DISCUSSION PAPER NO. 206

DUALITY IN UNCONSTRAINED OPTIMIZATION

by

Avinoam Perry

March 1976

Graduate School of Management
Northwestern University
Evanston, Illinois 60201
Duality in unconstrained nonlinear optimization was first introduced by Fletcher [3], who referred to the inverse transformation of the DFP update (Davidon [2], Fletcher and Powell [4]) as a dual update. Crem and Spedicato [8] applied the duality concept to the self-scaling variable metric procedure [6, 7] and showed that dual updates of either the DFP or the BFGS (Broyden [1], Goldfarb [5], Fletcher [3], Shanno [9]) procedures can be obtained directly upon replacing $p_k$ by $q_k$, $S_k$ by $T_k$, $\delta = 0$ by $\delta = 1$ and vice versa, where

$$
P_k = x_{k+1} - x_k
$$

$$
q_k = \nabla f(x_{k+1}) - \nabla f(x_k) = x_{k+1} - x_k
$$

$$
S_k = \nabla^2 f(x_k)
$$

More specifically, it was shown that the dual update of

$$
S_{k+1} = S_k - \frac{S_k q_k q_k^T S_k}{q_k^T S_k q_k} + \frac{S_k p_k p_k^T S_k}{p_k^T S_k p_k} + 8 v_k v_k^T
$$

(1)

can be formed by

$$
T_{k+1} = T_k - \frac{T_k p_k p_k^T T_k}{p_k^T T_k p_k} + \frac{T_k q_k q_k^T T_k}{q_k^T T_k q_k} + \frac{8 w_k w_k^T}{1 - \delta}
$$

(2)

where $\delta$ is either zero or one and $\delta = 1 - \delta$

$$
v_k = (q_k^T S_k q_k)^{-1} \left[ \begin{array}{c} \frac{T_k p_k}{p_k^T S_k p_k} - \frac{S_k q_k}{q_k^T S_k q_k} \\ \frac{T_k q_k}{q_k^T S_k q_k} - \frac{S_k p_k}{p_k^T S_k p_k} \end{array} \right]
$$

$$
w_k = (p_k^T T_k p_k)^{-1} \left[ \begin{array}{c} \frac{q_k}{q_k^T S_k q_k} - \frac{T_k p_k}{p_k^T S_k p_k} \\ \frac{q_k}{q_k^T S_k q_k} - \frac{T_k q_k}{q_k^T S_k q_k} \end{array} \right]
$$

The purpose of this communication is to show that the existence of dual updates in quasi-Newton algorithms implies the existence of a sequence of equivalent programs which are dual to the original unconstrained quadratic problem

$$
\text{minimize } f(x)
$$

$$
x \in \mathbb{R}^n
$$
Definition 1: The quadratic problem

\[ \text{minimize } b(y) \]
\[ y \in \mathbb{E}^n \]

is a dual of (3) if there exist positive scalars \( \alpha \) and \( \epsilon \) where \( 0 < \alpha \leq \epsilon \) such that the one-dimensional program in \( \alpha \)

\[ \text{minimize } b(y_k - \alpha T_k v(y_k)) \]

is equivalent to the one-dimensional program in \( \alpha \)

\[ \text{minimize } f(x_k - \alpha S_k v(x_k)) \]

where \( T_k \) and \( S_k \) are quasi-Newton updates in the \( y \) space and the \( x \) space respectively, and \( T_k \) is the dual update of \( S_k \).

Theorem 1: Given a quadratic function \( f(x), x \in \mathbb{E}^n \), there exists a one-to-one mapping from \( \mathbb{E}^n \) onto \( \mathbb{E}^n \) which carries a point \( x \in \mathbb{E}^n \) onto a point \( y \in \mathbb{E}^n \) in its dual space such that the conditions specified by Definition 1 are satisfied.

Proof: Denote

\[ x = S_k y \]
\[ y = T_k x \]

where \( T_k = S_k^{-1} \)

Then: \( f(x) = f(S_k y) = h_k(y) \) and \( \nabla h_k(y) = S_k \nabla f(x) \). It follows that

\[ y_{k+1} = y_k - \alpha T_k v(y_k) = T_k x_k - \alpha \nabla f(x_k) = T_k x_{k+1} \]

Multiplying both sides of (10) by \( S_k \) yields

\[ x_{k+1} = x_k - \alpha S_k \nabla f(x_k) \]

which implies the equivalence of (5) and (6).
Another conclusion which follows from (8), (9) is that

\[ y_k - y_{k-1} = T_k (x_k - x_{k-1}) = T_k h_{k-1} = S_{k-1} \]

and

\[ \nabla h_k (y_k) - \nabla h_k (y_{k-1}) = S_k (\nabla f (x_k) - \nabla f (x_{k-1})) = S_k S_{k-1} = P_{k-1} \]

and by induction we have

\[ y_{j+1} - y_j = T_k (x_{j+1} - x_j) = T_k P_j = \nabla f (x_{j+1}) - \nabla f (x_j) \]

\[ \nabla h_k (y_{j+1}) - \nabla h_k (y_j) = S_k (\nabla f (x_{j+1}) - \nabla f (x_j)) = S_k S_j = x_{j+1} - x_j \]

for all \( j < k \).

This result implies that \( T_k \) is a quasi-Newton update of \( h_k (y) \).

It is important to note here that if \( S_k \) is a BFGS update in \( x \) space then
\( T_k \) is a DFP update in its dual space.

The economic interpretation of the dual in unconstrained optimization is compatible with the interpretation given to duality in the constrained case.

From (12)–(15) it becomes evident that \( \Delta x_k \) can be interpreted as the change in the instantaneous rate of change in \( f (x) \) at the new point \( x_k \), while \( \Delta x_k \) is interpreted as the change in the instantaneous rate of change in the dual program.

Note that \( \Delta x_k \) is the change in dual of the dual which is identical to the change in the primal variables.

The conclusion of this communication is that the existence of quasi-Newton dual updates implies the existence of a sequence of dual programs in the unconstrained optimization case. This sequence of programs can be generated by applying a sequence of linear transformations which carry \( x \in \mathbb{R}^n \) onto \( y \in \mathbb{R}^n \) so that the interpretation of \( \Delta y_k \) is compatible with the common interpretation of duality.
References


