are given by Kuhn and Tucker:

$$\left.\begin{array}{ll}\hat{\mathbf{v}} \geqslant 0 & (2)\\ \hat{\mathbf{v}}^T \widehat{\frac{d\mathbf{a}}{d\mathbf{x}}} + \widehat{\frac{df}{d\mathbf{x}}} = 0 & (3)\\ \hat{\mathbf{v}}^T \hat{\mathbf{a}} = 0 ; & (4)\end{array}\right\} \begin{array}{l}\text{Necessary and sufficient}\\ K\text{-}T \text{ conditions relative}\\ \text{to Problem I.}\end{array}$$

where $\mathbf{v}$ is called the *dual variable* (vector whose components have the same indexing set as the components of $\mathbf{a}$).

We call *Problem II* the following program:

$$\left.\begin{array}{ll}\text{To maximize } g(\mathbf{x}, \mathbf{v}) \equiv f(\mathbf{x}) + \mathbf{v}^T \mathbf{a}(\mathbf{x}) & \\ \text{under the conditions} & \\ \mathbf{v}^T \frac{d\mathbf{a}}{d\mathbf{x}} + \frac{df}{d\mathbf{x}} = 0 & (5)\\ \mathbf{v} \geqslant 0 . & (6)\end{array}\right\} \text{Problem II}$$

Problem II is said to be a *dual program* of Problem I.

Duality theorem

1) *If there exists a vector* $\hat{\mathbf{x}}$ *that minimizes* $f(\mathbf{x})$ *in Problem I, then the vectors* $\hat{\mathbf{x}}$ *and* $\hat{\mathbf{v}}$ *maximize* $g(\mathbf{x}, \mathbf{v})$ *in Problem II.*

*Electricité de France, Paris, France.

2) *Conversely, if* $\mathbf{x} = \bar{\mathbf{x}}$ *and* $\mathbf{v} = \bar{\mathbf{v}}$ *maximize* $g(\mathbf{x}, \mathbf{v})$ *in Problem II and if in addition the matrix*

$$\overline{\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{v}^T \frac{d\mathbf{a}}{d\mathbf{x}}\right)} + \overline{\frac{d^2 f}{d\mathbf{x}^2}}$$

has an inverse, then the vector $\mathbf{x} = \bar{\mathbf{x}}$ *minimizes* $f(\mathbf{x})$ *in Problem I. In both cases, we have:*

minimum $f(\mathbf{x}) =$ *maximum* $g(\mathbf{x}, \mathbf{v})$.

Proof

1) Let us first observe that from (2) and (3) we have $\mathbf{x} = \hat{\mathbf{x}}$ and $\mathbf{v} = \hat{\mathbf{v}}$ as feasible solutions of Problem II. On the other hand if $\mathbf{x}$, $\mathbf{v}$ are any feasible solutions of Problem II, we have

$$\begin{aligned} &g(\hat{\mathbf{x}}, \hat{\mathbf{v}}) - g(\mathbf{x}, \mathbf{v}) \\ &\quad = \hat{\mathbf{f}} - f(\mathbf{x}) + \hat{\mathbf{v}}^T \hat{\mathbf{a}} - \mathbf{v}^T \mathbf{a}(\mathbf{x}) \\ &\quad \geqslant \frac{df}{d\mathbf{x}}(\hat{\mathbf{x}} - \mathbf{x}) - \mathbf{v}^T \mathbf{a}(\mathbf{x}) && \text{from (4) and } f \text{ convex} \\ &\quad \geqslant \left(\frac{df}{d\mathbf{x}} + \mathbf{v}^T \frac{d\mathbf{a}}{d\mathbf{x}}\right)(\hat{\mathbf{x}} - \mathbf{x}) - \mathbf{v}^T \hat{\mathbf{a}} && \text{for } \mathbf{a} \text{ is convex and from (6) } \mathbf{v} \geqslant 0 \\ &\quad = -\mathbf{v}^T \hat{\mathbf{a}} && \text{from (5)} \\ &\quad \geqslant 0 \,. && \text{from (1) and (6).} \end{aligned}$$

Finally, we have

$$g(\hat{\mathbf{x}}, \hat{\mathbf{v}}) - g(\mathbf{x}, \mathbf{v}) \geqslant 0 \tag{7}$$

for all pair $\mathbf{x}$, $\mathbf{v}$ satisfying (5) and (6), which proves that the feasible solution $\hat{\mathbf{x}}$, $\hat{\mathbf{v}}$ of Problem II is optimal. In addition

$$g(\hat{\mathbf{x}}, \hat{\mathbf{v}}) = f(\hat{\mathbf{x}}) + \hat{\mathbf{v}}^T \hat{\mathbf{a}} = f(\hat{\mathbf{x}}) \tag{8}$$

from (4), which concludes the proof of the first half of the theroem.

2) Let the vectors $\bar{\mathbf{x}}$, $\bar{\mathbf{v}}$ be an optimal solution of Problem II and call $\bar{\mathbf{y}}$ the corresponding dual variable. We write the necessary[5] *K-T* conditions, relative to Problem II:

$$\left.\begin{aligned} &\overline{\frac{d\mathbf{a}}{d\mathbf{x}}}\,\bar{\mathbf{y}} - \bar{\mathbf{a}} \geqslant 0 && (9) \\ &\left[\overline{\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{v}^T \frac{d\mathbf{a}}{d\mathbf{x}}\right)} + \overline{\frac{d^2 f}{d\mathbf{x}^2}}\right]\bar{\mathbf{y}} = 0 && (10) \\ &\bar{\mathbf{v}}^T\left[\overline{\frac{d\mathbf{a}}{d\mathbf{x}}}\,\bar{\mathbf{y}} - \bar{\mathbf{a}}\right] = 0 \,. && (11) \end{aligned}\right\} \begin{array}{l}\text{necessary } K\text{-}T \\ \text{conditions relative} \\ \text{to Problem II.}\end{array}$$

We must introduce here an additional hypothesis: the square matrix which multiplies $\bar{\mathbf{y}}$ on the left in Eq. (10) is presumed to have an inverse;[6] this implies that:

$$\bar{\mathbf{y}} = 0 \,. \tag{12}$$

Therefore it is easy to observe that the Eqs. (5) and (6) for $\bar{\mathbf{x}}$ and $\bar{\mathbf{v}}$ and the Eqs. (9), (11) and (12) imply the Eqs. (1) to (4) for $\hat{\mathbf{x}} = \bar{\mathbf{x}}$ and $\hat{\mathbf{v}} = \bar{\mathbf{v}}$.

Thus:

(9) and (12) $\Rightarrow$ (1)
(6) $\Rightarrow$ (2)
(5) $\Rightarrow$ (3)
(11) and (12) $\Rightarrow$ (4),

and therefore $\hat{\mathbf{x}} = \bar{\mathbf{x}}$ is an optimal solution of Problem I, since the conditions (1) to (4) are sufficient (f and $\mathbf{a}$ being convex).

In addition, from (11) and (12) we obtain $f(\bar{\mathbf{x}}) = g(\bar{\mathbf{x}}, \bar{\mathbf{v}})$, which completes the second half of the proof.

The case where f and a are partially linear

By partially linear functions we mean functions such that:

$$f(\mathbf{x}) \equiv h(\mathbf{x}_K) + \mathbf{l}^T \mathbf{x}_{\bar{K}} \tag{13}$$

$$\mathbf{a}(\mathbf{x}) \equiv \mathbf{b}(\mathbf{x}_K) + C\mathbf{x}_{\bar{K}} - \mathbf{c} \,, \tag{14}$$

where K and $\bar{K}$ are complementary subsets of the set indexing the components of $\mathbf{x}$.

$h(\mathbf{x}_K)$ is a scalar function of $\mathbf{x}_K$
$\mathbf{b}(\mathbf{x}_K)$ is a vector function of $\mathbf{x}_K$
$\mathbf{l}$ is a constant vector
C is a constant matrix
$\mathbf{c}$ is a vector.

Under these conditions, the matrix in (10) becomes

$$\overline{\frac{\partial}{\partial \mathbf{x}}\left(\mathbf{v}^T \frac{d\mathbf{a}}{d\mathbf{x}}\right)} + \overline{\frac{d^2 f}{d\mathbf{x}^2}} \equiv \begin{bmatrix} \dfrac{\partial}{\partial \mathbf{x}_K}\left(\mathbf{v}^T \dfrac{d\mathbf{b}}{d\mathbf{x}_K}\right) + \dfrac{d^2 h}{d{\mathbf{x}_K}^2} & 0 \\ 0 & 0 \end{bmatrix}. \tag{15}$$

Since this matrix is clearly singular, the hypothesis which was introduced in the proof of the second half of the theorem fails to be satisfied.

Let us consider first the extreme case where $K = \phi$, i.e., a program which is purely linear: Problem II and its Kuhn and Tucker conditions become

To maximize $-\mathbf{v}^T \mathbf{c}$

$$\left.\begin{aligned} \mathbf{v}^T C + \mathbf{l}^T &= 0 && (16) \\ \mathbf{v} &\geqslant 0 && (17) \end{aligned}\right\} \text{Problem II}$$

$$C\bar{\mathbf{y}} - C\bar{\mathbf{x}} + \mathbf{c} \geqslant 0 \tag{18}$$

$$0\mathbf{x}\bar{\mathbf{y}} = 0 \tag{19}$$

$$\bar{\mathbf{v}}^T[C\bar{\mathbf{y}} - C\bar{\mathbf{x}} + \mathbf{c}] = 0 \,. \tag{20}$$

By observing that Problem II has thus become independent of $\mathbf{x}$, we could set $\bar{\mathbf{x}} = 0$ and consequently Eqs. (16) to (20) immediately imply that $-\mathbf{x} = \bar{\mathbf{y}}$ is an optimal solution of Problem I. In other words $-\mathbf{y}$ has replaced $\mathbf{x}$ in the reciprocal theorem.

In the case where $f(\mathbf{x})$ and $\mathbf{a}(\mathbf{x})$ are partially linear, and defined by (13) and (14), we can show easily that the reciprocal theorem in the next section follows as a consequence.

The modified theorem

Second half—Reciprocally, if $\mathbf{x} = \bar{\mathbf{x}}$ and $\mathbf{v} = \bar{\mathbf{v}}$ maximize $g(\mathbf{x}, \mathbf{v})$ in Problem II, and if in addition the matrix

$$\overline{\frac{\partial}{\partial \mathbf{x}_K}\left(\mathbf{v}^T \frac{d\mathbf{b}}{d\mathbf{x}_K}\right)} + \overline{\frac{d^2 h}{d\mathbf{x}_K{}^2}}$$

has an inverse, the vector $\bar{\mathbf{x}}$ defined by

$$\bar{\mathbf{x}} = \begin{pmatrix} \bar{\mathbf{x}}_K \\ -\bar{\mathbf{y}}_{\bar{K}} \end{pmatrix}$$

minimizes $f(\mathbf{x})$ in Problem I. (We have $\bar{\mathbf{y}}_K = 0$ and $\bar{\mathbf{x}}_{\bar{K}} = 0$).

Remarks

In order to prove the second half of the theorem (in the case where $\mathbf{a}(\mathbf{x})$ is linear) Dorn introduces the following additional hypothesis (modified notations):

Calling $\mathbf{t}^T$ *the derivative of* $f(\mathbf{x})$ *with respect to* $\mathbf{x}$, *let*

$$\mathbf{t}^T = \frac{df}{d\mathbf{x}}. \tag{21}$$

Let us consider the inverse function $\mathbf{x}(\mathbf{t})$ defined implicitly by (21) and suppose that the matrix $d\mathbf{x}/d\mathbf{t}$ exists. This hypothesis is equivalent to saying (as was observed by Dorn) that the matrix $d^2f/d\mathbf{x}^2$ has an inverse, which is precisely our supposition (in this case it is well understood that $\mathbf{a}(\mathbf{x})$ is linear). This equivalence could be shown in the following fashion:

From (21) we have

$$I = \frac{d\mathbf{t}}{d\mathbf{t}} = \frac{d^2f}{d\mathbf{x}^2}\cdot\frac{d\mathbf{x}}{d\mathbf{t}} \text{ (where } I \text{ is the identity matrix)}, \tag{22}$$

which in turn implies that

$$\frac{d\mathbf{x}}{d\mathbf{t}} = \left[\frac{d^2f}{d\mathbf{x}^2}\right]^{-1}$$

Note added in proof

I recently saw the report of P. Wolfe "A duality theorem for nonlinear programming,"[7] which I had not yet seen when I wrote the Letter. Taking into consideration the very full report of P. Wolfe, the contribution of my Letter is reduced to the second part of the theorem.

References

1. W. S. Dorn, "A Duality Theorem for Convex Programs", *IBM Journal*, **4**, 407–413 (1960).
2. J. B. Dennis, *Mathematical programming and electrical networks*, Technology Press and John Wiley and Son, New York, N.Y., 1959.
3. H. W. Kuhn and A. W. Tucker, "Nonlinear Programming". Second Berkeley Symposium, pp. 481–492.
4. P. Huard, "Conditions de Kuhn et Tucker. Programme dual. Coûts marginaux" Groupe de Travail "Mathématique des programmes économiques" (AFCALTI-SOFRO) Séances du 30 Janvier et du 13 Février 1961.
5. We suppose that $(\bar{\mathbf{x}}, \bar{\mathbf{v}})$ is not a singular point, in the sense of Kuhn and Tucker. (See Ref. 3, pp. 483–484 and Ref. 4, p. 2.)
6. We shall see later how to weaken this hypothesis; otherwise we could not apply the theorem to programs where $f(\mathbf{x})$ is partially linear.
7. P. Wolfe, "A duality theorem for nonlinear programming", RAND Corporation Report P-2028, June 30, 1960; revised February 7, 1961. To appear in the *Quarterly of Applied Mathematics*.

Received August 4, 1961