ON THE USE OF APPROXIMATIONS IN ALGORITHMS FOR OPTIMIZATION PROBLEMS WITH EQUALITY AND INEQUALITY CONSTRAINTS

by

H. Mukai and E. Polak

Memorandum No. ERL-489

November 1974

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

ON THE USE OF APPROXIMATIONS IN ALGORITHMS FOR OPTIMIZATION PROBLEMS WITH EQUALITY AND INEQUALITY CONSTRAINTS

H. Mukai and E. Polak

Department of Electrical Engineering and Computer Sciences
and the Electronics Research Laboratory
University of California, Berkeley, California 94720

ABSTRACT

This paper presents an efficient implementation scheme for optimization algorithms in the family of gradient projection, reduced gradient, and gradient restoration methods.

Research sponsored by Joint Services Electronics Program Contract F44620-71-C-0087 and National Science Foundation Grant GK-37672.

1. Introduction

Optimization problems with equality constraints contain the implicit difficulty arising from the need to solve a system of equations while optimizing a cost function. This difficulty tends to result in doubly infinite algorithms, i.e. algorithms constructing infinite sequences based (explicitly or implicitly) on an infinte number of "inner" iterations for each "outer" iteration. An obvious example of explicitly infinite "inner" iterations is found in all penalty function methods. (See, e.g. [5], [10].) Implicitly infinite "inner" iterations are found in gradient projection [7], [9], [11], [13], [19], [20], gradient restoration [14], [15], and reduced gradient algorithms [1], [2], [6], [12], [21]. An important question which has not been analyzed in the literature relating to gradient projection, gradient restoration and reduced gradient methods is that of how the inner iterations should be truncated, without destroying convergence.

In this paper we present a general scheme for implementing algorithms of the gradient projection, gradient restoration and reduced gradient type, i.e., for converting them into convergent algorithms with finite inner iterations. The scheme is computationally quite efficient, but it is moderately complex. To make it transparent, it is presented by taking the reader through three successively more elaborate algorithm models. A particular application to a gradient projection method is given so as to illustrate that the application of our results is reasonably straightforward.

2. Basic Algorithm Model

We begin by considering the problem in a general setting. Let X be a normed linear space† and let T be a closed subset of X. Suppose that T contains a nonempty subset Δ of desirable points, and our problem is to find a point in Δ.

We shall denote the norm on X by $\|\cdot\|$ and we shall use the notation $B(z,\rho) \triangleq \{x \in X \mid \|x - z\| \leq \rho\}$, $\mathbb{R}^+ = [0,\infty)$, $\mathbb{Z} = \{0,1,2,3,\ldots\}$. Any sequence $\{z_i\}_{i=0}^{\infty}$ which is contained in a compact set, will be called a compact sequence.

Quite commonly (see [16], [22]) an algorithm for finding a desirable point will make use of an iteration map $\bar{A} : T \to 2^T$ and of an abstract cost (or descent) function $\bar{c} : T \to \mathbb{R}^1$ and will have the following form:

Algorithm Model 1

Step 0: Compute an $x_0 \in T$ and set $i = 0$.

Step 1: Compute a $y \in \bar{A}(x_i)$.

Step 2: If $\bar{c}(y) - \bar{c}(x_i) < 0$, go to step 3; else stop.

Step 3: Set $x_{i+1} = y$, set $i = i + 1$ and go to step 1. ¤

The requirement that $\bar{A} : T \to 2^T$ means that the sequence $\{x_i\}$ constructed by Algorithm Model 1 must be feasible (i.e., $x_i \in T$ for all i). This causes no difficulty in the case of algorithms such as the methods of feasible directions [23] for solving problems of the form $\min\{f(x) \mid h^j(x) \leq 0,\ j = 1,2,\ldots\ell\}$, where $T = \{x \mid h^j(x) \leq 0,\ j = 1,2,\ldots\ell\}$ and has an interior, but it does cause implementation difficulties in gradient projection, gradient restoration and reduced gradient type methods (see, e.g. [14], [20], [1]) for solving problems of the

†It is easy to generalize our result to topological spaces which are not normed or metric by restating our results in terms of sequential convergence rather than in terms of balls.

form $\min\{f(x) | h^j(x) \leq 0,\ j = 1,2,\ldots,\ell;\ g^k(x) = 0,\ k = 1,2,\ldots,m\}$ where $T = \{x | h^j(x) \leq 0,\ j = 1,2,\ldots,\ell;\ g^k(x) = 0,\ k = 1,2,\ldots m\}$ has no interior. The main reason for this difficulty is that when $y \in \bar{A}(x)$ is approximated by a $\eta \notin T$, it is not possible to maintain the requirement that $\bar{c}(x_{i+1}) < \bar{c}(x_i)$ be satisfied. Consequently, a simple-minded implementation of Algorithm Model 1 which merely substitutes approximations η to $y \in A(x_i)$ for y may result in jamming at an $x_i \notin T$, or in false convergence, i.e., in convergence to a point not in Δ and possibly not even in T.

We now present an implementation of Algorithm Model 1, which has the same convergence properties as Algorithm Model 1 (see Theorem 1.3.1 in [16]) and which is specifically conceived for the case where T has no interior. In this new algorithm model we make use of an abstract cost function $c : X \to \mathbb{R}^1$ which is an extension of $\bar{c}$, an iteration function $A : X \times \mathbb{R}^+ \to 2^X$ which approximates $\bar{A}$, a proximity function $p : X \to \mathbb{R}^+$, which is used to provide a measure of closeness of a point x to the set T, and a restoration map $r : \mathbb{Z} \times X \to X$, which will be used to drive points into T. As we shall later see, frequently Newton's method can be used to define r.

Assumption 1

(i) $c : X \to \mathbb{R}^1$ is continuously Fréchet differentiable.

(ii) $p : X \to \mathbb{R}^+$ is continuous and satisfies $p^{-1}(0) = T$.

(iii) For any $x \in X$, (a) the sequence $\{r(k,x)\}_{k=0}^{\infty}$ converges to a point in T, and (b) $r(0,x) = x$.

(iv) For any compact subset C of X there exists an $M > 0$ and an $e > 0$ such that

(2.1) $\|r(k,x) - r(x)\| \leq M\varepsilon, \ \forall x \in C \cap P(\varepsilon), \ \forall \varepsilon \in [0,e], \ \forall k \in \mathbb{Z}$,

where

(2.2a) $P(\varepsilon) \triangleq \{y \in X \mid p(y) \leq \varepsilon\}$,

(2.2b) $r(x) \triangleq \lim_{k\to\infty} r(k,x)$.

(v) For all $x \in X$, and for all $\varepsilon > 0$, $A(x,\varepsilon) \subset P(\varepsilon)$.

(vi) For any $z \in T$, satisfying $z \notin \Delta$, there exist $\rho(z) > 0$, $\delta(z) < 0$ and $\varepsilon(z) > 0$ such that

(2.3) $c(y) - c(x) \leq \delta(z), \ \forall y \in A(x,\varepsilon), \ \forall x \in B(z,\rho(z)), \ \forall \varepsilon \in [0,\varepsilon(z)]$. ¤

We can now state our new implementation scheme. The reader will note that it has certain structural similarities with the implementation scheme (A.1.1) in [16]. (See Step 5 below.) The scheme (A.1.1) in [16] is directed towards the case where T has an interior, and requires the construction of feasible sequences, whereas the one below does not.

Algorithm Model 2

Parameters: $\varepsilon_0 > 0$, $\gamma > 0$, $\beta \in (0,1)$.

Data: $z_0 \in X$.

Step 0: Set $i = 0$, $j = 0$, $\varepsilon = \varepsilon_0$.

Step 1: Set $k = 0$.

Step 2: If $p(r(k,z_i)) \leq \varepsilon$, go to step 4; else go to step 3.

Step 3: Compute $r(k+1,z_i)$, set $k = k + 1$ and go to step 2.

Step 4: Compute a $y \in A(r(k,z_i),\varepsilon)$.

Step 5: If $c(y) - c(r(k,z_i)) \leq -\gamma\varepsilon$, go to step 6; else set $x_j = r(k,z_i)$, $(y_j = y, \varepsilon_j = \varepsilon)$, $\varepsilon = \beta\varepsilon$, $j = j + 1$ and go to step 2.

Comment: y_j and ε_j above and ξ_i, k_i and e_i below are defined only to facilitate proofs to follow.

Step 6: Set $z_{i+1} = y$, $(\xi_i = r(k,z_i)$, $e_i = \varepsilon$, $k_i = k)$, $i = i + 1$, and go to step 1. ¤

We now establish the convergence properties of Algorithm Model 2. First, it follows easily from Assumption 1 (ii), (iii) and (vi) that the following proposition is true.

Proposition 1: (a) Algorithm Model 2 cannot cycle indefinitely in the loop defined by steps 2 and 3. (b) If Algorithm Model 2 jams up at z_i, cycling indefinitely between steps 2 and 5, which results in an infinite sequence $\{x_j\}_{j=0}^{\infty}$, then $x_j \to x^* \in \Delta$ as $j \to \infty$. ¤

Next, we observe that if Algorithm Model 2 constructs infinite sequences $\{z_i\}_{i=0}^{\infty}$ and $\{e_i\}_{i=0}^{\infty}$ such that $e_i = e^* > 0$ for all $i \geq i_o$, for some i_o, then $k_i = 0$ for all $i \geq i_o$ and hence $\xi_i = r(k_i,z_i) = r(0,z_i) = z_i$ for all $i \geq i_o$. But this implies that $c(z_{i+1}) - c(z_i) \leq -\gamma e^*$ for all $i \geq i_o$, so that $c(z_i) \to -\infty$ as $i \to \infty$. Consequently, we get the following

Proposition 2: Suppose that Algorithm Model 2 constructs a compact infinite sequence $\{z_i\}$. Then the corresponding sequence $\{e_i\}$ converges to zero. ¤

The next result is nowhere near as obvious.

Proposition 3: Suppose that Algorithm Model 2 constructs a compact infinite sequence $\{z_i\}$. Then any accumulation point of $\{z_i\}$ is in Δ.

Proof: Suppose that Algorithm Model 2 has constructed infinite sequences $\{z_i\}$, $\{\xi_i\}$, $\{e_i\}$ and $\{k_i\}$, and that the sequence $\{z_i\}$ is contained in a compact set C. Then, by Proposition 2, $e_i \to 0$ as $i \to \infty$. Next, since by construction $\xi_i = r(k_i,z_i)$, $z_i = r(0,z_i)$ and $p(z_i) \leq e_{i-1}$ for all i, it follows from Assumption 1 (iv) that there exist an $M > 0$ and an integer i_o such that for all $i \geq i_o$,

$$\text{(2.4a)} \qquad \|\xi_i - r(z_i)\| \leq Me_{i-1},$$

$$\text{(2.4b)} \qquad \|z_i - r(z_i)\| \leq Me_{i-1}.$$

Hence there exists a compact set $C' \supset C$ which contains $\{z_i\}$, $\{r(z_i)\}$ and $\{\xi_i\}$. It now follows from Assumption 1(i) that $c(\cdot)$ is Lipschitz continuous on C', with constant L, and hence for any $i \geq i_o$, because of (2.4a), (2.4b),

$$\text{(2.5)} \qquad c(\xi_i) - c(z_i) \leq L\|\xi_i - z_i\| \leq 2LMe_{i-1}.$$

Note that if $e_i = e_{i-1}$, then since $z_i \in A(\xi_{i-1},e_{i-1})$, it follows from Assumption 1(v) that $p(z_i) \leq e_{i-1} = e_i$, and $z_{i+1} \in A(\xi_i,e_{i-1})$. Hence we must have $\xi_i = z_i$ whenever $e_i = e_{i-1}$.

Now, suppose that $\{z_i\}_{i\in I}$ is a subsequence converging to z^*. Since $p(z_i) \leq e_{i-1}$, for all i, it follows from Proposition 2 that $p(z_i) \to 0$ as $i \to \infty$, so that $p(z^*) = 0$, and hence, from Assumption 1(ii) that $z^* \in T$. Referring to (2.4a), (2.4b) we conclude that not only

$z_i \to z^*$ as $i \to \infty$, $i \in I$, but that also $\xi_i \to z^*$ and $r(z_i) \to z^*$ as $i \to \infty$, $i \in I$. Now suppose that $z^* \notin \Delta$. Then by Assumption 1(vi) there exist $\rho(z^*) > 0$, $\delta(z^*) < 0$ and $e(z^*) > 0$ such that (2.3) holds for $z = z^*$, and hence, since $e_i \to 0$ and $\xi_i \to z^*$, as $i \to \infty$, $i \in I$, and $z_{i+1} \in A(\xi_i, e_i)$, there exists an $i_1 \geq i_o$ such that

$$c(z_{i+1}) - c(\xi_i) \leq \delta(z^*) < 0 \quad \forall i \geq i_1,\ i \in I. \tag{2.6}$$

Combining all the results obtained so far, we conclude that for all $i \geq i_1$

$$c(z_{i+1}) - c(z_{i_1}) = \sum_{\ell=i_1}^{i} c(z_{\ell+1}) - c(z_\ell) \tag{2.7}$$

$$= \sum_{\substack{\ell=i_1 \\ \ell\in I \\ e_\ell=e_{\ell-1}}}^{i} (z_{\ell+1}) - c(z_\ell) + \sum_{\substack{\ell=i_1 \\ \ell\notin I \\ e_\ell=e_{\ell-1}}}^{i} c(z_{\ell+1}) - c(z_\ell)$$

$$+ \sum_{\substack{\ell=i_1 \\ \ell\in I \\ e_\ell\neq e_{\ell-1}}}^{i} [(c(z_{\ell+1}) - c(\xi_\ell)) + (c(\xi_\ell) - c(z_\ell))]$$

$$+ \sum_{\substack{\ell=i_1 \\ \ell\notin I \\ e_\ell\neq e_{\ell-1}}}^{i} [(c(z_{\ell+1}) - c(\xi_\ell)) + (c(\xi_\ell) - c(z_\ell))]$$

$$\leq \sum_{\substack{\ell=i_1 \\ \ell\in I}}^{i} \delta(z^*) + 2LM \sum_{\substack{\ell=i \\ e_\ell \neq e_{\ell-1}}}^{i} e_{\ell-1}$$

$$\leq \sum_{\substack{\ell=i_1 \\ \ell\in I}}^{i} \delta(z^*) + 2LM\varepsilon_0/(1-\beta),$$

since $\{e_{\ell-1}\}_{\ell=i}^{\infty}$, $e_\ell \neq e_{\ell-1}$ is a subsequence of $\{\varepsilon_0\beta^j\}_{j=0}^{\infty}$. Hence $c(z_i) \to -\infty$ as $i \to \infty$, $i \in I$. But this contradicts the continuity of $c(\cdot)$ at z^*, and hence we conclude that $z^* \notin \Delta$ is false, i.e., that $z^* \in \Delta$. ¤

We can now summarize our conclusions as follows.

<u>Theorem 1</u>: (i) If Algorithm Model 2 stops at a particular z_i and constructs an infinite sequence $\{x_j\}$, then $x_j \to x^* \in \Delta$ as $j \to \infty$.
(ii) If Algorithm Model 2 constructs a compact infinite sequence $\{z_i\}$, then any accumulation point z^* of $\{x_i\}$ is in Δ. ¤

3. A Structure for the Map A

In specific applications, such as gradient projection, gradient restoration and reduced gradient methods, the map $\bar{A}$, corresponding to the one in Algorithm Model 1, has a very specific structure, viz. it is a composite map made up of a direction finding map, a step size map, a restoration map and other bits and pieces. The appropriate map A will usually have to be considerably more complex. Because this specific structure occurs so frequently, it is possible to save considerable effort in applications by working out in advance the implications of Assumptions 1 (v) and (vi) on the composite parts of A for this type of structure. We shall therefore do it in this section.

We shall define the map A in terms of a first-order-cost-reduction estimate function $\phi : X \to \mathbb{R}^-$, a set-valued, descent direction function $D : X \to 2^X$ and a curvilinear Armijo type [3] step size function $\tilde{\lambda} : \mathbb{R}^+ \times X \times X \to \mathbb{R}^1$.

The following hypotheses on ϕ and D will ensure that Assumption 1 (v),(vi) are satisfied.

Assumption 2

(i) The function $\phi(\cdot)$ is upper semi-continuous.

(ii) $\phi^{-1}(0) \cap T$ is a nonempty subset of Δ.

(iii) Given fixed, but arbitrary constants $\lambda_s > 0$, $\alpha_1 \in (0,1)$, $\beta_2 \in (0,1)$, for any $z \in T$, $z \notin \Delta$, there exist a $\rho_z > 0$ and an integer $\ell_z \geq 0$ such that for any $x \in B(z,\rho_z)$, for any $d \in D(x)$ and for any $k \in \mathbb{Z}$,

$$(3.1) \qquad c(r(k,x+\lambda_s\beta_2^{\ell_z}d)) - c(x) \leq \alpha_1\lambda_s\beta_2^{\ell_z}\phi(x). \qquad \square$$

We now proceed to define the step size function. First we define $\tilde{k} : X \times (0,\infty) \to \mathbb{Z}$ by

(3.2) $$\tilde{k}(x,\varepsilon) = \min\{k \in \mathbb{Z} \mid r(k,x) \in P(\varepsilon)\}$$

Next we define $\tilde{y} : X \times \mathbb{R}^+ \to S$ by

(3.3) $$\tilde{y}(x,\varepsilon) = \begin{cases} r(\tilde{k}(x,\varepsilon),x), & \text{if } \varepsilon > 0, \\ r(x), & \text{if } \varepsilon = 0, \end{cases}$$

i.e. $\tilde{y}(x,\varepsilon)$ is the first element of the sequence $\{r(k,x)\}_{k=o}^{\infty}$ to satisfy $p(r(k,x)) \leq \varepsilon$. Proceeding, given λ_s, β_2 as in Assumption 2 (iii), and $\lambda_e > 0$, arbitrary, but fixed, we define a set-valued function $\Lambda : \mathbb{R}^+ \to 2^{\mathbb{R}^+}$ of step size candidates by

(3.4) $$\Lambda(\varepsilon) = \{\lambda_s\beta_2^{\ell} \mid \lambda_s\beta_2^{\ell} \geq \lambda_e\varepsilon,\ \ell \in \mathbb{Z}\}$$

Finally, given fixed, but arbitrary constants $\gamma_1 > 0$, $\alpha_1 \in (0,1)$, we define the step size function $\tilde{\lambda} : \mathbb{R}^+ \times X \times X \to \mathbb{R}^+$ by

(3.5) $$\tilde{\lambda}(\varepsilon,x,d) = \begin{cases} 0 \text{ if } \phi(x) \geq -\gamma_1\varepsilon, \\ 0 \text{ if } c(\tilde{y}(x+\lambda d,\varepsilon)) - c(x) > \alpha_1\lambda\phi(x),\ \forall\lambda \in \Lambda(\varepsilon), \\ \max\{\lambda \in \Lambda(\varepsilon) \mid c(\tilde{y}(x+\lambda d,\varepsilon)) - c(x) \leq \alpha_1\lambda\phi(x)\} \\ \qquad\qquad \text{otherwise.} \end{cases}$$

Combining all the pieces, we now define the iteration map $A : X \times \mathbb{R}^+ \to 2^X$ by

$$(3.6)\qquad A(x,\varepsilon) = \bigcup_{d\in D(x)} \{\tilde{y}(x+\tilde{\lambda}(\varepsilon,x,d)d,\varepsilon)\}$$

Proposition 4: The iteration map A defined by (3.6) satisfies Assumption 1 (v), (vi).

Proof: Since by construction $\tilde{y}(x+\tilde{\lambda}(\varepsilon,x,d)d,\varepsilon) \in P(\varepsilon)$ for any $x \in X$, $d \in X$ and $\varepsilon > 0$, Assumption 1 (v) is obviously satisfied.

We now show that Assumption 1 (vi) is satisfied. Suppose $z \in T$ and $z \notin \Delta$. Then, by Assumption 2 (iii), there exist a $\rho_z > 0$ and an integer $\ell_z \le 0$ such that for all $x \in B(z,\rho_z)$, for all $d \in D(x)$, for all $k \in \mathbb{Z}$,

$$(3.7)\qquad c(r(k,x+\lambda_s\beta_2^{\ell_z}d)) - c(x) \le \alpha_1\lambda_s\beta_2^{\ell_z}\phi(x).$$

Next, since $z \notin \Delta$, it follows from Assumption 2 (i), that there exists a $\rho(z) \in (0,\rho_z)$ such that

$$(3.8)\qquad \phi(x) \le \frac{1}{2}\phi(z) < 0, \quad \forall x \in B(z,\rho(z)).$$

Now, let $e(z) > 0$ be such that for all $\varepsilon \in [0,e(z)]$

$$(3.9)\qquad \frac{1}{2}\,\phi(z) \le -\gamma_1\varepsilon \text{ and } \lambda_s\beta_2^{\ell_z} \ge \lambda_e\varepsilon.$$

It now follows from (3.7), (3.8) and (3.9) that for all $x \in B(z,\rho(z))$, for all $d \in D(x)$, for all $\varepsilon \in [0,e(z)]$

$$(3.10)\qquad c(\tilde{y}(x+\lambda_s\beta_2^{\ell_z}d,\varepsilon)) - c(x) \le \alpha_1\lambda_s\beta_2^{\ell_z}\phi(x) \le \alpha_1\lambda_s\beta_2^{\ell_z}\phi(z)/2 < 0,$$

(3.11) $\phi(x) \le \frac{1}{2}\phi(z) \le -\gamma_1 \varepsilon$,

(3.12) $\lambda_s \beta_2^{\ell_z} \ge \lambda_e \varepsilon$.

Consequently, we must have $\tilde{\lambda}(\varepsilon,x,d) \ge \lambda_s \beta_2^{\ell_z} > 0$ for all $x \in B(z,\rho(z))$, for all $d \in D(x)$, for all $\varepsilon \in [0,e(z)]$. Therefore, for all $x \in B(z,\rho(z))$, for all $d \in D(x)$, for all $\varepsilon \in [0,\varepsilon(z)]$,

(3.13) $c(\tilde{y}(x+\tilde{\lambda}(\varepsilon,x,d)d,\varepsilon)) - c(x)$

$$\le \alpha_1 \tilde{\lambda}(\varepsilon,x,d)\phi(x) \le \frac{1}{2}\alpha_1 \lambda_s \beta_2^{\ell_z}\phi(z) \triangleq \delta(z) < 0 .$$

Thus, Assumption 1 (vi) is satisfied. ◻

Substituting for A from (3.6) into Algorithm Model 2, we obtain the following expanded version.

<u>Algorithm Model 3</u>

<u>Parameters</u>: $\alpha_1 \in (0,1)$, $\beta_1 \in (0,1)$, $\beta_2 \in (0,1)$, $\gamma_1 > 0$, $\gamma_2 > 0$, $\lambda_s > 0$, $\varepsilon_0 > 0$, $\lambda_e \in (0,\lambda_s/\varepsilon_0)$.

<u>Data</u>: $z_0 \in X$.

<u>Step 0</u>: Set $i = 0$, set $j = 0$, set $\varepsilon = \varepsilon_0$.

<u>Step 1</u>: Set $k = 0$.

<u>Step 2</u>: If $p(r(k,z_i)) \le \varepsilon$, go to step 4; else go to step 3.

<u>Step 3</u>: Compute $r(k+1,z_i)$, set $k = k + 1$ and go to step 2.

<u>Step 4</u>: Compute $\phi(r(k,z_i))$ and a $d_k \in D(r(k,z_i))$.

<u>Step 5</u>: If $\phi(r(k,z_i)) \le -\gamma_1\varepsilon$, go to step 6; else set $y = r(k,z_i)$ and go to step 12.

<u>Step 6</u>: Set $\lambda = \lambda_s$.

Step 7: If $\lambda \geq \lambda_e \varepsilon$, go to step 8; else set $y = r(k,z_i)$ and go to step 12.

Step 8: Set $m = 0$.

Step 9: If $p(r(m,(r(k,z_i) + \lambda d_k))) \leq \varepsilon$, go to step 11; else go to step 10.

Step 10: Compute $r(m+1, (r(k,z_i) + \lambda d_k))$, set $m = m + 1$ and go to step 9.

Step 11: If

$$c(r(m,(r(k,z_i) + \lambda d_k))) - c(r(k,z_i)) \leq \alpha_1 \lambda \phi(r(k,z_i)),$$

set $y = r(m,(r(k,z_i) + \lambda d_k))$ and go to step 12.; else set $\lambda = \lambda\beta_2$ and go to step 7.

Step 12: If $c(y) - c(r(k,z_i)) \leq -\gamma_2 \varepsilon$, go to step 13; else set $x_j = r(k,z_i)$, set $j = j + 1$, set $\varepsilon = \varepsilon\beta_1$, and go to step 2.

Step 13: Set $z_{i+1} = y$, set $i = i + 1$, and go to step 1. ⌑

We note that in programming an algorithm of the form of Algorithm Model 3, one would remove the obvious redundancies which appear in Algorithm Model 3 only so as to exhibit its exact correspondence to Algorithm Model 2. Thus, step 5 would be modified to read: "else, set $\varepsilon = \varepsilon\beta_1$, $x_j = r(k,z_i)$, $j = j + 1$ and go to step 2," since the outcome of the test in step 12 is obvious at this point and hence can be omitted.

We now present a specific application of our implementation method.

4. A Gradient Projection Algorithm

The algorithm we shall present in this section solves problems of the form

(4.1) $\min\{f(x) \mid g(x) = 0, h(x) \leqq 0\}$

where $f : \mathbb{R}^n \to \mathbb{R}^1$, is continuously differentiable and $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^\ell$ are twice continuously differentiable.

We shall denote the components of a vector by superscripts; the elements of a sequence by subscripts. Vector inequalities are to be interpreted componentwise. We shall use the Euclidean norm $\|\cdot\|$ throughout, and we shall denote the feasible set by $\Omega \triangleq \{x \mid g(x) = 0, h(x) \leqq 0\}$.

Our algorithm makes use of two special functions: $h^+, h^- : \mathbb{R}^n \to \mathbb{R}^\ell$ defined, componentwise, by $h^{+i}(x) = \max\{0, h^i(x)\}$ and $h^{-i}(x) = \min\{0, h^i(x)\}$, $i = 1, \ldots, \ell$ respectively. The restoration function r which we shall use necessitates a special condition on g, h.

Assumption 3: For any $x \in \mathbb{R}^n$, the pair of Jacobian matrices $\left(\frac{\partial g(x)}{\partial x}, \frac{\partial h(x)}{\partial x}\right)$ satisfies the LI condition [17]; viz. $\frac{\partial g(x)}{\partial x}^T \mu + \frac{\partial h(x)}{\partial x}^T \nu = 0$, and $\nu \geqq 0$ implies that $\mu = 0$, $\nu = 0$. ¤

Proposition 5: Under Assumption 3, the set Ω satisfies the Kuhn-Tucher constraint qualification. (See theorem 3.3.17 [4]) ¤

Proposition 6: The system $\frac{\partial g(x)}{\partial x} v = a$, $\frac{\partial h(x)}{\partial x} v \leqq b$ has a solution for any $x \in \mathbb{R}^n$, $a \in \mathbb{R}^m$, $b \in \mathbb{R}^\ell$, if and only if Assumption 3 holds. (See [17]).¤

The algorithm in this section finds Kuhn-Tucher points, which we define as follows.

Definition 1: We shall say that a point $x \in \mathbb{R}^n$ is a Kuhn-Tucher point (KTP) if $x \in \Omega$ and there exist multipliers $\mu \geqq 0$, ψ such that

$$\nabla f(x) + \frac{\partial g(x)}{\partial x}^T \psi + \frac{\partial h(x)}{\partial x}^T \mu = 0 \tag{4.2}$$

and $\langle \mu, h(x) \rangle = 0$. Let

$$\mathcal{K} \triangleq \{x \in \Omega | x \text{ is KTP}\}. \tag{4.3}$$ ¤

Given a $z \in \Omega$, $z \notin \mathcal{K}$, $-\nabla f(z)$ defines a direction from z in which the cost is reduced. However this direction does not, in general, generate points in Ω. A projection map $\pi : \mathbb{R}^n \to \Omega$ maps this half line into a feasible curve $\zeta(z) = \{\pi(z - \lambda \nabla f(x)) | \lambda \geq 0\} \subset \Omega$, along which the cost can also be reduced. This fact gives rise to the classical, conceptual gradient projection algorithm [9] which, given $z_i \in \Omega$, computes z_{i+1} according to $z_{i+1} = \arg\min\{f(y) | y \in (z_i)\}$. However, the work of projection can be reduced substantially by first computing the negative gradient projection direction $\tilde{d}(z)$, which is defined by

$$\tilde{d}(z) = \arg\min_{d \in \mathbb{R}^n}\{\|\nabla f(z) + d\|^2 | g(z) + \frac{\partial g(z)}{\partial x} d = 0, \; h(z) + \frac{\partial h(z)}{\partial x} d \leqq 0\} \tag{4.4}$$

for a given $z \in \Omega$, and then computing $\pi(z + \lambda d(z))$, because the direction $\tilde{d}(z)$ satisfies the constraints to first order. This idea yields a somewhat more complex, conceptual gradient projection algorithm, as follows. For any $z \in \Omega$, let $\tilde{\zeta}(z) = \{\pi(z + \lambda \tilde{d}(z)) | \lambda \geq 0\}$. Then, given any $z_i \in \Omega$, $z_{i+1} = \arg\min\{f(z) | y \in \tilde{\zeta}(z_i)\}$. To obtain an implementable algorithm from this conceptual one, we replace the minimization scheme for selecting λ by an Armijo type scheme, and we approximate the projection map π by a restoration map r. In constructing the implementation we are guided by Algorithm Model 3.

We identify X with $\mathbb{R}^n$, T with Ω, Δ with $\mathcal{K}$ and $c(\cdot)$ with $f(\cdot)$. We define $\bar{p} : \mathbb{R}^n \to \mathbb{R}^+$ by

$$\bar{p}(x) = \frac{1}{2}\left\|\begin{matrix} g(x) \\ h^+(x) \end{matrix}\right\|^2 = \frac{1}{2}\|g(x)\|^2 + \frac{1}{2}\|h^+(x)\|^2, \tag{4.5}$$

and the proximity function $p : \mathbb{R}^n \to \mathbb{R}^+$ by

$$p(x) = \sqrt{2\bar{p}(x)}\ . \tag{4.6}$$

The restoration function $r : \mathbb{Z} \times X \to X$ is defined recursively as follows:

$$r(k+1,z) = a(r(k,z)),\ k = 0,1,2,\ldots, \tag{4.7a}$$

$$r(0,z) = z\ , \tag{4.7b}$$

where $a : \mathbb{R}^n \to \mathbb{R}^n$ is defined below.

<u>Restoration Iteration Function a</u>:

<u>Parameters</u>: $\alpha_2 \in (0,1/2)$, $\beta_3 \in (0,1)$ to be specified by user.

<u>Data</u>: $x \in \mathbb{R}^n$.

<u>Step 1</u>: If $\bar{p}(x) = 0$, set $a(x) = x$ and stop; else go to step 2.

<u>Step 2</u>: Compute $v(x)$ as a solution of

$$\min\{\|v\|^2 \,|\, g(x) + \frac{\partial g(x)}{\partial x} v = 0,\ h(x) + \frac{\partial h(x)}{\partial x} v \leqq 0\}. \tag{4.8}$$

<u>Step 3</u>: Compute the smallest integer $k \geq 0$ satisfying

$$\bar{p}(x+\beta_3^k v(x)) \leq (1-2\alpha_2\beta_3^k)\bar{p}(x). \tag{4.9}$$

<u>Step 4</u>: Set $a(x) = x + \beta_3^k v(x)$ and stop. ¤

<u>Proposition 7</u>: The function a is well defined.

Proof: It follows from Assumption 3 and Proposition 6 that $v(x)$ is well defined. Next, if $\bar{p}(x) \neq 0$, since $\nabla\bar{p}(x) = \frac{\partial g(x)}{\partial x}^T g(x) + \frac{\partial h(x)}{\partial x}^T h^+(x)$, we have, because of (4.8),

$$(4.10) \qquad \langle \nabla\bar{p}(x), v(x)\rangle = \langle g(x), \frac{\partial g(x)}{\partial x} v(x)\rangle + \langle h^+(x), \frac{\partial h(x)}{\partial x} v(x)\rangle$$

$$\leq -\|g(x)\|^2 - \|h^+(x)\|^2 = -2\bar{p}(x) < 0.$$

It now follows from the mean value theorem that a finite k, satisfying (4.9) exists. ¤

The restoration function a is due to Huang [8] and is based on Robinson's extension of Newton's method [17]. In [8] Huang established the global convergence of this algorithm, but he did not give the bounds and rate of convergence stated in theorem 2, below.

Theorem 2: With r defined as in (4.7),

(i) Given any x, either $\{r(k,x)\}_{k=0}^{\infty}$ has no accumulation points, or it has a unique limit point $x^* \in \Omega$ and $r(k,x) \to x^*$ quadratically.

(ii) Given any compact set $U \subset \mathbb{R}^n$, there exist $\varepsilon > 0$, $M \in (0,\infty)$ such that for all $x \in P(\varepsilon) \cap U$, for all $k \in \mathbb{Z}$, $r(k,x) \in B(x, Mp(x))$, where P is defined as in (2.2a) with p as in (4.6). ¤

The proof of this theorem is given in the Appendix.

We define $d : \mathbb{R}^n \to \mathbb{R}^n$ by

$$(4.11) \qquad d(x) = \underset{d' \in \mathbb{R}^n}{\arg\min}\{\|\nabla f(x) + d'\|^2 \;\Big|\; \frac{\partial g(x)}{\partial x} d' = 0,\ h^-(x) + \frac{\partial h(x)}{\partial x} d' \leqq 0\},$$

and associate it with the map $D : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ appearing in Algorithm Model 3 by setting

(4.12) $\quad D(x) = \{d(x)\}$

We note that d is an extension of $\tilde{d}$ as defined in (4.4).

We shall prove in the appendix the following

Proposition 8: The function $d : \mathbb{R}^n \to \mathbb{R}^n$ defined by (4.11) is continuous. ¤

Proposition 9: A point $x^* \in \mathbb{R}^n$ is a Kuhn-Tucker point if and only if $p(x^*) = 0$ and $d(x^*) = 0$.

Proof: $\Rightarrow$ If x^* is a Kuhn-Tucker point, then $p(x^*) = 0$ and $d' = 0$ satisfies the Kuhn-Tucker conditions for (4.11). Since the Kuhn-Tucker conditions are both necessary and sufficient for (4.11), $d(x^*) = 0$.
$\Leftarrow$ Suppose $p(x^*) = 0$ and $d(x^*) = 0$, then applying the Kuhn-Tucker conditions to (4.11), we immediately obtain that x* is a Kuhn-Tucker point. ¤

Proposition 10: For any $x \in \mathbb{R}^n$, (i) $\|d(x)\| \leq 2\|\nabla f(x)\|$, and (ii) $\langle \nabla f(x), d(x) \rangle \leq -\frac{1}{2}\|d(x)\|^2$.

Proof: Since the zero vector is feasible for the min problem in (4.11), $\|d(x) + \nabla f(x)\|^2 \leq \|\nabla f(x)\|^2$. Hence

(4.13) $\quad \|d(x)\| \leq \|d(x) + \nabla f(x)\| + \|\nabla f(x)\| \leq 2\|\nabla f(x)\|.$

Hence (ii) follows from

(4.14) $\quad \|\nabla f(x)\|^2 + \|d(x)\|^2 + 2\langle \nabla f(x), d(x) \rangle \leq \|\nabla f(x)\|^2.$ ¤

We now define $\phi : \mathbb{R}^n \to \mathbb{R}^-$ by

(4.15) $\quad \phi(x) = \langle \nabla f(x), d(x) \rangle .$

Proposition 11: The function ϕ defined by (4.15) is continuous. ¤

In what follows we depend on the following assumption, which will generally be satisfied.

Assumption 4: For any $x \in \mathbb{R}^n$, the sequence $\{r(k,x)\}_{k=0}^{\infty}$, with r defined as in (4.7) has a unique limit point. ¤

Lemma 1: Given a point $z \in \Omega$, such that $z \notin \mathcal{K}$, there exists a $\rho_z > 0$ and an integer $\ell_z \geq 0$ such that for any $k \in \mathbb{Z}$ and any $x \in B(z,\rho_z)$,

$$(4.16) \qquad f(r(k,x+\beta_2^{\ell_z}\lambda_s d(x))) - f(x) \leq \alpha_1\beta_2^{\ell_z}\lambda_s\phi(x).$$

Proof: Let $z \in \Omega$, with $z \notin \mathcal{K}$, and let U be a compact neighborhood of z. Then, because of Theorem 2, there exist $\varepsilon > 0$, and $M \in (0,\infty)$ such that $r(k,x) \in B(x,M\sqrt{2p(x)})$, $\forall x \in U \cap P(\varepsilon)$, $\forall k \in \mathbb{Z}$. Next, because of the continuity of $d(\cdot)$ (Proposition 8) and because $U \cap P(\varepsilon)$ is a neighborhood of z, there exist a $\rho_1 > 0$ and a $\lambda_1 \in (0,1)$ such that $(x+\lambda d(x)) \in U \cap P(\varepsilon)$, $\forall x \in B(z,\rho_1)$, $\forall \lambda \in [0,\lambda_1]$. Now the functions f, $\frac{\partial g}{\partial x}$, $\frac{\partial h}{\partial x}$ are continuously differentiable, and hence there exists a Lipschitz constant $L > 0$ as used below.

Let $x \in B(z,\rho_1)$, $\lambda \in (0,\lambda_1]$ and $k \in \mathbb{Z}$. Then, making use of Proposition 10,

$$(4.17) \qquad f(r(k,x+\lambda d(x))) - f(x) - \lambda\alpha_1\langle \nabla f(x),d(x)\rangle$$

$$= [f(r(k,x+\lambda d(x))) - f(x+\lambda d(x))] + f(x+\lambda d(x)) - f(x)$$

$$- \lambda\alpha_1\langle \nabla f(x),d(x)\rangle$$

$$\leq LM\sqrt{2p(x+\lambda d(x))} + \lambda \sup_{s\in[0,1]} \|\nabla f(x+s\lambda d(x))-\nabla f(x)\| \, \|d(x)\|$$

$$- \lambda(1-\alpha_1)\|d(x)\|^2/2 .$$

Let θ: $\mathbb{R}^n \times \mathbb{R}^1 \to \mathbb{R}^1$ be defined by

$$\theta(x,\lambda) \triangleq LM\sqrt{2\bar{p}(x+\lambda d(x))} + \lambda \sup_{s\in[0,1]} \|\nabla f(x+s\lambda d(x)) - \nabla f(x)\| \, \|d(x)\| - (1-\alpha_1) \|d(x)\|^2/2. \tag{4.18}$$

Now, for any $\lambda \in [0,\lambda_1]$, it follows from (4.11), since $g(z) = 0$ and $h(z) = h^-(z)$, that $g(z) + \lambda\frac{\partial g(z)}{\partial x} d(z) = 0$ and $h(z) + \lambda\frac{\partial h(z)}{\partial x}d(z) \leqq 0$. Hence for any $\lambda \in [0,\lambda_1]$,

$$\|g(z+\lambda d(z))\| = \|\lambda\int_0^1 (\frac{\partial g(z+t\lambda d(z))}{\partial x} - \frac{\partial g(z)}{\partial x})d(z)dt\| \leq L\lambda^2\|d(z)\|^2, \tag{4.19}$$

$$\|h(z+\lambda d(z))\| = \|\lambda\int_0^1 (\frac{\partial h(z+t\lambda d(z))}{\partial x} - \frac{\partial h(z)}{\partial x})d(z)dt \| \leq L\lambda^2\|d(z)\|^2 \tag{4.20}$$

and hence

$$\sqrt{2\bar{p}(z+\lambda d(z))} \leq \sqrt{2}\, L\lambda^2 \|d(z)\|^2 \tag{4.21}$$

Hence, for all $\lambda \in [0,\lambda_1]$,

$$\theta(z,\lambda) \leq \lambda[\sqrt{2}\, L^2M\lambda\|d(z)\| + \sup_{s\in[0,1]} \|\nabla f(z+s\lambda d(z)) - \nabla f(z)\| - (1-\alpha_1)\|d(z)\|/2]\|d(z)\|. \tag{4.22}$$

and there exists a $\lambda_2 \in (0,\lambda_1]$ such that for all $\lambda \in [0,\lambda_2]$, $\theta(z,\lambda) \leqq -\frac{1}{4}\lambda(1-\alpha_1)\|d(z)\|^2$. It now follows from the uniform continuity of θ on $B(z,\rho_1) \times [0,\lambda_2]$ that there exists a $\rho_z \in (0,\rho_1]$ and an $\ell_z \geq 0$, with $\lambda_s\beta_2^{\ell_z} \in (0,\lambda_2]$ such that

$$\theta(x,\beta_2^{\ell_z}\lambda_s) \leq -\beta_2^{\ell_z}\lambda_s(1-\alpha_1)\|d(z)\|^2/8, \quad \forall x \in B(z,\rho_z), \tag{4.23}$$

which completes our proof. ¤

We can now show that Assumption 1 is satisfied. Thus, the functions f, g and h are continuously differentiable, Assumption 1 (i) is satisfied and it follows from (4.5) and (4.6) that Assumption 1 (ii) is satisfied. Assumption 1 (iii), (iv) follows from Assumption 4 and Theorem 2. Assumption 2 follows from Lemma 1 and Propositions 9 and 11, since Proposition 10 implies that $d(x) = 0$ if and only if $\phi(x) = 0$. It now follows from Proposition 4 that Assumptions 1 (v), (vi) are satisfied. Consequently, Theorem 1 implies the following.

Theorem 3: (i) If the gradient projection algorithm jams up at a point x_ν, then $x_\nu \in \mathcal{K}$. (ii) If the gradient projection algorithm constructs either an infinite sequence $\{x_k\}$ or an infinite sequence $\{z_j\}$ then any accumulation point of such a sequence is in $\mathcal{K}$. (iii) If neither (i) nor (ii) takes place, then the gradient projection algorithm constructs an infinite unbounded sequence $\{z_i\}$. ¤

This concludes our analysis of the gradient projection algorithm.

Appendix

In the next two proofs we shall need the following result: For any $x \in \mathbb{R}^n$, we define

$$\text{(A1)} \qquad \mu(x) \triangleq \max \left\{ \min\left\{ \|v\| \,\Big|\, \frac{\partial g(x)}{\partial x} v = a,\ \frac{\partial h(x)}{\partial x} v \leqq b \right\} \,\Big|\, \left\| \begin{matrix} a \\ b \end{matrix} \right\| \le 1 \right\}.$$

Because of Assumption 3, it follows from [17] and [18] that $\mu(\cdot)$ is well defined and upper semicontinuous.

Proof of Theorem 2

First, since any v satisfying $g(x) + \frac{\partial g(x)}{\partial x} v = 0$, $h^+(x) + \frac{\partial h(x)}{\partial x} v \leqq 0$ also satisfies the constraints in (4.8), we must have

$$\text{(A2)} \qquad \|v(x)\| \le \mu(x) \left\| \begin{matrix} g(x) \\ h^+(x) \end{matrix} \right\| = \mu(x)\ \sqrt{2\bar{p}(x)} = \mu(x) p(x).$$

Then, making use of (4.10), the continuity of $\bar{p}(\cdot)$, the fact that $\nabla\bar{p}(x) = 0$ if and only if $\bar{p}(x) = 0$ and Theorem 1.3.10 in [16], we conclude that for any $x \in \mathbb{R}^n$, any accumulation point x^* of $\{r(k,x)\}_{k=0}^{\infty}$ satisfies $p(x^*) = 0$, i.e., $x^* \in \Omega$.

Next, if $\{r(k,x)\}_{k=0}^{\infty}$ has accumulation points, then $p(r(k,x)) \to 0$ as $k \to \infty$. It now follows from (A2) and [17] that there exists a k' such that the sequence $\{\xi_j\}_{j=0}^{\infty}$ defined by

$$\text{(A3)} \qquad \xi_{j+1} = \xi_j + v(\xi_j),\ j = 0,1,2,\ldots,\ \xi_0 = r(k',x),$$

converges to an $x^* \in \Omega$. Furthermore, there exists an $M \in (0,\infty)$ such that

$$\text{(A4)} \qquad \|\xi_{j+1} - \xi_j\| \le M\ \|\xi_j - \xi_{j-1}\|^2,\ j = 1,2,\ldots\ .$$

Now, from (4.8) we have that $\|g(\xi_j)\| \le \left\|\frac{\partial g(\xi_j)}{\partial x}\right\|\ \|v(\xi_j)\|$, and,

$h^k(\xi_j) \leq \left[-\left(\frac{\partial h(\xi_j)}{\partial x}\right)v(\xi_j)\right]^k$, $k = 1,2,\ldots,\ell$. Hence $\|h^+(\xi_j)\| \leq \|\frac{\partial h(\xi_j)}{\partial x}\| \|v(\xi_j)\|$. Therefore

$$\text{(A.5)} \qquad \bar{p}(\xi_j) \leq M' \ \|v(\xi_j)\|^2,$$

where $M' = \frac{1}{2} \sup_j (\|\frac{\partial g(\xi_j)}{\partial x}\|^2 + \|\frac{\partial h(\xi_j)}{\partial x}\|^2)$. Combining (A2), (A3) and (A4), we get

$$\text{(A6)} \qquad \bar{p}(\xi_{j+1}) \leq M'\|v(\xi_{j+1})\|^2 \leq M'M^2\|v(\xi_j)\|^4 \leq 4\mu(\xi_j)^2 M'M^2\bar{p}(\xi_j)^2$$

Since $\mu(\cdot)$ is upper semicontinuous and $\xi_j \to x^*$, there exists an $M^* \in (0,\infty)$ such that

$$\text{(A7)} \qquad \bar{p}(\xi_{j+1}) \leq (M^*\bar{p}(\xi_j))\bar{p}(\xi_j), \ j = 0,1,2,..$$

It can be shown that k' can be chosen sufficiently large to ensure that $M^*\bar{p}(\xi_j) \leq (1-2\alpha_2)$ for $j = 0,1,2,\ldots$. Hence, in the restoration algorithm, we must have $r(k'+j,x) = \xi_j$ for $j = 0,1,2,\ldots$. Thus part (i) of Theorem 2 is true. Part (ii) of Theorem 2 now follows directly from [17] ¤

Proof of Proposition 8

Suppose that $d(\cdot)$ is not continuous at x^*, then, because of Proposition 10, there exists a sequence $\{x_i\}_{i=0}^{\infty}$ such that $x_i \to x^*$, but $d(x_i) \to d^* \neq d(x^*)$. Since $\frac{\partial g(\cdot)}{\partial x}$, $\frac{\partial h(\cdot)}{\partial x}$ and $h^-(\cdot)$ are continuous, it follows that $\frac{\partial g(x^*)}{\partial x}d^* = 0$, $h^-(x^*) + \frac{\partial h(x^*)}{\partial x}d^* \leqq 0$, and hence $\|\nabla f(x^*) + d(x^*)\|^2 < \|\nabla f(x^*) + d^*\|^2$.

Now there exists a sequence $\{d_i\}_{i=0}^{\infty}$ such that d_i satisfies the constraints in (4.11) and for $i = 0,1,2,\ldots$,

$$\text{(A8)} \qquad \|d_i - d(x^*)\| \leq \mu(x_i) \left\| \begin{matrix} \left(\frac{\partial g(x^*)}{\partial x} - \frac{\partial g(x_i)}{\partial x}\right) d(x^*) \\ h^-(x^*) - h^-(x_i) + \left(\frac{\partial h(x^*)}{\partial x} - \frac{\partial h(x_i)}{\partial x}\right) d(x^*) \end{matrix} \right\|$$

Hence $d_i \to d(x^*)$, $\|\nabla f(x_i) + d(x_i)\|^2 \leq \|\nabla f(x_i) + d_i\|^2$, and $\|\nabla f(x_i) + d_i\|^2 \to \|\nabla f(x^*) + d(x^*)\|^2$. But by assumption $\|\nabla f(x_i) + d(x_i)\|^2 \to \|\nabla f(x^*) + d^*\|^2 > \|\nabla f(x^*) + d(x^*)\|$ and hence we have a contradiction. ¤

References

[1] J. Abadie and J. Carpentier, "Generalization of the Wolfe Reduced Gradient Method for the Case of Nonlinear Constraints," in Optimization, ed. by R. Fletcher, Academic Press, 1969.

[2] J. Abadie and J. Guigou, "Numerical Experiments with the GRG Method," in Integer and Nonlinear Programming, ed. by J. Abadie, North-Holland Pub. Co., Amsterdam, 1970.

[3] L. Armijo, "Minimization of Functions Having Continuous Partial Derivatives," Pacific J. Math., vol. 16, pp. 1-3, 1966.

[4] M. D. Canon, C. D. Cullum, Jr. and E. Polak, Theory of Optimal Control and Mathematical Programming, McGraw-Hill, 1970.

[5] A. V. Fiacco and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley, New York, 1968.

[6] D. Gabay and D. G. Luenberger, "Efficiently Converging Minimization Methods Based on the Reduced Gradient," SIAM Journal on Control, in press.

[7] A. A. Goldstein, "Convex Programming in Hilbert Space," Bull. of AMS, vol. 70, no. 5, pp. 709-710, 1964.

[8] T. J. Huang, "Algorithms for Solving Systems of Equations and Inequalities with Applications in Nonlinear Programming," Computer Sciences Tech. Rept. #191, University of Wisconsin, Sept. 1973.

[9] E. S. Levitin and B. T. Polyak, "Constrained Minimization Methods," USSR Comp. Math. and Math. Physics, vol. 6, no. 5, pp. 1-50, 1966.

[10] F. A. Lootsma, Boundary Properties of Penalty Functions for Constrained Minimization, Doc. Thesis, Technische Hogeschool Eindhoven, 1970.

[11] D. G. Luenberger, "The Gradient Projection Method Along Geodesics," Management Science, vol. 18, no. 11, pp. 620-631, 1972.

[12] D. G. Luenberger, Introduction to Linear and Nonlinear Programming, Addison-Wesley Pub. Co., 1973.

[13] G. P. McCormick and R. A. Tapia, "The Gradient Projection Method under Mild Differentiability Conditions," SIAM J. Control, vol. 10, no. 1, pp. 93-98, 1972.

[14] A. Miele, H. Y. Huang and J. C. Heideman, "Sequential Gradient-Restoration Algorithm for the Minimization of Constrained Functions - Ordinary and Conjugate Gradient Versions," JOTA, vol. 4, no. 4, pp. 213-243, 1969.

[15] A. Miele, A. V. Levy and E. E. Gragg, "Modifications and Extensions of the Conjugate Gradient - Restoration Algorithms for Mathematical Programming Problems," JOTA, vol. 7, no. 6, pp. 450-472, 1971.

[16] E. Polak, Computational Methods in Optimization, Academic Press, 1971.

[17] S. M. Robinson, "Extension of Newton's Method to Mixed Systems of Nonlinear Equations and Inequalities," Tech. Sum. Rept. no. 1161, Mathematical Research Center, University of Wisconsin, 1971.

[18] S. M. Robinson, "Bounds for Error in the Solution Set of a Perturbed Linear Program," Linear Algebra and Its Applications, vol. 6, pp. 69-81, 1973.

[19] J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming. Part I. Linear Constraints," J. of SIAM, vol. 8, pp. 181-217, 1960.

[20] J. B. Rosen, "The Gradient Projection Method for Nonlinear Programming. Part II. Nonlinear Constraints," J. of SIAM, vol. 9, no. 4,

pp. 514-532, 1961.

[21] P. Wolfe, "Methods for Nonlinear Constraints," in Nonlinear Programming, ed. by J. Abadie, North-Holland Pub. Co., Amsterdam, 1967.

[22] W. I. Zangwill, Nonlinear Programming: A Unified Approach, Prentice-Hall, Englewood Cliffs, New Jersey, 1969.

[23] G. Zoutendijk, Methods of Feasible Directions, Elsevier, Amsterdam, 1960.