JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **81**, 8–19 (1981)

An Algorithm for Linearly Constrained Nonlinear Programming Problems

MOKHTAR S. BAZARAA*

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

AND

JAMIE J. GOODE

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332

Submitted by Ky Fan

In this paper an algorithm for solving a linearly constrained nonlinear programming problem is developed. Given a feasible point, a correction vector is computed by solving a least distance programming problem over a polyhedral cone defined in terms of the gradients of the "almost" binding constraints. Mukai's approximate scheme for computing the step size is generalized to handle the constraints. This scheme provides an estimate for the step size based on a quadratic approximation of the function. This estimate is used in conjunction with Armijo line search to calculate a new point. It is shown that each accumulation point is a Kuhn–Tucker point to a slight perturbation of the original problem. Furthermore, under suitable second order optimality conditions, it is shown that eventually only one trial is needed to compute the step size.

1. INTRODUCTION

This paper addresses the linearly constrained nonlinear programming problem

$$P: \text{ minimize } \quad f(x)$$
$$\text{subject to } \quad Ax \leqslant b,$$

where f is a twice continuously differentiable function on R^n, and A is an $l \times n$ matrix whose jth row is denoted by a_j^t, and where a superscript t denotes the transpose operation.

There are several approaches for solving this problem. The first one relies on partitioning the variables into basic, nonbasic, and superbasic variables. The values of the superbasic and basic variables are modified while the

* This author's work is supported under USAFOSR Contract F49620-79-C-0120.

0022-247X/81/050008-12$02.00/0

nonbasic variables are fixed at their current values. Examples of methods in this class are the convex simplex method of Zangwill [18], the reduced gradient method of Wolfe [17], the method of Murtagh and Saunders [12], and the variable-reduction method of McCormick [8],

Another class of methods is the extension of quasi-Newton algorithms from unconstrained to constrained optimization. Here, at any iteration, a set of active restrictions is identified, and then a modified Newton procedure is used to minimize the objective function on the manifold defined by these active constraints. See, for example, Goldfarb [6], and Gill and Murray [5].

Other approaches for solving problems with linear constraints are the gradient projection method and the method of feasible directions. The former computes a direction by projecting the negative gradient on the space orthogonal to the gradients of a subset of the binding constraints while the latter determines a search direction by solving a linear programming problem. For a review of these methods the reader may refer to Rosen [14], Zontendijk [19], Frank and Wolfe [4], and Topkis and Veinott [15].

In this paper, an algorithm for solving problem P is proposed. At each iteration a correction vector is computed by finding the minimum distance from a given point to a polyhedral cone defined in terms of the gradients of the "almost" binding constraints. An approximate line search procedure which extends those of Armijo [1] and Mukai [10, 11] for unconstrained optimization is developed for determining the step size. First, an estimate of the step size based on a quadratic approximation to the objective function is computed, and then adjusted if necessary.

In Section 2, we outline the algorithm. In Section 3, we show that accumulation points of the algorithm are Kuhn–Tucker points to a slight perturbation of the original problem. Finally, in Section 4, assuming that the algorithm converges, and under suitable second order sufficiency optimality conditions, we show that the step size estimates which are based on the quadratic approximation are acceptable so that only one functional evaluation is eventually needed for performing the line search.

2. Statement of the Algorithm

Consider the following algorithm for solving Problem P.

Step 0. Choose values for the parameters c, z, δ, and ε. Select a point x_0 such that $Ax_0 \leqslant b$ and let $\delta_0 = \delta$. Let $i = 0$ and go to Step 1.

Step 1. Let w_i be the optimal solution to Problem $D(x_i)$ given below:

$$D(x_i): \text{ minimize } \quad \nabla f(x_i)^t w + \tfrac{1}{2} z w^t w$$
$$\text{subject to } \quad a_j^t w \leqslant 0 \qquad \text{for } \quad j \in I(x_i),$$

where

$$I(x_i) = \{j: a_j^t x_i \geqslant b_j - c\}. \quad (2.1)$$

If $w_i = 0$, stop. Else, go to Step 2.

Step 2. Let

$$I^+(w_i) = \{j: a_j^t w_i > 0\}, \quad (2.2)$$

and let

$$\beta_i = \min\left\{1, \frac{b_j - a_j^t x_i}{a_j^t w_i} \text{ for } j \in I^+(w_i)\right\}. \quad (2.3)$$

Let

$$d_i = \beta_i w_i, \quad (2.4)$$

and go to Step 3.

Step 3. If

$$f(x_i + \varepsilon d_i) + f(x_i - \varepsilon d_i) - 2f(x_i) \geqslant \varepsilon^2 \delta_i \|d_i\|^2, \quad (2.5)$$

let

$$\lambda_i = \frac{-\varepsilon^2 \nabla f(x_i)^t d_i}{f(x_i + \varepsilon d_i) + f(x_i - \varepsilon d_i) - 2f(x_i)} \quad (2.6)$$

and let $\delta_{i+1} = \delta_i$, and go to Step 4. Otherwise, let $\lambda_i = 1$, $\delta_{i+1} = \frac{1}{2}\delta_i$, and go to Step 4.

Step 4. Let

$$\alpha_i = \min\{1, \lambda_i\} \quad (2.7)$$

and compute the smallest nonnegative integer k satisfying

$$f(x_i + (\tfrac{1}{2})^k \alpha_i d_i) - f(x_i) \leqslant \tfrac{1}{3}(\tfrac{1}{2})^k \alpha_i \nabla f(x_i)^t d_i. \quad (2.8)$$

Let $k_i = k$, $x_{i+1} = x_i + \alpha_i(\frac{1}{2})^{k_i} d_i$, $i = i + 1$, and go to Step 1.

The following remarks are helpful in interpreting the above algorithm.

1. A direction w_i is determined by solving Problem $D(x_i)$. This problem finds the point in the convex polyhedral cone $\{w: a_j^t w \leqslant 0 \text{ for } j \in I(x_i)\}$

which is closest to the vector $-(1/z)\nabla f(x_i)$. Methods of least-distance programming, as in the works of Bazaraa and Goode [2] and Wolfe [16], can be used for solving this problem. Special methods that take advantage of the structure of the cone constraints may prove quite useful in this regard.

2. The restrictions enforced in Problem $D(x_i)$ are the c-binding constraints at x_i, that is, those satisfying $b_j - c \leqslant a_j^t x_i \leqslant b_j$. If $w_i = 0$, then the algorithm is terminated with x_i. In this case, from the Kuhn–Tucker conditions for Problem $D(x_i)$, there exist u_j for $j \in I(x_i)$ such that:

$$\nabla f(x_i) + \sum_{j \in I(x_i)} u_j a_j = 0,$$
$$u_j \geqslant 0, \qquad \text{for} \quad j \in I(x_i).$$

These conditions imply that x_i is a Kuhn–Tucker point for the following problem:

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad a_j^t x \leqslant a_j^t x_i \qquad \text{for} \quad j \in I(x_i), \\ &\qquad\qquad\quad a_j^t x \leqslant b_j \qquad\quad \text{for} \quad j \notin I(x_i). \end{aligned}$$

Noting that $b_j - c \leqslant a_j^t x_i \leqslant b_j$ for $j \in I(x_i)$, if c is sufficiently small, it is clear that the algorithm is termined if x_i is a Kuhn–Tucker (KT) solution to a slightly perturbed version of Problem P. The following definition will thus be useful.

DEFINITION 2.1. Let x^* be a feasible point to Problem P. If the optimal solution to Problem $D(x^*)$ is equal to zero, then x^* is called a c-KT solution to Problem P.

3. If $x_i + w_i$ is feasible to Problem P, then the search vector d_i is taken as w_i. Otherwise, d_i is taken to be the vector of maximum length along w_i which maintains feasibility of $x_i + d_i$.

4. Steps 3 and 4 of the algorithm compute the step size along the vector d_i in order to form x_{i+1}. As proposed by Mukai [10, 11], first an estimate of the step size λ_i is calculated. When appropriate, λ_i is computed by utilizing a quadratic approximation of the function f at x_i, otherwise λ_i is taken equal to 1. In order to ensure feasibility to Problem P, the first trial step size α_i used in conjunction with Armijo line search [1] is the minimum of λ_i and 1. As will be shown in Section 4, under suitable assumptions, for large i, test (2.5) passes, $k_i = 0$, and $\alpha_i = \lambda_i < 1$. This confirms efficiency of the line search scheme where eventually only one trial is needed to compute k_i and two functional evaluations are required to calculate λ_i.

3. Accumulation Points of the Algorithm

Theorem 3.1 shows that each accumulation point of the proposed algorithm is a c-KT point. In order to prove this theorem, Lemmas 3.1 and 3.2 are needed. These two lemmas extend similar results of Mukai [10] for uncontrained problems.

In order to facilitate the development in this section, the following notation is used. Let $w(x)$ be the optimal solution to Problem $D(x)$ and let $\beta(x)$ be as given in (2.3) with x_i replaced with x. Finally, let $d(x) = \beta(x)\, w(x)$.

Lemma 3.1. *Suppose that x^* is not a c-KT point for Problem P. Then, there exist scalars μ and $s > 0$ so that $\mu \leqslant \alpha(x) \leqslant 1$ for each x with $\|x - x^*\| < s$.*

Proof. We will show that $\|w(x)\|$ and $\beta(x)$ are bounded above and below away from zero in a neighborhood of x^*. Note that:

$$\tfrac{1}{2}z\, \|w(x)\|^2 \leqslant -\nabla f(x)^t\, w(x) \leqslant \|\nabla f(x)\|\, \|w(x)\|,$$

so that $\|w(x)\| \leqslant (2/z)\, \|\nabla f(x)\|$. By continuous differentiability of f, it is clear that $\|w(x)\|$ is bounded above in a neighborhood of x^*. Next, note that $I(x) \subset I(x^*)$ in a neighborhood of x^* since otherwise there would exist a sequence x_l converging to x^* and an index j so that $a_j^t x_l \geqslant b_j - c$ and in the mean time $a_j^t x^* < b_j - c$, which is impossible. Consider Problem $D(x)$ and denote its optimal objective value $\nabla f(x)^t\, w(x) + \frac{1}{2}z\, \|w(x)\|^2$ by $g(x)$. By contradiction to the desired conclusion that $\|w(x)\|$ is bounded aways from zero in a neighborhood of x^*, suppose there exists a sequence x_l converging to x^* so that $g(x_l) \to 0$. Since $I(x_l) \subset I(x^*)$, then the optimal objective value $g'(x_l)$ to Problem $D'(x_l)$ given below satisfies $0 \geqslant g'(x_l) \geqslant g(x_l)$.

$$\begin{aligned} D'(x_l)\text{: minimize} \quad & \nabla f(x_l)^t w + \tfrac{1}{2} z w^t w \\ \text{subject to} \quad & a_j^t w \leqslant 0 \qquad \text{for} \quad j \in I(x^*). \end{aligned}$$

This in turn implies that $g'(x_l) \to 0$. By continuous differentiability of f, and since $x_l \to x^*$, it then follows that the optimal objective value to problem $D(x^*)$ is equal to 0, see for example Daniel [3]. This, however, contradicts the assumption that x^* is not a c-KT point. Now consider $\beta(x)$. If $j \in I^+(w(x))$, then $b_j - a_j^t x > c$. Thus:

$$\begin{aligned} \beta(x) &= \min \left\{ 1, \frac{b_j - a_j^t x}{a_j^t w(x)} \text{ for } j \in I^+(w(x)) \right\} \\ &\geqslant \min \left\{ 1, \frac{c}{\|a_j\|\, \|w(x)\|} \text{ for } j \in I^+(w(x)) \right\}. \end{aligned}$$

Since $\|w(x)\|$ is bounded above, then $\beta(x)$ is bounded away from zero in a neighborhood of x^*. Also $\beta(x) \leqslant 1$, so that it is bounded above.

To summarize, there exist scalars s, α, ξ, and $\gamma > 0$ so that

$$\begin{aligned} \alpha \geqslant \|w(x)\| \geqslant \gamma \qquad & \text{if} \quad \|x - x^*\| \leqslant s, \\ 1 \geqslant \beta(x) \geqslant \xi \qquad & \text{if} \quad \|x - x^*\| \leqslant s. \end{aligned} \tag{3.1}$$

Thus if $\|x - x^*\| \leqslant s$, we must have

$$\begin{aligned} -\nabla f(x)^t\, d(x) = -\beta(x)\, \nabla f(x)^t\, w(x) &\geqslant \tfrac{1}{2} z\beta(x)\, \|w(x)\|^2 \\ &\geqslant \tfrac{1}{2} z\xi\gamma^2 = y > 0. \end{aligned} \tag{3.2}$$

From (3.1), $\|d(x)\| = \beta(x)\, \|w(x)\| \leqslant \alpha$ if $\|x - x^*\| \leqslant s$. Thus from continuity of f, there exists a scalar $q > 0$ so that

$$f(x + \varepsilon d(x)) + f(x - \varepsilon d(x)) - 2f(x) < q \qquad \text{if} \quad \|x - x^*\| \leqslant s. \tag{3.3}$$

If test (2.5) passes, then from (3.2) and (3.3), the following lower bound on $\lambda(x)$ is at hand:

$$\lambda(x) = \frac{-\varepsilon^2 \nabla f(x)^t\, d(x)}{f(x + \varepsilon d(x)) + f(x - \varepsilon d(x)) - 2f(x)} \geqslant \frac{\varepsilon^2 y}{q} \qquad \text{if} \quad \|x - x^*\| \leqslant s.$$

If test (2.5) fails, then $\lambda(x) = 1$. Thus $\lambda(x) \geqslant \min\{1, \varepsilon^2 y/q\} = \mu$. Since $\alpha(x) = \min\{1, \lambda(x)\}$, the desired result follows.

LEMMA 3.2. *If x^* is not a* c-KT *point for Problem* P, *then there exist a number $s > 0$ and an integer m so that $k(x) \leqslant m$ if $\|x - x^*\| < s$, where $k(x)$ is the Armijo integer given by* (2.8) *with x_i and α_i replaced with x and $\alpha(x)$, respectively.*

Proof. As in the proof of Lemma 3.1 and by continuous differentiability of f, there exist scalars s, h, and $y > 0$ so that for $\|x - x^*\| < s$ the following hold:

$$\nabla f(x)^t\, d(x) \leqslant -y, \tag{3.4}$$

$$|\nabla f(x + g d(x))^t\, d(x) - \nabla f(x)^t\, d(x)| \leqslant \tfrac{2}{3} y \qquad \text{for each } g \in [0, h]. \tag{3.5}$$

Now let m be the smallest nonnegative integer so that $(\frac{1}{2})^m \leqslant h$ and let x be such that $\|x - x^*\| < s$. Then there exists $\theta \in [0, 1]$ such that:

$$\begin{aligned} & f(x + (\tfrac{1}{2})^m \alpha(x)\, d(x)) - f(x) - \tfrac{1}{3}(\tfrac{1}{2})^m \alpha(x)\, \nabla f(x)^t\, d(x) \\ &\quad = (\tfrac{1}{2})^m \alpha(x)\, \nabla f(x + \theta(\tfrac{1}{2})^m \alpha(x)\, d(x))^t\, d(x) - \tfrac{1}{3}(\tfrac{1}{2})^m \alpha(x)\, \nabla f(x)^t\, d(x) \\ &\quad = (\tfrac{1}{2})^m \alpha(x) [\{\nabla f(x + \theta(\tfrac{1}{2})^m \alpha(x)\, d(x))^t\, d(x) - \nabla f(x)^t\, d(x)\} \\ &\qquad + \tfrac{2}{3} \nabla f(x)^t\, d(x)]. \end{aligned} \tag{3.6}$$

Since $\theta(\frac{1}{2})^m \alpha(x) \leqslant h$, (3.4) and (3.5) imply that the right hand side of (3.6) is $\leqslant 0$, which is turn shows that $k(x) \leqslant m$, and the proof is complete.

THEOREM 3.1. *Either the algorithm terminates with a* c-KT *point for problem P or else generates an infinite sequence* $\{x_i\}$ *of which any accumulation point is a* c-KT *point for Problem* P.

Proof. Clearly the algorithm stops at x_i only if x_i is a c-KT point. Now, suppose that the algorithm generates the infinite sequence $\{x_i\}$. Suppose that x^* is an accumulation point so that $x_i \to^{\mathscr{K}} x^*$ for some infinite set $\mathscr{K}$ of positive integers. Since $f(x_i)$ is decreasing monotonically and since $f(x_i) \to^{\mathscr{K}} f(x^*)$ *then* $f(x_i) \to f(x^*)$. Suppose by contradiction to the desired conclusion that x^* is not a c-KT point. From Lemmas 3.1 and 3.2, there exist positive numbers μ and y and an integer m so that $\alpha_i \geqslant \mu$, $\nabla f(x_i)^t d_i \leqslant -y$, and $k_i \leqslant m$ for large i in $\mathscr{K}$. Therefore,

$$f(x_{i+1}) - f(x_i) \leqslant \tfrac{1}{3}(\tfrac{1}{2})^{k_i} \alpha_i \nabla f(x_i)^t d_i \leqslant -\tfrac{1}{3}\mu y(\tfrac{1}{2})^m$$

for large i in $\mathscr{K}$. This implies that $f(x_i) \to -\infty$, contradicting the fact that $f(x_i) \to f(x^*)$. This completes the proof.

4. EVENTUAL ACCEPTANCE OF THE STEP SIZE ESTIMATE

In the previous section, we showed that an accumulation point x^* of the sequence $\{x_i\}$ generated by the algorithm is a c-KT point to the perturbed problem P′ given below:

$$\begin{aligned} \text{P}': \text{ minimize } & f(x) \\ \text{subject to } & a_j^t x \leqslant a_j^t x^* \quad \text{for } j \in I(x^*), \\ & a_j^t x \leqslant b_j \quad \text{for } j \notin I(x^*). \end{aligned}$$

Here, we assume that the whole sequence $\{x_i\}$ converges to a point x^* which satisfies suitable second order sufficiency conditions. Under this assumption, we show that test (2.5) is eventually passed. Furthermore, we show that $\lambda_i < 1$ and that $k_i = 0$ for i large enough.

The second order condition is given in Definition 4.1. It is well-known that x^* satisfying this condition is a strong local minimum for problem P′. That is, there exists a number $\gamma > 0$ so that $f(x^*) < f(x)$ if x is feasible to problem P′ and $\|x - x^*\| < \gamma$; see for example McCormick [9] and Han and Mangasarian [7].

DEFINITION 4.1. Let x^* be such that $Ax^* \leqslant b$ and let $I(x^*) = \{j\colon a_j^t x^* \geqslant b_j - c\}$. x^* is said to satisfy the second order sufficiency optimality conditions for problem P′ if $a_j^t x^* > b_j - c$ for each $j \in I(x^*)$ and if there exist scalars $u_j \geqslant 0$ for $j \in I(x^*)$ and $\gamma > 0$ so that

$$\nabla f(x^*) + \sum_{j \in I(x^*)} u_j a_j = 0 \tag{4.1}$$

$$\nabla f(x^*)^t d \leqslant 0, \quad a_j^t d \leqslant 0 \text{ for } j \in I(x^*), \quad \|d\| = 1 \Rightarrow d^t H(x^*) d > \gamma.$$

Theorem 4.1 shows that test (2.5) will eventually be passed so that λ_i is given by (2.6). The following two intermediate results are needed to prove this theorem.

LEMMA 4.1. *If $Cd \leqslant 0$ and $\|d\| = 1$ imply that $d^t H d \geqslant \gamma > 0$ then there is a number $\theta > 0$ so that $Cd \leqslant \theta 1$* and *$\|d\| = 1$ imply that $d^t H d \geqslant \gamma/2$.*

Proof. Suppose, by contradiction, that for each integer k there is a vector d_k such that

$$\|d_k\| = 1, \; Cd_k \leqslant \frac{1}{k} 1, \qquad d_k^t H d_k < \gamma/2. \tag{4.2}$$

Since the sequence $\{d_k\}$ is bounded, it has an accumulation point d. From (4.2), $\|d\| = 1$, $Cd \leqslant 0$, and $d^t H d \leqslant \gamma/2$ which contradicts the assumption of the lemma.

LEMMA 4.2. *Suppose that $x_i \to x^*$ and suppose that $b_j - a_j^t x^* < c$ for each $j \in I(x^*)$. Then $d_i \to 0$.*

Proof. Since $0 \leqslant \beta_i \leqslant 1$ and $d_i = \beta_i w_i$, it suffices to show that $w_i \to 0$. By Theorem 3.1, x^* is a c-KT point. Thus the optimal solution w^* to Problem $D(x^*)$ must satisfy $w^* = 0$. The assumption $b_j - a_j^t x^* < c$ for $j \in I(x^*)$, implies that $I(x_i) = I(x^*)$ for i large enough. This shows that the feasible region for Problem $D(x_i)$ is equal to that of Problem $D(x^*)$ for i large enough. Since $x_i \to x^*$, it then follows that $w_i \to w^*$, see, for example, Daniel [3]. Thus $w_i \to 0$ and $d_i \to 0$, and the proof is complete.

Throughout the remainder of this section, the following notation will be used for any scalar γ:

$$H_i^\gamma = 2 \int_0^1 (1 - y) H(x_i + y\gamma d_i) \, dy. \tag{4.3}$$

We can integrate by parts to obtain

$$f(x_i + \gamma d_i) - f(x_i) = \gamma \nabla f(x_i)^t d_i + \tfrac{1}{2} \gamma^2 d_i^t H_i^\gamma d_i. \tag{4.4}$$

For futher details, the reader may refer to Polak [13, p. 293].

THEOREM 4.1. *Let $\{x_i\}$ be a sequence generated by the algorithm. Suppose that $x_i \to x^*$ and x^* satisfies the second order optimality conditions for problem P′. Then there exists an integer m so that test (2.5) passes for all $i \geqslant m$.*

Proof. From (4.3) and (4.4) we get

$$f(x_i + \varepsilon d_i) - f(x_i) = \varepsilon \nabla f(x_i)^t d_i + \tfrac{1}{2}\varepsilon^2 d_i^t H_i^\varepsilon d_i,$$
$$f(x_i - \varepsilon d_i) - f(x_i) = -\varepsilon \nabla f(x_i)^t d_i + \tfrac{1}{2}\varepsilon^2 d_i^t H_i^{-\varepsilon} d_i,$$

Adding, we obtain

$$f(x_i + \varepsilon d_i) + f(x_i - \varepsilon d_i) - 2f(x_i) = \tfrac{1}{2}\varepsilon^2 d_i^t (H_i^\varepsilon + H_i^{-\varepsilon})\, d_i. \tag{4.5}$$

Now for $j \in I(x^*)$, $a_j^t x^* > b_j - c$. Since $x_i \to x^*$ then for i large enough, $a_j^t x_i > b_j - c$ so that $j \in I(x_i)$. By Step 1 of the algorithm $a_j^t w_i \leqslant 0$ and so $a_j^t(d_i/\|d_i\|) \leqslant 0$ for i large enough and $j \in I(x^*)$. Likewise, from Step 1 of the algorithm $\nabla f(x_i)^t w_i \leqslant 0$ and hence $\nabla f(x_i)^t(d_i/\|d_i\|) \leqslant 0$. Since $x_i \to x^*$, then for any number $\theta > 0$, $\nabla f(x^*)^t(d_i/\|d_i\|) \leqslant \theta$ for i large enough. Thus, Lemma 4.1 and the second order conditions imply that

$$d_i^t H(x^*)\, d_i \geqslant (\gamma/2)\, \|d_i\|^2 \qquad \text{for large } i. \tag{4.6}$$

Now note that

$$\|H_i^\varepsilon - H(x^*)\| = \left\| 2\int_0^1 (1-y)[H(x_i + y\varepsilon d_i) - H(x^*)]\, dy \right\|$$
$$\leqslant 2\int_0^1 (1-y)\, \|H(x_i + y\varepsilon d_i) - H(x^*)\|\, dy. \tag{4.7}$$

Since $x_i \to x^*$, then by Lemma 4.2, $d_i \to 0$. Particularly, for i large enough, $\|H(x_i + y\varepsilon d_i) - H(x^*)\| < \gamma/4$ for all $y \in [0, 1]$. From (4.7), $\|H_i^\varepsilon - H(x^*)\| \leqslant \gamma/4$. This together with (4.6) yields:

$$d_i^t H_i^\varepsilon d_i = d_i^t H(x^*)\, d_i + d_i^t (H_i^\varepsilon - H(x^*))\, d_i$$
$$\geqslant (\gamma/2)\, \|d_i\|^2 - \|d_i\|^2\, \|H_i^\varepsilon - H(x^*)\|$$
$$\geqslant (\gamma/4)\, \|d_i\|^2 \qquad \text{for large } i. \tag{4.8}$$

Similarly,

$$d_i^t H_i^{-\varepsilon} d_i \geqslant (\gamma/4)\, \|d_i\|^2 \qquad \text{for large } i. \tag{4.9}$$

From (4.5), (4.8), and (4.9) it immediately follows that

$$f(x_i+\varepsilon d_i)+f(x_i-\varepsilon d_i)-2f(x_i)\geqslant \varepsilon^2(\gamma/4)\,\|d_i\|^2 \qquad \text{for large } i. \tag{4.10}$$

From (4.10), if test (2.5) fails for a large i, we must have

$$\varepsilon^2\delta_i\,\|d_i\|^2 > f(x_i+\varepsilon d_i)+f(x_i-\varepsilon d_i)-2f(x_i)\geqslant \varepsilon^2(\gamma/4)\,\|d_i\|^2;$$

that is, $\delta_i > \gamma/4$. If the conclusion of the lemma does not hold, then test (2.5) fails infinitely often and then $\delta_i \to 0$. This contradicts $\delta_i > \gamma/4$ for large i, and the proof is complete.

THEOREM 4.2. *Let $\{x_i\}$ be a sequence generated by the algorithm. Suppose that $x_i \to x^*$ and that x^* satisfies the second order optimality conditions for Problem* P′. *Then there exists an integer m so that $f(x_i+\alpha_i d_i)-f(x_i)\leqslant \frac{1}{3}\alpha_i \nabla f(x_i)^t d_i$ for all $i\geqslant m$, that is, $k_i=0$ for all $i\geqslant m$.*

Proof. By Theorem 4.1, test (2.5) passes for large i so that λ_i is given by

$$\lambda_i=\frac{-\varepsilon^2\nabla f(x_i)^t d_i}{f(x_i+\varepsilon d_i)+f(x_i-\varepsilon d_i)-2f(x_i)}=\frac{-\nabla f(x_i)^t d_i}{\frac{1}{2}d_i^t(H_i^\varepsilon+H_i^{-\varepsilon})d_i}. \tag{4.11}$$

If $\lambda_i\leqslant 1$ so that $\alpha_i=\lambda_i$, then from (4.4) and (4.11) we get:

$$\begin{aligned} f(x_i+\alpha_i d_i)-f(x_i)-\tfrac{1}{3}\alpha_i\nabla f(x_i)^t d_i &= \tfrac{1}{2}\lambda_i^2\, d_i^t H_i^{\lambda_i} d_i+\tfrac{2}{3}\lambda_i\,\nabla f(x_i)^t d_i \\ &= \tfrac{1}{2}\lambda_i^2\big[d_i^t H_i^{\lambda_i} d_i-\tfrac{1}{2}d_i^t(H_i^\varepsilon+H_i^{-\varepsilon})d_i\big]-\tfrac{1}{12}\lambda_i^2\, d_i^t(H_i^\varepsilon+H_i^{-\varepsilon})d_i. \end{aligned} \tag{4.12}$$

Since $x_i\to x^*$, then by Lemma 4.2, $d_i\to 0$. Thus $H_i^{\lambda_i}$, H_i^ε, and $H_i^{-\varepsilon}$ converge to $H(x^*)$ and the first term in (4.12) will be less than $(\gamma/24)\lambda_i^2\,\|d_i\|^2$ for i large enough. As in the proof of Theorem 4.1, $d_i^t(H_i^\varepsilon+H_i^{-\varepsilon})d_i\geqslant(\gamma/2)\,\|d_i\|^2$ for large i. Substituting in (4.12), the desired result holds.

Now suppose that $\lambda_i>1$ so that $\alpha_i=1$. Then

$$f(x_i+\alpha_i d_i)-f(x_i)-\tfrac{1}{3}\alpha_i\,\nabla f(x_i)^t d_i=\tfrac{1}{2}d_i^t H_i^1 d_i+\tfrac{2}{3}\nabla f(x_i)^t d_i. \tag{4.13}$$

Since $\lambda_i>1$, then from (4.11) we must have

$$\nabla f(x_i)^t d_i<-\tfrac{1}{2}d_i^t(H_i^\varepsilon+H_i^{-\varepsilon})d_i.$$

Substituting in (4.13) we get:

$$\begin{aligned} &f(x_i+\alpha_i d_i)-f(x_i)-\tfrac{1}{3}\alpha_i\,\nabla f(x_i)^t d_i \\ &\quad<\tfrac{1}{2}\big[d_i^t H_i^1 d_i-\tfrac{1}{2}d_i^t(H_i^\varepsilon+H_i^{-\varepsilon})d_i\big]-\tfrac{1}{12}d_i^t(H_i^\varepsilon+H_i^{-\varepsilon})d_i. \end{aligned} \tag{4.14}$$

That the right-hand side of (4.14) is $\leqslant 0$ for large i follows exactly in the same manner in which we proved that (4.12) is $\leqslant 0$. This completes the proof.

Finally, we state certain condition in Theorem 4.3 below which guarantee that $\lambda_i < 1$ so that $\alpha_i = \lambda_i$ for i large enough.

THEOREM 4.3. *Let* $\{x_i\}$ *be a sequence generated by the algorithm. Suppose that* $x_i \to x^*$ *and that* x^* *satisfies the second order optimality conditions for Problem* P′. *If* $z < \gamma/4$, *then there is an integer* m *so that* $\lambda_i < 1$ *for all* $i \geqslant m$, *that is,* $\alpha_i = \lambda_i$ *for all* $i \geqslant m$.

Proof. By Theorem 4.1 there is an integer m so that for $i \geqslant m$ we have

$$\lambda_i = \frac{-\varepsilon^2 \nabla f(x_i)^t d_i}{f(x_i + \varepsilon d_i) + f(x_i - \varepsilon d_i) - 2f(x_i)} = \frac{-\nabla f(x_i)^t d_i}{\frac{1}{2} d_i^t (H_i^\varepsilon + H_i^{-\varepsilon}) d_i}. \tag{4.15}$$

As in the proof of Theorem 4.1,

$$\tfrac{1}{2} d_i^t (H_i^\varepsilon + H_i^{-\varepsilon}) d_i \geqslant (\gamma/4) \|d_i\|^2 \qquad \text{for } i \text{ large enough.} \tag{4.16}$$

Since w_i solves Problem $D(x_i)$, then there exist scalars $u_{ij} \geqslant 0$ for $j \in I(x_i)$ such that

$$\nabla f(x_i) + z w_i + \sum_{j \in I(x_i)} u_{ij} a_j = 0, \tag{4.17}$$

$$u_{ij} a_j^t w_i = 0 \qquad \text{for } \; j \in I(x_i) \tag{4.18}$$

From (4.17) and (4.18) it follows that $\nabla f(x_i)^t w_i = -z \|w_i\|^2$. But by Theorem 3.1, x^* is a c-KT point and hence the optimal solution w^* to Problem $D(x^*)$ is $w^* = 0$. Since $x_i \to x^*$, by continuity of the optimal solution to Problem $D(\cdot)$, and since $b_j - a_j^t x_i > c$ for each $j \in I^+(w_i)$, it follows from (2.3) that $\beta_i = 1$ for large i. Thus $d_i = w_i$ so that

$$\nabla f(x_i)^t d_i = -z \|d_i\|^2 \qquad \text{for large } i. \tag{4.19}$$

Substituting (4.19) and (4.16) in (4.15), it is clear that $\lambda_i < 1$ for i large enough, and the proof is complete.

REFERENCES

1. L. ARMIJO, Minimization of functions having continuous partial derivatives, *Pacific J. Math.* **16** (1966), 1–3.
2. M. S. BAZARAA AND J. J. GOODE, An algorithm for finding the shortest element of a polyhedral set with applications to Lagrangian Duality, *J. Math. Anal. Appl.* **65** (1978), 278–288.

3. J. W. DANIEL, Stability of the solution of definite quadratic programs, *Math. Programming* **5** (1973), 41–53.
4. M. FRANK AND P. WOLFE, An algorithm for quadratic programming, *Naval Res. Logist. Quart.* **3** (1956), 95–110.
5. P. E. GILL AND W. MURRAY, Newton-type methods for unconstrained and linearly constrained optimization, *Math. Programming* **7** (1974), 311–350.
6. D. GOLDFARB, Extension of Davidson's variable metric method to maximization under linear inequality and equality constraints, *SIAM J. Appl. Math.* **17** (1969), 739–764.
7. S. P. HAN AND O. L. MANGASARIAN, Exact penalty functions in nonlinear programming, *Math. Programming* **17** (1979), 251–269.
8. G. P. MCCORMICK, The variable-reduction method for nonlinear programming, *Management Sci.* **17** (1970), 146–160.
9. G. P. MCCORMICK, Second order conditions for constrained minima, *SIAM J. Appl. Math.* **15** (1967), 641–652.
10. H. MUKAI, Readily implementable conjugate gradient methods, *Math. Programming* **17** (1979), 298–319.
11. H. MUKAI, A scheme for determining step sizes for unconstrained optimization methods, *IEEE Tran. Automat. Control* **AC23** (1978), 987–995.
12. B. A. MURTAGH AND M. A. SAUNDERS, Large scale linearly constrained optimization, *Math. Programming* **14** (1978), 41–72.
13. E. POLAK, "Computational Methods in Optimization: A Unified Approach," Academic Press, New York, 1971.
14. J. B. ROSEN, The gradient projection method for nonlinear programming. I. Linear constraints, *SIAM J. Appl. Math.* **9** (1961), 514–553.
15. D. M. TOPKIS AND A. F. VEINOTT, On the convergence of some feasible direction algorithms for nonlinear programming, *SIAM J. Control* **5** (1967), 263–279.
16. P. WOLFE, Algorithm for a least-distance programming problem, *Math. Programming Stud.* **1** (1974), 190–205.
17. P. WOLFE, Methods of nonlinear programming, *in* "Nonlinear Programming" (J. Abadie, Ed.), North-Holland, Amsterdam, 1967.
18. W. I. ZANGWILL, The convex simplex method, *Management Sci.* **14** (1967), 221–238.
19. G. ZOUTENDIJK, "Methods of Feasible Directions," Elsevier, Amsterdam, 1960.