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

Received: October 1974

UNCONSTRAINED METHODS IN OPTIMIZATION*

by

O. L. Mangasarian

Computer Sciences Technical Report #224

October 1974

*Presented at the Twelfth Annual Allerton Conference on Circuit and System Theory, University of Illinois at Urbana-Champaign, October 2-4, 1974.

UNCONSTRAINED METHODS IN OPTIMIZATION[1)]

O. L. MANGASARIAN[2)]
Computer Sciences Department, University of Wisconsin, Madison, Wisconsin 53706

ABSTRACT

Some simple ideas are presented which make stationary points of inequality and equality constrained optimization problems equivalent to unconstrained solutions of a system of nonlinear equations. These ideas generalize the methods of unconstrained or augmented Lagrangians. A parametrically superlinearly convergent class of algorithms that includes that of Powell and Hestenes is proposed.

We shall consider in this paper the problem

$$\text{minimize } f(x), \text{ subject to } g(x) \leqq 0 \tag{1}$$

where $f:R^n \to R$ and $g:R^n \to R^m$. For the sake of simplicity we have excluded equality constraints which can be handled by minor modifications of the methods described here. We shall be concerned with stationary points of (1), that is points x in R^n which, for some $u \in R^m$, satisfy the Kuhn-Tucker conditions

$$\nabla f(x) + \nabla g(x)u = 0 \tag{2a}$$

$$ug(x) = 0,\ g(x) \leqq 0,\ u \geqq 0 \tag{2b}$$

where $\nabla f(x)$ is the gradient of f at x and $\nabla g(x)$ is the $n \times m$ Jacobian of g at x. We observe immediately that if the conditions (2b) can be replaced by an equivalent system of m equations <u>without</u> an increase in the number of variables then the Kuhn-Tucker conditions become equivalent to a system of $n + m$ equations in $n + m$ unknowns. That this can be done follows from the following simple but key lemma which was first proved in [9, lemma 2.7]

1. <u>Lemma</u> Let $y \in R$, $g \in R$ and let θ be an injective function from R into R (that is $\theta(a) = \theta(b)$ implies $a = b$) with $\theta(0) = 0$. Then

$$yg = 0,\ g \leqq 0,\ y \geqq 0 \Leftrightarrow \theta(g+y)_+ - \theta(y) = 0$$

where
$$\theta(z)_+ = \begin{cases} \theta(z) & \text{if } z \geqq 0 \\ 0 & \text{if } z < 0 \end{cases}.$$

<u>Proof</u> $$\theta(g+y)_+ - \theta(y) = 0 \Leftrightarrow \left\langle \begin{matrix} g+y \geqq 0,\ g+y = y \\ \text{or} \\ g+y < 0,\ y = 0 \end{matrix} \right\rangle \Leftrightarrow \left\langle \begin{matrix} g=0,\ y\geqq 0 \\ \text{or} \\ g<0,\ y=0 \end{matrix} \right\rangle$$

$$\Leftrightarrow yg = 0,\ g \leqq 0,\ y \geqq 0. \qquad \text{QED}$$

The following theorem, a generalization of [9, theorem 2.5], establishes the equivalence of the Kuhn-Tucker conditions to a system of $n + m$ equations in $n + m$ unknowns.

2. <u>Theorem</u>: Let θ be an injective function from R into R with $\theta(0) = 0$, and let ϕ be a strictly increasing surjective function from

---

1) Research supported by NSF Grant GJ35292
2) Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, Wisconsin 53706

R into R (that is ϕ maps R onto R) with $\phi(0) = 0$. Then for any $\alpha > 0$

$$\left\langle \begin{array}{l} \nabla f(x) + \nabla g(x)u = 0 \\ ug(x)=0, g(x)\leqq 0, u\geqq 0 \end{array} \right\rangle \underset{j=1,\ldots,m}{\overset{u_j=\phi(y_j)}{\Longleftrightarrow}} \left\langle \begin{array}{l} \nabla f(x)+\sum_{j=1}^{m}\phi(\alpha g_j(x)+y_j)_+\nabla g_j(x)=0 \\ \theta(\alpha g_j(x)+y_j)_+ - \theta(y_j)=0 \\ j=1,\ldots,m \end{array} \right\rangle \quad (3)$$

Proof: By lemma 1 and $\phi(\alpha g_j(x)+y_j)_+ = \phi(y_j)_+ = \phi(y_j) = u_j$ for $g_j(x) = 0$ and $\phi(\alpha g_j(x)+y_j)_+ = \phi(\alpha g_j(x))_+ = 0 = \phi(y_j) = u_j$ for $g_j(x) < 0$. QED

If we let $\alpha\theta(z) = \phi(z) = z$ then equations (3) become the gradient of Rockafellar's augmented Lagrangian [12] which has been intensively investigated recently [1,2,6,9]. Unfortunately equations (3) are not differentiable globally if $\theta'(0) \neq 0$ as is the case for Rockafellar's Lagrangian, although they are differentiable near a solution $(\bar{x},\bar{y})$ if we assume that strict complementarity holds, that is $\bar{y}_j+|g_j(\bar{x})|>0, j=1,\ldots,m$. In [9] by using $\alpha\theta(z)=\phi(z)=|z|^t z$, $t > 0$, equations (3) become differentiable globally. However the Jacobian of (3) becomes singular at the solution if there are inactive constrants. This difficulty was avoided in [9] by further augmenting the Lagrangian. We propose another way here to make the equations (3) differentiable globally while maintaining nonsingularity of the Jacobian by symmetrizing the last m equations of (3) with respect to $g_j(x)$ and y_j . This will have the advantage over [9] in that the ingenious method of Powell [11] and Hestenes [8] can be extended to inequalities while still maintaining global differentiability and a rate of convergence that is faster than any linear rate. We first symmetrize lemma 1 by interchanging the roles of g and y and exploiting the antisymmetry of θ .

3 Lemma Let $y \in R$, $g \in R$ and let θ be a strictly increasing odd function from R into R with $\theta(0) = 0$. Then

$$yg = 0,\ g \leqq 0,\ y \geqq 0 \Leftrightarrow \theta(|g+y|) + \theta(g) - \theta(y) = 0 .$$

Proof: If we let $h = -g$, then it is sufficient to establish that

$$yh = 0,\ h \geqq 0\ \ y \geqq 0 \Leftrightarrow \theta(|h-y|) - \theta(h) - \theta(y) = 0$$

($\Rightarrow$) Either $y = 0$ or $h = 0$. If $y = 0$, then $\theta(|h-y|) - \theta(h) - \theta(y) = \theta(|h|) - \theta(h) = 0$. If $h = 0$ a similar argument goes through with h and y interchanged.

($\Leftarrow$) (a) To show that $h \geqq 0$, assume the contrary that $h < 0$. Then $0 \leqq \theta(|y-h|) = \theta(h) + \theta(y) < \theta(y)$, from which it follows that $y > 0$ and $y > |y-h| = y - h$. The last inequality contradicts $h < 0$.

(b) To show that $y \geqq 0$, interchange the roles of y and h in (a).

(c) To show that $yh = 0$, assume the contrary that $y > 0$ and $h > 0$. Without loss of generality assume that $h \geqq y$, then $\theta(|h-y|) = \theta(h-y) < \theta(h) < \theta(h) + \theta(y)$. This contradicts $\theta(|h-y|) - \theta(h) - \theta(y) = 0$. QED.

The following equivalence theorem follows from lemma 3 in the same way as theorem 2 followed from lemma 1.

4 <u>Theorem</u> Let θ be a strictly increasing odd function from R into R with $\theta(0) = 0$ and let ϕ be a strictly increasing surjective function from R into R with $\phi(0) = 0$. Then for any $\alpha > 0$

$$\left.\begin{array}{l}\nabla f(x)+\nabla g(x)u=0 \\ \\ ug(x)=0, g(x)\leqq 0, u\geqq 0\end{array}\right\rangle \Leftrightarrow \left\langle\begin{array}{l} L_1(x,y,\alpha)=\nabla f(x)+\sum_{j=1}^{m}\phi(\alpha g_j(x)+y_j)_+\nabla g_j(x)=0 \\ \\ L_2(x,y,\alpha)=\theta(|\alpha g_j(x)+y_j|)+\theta(\alpha g_j(x))-\theta(y_j)=0 \\ \qquad j=1,\ldots,m\end{array}\right\rangle \tag{4}$$

and $u_j = \phi(y_j), j=1,\ldots,m$.

Again note that if we set $2\alpha\theta(z) = \phi(z) = z$ then equations (4) become the gradient of Rockafellar's Lagrangian. However we shall be concerned here mainly with functions θ and ϕ such that $\theta'(0) = \phi'(0)= 0$ in order that equations (4) be globally differentiable. The key to computational algorithms for solving (4) is the nonsingularity and the structure of its Jacobian at the solution. Let $L(x,y,\alpha) = \begin{bmatrix} L_1(x,y,\alpha) \\ L_2(x,y,\alpha)\end{bmatrix} = 0$

denote the $m + n$ equations (4) and let $(\bar{x},\bar{y})$ be a solution of (4) at which strict complementarity holds, $\bar{y}_j+|g_j(\bar{x})|>0, j=1,\ldots,m$. If ϕ and θ are differentiable on R with derivatives that vanish at the origin only, and if f and g are differentiable at $\bar{x}$, then

$$\nabla L(\bar{x},\bar{y},\alpha) = \begin{bmatrix} [\nabla_{11}L^0(\bar{x},\bar{u})+\sum_{i\in I}\alpha\phi'(\bar{y}_i)\cdot & \nabla g_i(\bar{x})\phi'(\bar{y}_i) & 0 \\ \quad\cdot\nabla g_i(\bar{x})\nabla g_i(\bar{x})^T] & (i\in I) & \\ \alpha(\theta'(\bar{y}_i)+\theta'(0))\nabla g_i(\bar{x})^T & 0 & 0 \\ (i\in I) & & \\ 0 & 0 & -\theta'(-\alpha g_i(\bar{x}))-\theta'(0) \\ & & (i\in J)\end{bmatrix} \tag{5}$$

where $\bar{u}_j = \phi(\bar{y}_j), j=1,\ldots,m, L^0(x,u)=f(x)+\sum_{j=1}^{m}u_jg_j(x)$, the classical Lagrangian, $I = \{j|g_j(\bar{x})=0\}$ and $J = \{j|g_j(\bar{x})<0\}$. It follows that by Debreu's theorem [4] (which states that $N + \alpha M^TM$ is positive definite for $\alpha \geqq \bar{\alpha}$ for some $\bar{\alpha} > 0$ if $xNx > 0$ for $Mx = 0$ and $x \neq 0$), by strict complementarity, and by second order sufficiency at $(\bar{x},\bar{u})$ [5] that the upper left $n \times n$ submatrix of (5) which we denote by $\nabla_1 L_1(\bar{x},\bar{y},\alpha)$ is positive definite for $\alpha \geqq \bar{\alpha}$ for some $\bar{\alpha} > 0$. This together with the linear independence of $\nabla g_i(\bar{x})$, $i \in I$ and $-\theta'(-\alpha g_i(\bar{x})) - \theta'(0) < 0$, $i \in J$, insure the nonsingularity of $\nabla L(\bar{x},\bar{y},\alpha)$. Henceforth we shall assume in the paper that strict complementarity, second order sufficiency and linear independence of the gradients of the active constraints hold at $(\bar{x},\bar{u})$.

An immediate algorithm for solving (4) would be a Newton or quasi-Newton algorithm [10,3] which would be locally superlinearly convergent if $\alpha \geq \bar{\alpha}$ for some $\bar{\alpha} > 0$. A concrete realization of (4) to be solved by such methods can be obtained by setting

$$\theta(z) = \phi(z) = \frac{1}{2} z|z|, \; \theta'(z) = \phi'(z) = |z| \tag{6}$$

Note that $\theta'(0) = \phi'(0) = 0$ and hence equations (4) are globally differentiable. Probably a more interesting and effective method would be the following one which is an asymptotic Newton method in the y-space and which for a specific $\theta(z)$ and $\phi(z)$ becomes the extension to inequalities of Powell's method [11] .

5. Algorithm Choose $\alpha > 0$, $y^0 \in R^m$ and $x^0 \in R^n$ such that $L_1(x^0,y^0,\alpha) = 0$. Let (x^i,y^i) determine (x^{i+1},y^{i+1}) as follows:

(a) Determine x^{i+1} such that $L_1(x^{i+1},y^i) = 0$. If x^{i+1} is not unique, take a closest x^{i+1}, in some norm, to x^i .

(b) $y^{i+1} = y^i + \alpha L_2(x^{i+1},y^i,\alpha)$.

Note that this algorithm becomes Powell's algorithm extended to inequalities if we take $2\alpha\theta(z) = \phi(z) = z$. However we shall propose another choice for $\theta(z)$ and $\phi(z)$ to use in algorithm 5 which will make $L(x,y,\alpha)$ differentiable globally. Before doing that however, we consider the local convergence of algorithm 5 and its rate of convergence under the sim-simplifying assumption that x^{i+1} is unique in step (a) of the algorithm. (The non-unique case can be handled in a manner similar to that in [9, theorem 4.10].) Let ϕ and θ be continuously differentiable on R and let f and g be twice continuously differentiable in a neighborhood of $\bar{x}$. Then by the implicit function theorem, for y^i sufficiently close to $\bar{y}$ and $\alpha \geq \bar{\alpha} > 0$, algorithm 5 is equivalent to

$$y^{i+1} = y^i + \alpha L_2(e(y^i),y^i,\alpha) \tag{7}$$

where e is a differentiable function from R^m into R^n satisfying $L_1(e(y),y,\alpha) = 0$ for all y in a neighborhood of $\bar{y}$. Consider now the mapping $G(y) = y + \alpha L_2(e(y),y,\alpha)$ underlying the iteration (7). By using $\nabla_1 L_1(\bar{x},\bar{y},\alpha)\nabla e(\bar{y}) + \nabla_2 L_1(\bar{x},\bar{y},\alpha) = 0$ and (5) we obtain that

$$\begin{aligned}\nabla G(\bar{y}) &= I + \alpha\nabla_1 L_2(\bar{x},\bar{y},\alpha)\nabla e(\bar{y}) + \alpha\nabla_2 L_2(\bar{x},\bar{y},\alpha) \\ &= I - \alpha(\nabla_1 L_2(\bar{x},\bar{y},\alpha)\nabla_1 L_1(\bar{x},\bar{y},\alpha)^{-1}\nabla_2 L_1(\bar{x},\bar{y},\alpha) - \nabla_2 L_2(\bar{x},\bar{y},\alpha))\end{aligned}$$

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

where $\theta'(\bar{y}_I) + \theta'(0)$ denotes a diagonal matrix with diagonal elements $\theta'(y_i) + \theta'(0)$, $i \in I$. Similarly $\phi'(\bar{y}_I)$ and $\theta'(-\alpha g_J(\bar{x})) + \theta'(0)$ are also diagonal matrices. The matrix $\nabla g_I(\bar{x})$ has as its columns $\nabla g_i(\bar{x})$, $i \in I$. We now make use of the following extremely useful lemma which can be considered a key lemma for obtaining convergence rate results for unconstrained Lagrangians. In a lemma in [11] Powell proved a related result by using determinants.

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

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

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

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

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

$$\nabla G(\bar{y}) = I - \alpha\begin{bmatrix} \alpha(\theta'(\bar{y}_I)+\theta'(0))(\alpha\phi'(\bar{y}_I)+K)^{-1}\phi'(\bar{y}_I) & 0 \\ 0 & \theta'(-\alpha g_J(\bar{x}))+\theta'(0) \end{bmatrix} \tag{7}$$

If $2\alpha\theta(z) = \phi(z) = z$, as in Powell's method, then

$$\nabla G(\bar{y}) = I - \alpha\begin{bmatrix} (\alpha I+K)^{-1} & 0 \\ 0 & \frac{I}{\alpha} \end{bmatrix} = \begin{bmatrix} I - (I+\frac{K}{\alpha})^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{7a}$$

Hence for sufficiently large α, $\rho(\nabla G(\bar{y})) \leq \frac{2\|K\|}{\alpha}$

where $\rho(\nabla G(\bar{y}))$ is the spectral radius of $\nabla G(\bar{y})$. Hence by Ostrowski's point of attraction theorem [10, theorem 10.1.3] and [10,10.1.4] the sequence $\{y^i\}$ is locally convergent to $\bar{y}$ and the root convergence factor

$$R_1\{y^i\} = \limsup_{i\to\infty} \|y^i - \bar{y}\|^{\frac{1}{i}} \leq 2\frac{\|K\|}{\alpha}$$

As $\alpha \to \infty$ this factor approaches zero and hence we have parametric root-superlinear convergence, the parameter being α of course.

We propose now another realization of algorithm 5 based on the following following functions

$$\phi(z) = \frac{1}{2} z|z| \qquad \phi'(z) = |z| \tag{8}$$

$$\theta(z) = \begin{cases} \frac{z}{\alpha} + \frac{1}{2\alpha^2} & \text{if } z \leqq -\frac{1}{\alpha} \\ \frac{1}{2} z|z| & \text{if } |z| < \frac{1}{\alpha} \\ \frac{z}{\alpha} - \frac{1}{2\alpha^2} & \text{if } z \geqq \frac{1}{\alpha} \end{cases}, \quad \theta'(z) = \begin{cases} \frac{1}{\alpha} & \text{if } z \leqq -\frac{1}{\alpha} \\ |z| & \text{if } |z| < \frac{1}{\alpha} \\ \frac{1}{\alpha} & \text{if } z \geqq \frac{1}{\alpha} \end{cases} \tag{9}$$

Note that $\theta'(0) = 0$ and that for sufficiently large but finite α, $\frac{1}{\alpha} \leqq \min_{j\in I} \bar{y}_j$ and $\frac{1}{\alpha} \leqq \min_{j\in J} (-\alpha g_j(\bar{x}))$. Hence for α large enough

$$\left.\begin{aligned} \theta'(\bar{y}_j) &= \frac{1}{\alpha} \quad \text{for } j \in I \\ \theta'(-\alpha g_j(\bar{x})) &= \frac{1}{\alpha} \quad \text{for } j \in J \end{aligned}\right\rangle \tag{10}$$

Substitution of (8) to (10) in (7) gives

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

where $\bar{y}_I$ is a diagonal matrix with diagonal elements $\bar{y}_i$, $i \in I$. Hence for sufficiently large α, $\rho(\nabla G(\bar{y})) \leqq \frac{2\,\|\bar{y}_I^{-1}K\|}{\alpha}$ and again as in the extended Powell method, we have a locally convergent sequence $\{y^i\}$ with parametric root-superlinear convergence.

For concreteness we spell out algorithm 5 when the functions (8) and (9) are used.

7 Algorithm Choose $\alpha > 0$, $y^0 \in R^m$ and $x^0 \in R^n$ such that $\nabla f(x^0) + \frac{1}{2}\sum_{j=1}^{m} (\alpha g_j(x^0) + y_j^0)_+^2 \nabla g_j(x^0) = 0$. Let (x^i, y^i) determine (x^{i+1}, y^{i+1}) as follows:

(a) Determine x^{i+1} such that $\nabla f(x^{i+1}) + \frac{1}{2}\sum_{j=1}^{m} (\alpha g_j(x^{i+1}) +$

$$y_j^i)_+^2 \nabla g_j(x^{i+1}) = 0$$

or

$$f(x^{i+1}) + \frac{1}{6\alpha}\sum_{j=1}^{m} (\alpha g_j(x^{i+1}) + y_j^i)_+^3 = \min_{x\in R^n} f(x) + \frac{1}{6\alpha}\sum_{j=1}^{m} (\alpha g_j(x) + y_j^i)_+^3$$

If x^{i+1} is not unique find a closest x^{i+1} to x^i in some norm.

(b) $y_j^{i+1} = y_j^i + \alpha(\theta(|\alpha g_j(x^{i+1})+y_j^i|) + \theta(\alpha g_j(x^{i+1})) - \theta(y_j^i))$

$j = 1,\ldots,m$, where θ is given by (9)

Note that in step (a) of algorithm (7) the function to be minimized is twice globally differentiable in x. This is unlike the function to be minimized in the extension of Powell's method which is

$f(x) + \frac{1}{2\alpha}\sum_{j=1}^{m}(\alpha g_j(x)+y_j)_+^2$ and which is not twice differentiable globally.

A possible difficulty in our proposed method is that as $\alpha \to \infty$ the function $\theta(z)$ of (9) approaches the zero function. This however is also true of Powell's function $\theta(z) = \frac{z}{2\alpha}$. Another realization of algorithm 5 which for large enough α gives (7a), the same $\nabla G(\bar{y})$ as Powell's method, is obtained from $\theta(z)$ of (9) and $\phi(z) = \alpha\theta(z)$.

In conclusion we state that the celebrated complementarity problem [7] of finding an s in R^ℓ satisfying

$$sF(s) = 0 \qquad F(s) \leqq 0 \qquad s \geqq 0 \tag{11}$$

where $F:R^\ell \to R^\ell$ is, by lemma 3 equivalent, to solving the system of equations

$$\theta(|F_i(s)+s_i|) + \theta(F_i(s)) - \theta(s_i) = 0 \quad i = 1,\ldots,\ell \tag{12}$$

where θ is any strictly increasing odd function from R into R with $\theta(0) = 0$. In particular if we take $\theta(z) = z|z|$, equations (12) become

$$(F_i(s)+s_i)^2 + F_i(s)|F_i(s)| - s_i|s_i| = 0 \quad i = 1,\ldots,\ell \tag{13}$$

It can be shown that the Jacobian of (13) evaluated at a solution $\bar{s}$ which satisfies the strict complementarity condition $\bar{s}_i + |F_i(\bar{s})| > 0$, $i = 1,\ldots,\ell$, is nonsingular provided that the Jacobian $\nabla F(\bar{s})$ has nonsingular principal minors. This will be discussed elsewhere in more detail. Note that equations (13) are globally differentiable if $F(s)$ is globally differentiable.

ACKNOWLEDGEMENT

I am indebted to my student S. M. Chung for reading the manuscript and for a simplification in the proof of lemma 3.

REFERENCES

1. D. P. Bertsekas: "Combined primal dual and penalty methods for constrained minimization," SIAM J. Control, 13(3), August 1975.
2. 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.
3. C. G. Broyden, J. E. Dennis & J. J. Moré: "On the local and superlinear convergence of quasi-Newton methods," J. Inst. Maths. Applics. 1973, 12, 223-245.
4. G. Debreu: "Definite and semidefinite quadratic forms," Econometrica 20, 1952, 295-300.
5. A. V. Fiacco & G. P. McCormick: "Nonlinear programming: sequential unconstrained minimization techniques", Wiley, New York, 1968.

6. R. Fletcher: "An ideal penalty function for constrained optimization," U.K.A.E.A. Report C.S.S.2, December 1973.

7. F. J. Gould & J. W. Tolle: "A unified approach to complementarity in optimization," Discrete Mathematics, 7, 1974, 225-271.

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

9. O. L. Mangasarian: "Unconstrained Lagrangians in nonlinear programming" SIAM J. Control, 13(3) August 1975.

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

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

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