J. Austral Math. Soc. 19 (Series B), (1976), 333–342.

# A PROGRAMMING PROBLEM WITH AN $L_p$ NORM IN THE OBJECTIVE FUNCTION

BERTRAM MOND and MURRAY SCHECHTER

(Received 3 February 1976)

(Revised 30 June 1976)

**Abstract**

We consider a programming problem in which the objective function is the sum of a differentiable function and the $p$ norm of $Sx$, where $S$ is a matrix and $p > 1$. The constraints are inequality constraints defined by differentiable functions. With the aid of a recent transposition theorem of Schechter we get a duality theorem and also a converse duality theorem for this problem. This result generalizes a result of Mond in which the objective function contains the square root of a positive semi-definite quadratic function.

## Introduction

Consider the programming problem

$$(P) \qquad \text{Minimize } F(x) = f(x) + \| Sx \|_p \qquad (p > 1)$$

$$\text{subject to } g(x) \geqq 0$$

where $f$ and $g$ are differentiable functions from $R^n$ into $R$ and $R^m$ respectively, $S$ is a $k \times n$ matrix and the $p$ norm is given by

$$\| y \|_p = \left( \sum_1^k | y_i |^p \right)^{1/p}.$$

Here necessary and sufficient conditions are given for a point to be optimal for $(P)$. A dual problem involving the conjugate norm is formulated and appropriate duality theorems established. Since $F$ may not be differentiable at the optimal point the Kuhn-Tucker conditions for a differentiable problem may not be applicable. This point is taken care of by using a special case of the general solvability theorem appearing in [9]. This procedure is



https://doi.org/10.1017/S0334270000001193 Published online by Cambridge University Press

similar to that carried out in [6], where the objective function contains the square root of a positive semi-definite quadratic form. In fact we show that the result of [6] is a special case of ours.

In order to use the transposition theorem it is necessary to impose a constraint qualification. This contraint qualification may be described in terms of directional derivatives, as discussed in [7] and from this description it can be seen that a number of classical constraint qualifications, such as the generalized Slater constraint quălification, imply the one we introduce here.

## 1. The transposition theorem

The following theorem appears in [9]:

THEOREM. *Let X be a finite dimensional real inner product space, let K be a closed convex cone and C a closed convex set, both in X. Let s be the support function of C. Then*

$$< x, y > \leqq s(x) \text{ for all } x \text{ in } K \text{ if and only if}$$

$$y \in \mathrm{cl}\,(K^0 + C)$$

*where for any set A*

$$A^0 = \{y \mid < x, y > \leqq 1 \quad \text{for all} \quad x \in A\}.$$

We get our desired transposition theorem by an appropriate choice of $X$, $K$ and $C$. We will use the following standard notation and facts: $p$ and $q$ are called *conjugate exponents* if $1/p + 1/q = 1$. The conjugate exponent of 1 is $\infty$ and, on $R^k$,

$$\| y \|_\infty = \max\{| y_i |, \ i = 1, \dots, k\}.$$

If $p \geqq 1$ and $p$ and $q$ are conjugate exponents then the Hölder inequality [3] says

$$| x'y | \leqq \| x \|_p \, \| y \|_q. \tag{1}$$

LEMMA 1.1 *Let S be a $k \times n$ matrix, $1 < p \leqq \infty$ and p and q conjugate exponents. Let*

$$C = \{y \mid y = S'z \text{ for some } z \in R^k, \| z \|_q \leqq 1\}. \qquad \textit{Then}$$

$$C^0 = \{x \mid \| Sx \|_p \leqq 1\}.$$

https://doi.org/10.1017/S0334270000001193 Published online by Cambridge University Press

PROOF. Let

$$D = \{x \mid \| Sx \|_p \leqq 1\}$$

If $x$ is in $D$ and $y = S'z$ with $\| z \|_q \leqq 1$ then $x'y = x'S'z = (Sx)'z \leqq \| Sx \|_p \| z \|_q \leqq 1$ where we have used (1). This shows that $D \subseteq C^0$.

Conversely, suppose $x$ lies in $C^0$, let $\xi = Sx$. We want to prove that $\| \xi \|_p \leqq 1$, if $\xi = 0$ we are finished. If not we will produce a vector $z$ such that $\| z \|_q = 1$ and $\xi' z = \| \xi \|_p$. If $p < \infty$ then $z$ is given by

$$z_i = (\operatorname{sgn} \xi_i) \mid \xi_i \mid^{p-1} (\| \xi \|_p)^{1-p}$$

where sgn $t = +1$ if $t \geqq 0$ and $-1$ otherwise. If $p = \infty$ then let $r$ be an integer such that $\| \xi \|_p = |\xi_r|$. Define $z$ by $z_r = \operatorname{sgn} \xi_r$ and $z_i = 0$ if $i \neq r$. It is easily verified that $z$ has the desired properties in either case. Since $x$ lies in $C^0$ and $\| z \|_q = 1$ we have

$$\| \xi \|_q = \xi' z = (Sx)' z = x'(S'z) \leqq 1$$

therefore $x$ lies in $D$, hence $C^0 \subseteq D$.

Now we can get our desired transposition theorem immediately.

THEOREM 1.1. *Let $A$ be an $m \times n$ matrix, $S$ a $k \times n$ matrix, $1 \leqq p \leqq \infty$ and $1/p + 1/q = 1$. Then*

$$Ax \geqq 0 \qquad \textit{implies} \qquad c'x + \| Sx \|_p \geqq 0$$

*if and only if there exists $y$ and $v$ such that*

$$A'y = c + S'v, \qquad y \geqq 0, \qquad \| v \|_q \leqq 1.$$

PROOF. In the theorem from [9] quoted above put

$$K = \{x \mid Ax \geqq 0\}.$$

Then it is easily verified that

$$K^0 = \{-A'y,\ y \geqq 0\}.$$

Put

$$C = \{y \mid y = S'z \text{ for some } z \in R^k, \qquad \| z \|_q \leqq 1\}.$$

$C$ is closed and convex and contains the origin. The support function of $C$ is the gauge function of $C^0$ (see [8] or [9]) and from Lemma 1.1 we see that this function is exactly $\| Sx \|_p$. $C$ is compact hence the sum $K^0 + C$ is closed so the theorem follows by substitution in the theorem quoted above, replacing $y$ by $-c$.

The transposition theorem of Eisenberg [1, 2] follows as a special case of our Theorem 1.1 when $p = 2$. This may be shown by a technique which will be used in Section 4 to show that our duality theorem includes the one proved by Mond in [6].

## 2. Notation and preliminaries

We will be dealing with the gradient of the $p$ norm frequently in what follows, so we begin by listing a number of easily verified formulas. We take $1 < p < \infty$ throughout. If $x \neq 0$ then $\| x \|_p$ is a differentiable function of $x$ and $\nabla \| x \|_p$ has as its $i$'th component $\operatorname{sgn} x_i \, | x_i |^{p-1} (\| x \|_p)^{1-p}$. From this we easily get the following two facts:

$$x' \nabla (\| x \|_p) = \| x \|_p. \tag{2}$$

$$\| \nabla (\| x \|_p) \|_q = 1. \tag{3}$$

To make the next computation possible we must have a name for the $p$ norm function. Accordingly, we define the function $h_p$ by $h_p(x) = \| x \|_p$. Then with $S$ a matrix of appropriate dimensions we have

$$\nabla (h_p(Sx)) = S' \nabla h_p(Sx) \tag{4a}$$

or more precisely

$$\nabla (h_p \circ S) = S'(\nabla h_p) \circ S. \tag{4b}$$

We have in this last formula taken the gradient of a scalar valued function to be a column and we will consistently follow this convention. If $g$ is a vector valued function then $\nabla g$ will denote the matrix which has $\nabla g_i$ as its $i$'th column. From this it follows that if $f$ is a scalar valued function then $\nabla^2 f = \nabla(\nabla f)$ is the matrix of second partial derivatives of $f$. It also follows that if $g$ is a vector valued function and $y$ is a constant vector then $\nabla(y'g) = (\nabla g)y$.

Next, returning to the problem $(P)$ we want to evaluate the directional derivative of the objective function $F$. We denote the directional derivative of $F$ at the point $x_0$ in the direction $z$ by $F'(x_0; z)$. If $Sx_0 \neq 0$ then $F$ is differentiable at $x_0$ and so $F'(x_0; z) = [\nabla F(x_0)]'z$. This gives, using (4a)

$$F'(x_0; z) = [\nabla f(x_0)]'z + [S' \nabla h_p(Sx_0)]'z, \qquad (Sx_0 \neq 0). \tag{5a}$$

If $Sx_0 = 0$ we use the definition of directional derivative to evaluate $h'_p(Sx_0; z)$

$$h'_p(Sx_0; z) = \lim_{t \to 0^+} t^{-1} [\| Sx_0 + tSz \|_p - \| Sx_0 \|_p]$$

$$= \lim_{t \to 0^+} t^{-1} \| tSz \|_p = \| Sz \|_p,$$

hence

$$F'(x_0; z) = [\nabla f(x_0)]'z + \| Sz \|_p \qquad (Sx_0 = 0). \tag{5b}$$

https://doi.org/10.1017/S0334270000001193 Published online by Cambridge University Press

Now, for a given feasible point $x_0$ define the set $Z_0$ as follows:

$$z \in Z_0 \quad \text{if} \quad [\nabla g_i(x_0)]'z \geqq 0 \quad \text{when} \quad g_i(x_0) = 0$$
$$\text{and also} \quad F'(x_0; z) < 0.$$

We know from [7] that if the generalized Slater constraint qualification is satisfied then $Z_0$ is empty when $x_0$ is optimal.

## 3. Necessary and sufficient conditions

From this point on $p$ will be a fixed number $> 1$ and $q$ will be its conjugate exponent.

THEOREM 3.1. *Suppose $x_0$ is optimal for $(P)$ and $Z_0$ is empty. Then there exists $y$ in $R^m$ and $v$ in $R^k$ satisfying*

$$\nabla y'g(x_0) = \nabla f(x_0) + S'v \tag{6}$$

$$y \geqq 0 \tag{7}$$

$$y'g(x_0) = 0 \tag{8}$$

$$\| v \|_q \leqq 1 \tag{9}$$

$$v'Sx_0 = \| Sx_0 \|_p. \tag{10}$$

PROOF. Suppose first that $Sx_0 \neq 0$. Then at the optimal point $x_0$ the Kuhn-Tucker conditions [4] are applicable. According to these there exists $y \geqq 0$ such that $y'g(x_0) = 0$ and $\nabla F(x_0) - \nabla y'g(x_0) = 0$. Using (4a) or (4b) to evaluate $\nabla F(x_0)$ this gives

$$\nabla F(x_0) - \nabla y'g(x_0) = \nabla f(x_0) + S'\nabla h_p(Sx_0) - \nabla y'g(x_0) = 0$$

where $h_p(x) = \| x \|_p$.

Let $v = \nabla h_p(Sx_0)$. Then from (3), $\| v \|_q = 1$ and from (2), $v'Sx_0 = \| Sx_0 \|_p$, therefore this choice of $v$ and $y$ satisfies conditions (6)–(10).

Now suppose $Sx_0 = 0$. Let $A$ be the matrix with rows $[\nabla g_i(x_0)]'$ for those $i$'s for which $g_i(x_0) = 0$. Since $Z_0$ is empty we know that

$$Az \geqq 0 \text{ implies } F'(x_0; z) \geq 0.$$

From (5b) we therefore have

$$Az \geqq 0 \text{ implies } [\nabla f(x_0)]'z + \| Sz \|_p \geqq 0.$$

https://doi.org/10.1017/S0334270000001193 Published online by Cambridge University Press

Then Theorem 1.1 tells us that there exist $y_0 \geqq 0$ and $v$ with $\|v\|_q \leqq 1$ such that

$$A'y = \nabla f(x_0) + S'v.$$

Finally, defining $y_i = 0$ for $1 \leqq i \leq m$, $g_i(x_0) \neq 0$, we see that this choice of $y$ and $v$ satisfies (6)–(10).

This last theorem tells us that, if $Z_0$ is empty, conditions (6)–(10) are necessary for $x_0$ to be optimal. The next theorem gives sufficient conditions for $x_0$ to be optimal.

THEOREM 3.2. *If $f$ is convex and $g$ concave and there exists $(x_0, y, v)$ satisfying (6)–(10) with $g(x_0) \geqq 0$ then $x_0$ is optimal for the problem $(P)$.*

PROOF. Let $x$ be feasible for $(P)$. We will show that $F(x) \geqq F(x_0)$.

$$\begin{aligned} F(x) - F(x_0) &= f(x) - f(x_0) + \|Sx\|_p - \|Sx_0\|_p \\ &\geqq (x - x_0)'\nabla f(x_0) + \|Sx\|_p - \|Sx_0\|_p \qquad \text{(by the convexity of } f) \\ &= (x - x_0)'\nabla y'g(x_0) - (x - x_0)'S'v + \|Sx\|_p - \|Sx_0\|_p \qquad \text{by (6))} \\ &\geqq y'g(x) - y'g(x_0) - x'S'v + \|Sx\|_p \end{aligned}$$

(where we have used the concavity of $g$, (7) and (10))

$$\geqq y'g(x) - x'S'v + \|Sx\|_p\|v\|_q \geqq 0$$

where we have used (8), (9) and the Hölder inequality (1).

## 4. Duality

It will be assumed henceforth that $f$ is convex and $g$ is concave. Under this hypothesis we shall establish duality relationships between problem $(P)$ and the following problem:

$$(D) \text{ Maximize } G(y, u, v) = f(u) - y'g(u) + u'[\nabla y'g(u) - \nabla f(u)]$$

subject to

$$\nabla y'g(u) = \nabla f(u) + S'v \tag{11}$$

$$\|v\|_q \leqq 1 \tag{12}$$

$$y \geqq 0. \tag{13}$$

THEOREM 4.1. (*Weak Duality*) *If $x$ is feasible for $(P)$ and $(y, u, v)$ is feasible for $D$ then $F(x) \geqq G(y, u, v)$.*

PROOF. $F(x) - G(y, u, v)$

$= [f(x) - f(u)] + \| Sx \|_p + y'g(u) - u'[\nabla y'g(u) - \nabla f(u)]$

$\geqq x'\nabla f(u) + \| Sx\|_p + y'g(u) - u'\nabla y'g(u)$ (by the convexity of $f$)

$= x'[\nabla y'g(u) - S'v] + \| Sx \|_p + y'g(u) - u'\nabla y'g(u)$ (by (11))

$\geqq y'g(x) - x'S'v + \| Sx \|_p$ (by the concavity of $g$)

$\geqq 0 - \| Sx \|_p \| v \|_q + \| Sx \|_p$ (by the feasibility of $x$, (13) and (1))

$\geqq 0$ by(12).

THEOREM 4.2. (*Strong Duality*) *If $x_0$ is optimal for $(P)$ and $Z_0$ is empty then there exists $(y, u, v)$ with $u = x_0$ which is optimal for the dual and the extreme values are equal.*

PROOF. By Theorem 3.1 there exist $y$ and $v$ satisfying (6)–(10). From (6), (7) and (9), $(y, x_0, v)$ is feasible for the dual problem $(D)$. By weak duality (Theorem 4.1) $(y, x_0, v)$ will be optimal if $G(y, x_0, u) = F(x_0)$. Now

$$\begin{aligned} F(x_0) &= f(x_0) + \| Sx_0 \|_p = f(x_0) + v'Sx_0 \\ &= f(x_0) + v'Sx_0 - y'g(x_0) \\ &= f(x_0) - y'g(x_0) + x_0'[\nabla y'g(x_0) - \nabla f(x_0)] = G(y, x_0, v). \end{aligned}$$

Now we want to prove a theorem going in the opposite direction, i.e., showing how from an optimal solution of $(D)$ we can get an optimal solution of $(P)$. It will be convenient to note first the following computational facts: if $f$ and $g$ are vector valued functions then

$$\nabla(f'g) = (\nabla f)g + (\nabla g)(f). \tag{14}$$

Also, the gradient of the vector valued identity functions is the identity matrix; i.e., if $u$ is a vector variable then

$$\nabla_u(u) = I. \tag{15}$$

THEOREM 4.3. (*Converse Duality*) *If $(y_0, u_o, v_0)$ is an optimal solution of the dual problem $(D)$ and the matrix $\nabla^2 y_0'g(u_0) - \nabla^2 f(u_0)$ is non-singular then $u_0$ is optimal for the primal problem $(P)$ and the two extreme values are equal.*

PROOF. In the dual problem $(D)$ one of the constraints (12), may be non-differentiable at the optimal point; this will be the case if $v_0 = 0$. Note, however, that if we replace the constraint (12) by

$$h(v) = (\| v \|_q)^q \leqq 1 \tag{16}$$

https://doi.org/10.1017/S0334270000001193 Published online by Cambridge University Press

we have a problem which is clearly equivalent to $(D)$ and which has both objective function and constraints differentiable. Furthermore $(y_0, u_0, v_0)$ is optimal for this new problem and hence the generalized Fritz John conditions are satisfied, [5]. Accordingly let $\alpha$ and $\lambda$ be scalars, $x$ and $z$ vectors in $R^n$ and $R^m$ respectively and consider the function

$$J(y, u, v) = -\alpha G(y, u, v) - x'[S'v + \nabla f(u) - \nabla y' g(u)] + \\ + \lambda[(\| v \|_q)^q - 1] - z'y.$$

From [5] we may assume that $(\alpha,\lambda,x,z)$ has been chosen so that $\alpha$, $\lambda$ and $z \geqq 0$, $(\alpha,\lambda,x,z) \neq 0$ and the partial derivatives of $J$ all vanish at $(y_0, u_0, v_0)$. We evaluate these partial derivatives using (14) and (15).

$$\nabla_u J(y_0,u_0,v_0) = [\nabla^2 y_0' g(u_0) - \nabla^2 f(u_0)](x - \alpha u_0) = 0 \tag{17}$$

$$\nabla_y J(y_0,u_0,v_0) = \alpha g(u_0) + [\nabla g(u_0)]'(x - \alpha u_0) - z = 0 \tag{18}$$

$$\nabla_v J(y_0,u_0,v_0) = -Sx + \lambda \nabla h(v_0) = 0 \tag{19}$$

where $h(v) = (\| v \|_q)^q$.

In addition the Fritz John conditions give

$$\lambda[h(v_0) - 1] - z'y_0 = 0. \tag{20}$$

By our hypothesis the matrix appearing in (17) is non-singular therefore, (17) implies $x = \alpha u_0$. Thus if $\alpha = 0$ then $x = 0$ and by (18), $z = 0$. If $v_0 \neq 0$ then a simple calculation gives $\nabla h(v_0) \neq 0$ so that (19) implies $\lambda = 0$. If $v_0 = 0$ then we deduce from (20) that $\lambda = 0$, so that in either case $\lambda = 0$ and hence if $\alpha = 0$ then $(\alpha,\lambda,x,z) = 0$. This is contrary to our assumption, therefore $\alpha > 0$. By appropriate normalization we may therefore suppose that $\alpha = 1$. Then we get from (17), $x = u_0$ and from (18), $g(u_0) = z \geqq 0$. Hence $u_0$ is feasible for the primal problem $(P)$.

Since $(y_0, u_0, v_0)$ is feasible for $(D)$ we know from (12) or (16) that $h(v_0) \leqq 1$ and from (20) that $\lambda[h(v_0) - 1] \geqq 0$, hence either $\lambda = 0$ or $h(v_0) = 1$ or both. If $\lambda = 0$ then (19) tells us that $Sx = 0$, so $v_0'Sx = 0 = \| Sx \|_p$.

If, on the other hand, $h(v_0) = 1$, then from (19)

$$v_0'Sx = \lambda v_0' \nabla h(v_0) = \lambda q \qquad \text{and also}$$

$$\lambda \|\nabla h(v_0)\|_p = \| Sx \|_p = \lambda q.$$

Therefore in either case, $v_0'Sx = \| Sx \|_p$.

Now, we can show the optimality of $u = x_0$ as follows

https://doi.org/10.1017/S0334270000001193 Published online by Cambridge University Press

$$
\begin{aligned}
F(x) = f(x) + \| Sx \|_p &\leqq f(u_0) - y_0' g(u_0) + v_0' S u_0 \\
&\quad (\text{since } y_0'(u_0) \geqq 0) \\
&= f(u_0) - y_0' g(u_0) + u_0' [\nabla y_0' g(u_0) - \nabla f(u_0)] \quad (\text{by } (11)) \\
&= G(y_0, u_0, v_0).
\end{aligned}
$$

This inequality together with weak duality (Theorem 4.1) completes the proof.

Finally, we want to show how the results we have obtained contain those of [6] as a special case. To begin with we observe that if $B$ is an $n \times n$ positive semi-definite matrix then there exists an $n \times n$ matrix $S$ such that $B = S'S$. This follows from the fact that $B$ is orthogonally equivalent to a diagonal matrix with non negative entries. This diagonal matrix may be written $D^2$ where $D$ is another diagonal matrix and we can, therefore, write $B = P'D^2P$ where $P$ is some orthogonal matrix. The desired result follows with $S = DP$.

Now, in the primal problem $(P)$ put $p = 2$ and choose $S$ so that $S'S = B$, where $B$ is a given positive semi-definite matrix. Then, $(P)$ takes the form

$$
(P') \quad \text{minimize} \quad f(x) + (x'Bx)^{1/2}
$$
$$
\text{subject to} \quad g(x) \geqq 0.
$$

Let $(y_0, u_0, v_0)$ be feasible for the corresponding dual problem. Since $R^n$ is the direct sum of the range of $S$ and the null space of $S'$ we may write $v_0 = Sw + v_1$ for some $w$ and some $v_1$ satisfying $S'y_1 = 0$. Furthermore, $Sw$ and $v_1$ are orthogonal so that

$$
\| v_0 \|_2 = \| Sw \|_2 + \| v_1 \|_2 \geqq \| Sw \|_2
$$

hence the constraints (11) and (12) of $(D)$ imply

$$
\nabla y'g(u) = \nabla f(u) + S'Sw = \nabla f(u) + Bw \tag{21}
$$

$$
w'Bw = (Sw)'(Sw) = (\| Sw \|_2)^2 \leqq (\| v_0 \|_2)^2 \leqq 1. \tag{22}
$$

Conversely if $(y, u, w)$ satisfies (21) and (22) then $(y, u, Bw)$ satisfies (11) and (12). Noting that $v$ does not appear in the objective function of $(D)$ we have shown that $(D)$ is equivalent to the following:

$$
(D') \quad \text{Maximize} \quad f(u) - y'g(u) + u'[\nabla y'g(u) - \nabla f(u)]
$$
$$
\text{subject to} \quad \nabla y'g(u) = \nabla f(u) + Bw
$$
$$
w'Bw \leqq 1, \qquad y \geqq 0.
$$

$(P')$ and $(D')$ are exactly the dual pair discussed in [6] and our results contain the results of that paper.

**References**

[1] E. Eisenberg, 'Supports of a convex function', *Bull. Am. Math. Soc.* 68, (1962), 192–195.
[2] E. Eisenberg, 'A gradient inequality for a class of non-differentiable functions', *Oper. Res.* 14 (1966), 157–163.
[3] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, Macmillan Company, New York, (1971).
[4] H. W. Kuhn and A. W. Tucker, 'Nonlinear programming' in *Proceedings 2nd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 481–492, University of California Press, (1951).
[5] O. L. Mangasarian and S. Fromovitz, 'The Fritz John necessary optimality conditions in the presence of equality and inequality constraints', *J. Math. Anal. Appl.* 17 (1967), 37–47.
[6] B. Mond, 'Aclass of non-differentiable mathematical programming problems', *J. Math. Anal. Appl.* 46 (1974), 169–174.
[7] B. Mond and M. Schechter, 'On a constraint qualification in a non-differentiable programming problem', *Nav. Res. Log. Quart.* 23 (1976), 611–613.
[8] R. T. Rockafeller, *Convex Analysis*, Princeton University Press, Princeton, N.J., (1969).
[9] M. Schechter, 'A solvability theorem for homogeneous functions', *SIAM J. Math. Anal.*, 7 (1976), 696–701.

Department of Pure Mathematics,
Latrobe University,
Melbourne,
Australia

and

Department of Applied Mathematics,
Technion,
Israel Institute of Technology,
Haifa,
Israel.

Department of Mathematics,
Lehigh University,
Bethlehem, Penna,
U.S.A.

and

Department of Applied Mathematics,
Technion,
Israel Institute of Technology,
Haifa,
Israel.