COMPUTER SCIENCES DEPARTMENT

University of Wisconsin-Madison

UNCONSTRAINED METHODS IN NONLINEAR PROGRAMMING[1]

by

O. L. Mangasarian

Computer Sciences Technical Report #253

April 1975

The University of Wisconsin
Computer Sciences Department
1210 West Dayton Street
Madison, Wisconsin 53706

Received: April 25, 1975

UNCONSTRAINED METHODS IN NONLINEAR PROGRAMMING[1]

by

O. L. Mangasarian

[1] Invited talk at the AMS-SIAM Symposium on Nonlinear Programming, New York, March 23-24, 1975.

UNCONSTRAINED METHODS IN NONLINEAR PROGRAMMING

O. L. Mangasarian[1]

ABSTRACT

Optimality conditions and algorithms that are free from inequalities are given for inequality constrained optimization problems. It is shown that stationary points of nonlinear programs which involve n variables and m constraints can be obtained by solving a system of $n + m$ equations in $n + m$ unknowns. Algorithms are presented which exploit the structure of these equations in such a way that, at each iteration, only n equations in n unknowns are solved or an unconstrained function of n variables is minimized. It is shown that these algorithms are parametrically superlinearly convergent, that is the error at the ith iteration is proportional to α^{-i} where α is a finite penalty parameter that can be made large.

A fundamental problem in nonlinear programming is to

$$\begin{array}{ll} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & g(x) \leq 0 \end{array} \tag{1}$$

where $f:R^n \to R$ and $g:R^n \to R^m$. That is find an $\bar{x}$ in R^n such that $g(\bar{x}) \leq 0$ and $f(\bar{x}) \leq f(x)$ for all x satisfying $g(x) \leq 0$. In general one cannot even compute numerically a solution to this problem, but, assuming that f and g are differentiable on R^n, one looks for x in R^n such that together with a Lagrange multiplier u in R^m the following Karush-Kuhn-Tucker necessary optimality conditions [17,18,20] are satisfied

$$\nabla f(x) + u\nabla g(x) = 0,\ ug(x) = 0,\ g(x) \leq 0,\ u \geq 0 \tag{2}$$

AMS(MOS) subject classifications (1970). Primary 90C30; Secondary 65H10.

[1] Research supported by National Science Foundation under grants GJ35292 and DCR74-20584.

where $\nabla f(x)$ is the $1 \times n$ gradient vector[1] of f at x, $\nabla g(x)$ is the $m \times n$ Jacobian of g at x, and $ug(x)$ denotes the scalar product $\sum_{j=1}^{m} u_j g_j(x)$. In terms of the classical Lagrangian

$$L_0(x,u) = f(x) + ug(x) \tag{3}$$

the Karush-Kuhn-Tucker conditions (2) are equivalent to

$$\nabla_1 L_0(x,u) = 0, \ \nabla_2 L_0(x,u)u = 0, \ \nabla_2 L_0(x,u) \leqq 0, \ u \geqq 0 \tag{4}$$

where $\nabla_1 L(x,u)$ and $\nabla_2 L(x,u)$ denote gradients with respect to the first argument x and the second argument u, respectively.

Both the original minimization problem (1) and the optimality conditions (2) or (4) involve inequalities. An inequality constrained problem is inherently different and often more difficult then one that involves equalities only. Inequalities introduce boundaries and combinatorial features which are absent when dealing with equalities alone. Unconstrained methods in nonlinear programming are those which convert an inequality constrained optimization problem (1) or its optimality conditions (2) to a problem involving equalities only. There are many ways for doing this. In 1937 Valentine [33] proposed replacing (1) by

$$\underset{x,z}{\text{minimize}} \ f(x) \text{ subject to } g_i(x) + z_i^2 = 0, \ i = 1,\ldots,m .$$

Unfortunately this transformation is not useful computationally and furthermore it destroys any convexity the original problem may have. In 1943 Courant [9] proposed an exterior penalty method for equality constraints. For the inequality constrained problem (1) this method becomes [11]

$$\underset{x}{\text{minimize}} \ f(x) + \frac{\alpha}{2} \sum_{i=1}^{m} (g_i(x))_+^2$$

where α is a positive number that must approach infinity, and $(g_i(x))_+ = \text{maximum}\{0, g_i(x)\}$. The main difficulty with this approach is that as α approaches infninity the Hessian of the penalty function becomes ill-conditioned and it becomes increasingly difficult to minimize it [19]. To overcome this difficulty exact penalty methods have been considered by

[1] To simplify notation all vectors, except gradients, are either row or column vectors depending on the context. However gradient vectors such as $\nabla f(x)$ are row vectors unless transposed to a column $\nabla f(x)^T$.

Ablow and Brigham in 1955, Zangwill in 1967 and Pietrzykowski in 1969 [1,35,27] in which the penalty parameter α remains finite and problem (1) is replaced by

$$\underset{x}{\text{minimize}}\ f(x) + \alpha \sum_{i=1}^{m} (g_i(x))_+$$

where α is a positive number that must be sufficiently large but finite. The main difficulty here is that the penalty function is not differentiable. A method which avoids most of the difficulties associated with the unconstrained methods described so far was introduced by Rockafellar in 1970 [30] under the name of the augmented Lagrangian method, where a function that serves both as a Lagrangian and as an exterior penalty function is used. This method originally introduced, independently of each other, by Hestenes and Powell in 1969 [29,14] for equality constrained minimization problems, has been intensely investigated recently [2,3,4,5,7,12,13,21,23, 24,28,30,31,32,34] and is known under a variety of names such as: method of multipliers, shifted penalty method, penalty Lagrangian method, and unconstrained Lagrangian method. We shall use the name augmented Lagrangian method here. In this method problem (1) is replaced by

$$F(x,y,\alpha) = \begin{bmatrix} F_1(x,y,\alpha) \\ \\ F_2(x,y,\alpha) \end{bmatrix} = 0 \qquad (5)$$

where $F: R^{n+m+1} \to R^{n+m}$, $F_1: R^{n+m+1} \to R^n$, $F_2: R^{n+m+1} \to R^m$, y is a one-to-one map of the Lagrange multiplier u, and α is a sufficiently large, but finite positive number. We shall give later below some specific choices for the function F. Some of the key properties of (5) are

$$\left.\begin{array}{ll}
\text{(a)} & F \text{ is locally or globally differentiable} \\
\text{(b)} & x \text{ and } y \text{ are not constrained by inequalities} \\
\text{(c)} & \text{The penalty parameter } \alpha \text{ remains finite} \\
\text{(d)} & \text{For large but finite values of } \alpha,\ \nabla_1 F_1(\bar{x},\bar{y},\alpha) \text{ is} \\
 & \text{positive definite at appropriate stationary points} \\
 & (\bar{x},\bar{y}). \\
\text{(e)} & \text{The system (5) can be solved by algorithms with a con-} \\
 & \text{vergence rate factor which is inversely proportional to} \\
 & \text{to the penalty parameter } \alpha\,.
\end{array}\right\} \qquad (6)$$

A typical augmented Lagrangian algorithm associated with (5) is the following. Given (x^i,y^i) determine (x^{i+1},y^{i+1}) as follows

1 ALGORITHM

Step 1: $F_1(x^{i+1},y^i,\alpha)=0$

Step 2: $y^{i+1} = y^i + \alpha F_2(x^{i+1},y^i,\alpha)$

Sometimes it is also happens that

$$F(x,y,\alpha) = \begin{bmatrix} F_1(x,y,\alpha) \\ \\ F_2(x,y,\alpha) \end{bmatrix} = \begin{bmatrix} \nabla_1 L(x,y,\alpha)^T \\ \\ \nabla_2 L(x,y,\alpha)^T \end{bmatrix} = \nabla L(x,y,\alpha)^T \tag{7}$$

where $L:R^{n+m+1} \to R$, then $L(x,y,\alpha)$ is called the <u>augmented Lagrangian</u> and the algorithm takes the form

2 ALGORITHM

Step 1: $L(x^{i+1},y^i,\alpha) = \min_x L(x,y^i,\alpha)$

Step 2: $y^{i+1} = y^i + \alpha\nabla_2 L(x^{i+1},y^i,\alpha)$

The simplest such augmented Lagrangian is due to Rockafellar [31] and is given by

$$L(x,y,\alpha) = f(x) + \frac{1}{2\alpha}\sum_{j=1}^{m}((\alpha g_j(x)+y_j)_+^2 - y_j^{\,2}) \tag{8}$$

and for this specific choice of $L(x,y,\alpha)$ we have that

$$F(x,y,\alpha) = \nabla L(x,y,\alpha)^T = \begin{bmatrix} \nabla f(x)^T + \sum_{j=1}^{m}(\alpha g_j(x)+y_j)_+\nabla g_j(x)^T \\ \\ \frac{1}{\alpha}((\alpha g_j(x)+y_j)_+ - y_j),\ j = 1,\ldots,m \end{bmatrix} \tag{9}$$

In the rest of this paper we shall try to justify the use of Rockafellar's augmented Lagrangian (8) by establishing for it and for a wider class of functions $F(x,y,\alpha)$ the key properties (6) stated above.

We begin by going back to the Karush-Kuhn-Tucker conditions (2), and attempting to replace the $1 + 2m$ conditions: $ug(x) = 0$, $g(x) \leq 0$, $u \geq 0$, by an equivalent system of m equations <u>without</u> an increase in the number of variables. If we can do this, then by maintaining the m equations $\nabla f(x) + u\nabla g(x) = 0$ we can replace the Karush-Kuhn-Tucker conditions by a system of $n + m$ equations in $n + m$ unknowns. That this can be done follows from the following simple but key lemma which can be obtained by symmetrizing the key Lemma 2.7 of [23] or from Lemma 3 of [21].

3 LEMMA

Let h, g and y be real numbers, let θ be a strictly increasing function from R into R, that is $\theta(a) < \theta(b) \Longleftrightarrow a < b$, and let $\theta(0) = 0$. Then

$$\left\langle \theta(|h-y|) - \theta(h) - \theta(y) = 0 \right\rangle \Longleftrightarrow \left\langle hy = 0,\ h \geqq 0,\ y \geqq 0 \right\rangle \tag{10}$$

or equivalently

$$\left\langle \theta(|g+y|) - \theta(-g) - \theta(y) = 0 \right\rangle \Longleftrightarrow \left\langle gy = 0,\ g \leqq 0,\ y \geqq 0 \right\rangle \tag{11}$$

Proof

The equivalence of (10) and (11) is evident once we set $g = -h$. We establish (10) now. The backward implication of (10) is trivial because either $h = 0,\ y \geqq 0$ or $h \geqq 0,\ y = 0$; in the first case we have $\theta(|h-y|) - \theta(h) - \theta(y) = \theta(y) - \theta(y) = 0$ and in the second case we have $\theta(|h-y|) - \theta(h) - \theta(y) = \theta(h) - \theta(h) = 0$. We prove now the forward implication of (10) by looking at two cases.

Case I: $\underline{h - y \geqq 0}$ $\quad \theta(h-y) = \theta(h) + \theta(y)$

$\underline{y > 0} \Rightarrow \theta(h-y) > \theta(h) \Rightarrow h - y > h \Rightarrow y < 0$ (Contradiction)

$\underline{y < 0} \Rightarrow \theta(h-y) < \theta(h) \Rightarrow h - y < h \Rightarrow y > 0$ (Contradiction)

Hence we must have that

$$y = 0,\ h \geqq 0 \ \text{ and } \ hy = 0.$$

Case II: $\underline{y - h \geqq 0}$

By symmetry we obtain from case I that

$$h = 0,\ y \geqq 0 \ \text{ and } \ hy = 0.$$

Hence in either case we have that $h \geqq 0,\ y \geqq 0$ and $hy = 0$. $\square$

The simplest choices for θ are $\theta(z) = z$ and $\theta(z) = z|z|$. The first choice $\theta(z) = z$ which also underlies Rockafellar's Lagrangian (8) leads to equations such as $|h-y| - h - y = 0$ which are not differentiable globally. The second choice $\theta(z) = z|z|$ however leads to globally differentiable equations such as $(h-y)^2 - h|h| - y|y| = 0$. As an aside we note that by using $\theta(z) = z|z|$ in Lemma 3 above it follows immediately that the classical complementarity problem [8,15] of finding an $x \in R^n$ such that $G(x) \geqq 0,\ x \geqq 0,\ xG(x) = 0$, where $G: R^n \to R^n$, or its equivalent variational inequality formulation of finding an $x \geqq 0$ such that

$(y-x)G(x) \geqq 0$ for all $y \geqq 0$ [16,25], are both completely equivalent to solving the following system of n equations in n unknowns

$$(G_j(x)-x_j)^2 - G_j(x)|G_j(x)|-x_j|x_j| = 0 \quad j = 1,\dots,n \tag{12}$$

Note that in the system (12) x is not constrained by any inequalities and that the equations are globally differentiable if G is globally differentiable. Furthermore, it can be shown [22] that the Jacobian of (12) is nonsingular at solutions $\bar{x}$ for which the nondegeneracy condition $\bar{x} + G(\bar{x}) > 0$ holds and $\nabla G(\bar{x})$ has nonsingular principal minors.

Going back to the Karush-Kuhn-Tucker conditions (2) we can immediately establish the following equivalence between (2) and a system of $n + m$ equations in $n + m$ unknowns.

4 THEOREM

$$\left\langle \begin{array}{l} \nabla f(x) + u\nabla g(x) = 0 \\ ug(x) = 0,\ g(x)\leqq 0,\ u\geqq 0 \end{array} \right\rangle \iff \left\langle F(x,u) = \begin{bmatrix} (\nabla f(x)+u\nabla g(x))^T \\ (g_j(x)+u_j)^2+g_j(x)|g_j(x)|-u_j|u_j| \\ j = 1,\dots,m \end{bmatrix} = 0 \right. \tag{13}$$

Furthermore the Jacobian $\nabla F(\bar{x},\bar{u})$ is nonsingular at any solution $(\bar{x},\bar{u})$ which satisfies the Jacobian nonsingularity conditions

$$\left.\begin{array}{ll} \text{(a)} & \bar{u}_j > 0 \text{ for } j \in I = \{j|g_j(\bar{x}) = 0\} \text{ (Strict complementarity)} \\ \text{(b)} & \nabla g_j(\bar{x}),\ j \in I, \text{ are linearly independent} \\ \text{(c)} & f \text{ and } g \text{ are twice differentiable at } \bar{x} \text{ and} \\ & \nabla g(\bar{x})_{j\in I}x = 0,\ x \neq 0 \implies x\nabla_{11}L_0(\bar{x},\bar{u})x > 0 \text{ where} \\ & \nabla_{11}L_0(\bar{x},\bar{u}) \text{ is the Hessian of the classical Lagrangian (3)} \\ & \text{with respect to } x \text{ (Second order sufficiency)} \end{array}\right\rangle \tag{14}$$

Proof.

The equivalence (13) follows from Lemma 3 by setting $\theta(z) = z|z|$, $y = u_j$, $j = 1,\dots,m$ in (11). To establish the nonsingularity of $\nabla F(\bar{x},\bar{y})$ under (14) note that

$$\nabla F(\bar{x},\bar{u}) = \begin{bmatrix} \nabla_{11}L_0(\bar{x},\bar{u}) & \nabla g_I(\bar{x})^T & \nabla g_J(\bar{x})^T \\ 2\bar{u}_I\nabla g_I(\bar{x}) & 0 & 0 \\ 0 & 0 & 2g_J(\bar{x}) \end{bmatrix}$$

where $J = \{j|g_j(\bar{x})<0\}$, $\bar{u}_I$ is a diagonal matrix with elements $\bar{u}_{j\in I}$ and

$\nabla g_I(\bar{x})$ and $\nabla g_J(\bar{x})$ are matrices with rows $\nabla g_j(\bar{x})$, $j \in I$ and $\nabla g_j(\bar{x})$, $j \in J$ respectively. Evidently the Jacobian $\nabla F(\bar{x},\bar{u})$ is nonsingular if $\begin{bmatrix} \nabla_{11}L_0(\bar{x},\bar{u}) & \nabla g_I(\bar{x})^T \\ 2\bar{u}_I \nabla g_I(\bar{x}) & 0 \end{bmatrix}$ is nonsingular. This latter fact follows from the implications

$$\begin{bmatrix} \nabla_{11}L_0(\bar{x},\bar{u}) & \nabla g_I(\bar{x})^T \\ 2\bar{u}_I \nabla g_I(\bar{x}) & 0 \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = 0 \iff \begin{cases} \nabla_{11}L_0(\bar{x},\bar{u})x + \nabla g_I(x)^T v = 0 \\ \nabla g_I(\bar{x})x = 0 \qquad \text{(By (14a))} \end{cases}$$

$$\implies x\nabla_{11}L_0(\bar{x},\bar{u})x = 0, \quad \nabla g_I(\bar{x})x = 0$$

$$\implies x = 0, \ \nabla g_I(\bar{x})^T v = 0 \quad \text{(By (14c))}$$

$$\implies x = 0, \ v = 0 \quad \text{(By (14b))}. \ \square$$

Note that in the equivalence (13) all inequalities have disappeared from the system $F(x,u) = 0$. In addition F is globally differentiable if f and g are twice differentiable globally. Furthermore at isolated local minima of (1) satisfying (14), $\nabla F(\bar{x},\bar{u})$ is nonsingular and hence locally superlinearly convergent quasi-Newton methods [6] and other methods for solving nonlinear equations [26] can be used for finding points that satisfy the Karush-Kuhn-Tucker conditions (2). Note also that there is no parameter in the equations $F(x,u) = 0$ of (13). There are certain disadvantages however to solving $F(x,u) = 0$ directly. The main disadvantage is that $n + m$ may be much larger than n, which is the case when there are many $g_j(\bar{x}) < 0$, and hence in solving $F(x,u) = 0$ we are working in a high dimensional space R^{n+m} when it is preferable to work in a lower dimensional space such as R^n. Also $L_0(x,\bar{u})$ may not be convex with respect to x near $\bar{x}$, even if (14) is satisfied. This convexity is useful in solving $\nabla_1 L_0(x,u) = 0$, the first n equations of (13). To avoid these difficulties we give another equivalence between the Karush-Kuhn-Tucker conditions (2) and another system of $n + m$ equations which is the underlying relation for all augmented Lagrangian methods.

5 THEOREM

Let θ and ϕ be strictly increasing functions from R into R with $\theta(0) = \phi(0) = 0$ and let ϕ be surjective (that is ϕ maps R onto R). Then for any $\alpha > 0$

$$\left.\begin{array}{l}\nabla f(x) + u\nabla g(x) = 0 \\ ug(x) = 0,\ g(x) \leqq 0,\ u \geqq 0\end{array}\right\rangle \Longleftrightarrow$$

$$\left\langle F(x,y,\alpha) = \begin{bmatrix} F_1(x,y,\alpha) \\ F_2(x,y,\alpha)\end{bmatrix} = \begin{bmatrix} (\nabla f(x) + \sum_{j=1}^{m} \phi(\alpha g_j(x)+y_j)_+ \nabla g_j(x))^T \\ \theta(|\alpha g_j(x)+y_j|) - \theta(-\alpha g_j(x)) - \theta(y_j) \\ j = 1,\ldots,m \end{bmatrix} = 0 \right. \tag{15}$$

where $u_j = \phi(y_j)$, $j = 1,\ldots,m$, and $\phi(z)_+ = \left\langle \begin{array}{ll} \phi(z) & \text{if } z \geqq 0 \\ 0 & \text{if } z < 0 \end{array}\right.$.

Proof

The equivalence (15) follows from Lemma 3 and

$$\phi(\alpha g_j(x)+y_j)_+ = \phi(y_j)_+ = \phi(y_j) = u_j \quad \text{for} \quad g_j(x) = 0$$

$$\phi(\alpha g_j(x)+y_j)_+ = \phi(\alpha g_j(x))_+ = 0 = \phi(y_j) = u_j \quad \text{for} \quad g_j(x) < 0 \quad \square$$

REMARKS

(a) The equivalence of (15) holds for any positive value of α no matter how small.

(b) $F(x,y,\alpha)$ is globally differentiable as a function of (x,y) provided that f and g are twice globally differentiable, θ and ϕ differentiable on R, and $\theta'(0) = \phi'(0) = 0$.

(c) The simplest choice of $2\alpha\theta(z) = \phi(z) = z$ leads to $F(x,y,\alpha) = \nabla L(x,y,\alpha)^T$ where $L(x,y,\alpha)$ is Rockafellar's augmented Lagrangian (8). Note however that $\nabla L(x,y,\alpha)$ is not differentiable globally, and in particular it is not differentiable whenever $\alpha g_j(x) + y = 0$, for some $j = 1,\ldots,m$. However if strict complementarity (14a) holds at $(\bar{x},\bar{u})$ then $\alpha g_j(\bar{x}) + \bar{y}_j \neq 0$ for $\bar{u}_j = \phi(\bar{y}_j)$, and $\nabla L(x,y,\alpha)$ is differentiable at $(\bar{x},\bar{y})$. We shall give below $F(x,y,\alpha)$ which is globally differentiable.

We proceed now to the other key properties (6c) and (6d) mentioned aboved by means of the following theorem.

6 THEOREM

(Positive definiteness of $\nabla_1 F_1(\bar{x},\bar{y},\alpha)$) Let F,θ and ϕ be as in Theorem 5, let θ and ϕ be differentiable on R, let $\phi'(z) > 0$ for $z > 0$, let $(\bar{x},\bar{u})$ satisfy the Karush-Kuhn-Tucker conditions (2), and conditions (14a) and (14c). Then $\nabla_1 F_1(\bar{x},\bar{y},\alpha)$ is positive definite for

$\alpha \geqq \bar{\alpha}$ for some $\bar{\alpha} \geqq 0$, where $\bar{u}_j = \phi(\bar{y}_j)$, $j = 1,\ldots,m$.

Proof

By (14a) we have that

$$\nabla_1 F_1(\bar{x},\bar{y},\alpha) = \nabla_{11} L_0(\bar{x},\bar{u}) + \sum_{j\in I} \alpha\phi'(\bar{y}_j)\nabla g_j(\bar{x})^T \nabla g_j(\bar{x})$$

where $\bar{u}_j = \phi(\bar{y}_j)$, $j = 1,\ldots,m$, and by (14c) we have have

$$\left\langle \phi'(\bar{y}_j)^{\frac{1}{2}} \nabla g_{j\in I}(\bar{x})x = 0,\ x \neq 0 \right\rangle \Longrightarrow \left\langle x\nabla_{11} L_0(\bar{x},\bar{u})x > 0 \right\rangle$$

Hence by Debreu's theorem [10] which states that for $m_1 \times n$ and $n \times n$ matrices M and N

$$\left\langle Mx = 0,\ x \neq 0 \Longrightarrow xNx > 0 \right\rangle \Longleftrightarrow \left\langle \begin{array}{l} N + \alpha M^T M \text{ is positive definite} \\ \\ \text{for all } \alpha \geqq \bar{\alpha} \text{ for some } \bar{\alpha} \geqq 0 \end{array} \right\rangle$$

it follows that $\nabla_1 F_1(\bar{x},\bar{y},\alpha)$ is positive definite for all $\alpha \geqq \bar{\alpha}$ for some $\bar{\alpha} \geqq 0$. □

We finally establish the key property (6e) under the following simplifying assumption.

7 ASSUMPTION

Let the assumptions of Theorem 6 hold. For each $\alpha \geqq \bar{\alpha}$ let $F(x,y,\alpha) = 0$ have a unique solution x for each y in some open neighborhood of $\bar{y}$.

Although this assumption is not essential and can be gotten around [23, Theorem 4.10], without it one must append to step 1 of Algorithms 1 and 2, the additional computational procedure of finding a closest x^{i+1} to x^i in some norm. In practice because x^{i+1} is obtained by an iterative subprocedure from x^i, it is not an unrealistic supposition that a closest x^{i+1} will be obtained as a matter of course.

8 THEOREM

(Parametric superlinear convergence of Algorithms 1 and 2) Let the assumptions of Theorem 6 hold, let $\alpha \geqq \bar{\alpha}$, let (14b) and hence (14) hold, let Assumption 7 hold, let ϕ and θ be continuously differentiable on R and let f and g be twice continuously differentiable on an open neighborhood of $\bar{x}$. Then for sufficiently large but finite α there exists an open neighborhood $N(\bar{y})$ of $\bar{y}$ such that for y^0 in $N(\bar{y})$ there exists x^0 such that $F_1(x^0,y^0,\alpha) = 0$, and the iterates (x^i,y^i) generated by Algorithms 1 and 2 exist and converge to $(\bar{x},\bar{y})$ at the root rate

$$\|x^i-\bar{x},y^i-\bar{y}\| \leqq \left(\frac{c}{\alpha}\right)^i \quad \text{for } i \geqq \bar{i} \tag{16}$$

for some integer $\bar{i}$ and some positive constant c, where $\| \ \|$ is the ℓ_2 norm, provided that θ satisfies one of the two following conditions

$$\alpha(\theta'(z)+\theta'(0)) = 1 \quad \text{for all} \quad z > 0 \tag{17}$$

or

$$\left.\begin{array}{l} \alpha(\theta'(z)+\theta'(0)) = 1 \quad \text{for all} \quad z \geqq \frac{1}{\alpha} \\ \text{and} \\ \alpha \geqq \max\{\bar{y}_{j\epsilon I}^{-1}, \ (-g_{j\epsilon J}(\bar{x}))^{-\frac{1}{2}}\} \end{array}\right\rangle \tag{18}$$

Proof

By Theorem 6, $\nabla_1 F_1(\bar{x},\bar{y},\alpha)$ is positive definite for $\alpha \geqq \bar{\alpha}$. Hence by the implicit function theorem there exists a continuously differentiable function $e:R^n \to R^m$ on an open neighborhood of $\bar{y}$ such that

$$F_1(e(y),y,\alpha) = 0 \quad \text{for all} \quad y \in N(\bar{y}) \tag{19}$$

It follows that Algorithms 1 and 2 are equivalent to

$$y^{i+1} = y^i + \alpha F_2(e(y^i),y^i,\alpha) \tag{20}$$

Consider now the mapping

$$G(y,\alpha) = y + \alpha F_2(e(y),y,\alpha) \tag{21}$$

underlying the iteration (20). From (19) we have that

$$\nabla_1 F_1(\bar{x},\bar{y},\alpha)\nabla e(\bar{y}) + \nabla_2 F_1(\bar{x},\bar{y},\alpha) = 0$$

and so

$$\begin{aligned} \nabla G(\bar{y},\alpha) &= I + \alpha\nabla_1 F_2(\bar{x},\bar{y},\alpha)\nabla e(\bar{y}) + \alpha\nabla_2 F_2(\bar{x},\bar{y},\alpha) \\ &= I - \alpha(\nabla_1 F_2(\bar{x},\bar{y},\alpha)\nabla_1 F_1(\bar{x},\bar{y},\alpha)^{-1}\nabla_2 F_1(\bar{x},\bar{y},\alpha) - \nabla_2 F_2(\bar{x},\bar{y},\alpha)) \end{aligned}$$

By making use now of

$$\nabla F(\bar{x},\bar{y},\alpha) = \begin{bmatrix} \begin{array}{l}[\nabla_{11}L_0(\bar{x},\bar{u})+\sum_{j\epsilon I}\alpha\phi'(\bar{y}_j)\cdot \\ \quad \cdot\nabla g_j(\bar{x})^T\nabla g_j(\bar{x})]\end{array} & \underset{(j\epsilon I)}{\nabla g_j(\bar{x})^T\phi'(\bar{y}_j)} & 0 \\ \underset{(j\epsilon I)}{\alpha(\theta'(\bar{y}_j)+\theta'(0))\nabla g_j(\bar{x})} & 0 & 0 \\ 0 & 0 & \underset{(j\epsilon J)}{-\theta'(-\alpha g_j(\bar{x}))-\theta'(0)} \end{bmatrix}$$

where $\bar{u}_j = \phi(\bar{y}_j)$, $j = 1,\ldots,m$, we obtain

$$\nabla G(\bar{y},\alpha) = I - \alpha \begin{bmatrix} \alpha(\theta'(\bar{y}_I)+\theta'(0))\nabla g_I(\bar{x})[\nabla_{11}L_0(\bar{x},\bar{u})+ \\ \quad \nabla g_I(\bar{x})^T(\alpha\phi'(\bar{y}_I))\nabla g_I(\bar{x})]^{-1}\nabla g_I(\bar{x})^T\phi'(\bar{y}_I) & 0 \\ 0 & \theta'(-\alpha g_J(\bar{x}))+\theta'(0) \end{bmatrix}$$

where $\theta'(\bar{y}_I) + \theta'(0)$ denotes a diagonal matrix with diagonal elements $\theta'(\bar{y}_j) + \theta'(0)$, $j \in I$. Similarly $\phi'(\bar{y}_I)$ and $\theta'(-\alpha g_J(\bar{x})) + \theta'(0)$ are also diagonal matrices with elements $\phi'(\bar{y}_j)$, $j \in I$, and $\theta'(-\alpha g_j(\bar{x})) + \theta'(0)$, $j \in J$, respectively. The matrix $\nabla g_I(\bar{x})$ has as its rows, $\nabla g_j(\bar{x})$, $j \in I$. We now make use of the following extremely useful lemma which can be considered a key lemma for obtaining convergence rate results for augmented Lagrangians. In [29] Powell proved a related result by using determinants.

9 LEMMA

Let $C(\alpha) = B(A+B^TQ(\alpha)B)^{-1}B^T$ where B is a given $m \times n$ matrix of rank m, A is an $n \times n$ matrix, $Q(\alpha)$ is a differentiable $m \times m$ matrix function on R and $A + B^TQ(\alpha)B$ is positive definite for $\alpha \geq \bar{\alpha}$ for some $\bar{\alpha}$. Then $C(\alpha) = (Q(\alpha)+K)^{-1}$ for all $\alpha \geq \bar{\alpha}$ for some $m \times m$ matrix K which is independent of α.

Proof

Recall that the formula for differentiating the inverse of a matrix is given by $\frac{dC(\alpha)^{-1}}{d\alpha} = -C(\alpha)^{-1}\frac{dC(\alpha)}{d\alpha}C(\alpha)^{-1}$. Hence from the definition of $C(\alpha)$ we have that

$$\frac{dC(\alpha)}{d\alpha} = -B(A+B^TQ(\alpha)B)^{-1}B^T\frac{dQ(\alpha)}{d\alpha}B(A+B^TQ(\alpha)B)^{-1}B^T = -C(\alpha)\frac{dQ(\alpha)}{d\alpha}C(\alpha)$$

Hence $\frac{dC(\alpha)^{-1}}{d\alpha} = \frac{dQ(\alpha)}{d\alpha}$ and $C(\alpha)^{-1} = Q(\alpha) + K$. $\square$

By using this lemma in the last expression for $\nabla G(\bar{y},\alpha)$ we obtain

$$\nabla G(\bar{y},\alpha) = I - \alpha\begin{bmatrix} \alpha(\theta'(\bar{y}_I)+\theta'(0))(\alpha\phi'(\bar{y}_I)+K)^{-1}\phi'(\bar{y}_I) & 0 \\ 0 & \theta'(-\alpha g_J(\bar{x}))+\theta'(0) \end{bmatrix}$$

$$= \begin{bmatrix} I-\alpha(\theta'(\bar{y}_I)+\theta'(0))\left(I+\frac{\phi'(\bar{y}_I)^{-1}K}{\alpha}\right)^{-1} & 0 \\ 0 & I - \alpha(\theta'(-\alpha g_J(\bar{x}))+\theta'(0)) \end{bmatrix}$$

which upon using condition (17) or (18) gives

$$\nabla G(\bar{y},\alpha) = \begin{bmatrix} I - (I + \frac{\phi'(\bar{y}_I)^{-1}K}{\alpha})^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

By choosing α large enough such that $\|\phi'(\bar{y}_I)^{-1}K\| \leqq \frac{\alpha}{2}$, it follows that

$$\|I - (I + \frac{\phi'(\bar{y}_I)^{-1}K)}{\alpha})^{-1}\| \leq \frac{2}{\alpha} \|\phi'(\bar{y}_I)^{-1}K\|$$

Hence the spectral radius $\rho(\nabla G(\bar{y},\alpha))$ of $\nabla G(\bar{y},\alpha)$ satisfies the inequality

$$\rho(\nabla G(\bar{y},\alpha)) \leqq \frac{\bar{c}}{\alpha} \tag{22}$$

where $\bar{c} = 2\|\phi'(\bar{y}_I)^{-1}K\|$. It follows from (22) and Ostrowski's point of attraction theorem [26, Theorem 10.1.3] and [26, Theorem 10.1.4] that the sequence $\{y^i\}$ is locally convergent to $\bar{y}$ and the root convergence factor $R_1\{y^i\} = \limsup_{i\to\infty} \|y^i-\bar{y}\|^{\frac{1}{i}}$ satisfies the inequality

$$\limsup_{i\to\infty} \|y^i-\bar{y}\|^{\frac{1}{i}} \leqq \frac{\bar{c}}{\alpha}$$

and hence for some $\bar{i}$

$$\|y^i-\bar{y}\| \leqq (\frac{2\bar{c}}{\alpha})^i, \quad i \geqq \bar{i}$$

Since $x^i = e(y^i)$ and e is continuously differentiable on $N(\bar{y})$ it follows that for some constant $\hat{c} \geqq 1$

$$\|x^i-\bar{x}\| = \|e(y^i)-e(\bar{y})\| \leqq \hat{c}\,\|y^i-\bar{y}\| \leqq \hat{c}(\frac{2\bar{c}}{\alpha})^i \leqq (\frac{2\bar{c}\hat{c}}{\alpha})^i, \quad i \geqq \bar{i}$$

and hence

$$\|x^i-x, y^i-\bar{y}\| \leqq (\frac{4\bar{c}\hat{c}}{\alpha})^i = (\frac{c}{\alpha})^i, \quad i \geqq \bar{i}. \qquad \square$$

We call the convergence rate (16) parametricaly superlinear, because by increasing α it can be made better than any linear root rate convergence: $\|x^i-\bar{x}, y^i-\bar{y}\| \leqq (\gamma)^i$, where γ is some fixed number less than one.

We remark also that the choice $2\alpha\theta(z) = \phi(z) = z$ which leads to Rockafellar's Lagrangian (8) satisfies (17) and hence the root rate convergence (16) holds for Algorithm 2 with $L(x,y,\alpha)$ given by (8).

Bertsekas [3,4,5] has given a parametricaly superlinear quotient rate for this algorithm also. As noted earlier for this choice of $L(x,y,\alpha)$, $F(x,y,\alpha) = \nabla L(x,y,\alpha)^T$ is not globally differentiable. To avoid possible convergence difficulties that may arise from such nondifferentiability, we propose here another choice of $F(x,y,\alpha)$ which is globally differentiable but which satisfies assumption (18) in Theorem 8 instead of (17), and hence the convergence rate result (16) holds for it also. In particular we propose

$$\theta(z) = \frac{1}{\alpha}\phi(z) = \begin{cases} \frac{z}{\alpha} + \frac{1}{2\alpha^2} & \text{if} \quad z \leq -\frac{1}{\alpha} \\ \frac{1}{2}z|z| & \text{if} \quad |z| < \frac{1}{\alpha} \\ \frac{z}{\alpha} - \frac{1}{2\alpha^2} & \text{if} \quad z \geq \frac{1}{\alpha} \end{cases} \tag{23}$$

For the choice of θ and ϕ we have that

$$\theta'(z) = \frac{1}{\alpha}\phi'(z) = \begin{cases} \frac{1}{\alpha} & \text{if} \quad z \leq -\frac{1}{\alpha} \\ |z| & \text{if} \quad |z| < \frac{1}{\alpha} \\ \frac{1}{\alpha} & \text{if} \quad z \geq \frac{1}{\alpha} \end{cases} \tag{24}$$

Hence for $z \geq \frac{1}{\alpha}$, $\alpha(\theta'(z)+\theta'(0)) = \alpha(\frac{1}{\alpha}+0) = 1$, and (18) is satisfied. <u>Thus the convergence rate result (16) holds for Algorithm 1 with $F(x,y,\alpha)$ given by (15) and (23).</u> Note that $F(x,y,\alpha)$ as defined by (15) and (23) is globally differentiable on R^{n+m} if f and g are twice globally differentiable on R^n. This is true because $\theta'(0) = \phi'(0) = 0$. But, our function $F(x,y,\alpha)$ as defined by (15) and (23) is not the gradient of some augmented Lagrangian. However, $F_1(x,y,\alpha) = \nabla_1 L(x,y,\alpha)^T$ where

$L(x,y,\alpha) = f(x) + \sum_{j=1}^{m} \psi(\alpha g_j(x)+y_j)_+$ and

$$\psi(z) = \begin{cases} \frac{z^2}{2\alpha} + \frac{z}{2\alpha^2} + \frac{1}{6\alpha^3} & \text{if} \quad z \leq -\frac{1}{\alpha} \\ \frac{1}{6}|z|^3 & \text{if} \quad |z| < \frac{1}{\alpha} \\ \frac{z^2}{2\alpha} - \frac{z}{2\alpha^2} + \frac{1}{6\alpha^3} & \text{if} \quad z \geq \frac{1}{\alpha} \end{cases}$$

But $F_2(x,y,\alpha) \neq \nabla_2 L(x,y,\alpha)^T$. It is interesting then to pose the following proposal for further investigation.

10 PROPOSAL

Find an $F(x,y,\alpha)$ which satisfies the assumptions of Theorem 8, which is globally differentiable on R^{n+m} and such that $F(x,y,\alpha) = \nabla L(x,y,\alpha)^T$ for some $L(x,y,\alpha)$.

ACKNOWLEDGEMENT

I am indebted to my colleagues R. R. Meyer and S. M. Robinson for valuable comments on this paper and G. P. McCormick for reference [1].

REFERENCES

1. C. M. Ablow & G. Brigham: "An analog solution of programming problems", Operations Research, 3, 1955, 388-394.

2. K. J. Arrow, F. J. Gould & S. M. Howe: "A general saddle point result for constrained optimization", Mathematical Programming 5, 1973, 225-234.

3. D. P. Bertsekas: "Convergence rate of penalty and multiplier methods," Proceedings of 1973 IEEE Conference on Decision and Control, San Diego, California, December 1973, 260-264.

4. D. P. Bertsekas: "Combined primal dual and penalty methods for constrained minimization," SIAM J. Control, 13(3), 1975.

5. D. P. Bertsekas: "On penalty and multiplier methods for constrained optimization", in "Nonlinear Programming 2", O. L. Mangasarian, R. R. Meyer and S. M. Robinson (eds), Academic Press, New York, 1975, 165-191.

6. C. G. Broyden, J. E. Dennis & J. J. Moré: "On the local and superlinear convergence of quasi-Newton methods," J. Inst. Math. Applics. 12, 1973, 223-245.

7. J. D. Buys: "Dual algorithms for constrained optimization problems", Doctorate Dissertation, University of Leiden, June 1972.

8. R. W. Cottle: "Nonlinear programs with positively bounded Jacobians", SIAM J. Appl. Math. 12, 1964, 663-665.

9. R. Courant: "Variational methods for the solution of problems of equilibrium and vibrations", Bull. Amer. Math. Soc. 49, 1943, 1-23.

10. G. Debreu: "Definite and semidefinite quadratic forms," Econometrica 20, 1952, 295-300.

11. A. V. Fiacco & G. P. McCormick: "Nonlinear programming: sequential unconstrained minimization techniques", Wiley, New York, 1968.

12. R. Fletcher: "An ideal penalty function for constrained optimization", in Nonlinear Programming 2", O. L. Mangasarian, R. R. Meyer & S. M. Robinson (eds), Academic Press, New York, 1975, 121-163.

13. E. G. Golshtein & N. V. Tretyakov: "Modified Lagrangian functions", (Russian) Economic and Mathematical Methods 10(3), 1974, 568-591.

14. M. R. Hestenes: "Multiplier and gradient methods," J. Optimization Theory Appl. 4, 1969, 303-320.

15. S. Karamardian: "The nonlinear complementarity problem with applications" parts 1 and 2, J. Optimization Theory Appl. 4, 1969, 87-98, 167-181.

16. S. Karamardian: "The complementarity problem", Math. Programming 2, 1972, 107-129.

17. W. Karush: "Minima of functions of several variables with inequalities as side conditions", Master of Science Dissertation, Dept. Math., Univ. Chicago, Dec. 1939.

18. H. W. Kuhn & A. W. Tucker: "Nonlinear programming", Proc. 2nd Berkeley Symp. on Math. Statist. & Probability, J. Neyman (ed), Univ. Calif. Press, Berkeley, 1951, 481-492.

19. F. A. Lootsma: "A survey of methods for solving constrained minimization problems via unconstrained minimization", in Numerical Methods for Nonlinear Optimization, F. A. Lootsma (ed), Academic Press, New York, 1972, 313-347.

20. O. L. Mangasarian: "Nonlinear programming", McGraw-Hill, New York, 1969.

21. O. L. Mangasarian: "Unconstrained optimization methods" in Proceedings Twelfth Annual Allerton Conference on Circuit and System Theory, Oct. 2-4, 1974, University of Illinois, Urbana-Champaign, 153-160.

22. O. L. Mangasarian: "Equivalence of the complementarity problem to a system of nonlinear equations", Univ. Wisconsin, Comp. Sci. Report 227, Nov. 1974.

23. O. L. Mangasarian: "Unconstrained lagrangians in nonlinear programming" SIAM J. Control, 13(4) 1975.

24. A. Miele, P. E. Moseley & E. E. Cragg: "A modification of the method of multipliers for mathematical programming problems," in "Techniques of optimization," A. V. Balakrishnan (ed.), Academic Press, New York, 1972, 247-260.

25. J. J. Moré: "Coercivity conditions in nonlinear complementarity problems", SIAM Review 16, 1974, 1-16.

26. J. M. Ortega & W. C. Rheinboldt: "Iterative solution of nonlinear equations in several variables," Academic Press, New York, 1970.

27. T. Pietrzykowski: "An exact potential method for constrained maxima", SIAM J. Numer. Anal. 6, 1969, 299-304.

28. B. T. Polyak & N. V. Tretyakov: "Penalty method rate for constrained optimization problems", (Russian), USSR Computational Math. & Mathematical Physics, 13 (1), 1973, 34-46.

29. M. J. D. Powell: "A method for nonlinear constraints in minimization problems", in "Optimization", R. Fletcher, ed., Academic Press, New York, 1969, 283-298.

30. R. T. Rockafellar: "New applications of duality in convex programming", Seventh International Symposium on Mathematical Programming, The Hague 1970, published in Proceedings of the Fourth Conference on Probability, Brasov, Romania, 1971.

31. R. T. Rockafellar: "Augmented Lagrange multiplier functions and duality in nonconvex programming," SIAM J. Control 12, 1974, 268-285.

32. N. V. Tretyakov: "Penalty method rate for convex programming problems", (Russian) Economic and Mathematical Methods, 9(3), 1973, 526-540.

33. F. A. Valentine: "The problem of Lagrange with differential inequalities as added side conditions", in Contributions to the Calculus of Variations 1933-37, Univ. Chicago Press, Chicago, 1937, 407-448.

34. A. P. Wierzbicki: "A penalty function shifting method in constrained static optimization and its convergence properties", Archiwum Automatyki i Telemechaniki 16, 1971, 395-416.

35. W. I. Zangwill: "Nonlinear programming via penalty functions", Management Science, 13, 1967, 344-358.

UNIVERSITY OF WISCONSIN - MADISON