JOURNAL OF APPROXIMATION THEORY 23, 401–411 (1978)

A Class of Programming Problems Whose Objective Function Contains a Norm

G. A. WATSON

Department of Mathematics, University of Dundee, Dundee, Scotland

Communicated by E. W. Cheney

Received February 22, 1977

1. INTRODUCTION

Consider the problem

(P) Find $\mathbf{x}$ to minimize $f(\mathbf{x}) + \|\mathbf{r}(\mathbf{x})\|$ subject to $\mathbf{g}(\mathbf{x}) \geqslant 0$,

where $f(\mathbf{x})$, $\mathbf{r}(\mathbf{x})$, and $\mathbf{g}(\mathbf{x})$ are differentiable functions from R^n into R, R^m, and R^t, respectively, and R^m is equipped with an abstract norm $\|\cdot\|$. We give here necessary conditions for $\mathbf{x}$ to be a solution of the above problem. Under certain additional assumptions, we show that these conditions are also sufficient, and subsequently formulate a dual problem and establish appropriate relationships between the solution to this and the original problem. Our results specifically generalize those of Mond and Schechter [7] and also include the results of, for example, Mond [6] and Sreedharan [10] as special cases. The theorems given here are particularly relevant to the problem of (constrained) best approximation in abstract norms. For example, in the unconstrained linear case, a number of algorithms have been proposed which solve the dual, rather than the original primal problem [1, 2, 11]. A primal-dual pair for a related class of linear problems is given by Oettli [8].

Throughout, we use the notation that $\mathbf{d}(\mathbf{x})$ denotes the vector of partial derivatives of f with respect to $\mathbf{x}$, $A(\mathbf{x})$ denotes the $m \times n$ matrix of partial derivatives of $\mathbf{r}$ with respect to $\mathbf{x}$, and $G(\mathbf{x})$ is the $t \times n$ matrix of partial derivatives of $\mathbf{g}$ with respect to $\mathbf{x}$. (The dependence on $\mathbf{x}$ is often suppressed in the notation, when no confusion is likely.) If $\mathbf{x}$ satisfies the constraints $\mathbf{g}(\mathbf{x}) \geqslant 0$, $\mathbf{x}$ is said to be *feasible*, and, following Fiacco and McCormick [4], we can define a *feasible direction* at $\mathbf{x}$ as a direction which is tangent at $\mathbf{x}$ to a once differentiable arc, the arc emanating from $\mathbf{x}$ and contained in the feasible region (which we will assume has a nonempty interior).

If a regularity condition, for example the first-order constraint qualification of Kuhn and Tucker (see [4]) is assumed satisfied, the set of feasible directions may be identified with the set of feasible directions for the constraints

0021-9045/78/0234–0401$02.00/0

640/23/4-8

linearized at $\mathbf{x}$, and therefore may be conveniently described as follows. Let G_1 be the matrix obtained from G by deleting the rows corresponding to constraints holding with strict inequality. Then a direction γ is said to be feasible at $\mathbf{x}$ if γ satisfies

$$G_1\gamma \geqslant 0. \tag{1.1}$$

We will assume throughout that the first-order constraint qualification holds at a solution to the problem.

It will be necessary to make use of some properties which are possessed by abstract norms (for details of these, see Householder [5]). In particular, it is natural to define the dual norm $\|\cdot\|^*$ on R^m by means of the relation

$$\|\mathbf{v}\|^* = \max_{\|\mathbf{u}\|=1} \mathbf{u}^T\mathbf{v}, \tag{1.2}$$

where $\mathbf{u}, \mathbf{v} \in R^m$. We can also write

$$\|\mathbf{v}\| = \max_{\|\mathbf{u}\|^*=1} \mathbf{u}^T\mathbf{v}. \tag{1.3}$$

Examples of dual pairs of norms are readily obtained from the class of L_p norms, $1 \leqslant p \leqslant \infty$: the norms L_p and L_q are dual when $1/p + 1/q = 1$.

We also require the *subdifferential* (or set of subgradients) of a convex function $h(\mathbf{w})$ at $\mathbf{w}$, which we denote by $\partial h(\mathbf{w})$. A vector $\mathbf{v}$ is a subgradient of $h(\mathbf{w})$ at $\mathbf{w}$ if it satisfies the subgradient inequality

$$h(\mathbf{u}) \geqslant h(\mathbf{w}) + \mathbf{v}^T(\mathbf{u} - \mathbf{w}), \qquad \forall\, \mathbf{u} \tag{1.4}$$

(for a full discussion of the properties of subgradients, see Rockafellar [9]). In particular, we make frequent use of the subdifferential of $\|\mathbf{w}\|$ at $\mathbf{w}$, which, using (1.4), may be shown to be given by

$$\partial\|\mathbf{w}\| = \{\mathbf{v} : \|\mathbf{w}\| = \mathbf{w}^T\mathbf{v}, \|\mathbf{v}\|^* \leqslant 1\}. \tag{1.5}$$

If $h(\mathbf{w})$ is differentiable at $\mathbf{w}$, then $\partial h(\mathbf{w})$ consists of a single vector $\mathbf{v}$, which is just the gradient of $h(\mathbf{w})$ at $\mathbf{w}$.

Finally, we require the set of feasible directions with respect to a constraint of the form

$$\|\mathbf{w}\| \leqslant 1. \tag{1.6}$$

By analogy with the definition given above for a differentiable constraint, we say that γ is a feasible direction for (1.6) at $\mathbf{w}$ if, when (1.6) holds with equality,

$$\gamma^T\mathbf{v} \leqslant 0 \qquad \text{for all } \mathbf{v} \in \partial\|\mathbf{w}\|. \tag{1.7}$$

2. Necessary and Sufficient Conditions

Without imposing further assumptions on problem (P), it is possible to give necessary conditions for a vector $\mathbf{x}$ to be a solution. We require the following preliminary lemma.

Lemma 1. *Given a vector* $\mathbf{x} \in R^n$, let $\mathbf{x}(\theta)$ *define a once differentiable arc, parameterized by* $\theta \geqslant 0$ *in an interval* $[0, T]$, *where* $T > 0$, *and emanating from* $\mathbf{x} \equiv \mathbf{x}(0)$, *and let* $\mathbf{w}(\theta) \in \partial \| \mathbf{r}(\mathbf{x}(\theta))\|$. *Then the limit points as* $\theta \to 0$ *of the sequence* $\{\mathbf{w}(\theta)\}$ *all lie in* $\partial \| \mathbf{r}(\mathbf{x})\|$.

Proof. Writing $\mathbf{w} \equiv \mathbf{w}(0)$, for all θ satisfying $0 \leqslant \theta \leqslant T$, we have

$$\begin{aligned}\mathbf{w}^T\mathbf{r}(\mathbf{x}(\theta)) &\leqslant \| \mathbf{r}(\mathbf{x}(\theta))\| \qquad \text{using Eq. (1.3)}\\ &= \mathbf{w}(\theta)^T\, \mathbf{r}(\mathbf{x}(\theta))\\ &= \mathbf{w}(\theta)^T\, \mathbf{r}(\mathbf{x}) + \theta \mathbf{w}(\theta)^T A\mathbf{z}(\theta) + O(\theta^2),\end{aligned}$$

where $z_i(\theta) = dx_i(\theta)/d\theta$, $i = 1, 2, \ldots, n$.

Thus $\| \mathbf{r}(\mathbf{x})\| + \theta \mathbf{w}^T A\mathbf{z}(\theta) \leqslant \mathbf{w}(\theta)^T\mathbf{r}(\mathbf{x}) + \theta \mathbf{w}(\theta)^T A\mathbf{z}(\theta) + O(\theta^2)$, i.e.,

$$\theta(\mathbf{w}(\theta)^T A\mathbf{z}(\theta) - \mathbf{w}^T A\mathbf{z}(\theta)) + O(\theta^2) \geqslant \| \mathbf{r}(\mathbf{x})\| - \mathbf{w}(\theta)^T\, \mathbf{r}(\mathbf{x}) \geqslant 0.$$

The result follows on letting $\theta \to 0$. Q.E.D.

Theorem 1. *Let* $\mathbf{x}$ *solve* (*P*), *and let* $V \equiv \partial \| \mathbf{r} \|$. *Then* $\exists\, \mathbf{v} \in V, \boldsymbol{\lambda} \geqslant 0$ such that

$$\boldsymbol{\lambda}^T\mathbf{g} = 0,$$

$$\mathbf{v}^T A + \mathbf{d}^T = \boldsymbol{\lambda}^T G.$$

Proof. Assume $\mathbf{x}$ is a solution, and that vectors $\mathbf{v} \in V$, $\boldsymbol{\lambda} \geqslant 0$ satisfying the given conditions do not exist.

Thus

$$0 \notin \operatorname{conv} \left\{ [\mathbf{v}^T, \boldsymbol{\lambda}^T] \begin{bmatrix} A & 0 \\ -G & \mathbf{g} \end{bmatrix} + [\mathbf{d}^T\ 0],\ \mathbf{v} \in V, \boldsymbol{\lambda} \geqslant 0 \right\},$$

and by the theorem on linear inequalities (Cheney [3, p. 19]), $\exists\, \tilde{\mathbf{z}} = \left[\begin{smallmatrix} \mathbf{z} \\ \alpha \end{smallmatrix}\right]$ such that

$$\mathbf{v}^T A\mathbf{z} + \mathbf{d}^T\mathbf{z} + \boldsymbol{\lambda}^T(\alpha \mathbf{g} - G\mathbf{z}) < 0 \qquad \forall\, \mathbf{v} \in V, \boldsymbol{\lambda} \geqslant 0. \tag{2.1}$$

Taking $\boldsymbol{\lambda} = \mathbf{0}$, it is clear that

$$\mathbf{v}^T A\mathbf{z} + \mathbf{d}^T\mathbf{z} < 0 \qquad \text{for all } \mathbf{v} \in V. \tag{2.2}$$

In addition we must have

$$Gz \geqslant \alpha g, \tag{2.3}$$

otherwise we can violate (2.1) by a suitable choice of $\boldsymbol{\lambda}$. Thus, if G is partitioned as before so that G_1 consists of the rows of G corresponding to zero components of $\mathbf{g}$, from (2.3) we have

$$G_1\mathbf{z} \geqslant 0$$

and it follows that $\mathbf{z}$ is a feasible direction.

Now let $\mathbf{x}(\theta)$ define a once differentiable arc, parametrized by θ in the interval $0 \leqslant \theta \leqslant T$, $T > 0$, emanating from $\mathbf{x} \equiv \mathbf{x}(0)$, and contained in the feasible region. Then, since $\mathbf{x}$ is a solution,

$$\| \mathbf{r}(\mathbf{x}(\theta))\| + f(\mathbf{x}(\theta)) - \| \mathbf{r}(\mathbf{x})\| - f(\mathbf{x}) \geqslant 0, \quad 0 \leqslant \theta \leqslant T.$$

Thus if $\mathbf{w}(\theta) \in \partial \| \mathbf{r}(\mathbf{x}(\theta))\|$, we have

$$\mathbf{w}(\theta)^T \mathbf{r}(\mathbf{x}(\theta)) + f(\mathbf{x}(\theta)) - \| \mathbf{r}(\mathbf{x})\| - f(\mathbf{x}) \geqslant 0, \quad 0 \leqslant \theta \leqslant T.$$

$$\therefore \mathbf{w}(\theta)^T \mathbf{r}(\mathbf{x}) + \theta \mathbf{w}(\theta)^T A\, \mathbf{z}(\theta) + \theta\, \mathbf{d}^T\mathbf{z}(\theta) - \| \mathbf{r}(\mathbf{x})\| + O(\theta^2) \geqslant 0 \quad 0 \leqslant \theta \leqslant T, \tag{2.4}$$

where $z_i(\theta) = dx_i(\theta)/d\theta$, $i = 1, 2,..., n$. Now, using Lemma 1, there exists a sequence $\theta_1, \theta_2,...$, in $[0, T]$, tending to zero, such that

$$\mathbf{w}(\theta_k)^T A\mathbf{z}(\theta_k) \to \mathbf{v}^T A\mathbf{z}, \qquad \text{as } k \to \infty,$$

for some $\mathbf{v} \in V$. It follows from (2.4) that

$$\mathbf{v}^T A\mathbf{z} + \mathbf{d}^T\mathbf{z} \geqslant 0$$

for some $\mathbf{v} \in V$, which contradicts (2.2) and proves the result. Q.E.D.

Now let $f(\mathbf{x})$, $\| \mathbf{r}(\mathbf{x})\|$ be convex functions of $\mathbf{x}$, and let $\mathbf{g}$ be concave. Then problem (P) becomes a convex programming problem, which we refer to as *problem* (PC). We now show that the conditions of Theorem 1, together with the satisfaction of the constraints, are *sufficient* for $\mathbf{x}$ to be a solution to problem (PC). Since $\| \mathbf{r}(\mathbf{x})\|$ is a convex function of $\mathbf{x}$, we can define the subdifferential of $\| \mathbf{r}(\mathbf{x})\|$ at $\mathbf{x}$ *as a function of* $\mathbf{x}$, and this we denote by U. Letting $V = \partial \| \mathbf{r} \|$, we have

LEMMA 2. *$\mathbf{u} \in U$ iff $\mathbf{v} \in V$ with $\mathbf{u} = A^T\mathbf{v}$.*

Proof. Let $\mathbf{u} \in U$. Then the subgradient inequality (1.4) gives that

$$\| \mathbf{r}(\mathbf{z})\| \geqslant \| \mathbf{r}(\mathbf{x})\| + (\mathbf{z} - \mathbf{x})^T \mathbf{u}, \qquad \forall\, \mathbf{z}.$$

Thus the function $F(\mathbf{z}) = \|\mathbf{r}(\mathbf{z})\| - \mathbf{z}^T\mathbf{u}$ is minimized at $\mathbf{z} = \mathbf{x}$, and by Theorem 1 $\exists\, \mathbf{v} \in V$ such that $\mathbf{v}^T A = \mathbf{u}$.

Now let $\mathbf{v} \in V$. Since $\|\mathbf{r}(\mathbf{x})\|$ is a convex function of $\mathbf{x}$, we must have, for any $\mathbf{z}$,

$$\begin{aligned}
\|\mathbf{r}(\mathbf{z})\| - \|\mathbf{r}(\mathbf{x})\| &\geqslant (1/\mu)(\|\mathbf{r}(\mathbf{x} + \mu(\mathbf{z} - \mathbf{x}))\| - \|\mathbf{r}(\mathbf{x})\|) && 0 < \mu \leqslant 1 \\
&\geqslant (1/\mu)(\mathbf{v}^T\mathbf{r}(\mathbf{x} + \mu(\mathbf{z} - \mathbf{x})) - \|\mathbf{r}(\mathbf{x})\|) && 0 < \mu \leqslant 1 \\
&= \mathbf{v}^T A(\mathbf{z} - \mathbf{x}) + O(\mu) && 0 < \mu \leqslant 1.
\end{aligned}$$

Letting $\mu \to 0$, it follows that we must have $\mathbf{v}^T A \in U$, since $\mathbf{z}$ is arbitrary.
Q.E.D.

THEOREM 2. $\mathbf{x}$ solves (PC) iff $\exists\, \mathbf{v} \in V$, $\boldsymbol{\lambda} \geqslant 0$ such that

$$\mathbf{g} \geqslant 0,$$
$$\boldsymbol{\lambda}^T\mathbf{g} = 0,$$
$$\mathbf{v}^T A + \mathbf{d}^T = \boldsymbol{\lambda}^T G.$$

Proof. Necessity follows from Theorem 1. Let the conditions be satisfied at $\mathbf{x}$, and let $\mathbf{z}$ be any other feasible vector. Then

$$\begin{aligned}
&f(\mathbf{z}) + \|\mathbf{r}(\mathbf{z})\| - f(\mathbf{x}) - \|\mathbf{r}(\mathbf{x})\| \\
&\quad \geqslant \mathbf{d}(\mathbf{x})^T(\mathbf{z} - \mathbf{x}) + \mathbf{v}^T A(\mathbf{x})(\mathbf{z} - \mathbf{x}) && \text{by the convexity of } f \text{ and Lemma 2} \\
&\quad = \boldsymbol{\lambda}^T G(\mathbf{x})(\mathbf{z} - \mathbf{x}) \\
&\quad \geqslant \boldsymbol{\lambda}^T\mathbf{g}(\mathbf{z}) - \boldsymbol{\lambda}^T\mathbf{g}(\mathbf{x}) && \text{by the concavity of } \mathbf{g} \\
&\quad \geqslant 0.
\end{aligned}$$

Q.E.D.

Remark. This result may also be obtained as a consequence of Rockafellar [9, Theorem 28.3] and Lemma 2.

A particular application of Theorems 1 and 2 is in the provision of characterization results for nonlinear constrained best approximation problems ($f \equiv 0$). As an example, consider the case of L_∞ (Chebyshev) approximation. Here, if $\mathbf{r} \neq \mathbf{0}$,

$$V = \operatorname{conv}\{\operatorname{sgn}(r_j)\, \mathbf{e}_j,\, j \in J\},$$

where $J = \{j : |r_j| = \|\mathbf{r}\|\}$, and $\mathbf{e}_j$ is the jth coordinate vector. For a given $\mathbf{x}$, we will assume for convenience that the components of $\mathbf{r}$ are ordered so that $J = \{1, 2, \ldots, k\}$. Let G_1 be the matrix obtained from G by deleting the rows corresponding to constraints holding with strict inequality, and let A_1 be the matrix formed by the first k rows of A. Then, the conditions of Theorem 1 correspond to the existence of a vector $\boldsymbol{\alpha}$ satisfying

$$\boldsymbol{\alpha}^T \begin{bmatrix} A_1 \\ -G_1 \end{bmatrix} = 0,$$

$$\alpha_i \operatorname{sgn}(r_i) \geqslant 0, \qquad i = 1, 2, \ldots, k,$$

$$\alpha_i \geqslant 0, \qquad i > k,$$

$$\sum_{i=1}^{k} | \alpha_i | = 1.$$

In fact, using Carathéodory's theorem, an appropriate vector $\boldsymbol{\alpha}$ exists with at most $(n + 1)$ nonzero components.

3. Duality Theory

For certain (nondifferentiable) convex programming problems, the existence and nature of the dual problem is considered in general terms by Rockafellar [9, Section 30]. Because of the special nature of the class of problems with which we are concerned, it is easy to show directly that problem (PC) is dual to the following problem, problem (DC):

find $\mathbf{y}$, $\mathbf{v}$, $\boldsymbol{\lambda}$ to maximize

$$F(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda}) = f(\mathbf{y}) + \mathbf{v}^T\mathbf{r}(\mathbf{y}) - \boldsymbol{\lambda}^T\mathbf{g}(\mathbf{y})$$

subject to

$$\text{(DC)} \qquad \mathbf{v}^T A(\mathbf{y}) + \mathbf{d}(\mathbf{y})^T = \boldsymbol{\lambda}^T G(\mathbf{y}),$$

$$\mathbf{v} \in \partial \| \mathbf{r}(\mathbf{y}) \|,$$

$$\boldsymbol{\lambda} \geqslant 0.$$

An argument similar to that used in the sufficiency proof of Theorem 2 gives:

THEOREM 3. *If* $\mathbf{x}$ *is feasible for* (PC) *and* $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ *is feasible for* (DC) *then*

$$\| \mathbf{r}(\mathbf{x}) \| + f(\mathbf{x}) \geqslant F(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda}).$$

An immediate consequence of Theorem 1 (or Theorem 2) is:

THEOREM 4. *If* $\mathbf{x}$ *is optimal for* (PC), $\exists\, (\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ *with* $\mathbf{y} = \mathbf{x}$ *which is optimal for* (DC) *with the objective functions equal.*

It might be presumed, from inspection of problem (DC), that the constraint $\mathbf{v} \in \partial \| \mathbf{r}(\mathbf{y}) \|$ could be relaxed to the simpler constraint

$$\| \mathbf{v} \|^* \leqslant 1.$$

However, this is not possible in general, as the following example shows.

EXAMPLE. $n = 1, m = 2, t = 0$ (no constraints), $f \equiv 0$, L_1 norm. $\mathbf{r} = (x^2, x - x^2)^T$, giving zero as the minimum value of the norm. The dual objective function is

$$y^2(v_1 - v_2) + yv_2$$

where $\mathbf{v} = (v_1, v_2)^T$, and the equality constraint is

$$2y(v_1 - v_2) + v_2 = 0.$$

This constraint is satisfied by $y = \frac{1}{2}$, $v_1 = 0$, $v_2 = 1$, and also $\| \mathbf{v} \|^* \leqslant 1$. However $\mathbf{v} \notin \partial \| \mathbf{r}(\mathbf{y}) \|$, and it is easily verified that Theorem 3 is violated.

An important subset of the class of problems (PC) for which the simpler form of the dual may be shown to be appropriate occurs when $\mathbf{r}(\mathbf{x})$ is a *linear* function of $\mathbf{x}$. We will now restrict consideration to this particular case which we refer to as *problem (PL)*. We can therefore write

$$\mathbf{r}(\mathbf{x}) = A\mathbf{x} + \mathbf{b},$$

where A is a (constant) $m \times n$ matrix, and $\mathbf{b} \in R^m$ is a constant vector. Consider now the problem:

find $\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda}$ to maximize

$$F(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda}) = f(\mathbf{y}) + \boldsymbol{\lambda}^T G(\mathbf{y})\, \mathbf{y} - \mathbf{d}(\mathbf{y})^T \mathbf{y} + \mathbf{v}^T \mathbf{b} - \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{y})$$

subject to

$$\text{(DL)}\quad \begin{aligned} \mathbf{v}^T A + \mathbf{d}(\mathbf{y})^T &= \boldsymbol{\lambda}^T G(\mathbf{y}), \\ \| \mathbf{v} \|^* &\leqslant 1, \\ \boldsymbol{\lambda} &\geqslant 0. \end{aligned}$$

THEOREM 5. *If* $\mathbf{x}$ *is feasible for* (PL) *and* $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ *is feasible for* (DL) *then*

$$\| \mathbf{r}(\mathbf{x}) \| + f(\mathbf{x}) \geqslant F(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda}).$$

Proof. $\| \mathbf{r}(\mathbf{x}) \| + f(\mathbf{x}) - f(\mathbf{y}) - \boldsymbol{\lambda}^T G(\mathbf{y})\, \mathbf{y} + \mathbf{d}(\mathbf{y})^T \mathbf{y} - \mathbf{v}^T \mathbf{b} + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{y}) \geqslant \mathbf{v}^T A \mathbf{x} + f(\mathbf{x}) - f(\mathbf{y}) - \boldsymbol{\lambda}^T G(\mathbf{y})\, \mathbf{y} + \mathbf{d}(\mathbf{y})^T \mathbf{y} + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{y}) \geqslant 0$ using the convexity of f, concavity of $\mathbf{g}$ and the constraints. Q.E.D.

Use of Theorem 1 (or Theorem 2) immediately gives:

THEOREM 6. *If* $\mathbf{x}$ *is optimal for* (PL), $\exists$ $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ *with* $\mathbf{y} = \mathbf{x}$ *which is optimal for* (DL) *with the objective functions equal.*

The remainder of this paper is devoted to the provision of a theorem going in the opposite direction to Theorem 6. To this end, we require necessary conditions for a solution to problem (DL), and as before, we assume that the first-order constraint qualification is satisfied there, so that the sets of feasible directions introduced in Section 1 are appropriate. Before proving the main theorem, we require some preliminary results, including the following lemma which is essentially a generalization of Farkas' lemma.

LEMMA 3. *Let* W *be a closed, bounded, convex set not containing the origin, let* $\mathbf{r}_i$, $i = 1, 2,..., s$ *be given vectors, and suppose that there are no vectors* $\boldsymbol{\delta} \geqslant 0$, $\mathbf{w} \in W$ *such that*

$$\mathbf{w} + \sum_{i=1}^{s} \delta_i \mathbf{r}_i = 0, \tag{3.1}$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_s)^T$. *Then, for a given vector* $\mathbf{q}$, *the set*

$$\{\boldsymbol{\gamma}: \boldsymbol{\gamma}^T\mathbf{q} > 0,\ \boldsymbol{\gamma}^T\mathbf{r}_i \leqslant 0,\ i = 1, 2,..., s,\ \boldsymbol{\gamma}^T\mathbf{w} \leqslant 0, \textit{for all } \mathbf{w} \in W\}$$

is empty iff $\exists\ \beta \geqslant 0$, $\boldsymbol{\alpha} \geqslant 0$, $\mathbf{w} \in W$ *such that*

$$\sum_{i=1}^{s} \alpha_i \mathbf{r}_i + \beta \mathbf{w} = \mathbf{q}. \tag{3.2}$$

Remark. The conditions of this lemma are actually stronger than is necessary. They are, however, appropriate for our purposes.

Proof. The sufficiency of the conditions is obvious. Suppose (3.2) cannot be satisfied. Let K_1 be the convex cone generated by the vectors $\mathbf{r}_i$, $i = 1, 2,..., s$, and let K_2 be the convex cone generated by the set W. Then both K_1 and K_2 are closed. Further, by (3.1) there is no $\mathbf{k}_1 \in K_1$, $\mathbf{k}_2 \in K_2$ satisfying $\mathbf{k}_1 + \mathbf{k}_2 = 0$ save for $\mathbf{k}_1 = \mathbf{k}_2 = 0$. Thus $K = \text{conv}(K_1 \cup K_2)$ is a closed, convex cone (Rockafellar [9, p. 75]). Let $\mathbf{h} \in K$ be such that

$$\| \mathbf{h} - \mathbf{q} \|_2 \leqslant \| \mathbf{k} - \mathbf{q} \|_2, \qquad \text{all } \mathbf{k} \in K,$$

which exists since K is closed. Then we must have

$$\mathbf{h}^T(\mathbf{h} - \mathbf{q}) = 0.$$

Let $\boldsymbol{\gamma} = \mathbf{q} - \mathbf{h}$. Then $\mathbf{q}^T\boldsymbol{\gamma} = \| \mathbf{h} - \mathbf{q} \|_2 > 0$. The definition of $\mathbf{h}$ also implies that for every $\mathbf{k} \in K$

$$(\mathbf{h} - \mathbf{q})^T (\mathbf{k} - \mathbf{h}) \geqslant 0$$
$$\therefore \boldsymbol{\gamma}^T\mathbf{k} \leqslant 0 \qquad \text{for all } \mathbf{k} \in K.$$

This shows the existence of a suitable vector $\boldsymbol{\gamma}$ and concludes the proof. Q.E.D.

Let us return now to problem (DL). Let

$$W = \partial \| \mathbf{v} \|^*, \tag{3.3}$$

and let $\boldsymbol{\gamma}^T = (\boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T, \boldsymbol{\gamma}_3^T)$, where $\boldsymbol{\gamma}_1 \in R^n$, $\boldsymbol{\gamma}_2 \in R^m$, and $\boldsymbol{\gamma}_3 \in R^t$. Writing

$$M \equiv M(\mathbf{y}, \boldsymbol{\lambda}) \equiv \nabla^2(\boldsymbol{\lambda}^T\mathbf{g}(\mathbf{y}) - f(\mathbf{y})),$$

where the partial differentiation is with respect to the components of $\mathbf{y}$, it follows that the triple $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \boldsymbol{\gamma}_3)$ is a feasible direction for the constraints of problem (DL) at $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ if

$$\boldsymbol{\gamma}_1^TM - \boldsymbol{\gamma}_2^TA + \boldsymbol{\gamma}_3^TG = 0, \tag{3.4}$$
$$\boldsymbol{\gamma}_2^T\mathbf{w} \leqslant 0 \qquad \text{for all } \mathbf{w} \in W, \text{ if } \| \mathbf{v} \|^* = 1, \tag{3.5}$$
$$-\boldsymbol{\gamma}_3^T\mathbf{e}_i \leqslant 0 \qquad \text{if } \lambda_i = 0,\ i = 1, 2, \ldots, t, \tag{3.6}$$

where $\mathbf{e}_i$ is the ith coordinate vector. Further, if $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ is a solution of problem (DL), any feasible direction $\boldsymbol{\gamma}$ must satisfy

$$\boldsymbol{\gamma}_1^TM\mathbf{y} + \boldsymbol{\gamma}_2^T\mathbf{b} + \boldsymbol{\gamma}_3^T(G\mathbf{y} - \mathbf{g}) \leqslant 0. \tag{3.7}$$

Now, let R be the $(n + m + t) \times (2n + t)$ matrix defined by

$$R = \begin{bmatrix} M & -M & 0 \\ -A & A & 0 \\ G & -G & -I \end{bmatrix}, \tag{3.8}$$

let

$$\mathbf{q} = \begin{bmatrix} M\mathbf{y} \\ \mathbf{b} \\ G\mathbf{y} - \mathbf{g} \end{bmatrix}, \tag{3.9}$$

and let $\hat{\mathbf{w}} = (\mathbf{O}^T, \mathbf{w}^T, \mathbf{O}^T)$, where the vector $\mathbf{w}$ occupies positions $(n + 1)$ to $(n + m)$ in $\hat{\mathbf{w}}$. Then if $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ solves problem (DL), the set

$$\{\boldsymbol{\gamma}: \boldsymbol{\gamma}^T\mathbf{q} > 0, \tag{3.10}$$
$$k\boldsymbol{\gamma}^T\hat{\mathbf{w}} \leqslant 0, \qquad \text{for all } \mathbf{w} \in W, \tag{3.11}$$
$$\boldsymbol{\gamma}^T\hat{R} \leqslant 0\} \tag{3.12}$$

is empty, where $k = 1$ if $\| \mathbf{v} \|^* = 1$ and zero otherwise, and $\hat{R}$ is the matrix obtained from R by deleting those columns from the last t which correspond to components of $\boldsymbol{\lambda}$ which are not equal to zero.

LEMMA 4. *Let M be nonsingular at* $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$. *Then if* $\| \mathbf{v} \|^* = 1$, $\nexists\, \hat{\boldsymbol{\delta}}, \mathbf{w} \in W$ such that

$$\hat{R}\hat{\boldsymbol{\delta}} + \hat{\mathbf{w}} = 0. \tag{3.13}$$

Proof. The existence of $\hat{\boldsymbol{\delta}}, \mathbf{w} \in W$ satisfying (3.13) implies that $\mathbf{w} = 0$, which contradicts the fact that $\| \mathbf{v} \|^* = 1$. Q.E.D.

LEMMA 5. *Let* $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ *solve problem* (DL) *with M nonsingular. Then*

(1) $\mathbf{g}(\mathbf{y}) \geqslant 0$,

(2) $\exists\, \beta \geqslant 0$, $\mathbf{w} \in W$ *such that*

$$\mathbf{r}(\mathbf{y}) = \beta \mathbf{w}$$

with

$$\beta(\| \mathbf{v} \|^* - 1) = \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{y}) = 0.$$

Proof. If $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ solves problem (DL) then $\nexists\, \boldsymbol{\gamma}$ satisfying (3.10)–(3.12). Let $\| \mathbf{v} \|^* = 1$. Then, using Lemma 4, Lemma 3 gives that $\exists\, \beta \geqslant 0$, $\hat{\boldsymbol{\alpha}} \geqslant 0$, $\mathbf{w} \in W$ such that

$$\hat{R}\hat{\boldsymbol{\alpha}} + \beta \hat{\mathbf{w}} = \mathbf{q},$$

i.e.,

$$R\boldsymbol{\alpha} + \beta \hat{\mathbf{w}} = \mathbf{q}$$

where $\boldsymbol{\alpha}$ is formed from $\hat{\boldsymbol{\alpha}}$ by adding zeros. Thus, writing $\boldsymbol{\alpha}^T = (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \boldsymbol{\alpha}_3^T)$, where $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in R^n$, we have

$$M(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) = M\mathbf{y} \tag{3.14}$$

$$-A(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) + \beta \mathbf{w} = \mathbf{b} \tag{3.15}$$

$$G(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) - \boldsymbol{\alpha}_3 = G\mathbf{y} - \mathbf{g}. \tag{3.16}$$

The nonsingularity of M shows that $\mathbf{y} = \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2$, and the result follows from (3.15) and (3.16).

When $\| \mathbf{v} \|^* < 1$, we may apply Farkas' lemma directly in the usual way (see, for example, Fiacco and McCormick [4]) to obtain conditions which are the required ones with $\beta = 0$. Q.E.D.

THEOREM 7. *Let* $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ *be optimal for* (DL) *with M nonsingular. Then* $\mathbf{y}$ *is optimal for* (PL).

Proof. An immediate consequence of Lemma 5 is that if $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ solves (DL) with M nonsingular, then $\mathbf{y}$ is feasible for (PL). It remains to show that $\mathbf{y}$ is also optimal for (PL), which by Theorem 2 and the dual feasibility of $(\mathbf{y}, \mathbf{v}, \boldsymbol{\lambda})$ reduces to showing that $\mathbf{v} \in \partial \| \mathbf{r}(\mathbf{y}) \|$.

Now Lemma 5 shows that $\exists \mathbf{w} \in W$ such that

$$\mathbf{r}(\mathbf{y}) = \beta \mathbf{w},$$

and so $\| \mathbf{r}(\mathbf{y}) \| = \beta \| \mathbf{w} \| = \beta$. Further

$$\mathbf{r}(\mathbf{y})^T \mathbf{v} = \beta \mathbf{v}^T \mathbf{w} = \beta \| \mathbf{v} \|^*.$$

Since $\beta(1 - \| \mathbf{v} \|^*) = 0$ by Lemma 5, the result follows.

Acknowledgments

I am grateful to Dr. M. R. Osborne for stimulating my interest in this kind of problem, and to Dr. R. Fletcher for a number of valuable comments on an earlier version of this paper.

References

1. H. Anton and C. S. Duris, On an algorithm for best approximate solutions to $A\mathbf{v} = \mathbf{b}$ in normed linear spaces, *J. Approximation Theory* **8** (1973), 133–141.
2. M. S. Borowsky, Algorithms for solving the dual problem for $A\mathbf{v} = \mathbf{b}$ with varying norms, *J. Approximation Theory* **17** (1976), 107–114.
3. E. W. Cheney, "Introduction to Approximation Theory," McGraw–Hill, New York, 1966.
4. A. V. Fiacco and G. P. McCormick, "Nonlinear Programming: Sequential Unconstrained Minimisation Techniques," Wiley, New York, 1968.
5. A. S. Householder, "The Theory of Matrices in Numerical Analysis," Blaisdell, New York, 1964.
6. B. Mond, A class of non-differentiable mathematical programming problems, *J. Math. Anal. Appl.* **46** (1974), 169–174.
7. B. Mond and M. Schechter, A programming problem with an L_p norm in the objective function, *J. Austral. Math. Soc. Ser. B.* **19** (1976), 333–342.
8. W. Oettli, Symmetric duality, and a convergent subgradient method for discrete, linear, constrained approximation problems with arbitrary norms appearing in the objective function and in the constraints, *J. Approximation Theory* **14** (1975), 43–50.
9. R. T. Rockafellar, "Convex Analysis," Princeton Univ. Press, Princeton, N.J., 1970.
10. V. P. Sreedharan, Solutions of overdetermined linear equations which minimise error in an abstract norm, *Numer. Math.* **13** (1969), 146–151.
11. V. P. Sreedharan, Least squares algorithms for finding solutions of overdetermined linear equations which minimise error in an abstract norm, *Numer. Math.* **17** (1971), 387–401.