A SECOND ORDER METHOD FOR THE GENERAL NONLINEAR PROGRAMMING PROBLEM

by

H. Mukai and E. Polak

Memorandum No. ERL-M562

14 August 1975

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

A Second Order Method for the General Nonlinear Programming Problem

H. Mukai and E. Polak

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

The work of the first author was supported by NSF RANN (National Science Foundation Research Applied to National Need) and JSEP Contract F44620-71-C-0087; the work of the second author was supported by the National Science Foundation grant ENG73-08214A01 and the U.S. Army Research Office Durham Contract DAHC04-73-C-0025.

1. Introduction

Although similar ideas were used in the study of second order conditions as far back as the 1930's (see e.g. Hestenes [11]), primal-dual methods, in their current form, are derived from more recent proposals by Hestenes [12], Powell [25], and somewhat later Haarhoff and Buys [10]. Specifically, in the case of problems of the form $\min\{f(x) \mid g(x) = 0\}$, they depend on an interesting property of the Lagrangian $f(x) + c\|g(x)\|^2 + \langle \lambda, g(x) \rangle$ of the equivalent problem $\min\{f(x) + c\|g(x)\|^2 \mid g(x) = 0\}$. Namely, for λ suitably chosen and c large enough the local minimizers of this Lagrangian are also local minimizers of the original problem. Because primal-dual methods reduce an equality constrained minimization problem to an unconstrained one, somewhat like penalty function methods, but without the accompanying ill conditioning of ordinary penalty function methods, they have attracted a great deal of attention (see Buys [5], Polyak and Tret'yakov [24], Miele et al. [17], [18], [19], [20], Tripathi and Narendra [30], Rupp [29], Bertsekas [2], [3], [4], Fletcher [6], [7], Fletcher and Lill [9], Martensson [16], and Mukai-Polak [21]). There are at present two types of primal-dual methods: Those that compute estimates of the multiplier λ discretely (e.g. as described by Hestenes [12] and Powell [25]), and those that use some continuous function $\lambda(x)$ for λ, as in Fletcher [6], and Mukai-Polak [21]. To avoid confusion, we shall refer to the latter as methods of multipliers and to the former as primal-dual method. Martensson [16] has established an important difference between primal-dual and multiplier type methods; viz. in multiplier methods a sufficiently large c ensures that a local minimizer of the original problem satisfies second order necessary conditions for a minimizer of the derived (augmented) Lagrangian, while in primal-dual type methods this is not always so. Thus, multiplier methods appear to have an advantage.

Although primal-dual and multiplier methods have also been proposed for problems of the form $\min\{f(x) \mid g(x) = 0,\ h(x) \leq 0\}$ (see Buys [5], Rockafellar [26], [27], [28], Arrow, Gould and Howe [1], Mangasarian [15], Wierzbicki [31], Fletcher [8], Lill [13]), none of these methods are entirely satisfactory, because they either fail to incorporate a scheme for automatically selecting a correct value for the penalty coefficient or they involve "inner" unconstrained minimization at each iteration, which is computationally quite costly. In this paper we present a quadratically convergent method which does not suffer from either of these two drawbacks. It is based on three elements: (i) the little known fact that (as is shown in the paper) the introduction of slack variables does not preserve Kuhn-Tucker points, but it does preserve points satisfying second order necessary conditions, (ii) an automatic scheme for selecting the penalty coefficient c in a multiplier method for problems with equality constraints, described in [21] and [23], and (iii) a new second order unconstrained minimization algorithm, described in [22], which permits us to "avoid" saddle and inflection points of the problem with slack variables. Our computational experience with this method is quite favorable.

2. Slack Variables and Convexified Lagrangians

Consider the following minimization problem:

$$\min\{f(x) \mid g(x) = 0,\ h(x) \leqq 0\} \tag{1}$$

where $f: \mathbb{R}^n \to \mathbb{R}^1$, $g: \mathbb{R}^n \to \mathbb{R}^m$ and $h: \mathbb{R}^n \to \mathbb{R}^p$, with $m \leq n$, are three times continuously differentiable and $h(x) \leqq 0$ is used to denote $h^j(x) \leq 0$, $j = 1,2,\dots,p$.

We begin by recalling a few standard results.

Definition 1: We shall say that $x^* \in \mathbb{R}^n$ is a feasible point if $g(x^*) = 0$ and $h(x^*) \leqq 0$, and we shall say that $x^* \in \mathbb{R}^n$ is a regular point if $\nabla g^j(x^*)$, $j = 1,2,\dots,m$, $\nabla h^i(x^*)$,[†] $i \in J(x^*) \triangleq \{j \mid h^i(x^*) = 0\}$ are linearly independent. ⌑

Note that as defined above, a regular point need not be a feasible point. Next, we need to reproduce the statements of second order conditions of optimality (see e.g. [14]). Let the Lagrangian ℓ: $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^1$ be defined by

$$\ell(x,\mu,\nu) = f(x) + \langle \mu, g(x) \rangle + \langle \nu, h(x) \rangle \quad (2)$$

with f, g, h as in (1). Then,

Lemma 1: Suppose that a regular feasible point x^* is a local minimizer for (1). Then there exist a $\mu^* \in \mathbb{R}^n$ and a $\nu^* \in \mathbb{R}^p$, $\nu^* \geqq 0$, such that

$$\frac{\partial \ell}{\partial x}(x^*,\mu^*,\nu^*) = 0 \quad (3)$$

$$\langle \nu^*, h(x^*) \rangle = 0 \quad (4)$$

and[††]

$$\frac{\partial^2 \ell(x^*,\mu^*,\nu^*)}{\partial x^2} \geq 0 \quad (5)$$

[†]We denote components of a vector by superscripts and we shall treat gradients as column vectors throughout: $\nabla g^j(x^*) = \frac{\partial g^j(x^*)^t}{\partial x}$ etc.

[††]We indicate the positive semidefiniteness of a matrix A by $A \geq 0$ and its positive definiteness by $A > 0$.

on the tangent subspace

$$T(x^*) \triangleq \{y \mid \frac{\partial g(x^*)}{\partial x} y = 0; \frac{\partial h^j(x^*)}{\partial x} h = 0, j \in J(x^*)\} \qquad (6)$$

Lemma 2: Suppose that x^* is a regular feasible point and that there exist a $\mu^* \in \mathbb{R}^m$ and a $\nu^* \in \mathbb{R}^p$, $\nu^* \geqq 0$ such that (3) and (4) are satisfied and $\frac{\partial^2 \ell}{\partial x^2}(x^*,\mu^*,\nu^*) > 0$ on the subspace

$$T'(x^*) \triangleq \{y \mid \frac{\partial g(x^*)}{\partial x} y = 0; \frac{\partial h^j(x^*)}{\partial x} h = 0, j \in J_1(x^*,\nu^*)\} \qquad (6)$$

with $J_1(x^*,\nu^*) \triangleq \{j \in J(x^*) \mid \nu^{*j} > 0\}$, then x^* is a strong local minimizer for (1). ¤

Definition 2: We shall say that a regular feasible point $x^* \in \mathbb{R}^n$ satisfies SONC† if it satisfies the conclusions in Lemma 1; i.e., for some μ^*, $\nu^* \geqq 0$, (3)-(6) holds. We shall say that a regular feasible point $x^* \in \mathbb{R}^n$ satisfies NSOSC if it satisfies the conditions in Lemma 2, and is nondegenerate in the sense that $T(x^*) = T'(x^*)$. ¤

Next, we turn to the use of slack variables. Let $\bar{f}$: $\mathbb{R}^{n+p} \to \mathbb{R}^1$, $\bar{g}$: $\mathbb{R}^{n+p} \to \mathbb{R}^{m+p}$ and $\bar{\ell}$: $\mathbb{R}^{n+p} \times \mathbb{R}^{m+p} \to \mathbb{R}^1$ be defined by

$$\bar{f}(z) = f(x) \qquad (7)$$

$$\bar{g}(z) = \begin{pmatrix} g(x) \\ h(x) + s(y) \end{pmatrix} \qquad (8)$$

$$\bar{\ell}(z,\lambda) = \bar{f}(z) + \langle \lambda, \bar{g}(z) \rangle \qquad (9)$$

†SONC stands for second order necessary conditions and NSOSC stands for nondegenerate second order sufficiency conditions.

where $z = (x,y)$ ($x \in \mathbb{R}^n$, $y \in \mathbb{R}^p$) and $s: \mathbb{R}^p \to \mathbb{R}^p$ is defined by $s^i(y) = (y^i)^2$, $i = 1,2,\ldots,p$. Now consider the <u>derived problem</u>.

$$\min\{\bar{f}(z) \mid \bar{g}(z) = 0\} \tag{10}$$

First we note that Definition 1 and Definition 2 as well as Lemmas 1 and 2 apply to Problem (10) as well (replace g by $\bar{g}$ and remove h from (1)). Hence we shall use them in conjunction with both of these problems. Next, we state an obvious result.

<u>Proposition 1</u>: (i) If $z^* = (x^*,y^*)$ is, respectively, feasible, regular, or optimal for Problem (10), then x^* is, respectively, feasible, regular, or optimal for Problem (1). (ii) If x^* is, respectively, feasible, regular, or optimal for Problem (1), then $z^* = (x^*,y^*)$, with $y^{*j} = \sqrt{|h^j(x^*)|}$, $j = 1,2,\ldots,p$, is, respectively, feasible, regular, or optimal for Problem (10). ¤

Now, suppose that x^* is a feasible Kuhn-Tucker point for (1), i.e. for some multipliers μ^* and $\nu^* \geqq 0$, $\nabla_x \ell(x^*,\mu^*,\nu^*) = 0$, and $\langle \nu^*,h(x^*) \rangle = 0$. Then, setting $\lambda^* = (\mu^*,\nu^*)$, $y^{*j} = \sqrt{-h^j(x^*)}$, $j = 1,2,\ldots,p$, and $z^* = (x^*,y^*)$, we get $\nabla_z \bar{\ell}(z^*,\lambda^*) = 0$. Next, suppose that $z^* = (x^*,y^*)$ is a feasible point satisfying the Lagrange condition for (10), i.e., for some multiplier $\lambda^* = (\mu^*,\nu^*)$, $\nabla_z \bar{\ell}(z^*,\lambda^*) = 0$. It is easy to see that this implies that $\langle \nu^*,h(x^*) \rangle = 0$, but we cannot conclude that $\nu^* \geqq 0$. Hence, x^* is not necessarily a Kuhn-Tucker point for (1).

However, the following results do hold.

<u>Lemma 3</u>: A point x^* is a regular feasible point satisfying SONC for problem (1) if and only if $z^* = (x^*,y^*)$, with $y^{*j} = \sqrt{-h^j(x^*)}$, $j = 1,2,\ldots,p$, is a regular feasible point satisfying SONC for problem (10).

Proof: First, the fact that x^* is a regular feasible point for (1) if and only if z^* (as defined) is a regular feasible point for (10) was established in Proposition 1.

Next, suppose that a regular feasible point x^* satisfies SONC for (1), with multipliers $\mu^*, \nu^* \geq 0$. Then, setting $y^{*j} = \sqrt{-h^j(x^*)}$, $j = 1,2,..,p$, $z^* = (x^*,y^*)$ and $\lambda^* = (\mu^*,\nu^*)$, we find (cf (12) below) that $\partial\bar{\ell}(z^*,\lambda^*)/\partial z = 0$, since $\nu^{*j}y^{*j} = 0$, $j = 1,2,...,p$, and that

$$\frac{\partial^2\bar{\ell}(z^*,\lambda^*)}{\partial z^2} = \left(\begin{array}{c|c} \frac{\partial^2 \ell(x^*,\mu^*,\nu^*)}{\partial x^2} & 0 \\ \hline 0 & 2N^* \end{array}\right), \tag{11}$$

where $N^* = \text{diag}\,(\nu^{*1}, \nu^{*1},..., \nu^{*p})$, is positive semidefinite on $\bar{T}(z^*) = \{\zeta | \frac{\partial\bar{g}(z^*)}{\partial z}\zeta = 0\}$.

We now turn to the more difficult part of the proof. Suppose that $z^* = (x^*,y^*)$ is a regular feasible point satisfying SONC for (10), with a multiplier $\lambda^* = (\mu^*,\nu^*)$. As we have already established, x^* is a regular feasible point for (1). Next,

$$\nabla_z\bar{\ell}(z^*,\lambda^*) = \begin{pmatrix} \nabla f(x^*) \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{\partial g(x^*)}{\partial x}^T & \frac{\partial h(x^*)}{\partial x}^T \\ 0 & \frac{\partial s(y^*)}{\partial y} \end{pmatrix}\begin{pmatrix} \mu^* \\ \nu^* \end{pmatrix} = 0 \tag{12}$$

Hence we obtain that $\nabla_x \ell(x^*,\mu^*,\nu^*) = 0$ and that $\langle y^*,\nu^* \rangle = 0$, and therefore, that $\langle \nu^*,h(x^*) \rangle = 0$. Also, the matrix

$$\frac{\partial^2\bar{\ell}(z^*,\lambda^*)}{\partial z^2} = \left(\begin{array}{c|c} \frac{\partial^2 \ell(x^*,\mu^*,\nu^*)}{\partial x^2} & 0 \\ \hline 0 & 2N^* \end{array}\right). \tag{13}$$

is positive semidefinite on $\bar{T}(z^*) = \{\zeta | \frac{\partial \bar{g}(z^*)}{\partial z} \zeta = 0\}$. Since N^* is diagonal and, with $\zeta = (\xi,\eta)$, since the vectors $(0,\ldots0,\eta^j,0\ldots0)^T \in \bar{T}(z^*)$ for all j such that $\nu^{*j} \neq 0$, we must have $\nu^* \geq 0$. Finally, setting $\zeta = (\xi,\eta)$, we see that $\frac{\partial \bar{g}(z^*)}{\partial z} \zeta = 0$, implies that

$$\frac{\partial g(x^*)}{\partial x} \xi = 0 \tag{14a}$$

$$\frac{\partial h(x^*)}{\partial x} \xi + \frac{\partial s(y^*)}{\partial y} \eta = 0 \tag{14b}$$

Now (14b) implies that $(\partial h^j(x^*)/\partial x)\xi = 0$ for all $j \in J(x^*)$, and therefore (14a,b) imply $\xi \in T(x^*)$, for all $\zeta \in \bar{T}(z^*)$, and hence $\partial^2 \ell(x^*,\mu^*,\nu^*)/\partial x^* \geq 0$ on $T(x^*)$, so that x^* satisfies SONC for (1). This concludes our proof. ¤

The following result is obvious in the light of the arguments used to prove Lemma 3.

Lemma 4: A point x^* is a regular feasible point satisfying NSOSC for problem (1), if and only if $z^* = (x^*,y^*)$, with $y^{*j} = \sqrt{-h^j(x^*)}$, $j = 1,2,..,p$, is a regular, feasible point satisfying NSOSC† for Problem (10). ¤

This concludes our investigation of the relationships between Problems (1) and (10).

3. The Modified Lagrangian.

As is customary in primal-dual methods, we substitute for the Problem (10), the family of equivalent problems P_c: $\min\{\bar{f}(z) + \frac{1}{2} c\|\bar{g}(z)\|^2 | \bar{g}(z) = 0\}$, where $c \geq 0$, and whose Lagrangian is

---

†The nondegeneracy part of NSOSC is obviously satisfied trivially for Problem (10), since it has no inequality constraints.

$$L_c(z,\lambda) = \bar{\ell}(z,\lambda) + \frac{1}{2} c\|\bar{g}(z)\|^2 \tag{15}$$

We recall [14] that if $\hat{z}$ is a regular optimal point for (10) and $\hat{\lambda}$ is the corresponding Lagrange multiplier, then $\hat{z}$ is regular and optimal for P_c and $\hat{\lambda}$ is the corresponding multiplier, for any $c \geq 0$. Furthermore while $\frac{\partial^2 \bar{\ell}(\hat{z},\hat{\lambda})}{\partial z^2}$ need not be positive definite, $\frac{\partial^2 L_c(\hat{z},\hat{\lambda})}{\partial z^2} > 0$ for all c sufficiently large. This convexifying property, as we shall later see, can be utilized both to ensure satisfactory convergence and to obtain quadrate convergence of an algorithm. First, however, we make the following

Assumption 1: All the feasible points for Problem 1 are regular. ¤

From now on, we shall always assume that Assumption 1 is satisfied.

We now define $\mathcal{R} \subset \mathbb{R}^{n+p}$ to be the set of regular points for the Problem (10), i.e. $\mathcal{R} = \{z| \frac{\partial \bar{g}(z)}{\partial z}$ has maximum rank$\}$. It is clear that $\mathcal{R}$ is an open set containing all the feasible points for the Problem (10) (see Proposition 1).

As was also done in [7], [16] and [21], for all $z \in \mathcal{R}$ we shall make λ in (15) a well defined function of z, as follows:

$$\lambda(z) = \arg\min\{\|\nabla_z \bar{\ell}(z,\lambda)\|^2 | \lambda \in \mathbb{R}^{n+p}\}$$

$$= -\left(\frac{\partial \bar{g}(z)}{\partial z}\frac{\partial \bar{g}(z)}{\partial z}^T\right)^{-1} \frac{\partial \bar{g}(z)}{\partial z} \nabla \bar{f}(z) \tag{16}$$

Proposition 2: The function $\lambda: \mathcal{R} \to \mathbb{R}^{m+p}$ is twice continuously differentiable and for all $z \in \mathcal{R}$,

$$\frac{\partial \lambda(z)}{\partial z} = -\left(\frac{\partial \bar{g}(z)}{\partial z}\frac{\partial \bar{g}(z)}{\partial z}^T\right)^{-1} \left[\frac{\partial \bar{g}(z)}{\partial z}\frac{\partial^2 \bar{\ell}(z,\lambda(z))}{\partial z^2} + \right.$$

$$\sum_{j=1}^{m+p} e_j \nabla_z \bar{\ell}(z,\lambda(z))^T \frac{\partial^2 \bar{g}(z)}{\partial z^2}\Bigg] \tag{17}$$

where e_j is the jth column of the $(m+p)\times(m+p)$ identity matrix.

<u>Proof</u>: By assumption, $\bar{f}$ and $\bar{g}$ are three times continuously differentiable. It therefore follows from (16) that λ is twice continuously differentiable. To obtain (17), we note that

$$\frac{\partial \bar{g}(z)}{\partial z} \nabla_z \bar{\ell}(z,\lambda(z)) = \sum_{j=1}^{m+p} e_j \frac{\partial \bar{g}^j(z)}{\partial z} \nabla_z \bar{\ell}(z,\lambda(z)) = 0 \tag{18}$$

Differentiating the right hand side of (18) and making use of the fact that $\frac{\partial^2 \bar{\ell}(z,\lambda(z))}{\partial \lambda^2} = \frac{\partial \bar{g}(z)}{\partial z}^T$, we obtain (17). ¤

As was also done in [7], [16] and [21], with λ defined by (16), for any $c \geq 0$, we define ψ_c: $\mathbb{R}^n \to \mathbb{R}^1$ by $\psi_c(z) = L_c(z,\lambda(z))$, i.e.,

$$\psi_c(z) \triangleq \bar{\ell}(z,\lambda(z)) + \frac{1}{2} c\|\bar{g}(z)\|^2 \tag{19}$$

We note that

$$\nabla\psi_c(z) = \nabla_z \bar{\ell}(z,\lambda(z)) + \frac{\partial \lambda(z)}{\partial z}^T \bar{g}(z) + c \frac{\partial \bar{g}(z)}{\partial z}^T \bar{g}(z) \tag{20}$$

$$\frac{\partial^2 \psi_c(z)}{\partial z^2} = \frac{\partial^2 \bar{\ell}(z,\lambda(z))}{\partial z^2} + \frac{\partial \bar{g}(z)}{\partial z}^T \frac{\partial \lambda(z)}{\partial z} + \frac{\partial \lambda(z)}{\partial z}^T \frac{\partial \bar{g}(z)}{\partial z}$$

$$+ c \frac{\partial \bar{g}(z)}{\partial z}^T \frac{\partial \bar{g}(z)}{\partial z} + \sum_{j=1}^{m+p} \bar{g}^j(z) \left[\frac{\partial^2 \lambda(z)}{\partial z^2} + c \frac{\partial^2 \bar{g}^j(z)}{\partial z^2}\right] \tag{21}$$

Finally, we establish a number of relationships between Problem (10) and the family of unconstrained problems, parametrized by $c \geq 0$,

$$\min\{\psi_c(z) \mid z \in \mathcal{R}\} \tag{22}$$

The following result is obvious in view of (16) and (20).

Proposition 3: If $z^* \in \mathbb{R}^{n+p}$ is a regular feasible point for Problem (10) satisfying, for some $\lambda^*, \nabla_z \bar{\ell}(z^*,\lambda^*) = 0$, then $\lambda^* = \lambda(z^*)$ and $\nabla\psi_c(z^*) = 0$ for all $c \geq 0$. ¤

Proposition 4: If $z^* \in \mathbb{R}^{n+p}$ is a regular feasible point for (10) satisfying, respectively, SONC or NSOSC, with multiplier λ^*, then $\lambda^* = \lambda(z^*)$ and there exists a $c^* \geq 0$ such that z^* satisfies, respectively, SONC or NSOSC for Problem (22), for all $c \geq c^*$.

Proof: Since by Proposition 3, $\lambda^* = \lambda(z^*)$ and $\nabla\psi_c(z^*) = 0$, we only need to show that there exists a c^* such that $\frac{\partial^2\psi_c(z^*)}{\partial z^*} \geq 0$ (> 0, respectively) for all $c \geq c^*$. Thus, since $\bar{g}(z^*) = 0$, we need to show that

$$\frac{\partial^2\psi_c(z^*)}{\partial z^2} = \frac{\partial^2\bar{\ell}(z^*,\lambda(z^*))}{\partial z^2} + \frac{\partial\bar{g}(z^*)^T}{\partial z}\frac{\partial\lambda(z^*)}{\partial z}$$

$$+ \frac{\partial\lambda(z^*)^T}{\partial z}\frac{\partial\bar{g}(z^*)}{\partial z} + c\,\frac{\partial\bar{g}(z^*)^T}{\partial z}\frac{\partial\bar{g}(z^*)}{\partial z} \geq 0 \ (> 0) \tag{23}$$

for all c sufficiently large. Since this result has already been established by Martensson [16], we are done. ¤

Proposition 5: For every compact subset $\mathcal{S} \subset \mathcal{R}$, there exists a $c_s \in \mathbb{R}$ such that for all $c \geq c_s$, if $z^* \in \mathcal{S}$ satisfies, respectively, SONC, or NSOSC, for Problem (22), then z^* is a regular feasible point for (10), satisfying, for (10), respectively, SONC, or NSOSC.

Proof: Let $\mathcal{S}$ be a compact subset of $\mathcal{R}$. Since all the matrices, below, are continuous, there exists a $c_s \geq 0$ such that for all $z \in \mathcal{S}$, for all $c \geq c_s$,

$$\det\left[cI + \left(\frac{\partial\bar{g}(z)}{\partial z}\frac{\partial\bar{g}(z)}{\partial z}^T\right)^{-1}\frac{\partial\bar{g}(z)}{\partial z}\frac{\partial\lambda(z)}{\partial z}^T\right] \neq 0 \tag{24}$$

Now suppose that $c \geq c_s$ and $z^* \in \mathcal{S}$ is such that $\nabla\psi_c(z^*) = 0$. Then, since $\frac{\partial\bar{g}(z)}{\partial z}\nabla_z\bar{\ell}(z,\lambda(z)) = 0$ for all $z \in \mathcal{R}$, it follows from (20) that

$$0 = \frac{\partial\bar{g}(z^*)}{\partial z}\nabla\psi_c(z^*) = \frac{\partial\bar{g}(z^*)}{\partial z}\left[\frac{\partial\lambda(z^*)}{\partial z}^T\bar{g}(z^*) + c\,\frac{\partial\bar{g}(z^*)}{\partial z}^T\bar{g}(z^*)\right] \tag{25}$$

Hence,

$$\left[cI + \left(\frac{\partial\bar{g}(z^*)}{\partial z}\frac{\partial\bar{g}(z^*)}{\partial z}^T\right)^{-1}\frac{\partial\bar{g}(z^*)}{\partial z}\frac{\partial\lambda(z^*)}{\partial z}^T\right]\bar{g}(z^*) = 0 \tag{26}$$

It now follows from (24) that $\bar{g}(z^*) = 0$ and hence, (20) implies that $\nabla_z\bar{\ell}(z^*,\lambda(z^*)) = 0$. Finally, suppose that $\frac{\partial^2\psi_c(z^*)}{\partial z^2} \geq 0$ (> 0). Then from (23), we conclude that $\frac{\partial^2\bar{\ell}(z^*,\lambda(z^*))}{\partial z^2} \geq 0$ (> 0) on $\bar{T}(z^*) = \{\zeta \mid \frac{\partial\bar{g}(z^*)}{\partial z}\zeta = 0\}$. This completes our proof. ¤

Lemmas 3 and 4 enable us to translate the above results into a relationship between Problems (1) and (22), as follows.

Theorem 1: (i) If $x^* \in \mathbb{R}^n$ is a feasible, regular point satisfying, respectively, SONC, or NSOSC for Problem (1), then $z^* = (x^*,y^*)$, with $y^{*j} = \sqrt{-h^j(x^*)}$, $j = 1,2,..,p$, is in $\mathcal{R}$ and there exists a $c^* \geq 0$ such that z^* satisfies, respectively, SONC, or NSOSC, for Problem (22) for all $c \geq c^*$.

(ii) For every compact subset $\mathcal{S} \subset \mathcal{R}$ there exists a $c_s \geq 0$ such that for all $c \geq c_s$, if a $z^* = (x^*, y^*) \in \mathcal{S}$ satisfies, respectively, SONC, or NSOSC, for Problem (22), then x^* is a feasible regular point satisfying, respectively, SONC, or NSOSC, for Problem (1). ¤

The above result shows that, provided we succeed in producing c large enough, we can obtain a solution to (1) by solving (22). An algorithm which achieves this will now be described.

4. The Algorithm:

Our algorithm is based on an Algorithm Model, first presented in [23]. Let $\{c_j\}_{j=0}^{\infty}$ be any strictly monotonically increasing sequence such that $c_j > 0$ and $c_j \to \infty$ as $j \to \infty$. Let $\theta_j(\cdot) \triangleq \psi_{c_j}(\cdot)$; let Δ be the set of all $z = (x,y)$, $\subset \mathbb{R}^{n+p}$ such that the x are feasible points for Problem (1) satisfying SONC and $y^j = \sqrt{-h^j(x)}$, $j = 1,2,..,p$; and let Δ_j, $j = 0,1,2,\ldots$, be the set of points in $\mathbb{R}^{n+p}$ satisfying SONC for Problem (22) with $c = c_j$. The Algorithm Model below makes use of a sequence of testing functions t_j: $\mathbb{R}^{n+p} \to \mathbb{R}^1$ and of iteration maps A_j: $\mathbb{R}^{n+p} \to \mathbb{R}^{n+p}$.

Algorithm Model

Data: $z_0 \in \mathbb{R}^{n+p}$.

Step 0: Set $i = 0$, $j = 0$.

Step 1: If $t_j(z_i) \leq 0$, go to step 2; else go to step 4.

Step 2: Compute $\zeta = A_j(z_i)$.

Step 3: If $\theta_j(\zeta) < \theta_j(z_i)$, set $z_{i+1} = \zeta$, $i = i+1$ and go to step 1; else stop.

Step 4: Set $w_j = z_i$, set $j = j+1$ and go to step 1. ¤

We find in [23] the following result.

Theorem 2: (i) Suppose that for each j, $j = 0, 1, 2, \ldots$, and any $z \notin \Delta_j$, there exist an $\varepsilon(z) > 0$ and a $\delta(z) < 0$ such that for all z' satisfying $\|z'-z\| \leq \varepsilon(z)$ and $z'' = A_j(z')$,

$$\theta_j(z'') - \theta_j(z') \leq \delta(z) \tag{27}$$

(ii) The functions $t_j(\cdot)$ are continuous for $j = 0,1,2,\ldots$.

(iii) For $j = 0,1,2,\ldots$, $\{z \in \Delta_j \mid t_j(z) \leq 0\} \subset \Delta$.

(iv) For every $z^* \in \mathcal{R}$ there exists a j^* and an $\varepsilon^* > 0$ such that $t_j(z) \leq 0$ for all $j \geq j^*$ for all z such that $\|z-z^*\| \leq \varepsilon^*$.

(v) The sequence $\{z_i\}$ constructed by the Algorithm Model is contained in a closed set $Q \subset \mathcal{R}$.

Under these assumptions, (i) if the algorithm model constructs a finite sequence $\{w_j\}$ and $\{z_i\}$ is infinite, then every accumulation point of $\{z_i\}$ is in Δ; (i) if $\{z_i\}$ is finite, then the last element of $\{z_i\}$ is an Δ; (iii) if $\{w_j\}$ is infinite, then $\{w_j\}$ has no accumulation points.

Thus, to construct an algorithm, we must invent a sequence of testing function $\{t_j(\cdot)\}$ which can then be used in conjunction with any convergent[†] iteration function A_j for solving Problem (22) with $c = c_j$. Although the choice is not unique, we propose to use $t_j(\cdot)$ defined as follows (cf. [21]):

$$t_j(z) \triangleq -\left\langle \frac{\partial\bar{g}(z)}{\partial z}^T \left(\frac{\partial\bar{g}(z)}{\partial z}\frac{\partial\bar{g}(z)}{\partial z}^T\right)^{-1} \bar{g}(z),\ \nabla\theta_j(z)\right\rangle + \gamma\|\bar{g}(z)\|^2, \tag{28}$$

[†]That is, satisfying condition (i) of Theorem 2.

where $\gamma > 0$ is a preselected constant. Thus, $t_j(z)$ tests the angle between $\nabla\theta_j(z)$ and the Newton direction for solving $\bar{g}(z) = 0$. Obviously, the $t_j(\cdot)$ are continuous, so hypothesis (ii) of Theorem 2 is satisfied. Next, suppose that $z \in \Delta_j$ and $t_j(z) \leq 0$. Then $\nabla\theta_j(z) = 0$ and hence, from (28) $\bar{g}(z) = 0$, i.e. z is a regular feasible point for (10). Furthermore, from the arguments used in the proof of Proposition 4, we conclude that z satisfies SONC for (10). In view of the established relationships between (1) and (10), we now conclude that $z \in \Delta$, i.e., with $t_j(\cdot)$ defined as in (28), assumption (iii) of Theorem 2 holds. Next, expanding (28), since $\frac{\partial\bar{g}(z)}{\partial z} \nabla_z \bar{\ell}(z,\lambda(z)) = 0$, we obtain

$$t_j(z) = \gamma\|\bar{g}(z)\|^2 - \langle \left(\frac{\partial\bar{g}(z)}{\partial z} \frac{\partial\bar{g}(z)}{\partial z}^T\right)^{-1}, \frac{\partial\bar{g}(z)}{\partial z} \nabla\psi_{c_j}(z) \rangle$$

$$= \langle g(z), \left[-(c_j-\gamma)I + \left(\frac{\partial\bar{g}(z)}{\partial z} \frac{\partial\bar{g}(z)}{\partial z}^T\right)^{-1} \frac{\partial\bar{g}(z)}{\partial z} \frac{\partial\lambda(z)}{\partial z}^T\right] g(z) \rangle \qquad (29)$$

and, hence, clearly, given any $z^* \in \mathcal{R}$, there exists a $j^* \geq 0$ and an $\varepsilon^* > 0$ such that $t_j(z) \leq 0$ for all $j \geq j^*$ and all $z \in \mathcal{R}$ such that $\|z-z^*\| \leq \varepsilon^*$. Thus, the functions $t_j(\cdot)$ defined in (28), satisfy all the appropriate assumptions of Theorem 2. For the maps A_j we propose to use the iteration function of the extended Newton Method developed in [22]. It can be concluded from the results in [22] that the A_j as will be defined below, and the θ_j and Δ_j satisfy assumption (i) of Theorem 2. Consequently, the conclusions of Theorem 2 apply to the algorithm below.

Algorithm:

Parameters: $\alpha \in (0,1)$, $\beta \in (0,1)$, $0 < \varepsilon_0 << 1$, a sequence $\{c_j\}_{j=0}^{\infty}$, ($c_{j+1} > c_j \ \forall j$, $c_j \to \infty$ as $j \to \infty$), and an initial guess z_0.

Step 0: Set i=0, j=0.

Step 1: ($t_j(\cdot)$ is defined as in (28).) If $t_j(z_i) \leq 0$, go to step 2; else go to step 11.

Comment: The map A_j is defined by steps 2-10, below.

Step 2: Solve the following direction finding problem for a minimizer v_i:

$$\phi_j(z_i) \triangleq \min\{ \tfrac{1}{2}\langle v, H_j(z_i)v \rangle \mid \langle \nabla\theta_j(z_i), v \rangle \leq 0,\ \|v\| \leq 1\} \tag{30}$$

where

$$H_j(z_i) \triangleq \frac{\partial^2 \bar{\ell}(z_i,\lambda(z_i))}{\partial z^2} + \frac{\partial \bar{g}(z_i)^T}{\partial z}\frac{\partial \lambda(z_i)}{\partial z} + \frac{\partial \lambda(z_i)^T}{\partial z}\frac{\partial \bar{g}(z_i)}{\partial z}$$
$$+ c_j \left[\frac{\partial \bar{g}(z_i)^T}{\partial z}\frac{\partial \bar{g}(z_i)}{\partial z} + \sum_{k=1}^{m+p} \bar{g}^k(z_i) \frac{\partial^2 \bar{g}^k(z_i)}{\partial z^2} \right] \tag{31}$$

Step 3: If $\phi_j(z_i) < 0$, go to step 7; else go to step 4.

Step 4: If $\nabla\theta_j(z_i) = 0$, stop; else go to step 5.

Step 5: If $|\det H_j(z_i)| < \varepsilon_0$ go to step 7; else, go to step 6.

Step 6: Set $u_i = -H_j(z_i)^{-1}\nabla\theta_j(z_i)$ and go to step 8.

Step 7: Set $u_i = -\nabla\theta_j(z_i) + v_i$.

Step 8: If $\langle u_i, H_j(z_i)u_i \rangle \leq 0$, set $\lambda_0 = 1$; else set $\lambda_0 = \beta^{k_i}$ where $k_i \geq 0$ is the smallest integer satisfying

$$\beta^{k_i} \leq -\langle \nabla\theta_j(z_i), u_i \rangle / \langle u_i, H_j(z_i)u_i \rangle \tag{32}$$

Step 9: Compute the smallest integer $\ell_i \geq 0$ such that

$$\theta_j(z_i + \lambda_0\beta^{\ell_i}u_i) - \theta_j(z_i) \leq \alpha\left[\lambda_0\beta^{\ell_i}\langle \nabla\theta_j(z_i),u_i\rangle + \frac{1}{2}\left(\lambda_0\beta^{\ell_i}\right)^2\langle u_i,H_j(z_i)u_i\rangle\right] \tag{33}$$

Step 10: Set $z_{i+1} = z_i + \lambda_0\beta^{\ell_i}u_i$, set $i = i+1$ and go to step 1.

Step 11: Set $w_{j+1} = z_i$, set $j = j+1$ and go to step 1. ¤

The following theorem follows immediately from Theorem 2.

Theorem 3: Suppose that the Algorithm does not jam up in step 2, i.e., the entire sequence it has constructed is in $\mathcal{R}$. Under this assumption, (a) (i) if $\{w_j\}$ is finite and $\{z_i\}$ is infinite, then every accumulation point of $\{z_i\}$ satisfies SONC for Problem (1); (ii) if $\{z_i\}$ is finite, then its last element satisfies SONC for Problem (1); (iii) if $\{w_j\}$ is infinite, then $\{w_j\}$ has no accumulation points.

(b) if $\{w_j\}$ is finite, $\{z_i\}$ is infinite and has an accumulation point z^* satisfying NSOSC for Problem (1), then $z_i \to z^*$ as $i \to \infty$, with $\|z_{i+1}-z^*\|/\|z_i-z^*\| \to 0$ as $i \to \infty$. Furthermore, if the functions f, g and h in (1) are three times Lipschitz continuously differentiable at z^*, then there exists an $M > 0$ and an i_0 such that

$$\|z_{i+1}-z^*\| \leq M\|z_i-z^*\|^2 \quad \text{for all } i \geq i_0 \tag{34}$$

Conclusion

All of the theoretical results in this paper are predicated upon the constructed sequences remaining within the regularity set $\mathcal{R}$. Thus, just like a number of other very successful methods, such as Newton's method, the Variable Metric method and conjugate directions methods, to mention a few, it may fail from time to time on a specific problem. However, our limited computational experience indicates that this will happen rather infrequently and that the excellent properties of our method, in the cases where it does not fail, certainly justify its use.

References

[1] K.J. Arrow, F.J. Gould and S.M. Howe, "A general saddle point result for constrained optimization", Mathematical Programming 5 (1973) 225-234.

[2] D.P. Bertsekas, "On penalty and multiplier methods, Dept. of Engrg.-Economic Syst. Working Paper, Stanford University (1973).

[3] ______________, "On the method of multipliers for convex programming", Dept. of Engrg.-Economic Syst. Working Paper, Stanford University (1973).

[4] ______________, "Convergence rate of penalty and multiplier methods, Proc. 1973 IEEE Conf. on Decision and Control, San Diego, Calif. (1973) 260-264.

[5] J.D. Buys, "Dual algorithms for constrained optimization", Ph.D. thesis, University of Leiden, the Netherlands (1972).

[6] R. Fletcher, "A class of methods for nonlinear programming with termination and convergence properties, in: <u>Integer and Nonlinear Programming</u>, Ed. J. Abadie (North-Holland, Amsterdam, 1970).

[7] ___________, "A class of methods for nonlinear programming. III: Rate of convergence", in <u>Numerical Methods for Nonlinear Optimization</u>, Ed. F.A. Lootsma (Academic Press, New York, 1973).

[8] ___________, "An exact penalty function for nonlinear programming with inequalities", <u>Mathematical Programming</u> 5(1973) 129-150.

[9] R. Fletcher and S.A. Lill, "A class of methods for nonlinear programming. II: Computational experience", in: <u>Nonlinear Programming</u>, Ed. J.B. Rosen, O.L. Mangasarian and K. Ritter (Academic Press, New York, 1971).

[10] P.C. Haarhoff and J.D. Buys, "A new method for the optimization of a nonlinear function subject to nonlinear constraints", Computer Journal, 13(1970), 178-184.

[11] M.R. Hestenes, "The Weierstrass E-function in the calculus of variations", Trans of AMS 60(1946) 51-71.

[12] ____________, "Multiplier and gradient methods", J. Optimization Theory and Applications. 4(1969) 303-320.

[13] S.A. Lill, "Generalization of an exact method for solving equality constrained problems to deal with inequality constraints", in: Numerical Methods for Nonlinear Optimization, Ed. F.A. Lootsma (Academic Press, New York, 1972).

[14] D.G. Luenberger, Introduction to Linear and Nonlinear Programming (Addison-Wesley, Reading, Mass., 1973).

[15] O.L. Mangasarian, "Unconstrained Lagrangians in nonlinear programming", SIAM J. on Control 13(1975) 772-791.

[16] K. Martensson, "A new approach to constrained function optimization", J. of Optimization Theory and Applications 12(1973) 531-554.

[17] A. Miele, P.E. Moseley, A.V. Levy and G.M. Coggins, "On the method of multipliers for mathematical programming problems", J. Optimization Theory and Applications 10(1972) 1-33.

[18] A. Miele, P.E. Moseley and E.E. Cragg, "A modification of the method of multipliers for mathematical programming problems," in: Techniques of Optimization, Ed. A.V. Balakrishnan (Academic Press, New York, 1972).

[19] A. Miele, E.E. Cragg, R.R. Iyer and A.V. Levy, "Use of the augmented penalty function in mathematical programming problems, Part I", J. Optimization Theory and Applications 8(1971) 115-130.

[20] A. Miele, E.E. Cragg and A.V. Levy, "Use of the augmented penalty function in mathematical programming problems, Part II, J. Optimization Theory and Applications 8(1971) 131-153.

[21] H. Mukai and E. Polak, "A quadratically convergent primal-dual algorithm with global convergence properties for solving optimization problems with equality constraints", Electronics Research Lab. Memo. No. 455 University of California, Berkeley (1974); to appear in Mathematical Programming.

[22] H. Mukai and E. Polak, "A second order method for unconstrained Optimization," Electronics Research Lab. Memo. No. M561 University of California, Berkeley (1975).

[23] E. Polak, "On the global stabilization of locally convergent algorithms for optimization and root finding", University of California, Berkeley (1974) also, Proc. 1975 IFAC Congress, Boston Mass., Aug. 25-29, 1975.

[24] V.T. Polyak and N.V. Tret'yakov, "The method of penalty estimates for conditional extremum problems", U.S.S.R. Comp. Math. and Math. Physics 13(1974) 42-58.

[25] M.J.D. Powell, "A method for nonlinear constraints in minimization problems", in: Optimization, Ed. R. Fletcher (Academic Press, New York, 1969) 283-298.

[26] R.R. Rockafellar, "A dual approach to solving nonlinear programming problems by unconstrained optimization", Mathematical Programming 5(1973) 354-373.

[27] ________________, "The multiplier method of Hestenes and Powell applied to convex programming", J. Optimization Theory and Applications 12(1973) 555-562.

[28] R.R. Rockafellar, "Augmented Lagrange multiplier functions and duality in nonconvex programming", SIAM J. on Control 12(1974) 268-285.

[29] R.D. Rupp, "On the combination of the multiplier method of Hestenes and Powell with Newton's method", J. Optimization Theory and Applications 15(1975) 167-187.

[30] S.S. Tripathi and K.S. Narendra, "Constrained optimization problems using multiplier methods", J. Optimization Theory and Applications 9(1972) 52-70.

[31] A.P. Wierzbicki, "A penalty function shifting method in constrained static optimization and its convergence properties", Archiwum Automatyki i Telemechaniki 16(1971) 395-416.