Lagrange Multiplier Methods for Minimization under Equality Constraints

Glad, Torkel

1973

3056

LAGRANGE MULTIPLIER METHODS FOR MINIMIZATION UNDER EQUALITY CONSTRAINTS

T. GLAD

Report 7323 August 1973
Lund Institute of Technology
Division of Automatic Control

LAGRANGE MULTIPLIER METHODS FOR MINIMIZATION UNDER EQUALITY CONSTRAINTS.

T. Glad

Abstract.

Different ways of converting the minimization problem under equality constraints into an unconstrained problem are considered. The methods of Hestenes [3], Powell [4], Fletcher [5] and Mårtensson [7], based on Lagrange multiplier theory, are discussed and some theoretical results concerning their properties are presented. These methods and a modification of the method of Hestenes and Powell are compared with the classical penalty function method on a number of numerical test problems.

TABLE OF CONTENTS. Page

2. PRESENTATION OF THE OPTIMIZATION PROBLEM.

This report deals with optimization under equality constraints. The problem can be formulated

(P) minimize $f(x)$ with respect to $x \in R^n$
subject to $h(x) = 0$
where $f: \mathbb{R}^n \to \mathbb{R}$
and $h: \mathbb{R}^n \to \mathbb{R}^m$

Apart from differentiability assumptions and the "constraint qualification" defined below $f(x)$ and $h(x)$ can be arbitrary functions.

To study this problem it is convenient to introduce the Lagrangian $L(x,\mu)$.

$$L(x,\mu) \hat{=} f(x) + \mu^T h(x) \qquad \mu \in R^m$$

The components of μ are called Lagrange multipliers.

The following theorems form the basis for the discussion of (P).

<u>Theorem 2.1</u>. Suppose that f and h are twice continuously differentiable and that $h_x(x^*)$ has full rank. A necessary condition for x^* to be a solution to (P) is then that there exists a μ^* such that (x^*,μ^*) satisfies

(i) $h(x^*) = 0$

(ii) $L_x(x^*,\mu^*) = 0$

(iii) for every $y \neq 0$ such that $h_x(x^*)y = 0$ it follows that $y^T L_{xx}(x^*,y^*)y \geq 0$.

1. INTRODUCTION.

In recent years great progress has been made in the development of numerical methods for the minimization of a function of several variables. A description of some of these methods can be found in references [9]-[12]. In practise the independant variables often can not vary freely, but are confined to a certain region by equality or inequality constraints. It is possible to develop special methods to deal with these constraints and this has been done to some extent. However it is very attractive, considering the advances in unconstrained minimization, to convert the problem to an unconstrained one. In this report some ways of doing this for equality constraints will be considered. The attention will be focused on methods that use the theory of Lagrange multipliers.

Proof: See [1].

Theorem 2.2. A sufficient condition for x^* to be an isolated local minimum of (P) is that there exists a μ^* such that

(i) $h(x^*) = 0$

(ii) $L_x(x^*,\mu^*) = 0$

(iii) for every y with $h_x(x^*)y = 0$ and $y \neq 0$ it follows that $y^T L_{xx}(x^*,\mu^*)y > 0$.

Proof: See [1].

The condition that $h_x(x^*)$ has full rank is called a constraint qualification. Since it will always be assumed from now on that this condition holds, the conditions (i) - (iii) of Theorem 1 will always hold at the minimum. If the stronger conditions of Theorem 2 are satisfied the problem is said to be nonsingular. The solution of (P) which is assumed to exist will in this report always be called x^* and the corresponding multiplier μ^*.

3. A SURVEY OF SOME NUMERICAL METHODS PROPOSED IN THE LITERATURE.

This survey contains methods that solve the minimization problem with equality constraints by converting it to an unconstrained optimization problem. A good reason for doing this is that very effective algorithms for unconstrained optimization have been developed, for instance some of the Quasi-Newton methods, see [9], [10], [11].

3.1. The Penalty Function Method.

This is the classical method of solving (P). The function

$$F(x) = f(x) + \frac{1}{r} h^T(x)h(x)$$

is minimized for a sequence of decreasing values of r. Let x_k be the minimum for $r = r_k$. Then it can be shown that under mild assumptions about f and h, $x_k \to x^*$ when $r_k \to 0$. The disadvantage of the method is that F becomes very ill-conditioned when r is small. The algorithm is often used together with Richardson extrapolation : The minimum can be thought of as a function of r, x(r) and can be represented by a power series.

$$x(r) = x(0) + a_1 r + a_2 r^2 + \ldots$$

When two points $x(r_1)$ and $x(r_2)$ have been computed an extrapolated value x(0) can be calculated from the equations which remain when the series is truncated after the second term :

$$\begin{cases} x(r_1) = x(0) + a_1 r_1 \\ x(r_2) = x(0) + a_1 r_2 \end{cases}$$

When three points are computed an extrapolation can be done using the first three terms and so on. In a similar way it is possible to predict where the minimum will be for the next value of r. This can be used in starting up the unconstrained optimization algorithm. Detailed computational schemes for the extrapolation are given in [1].

3.2. Methods Which Use the Lagrangian.

The purpose of these methods is to create a function which has an unconstrained minimum at the point which is a solution to (P). A good candidate seems to be the Lagrangian $L(x,\mu)$, since for $\mu = \mu^*$ it has a stationary point at $x = x^*$. Suppose that this stationary point is a minimum and also that $\min_x L(x,\mu)$ exists for all μ in a neighbourhood of μ^*. Then it is possible to define

$$G(\mu) = \min_x L(x,\mu).$$

It then follows that

$$G(\mu^*) = L(x^*,\mu^*) = f(x^*)$$

$$G(\mu) = \min_x L(x,\mu) \leq L(x^*,\mu) = f(x^*) = G(\mu^*).$$

This means that

$$\max_\mu G(\mu) = G(\mu^*) = f(x^*)$$

Consequently the optimization problem can be solved by minimizing with respect to x in an inner loop and maximizing with respect to μ in an outer loop. It is also possible to show that derivatives of G can be calculated. A survey of these ideas is given in [2].

Unfortunately this approach is only possible for certain functions f and h. In many cases $\min_x L(x,\mu)$ does not exist for any μ. This has been the incentive for the following modifications of the Lagrangian.

3.3. Methods That Use a Modified Lagrangian and Iterate on the Multipliers.

A suggestion of this kind of method has been made by Hestenes [3]. He studied the function

$$F(x,\mu) = f(x) + \mu^T h(x) + c h^T(x) h(x)$$

where c is a positive constant.

This can be thought of as a combination of the Lagrangian and a penalty function. In contrast to the penalty function methods the parameter c is not increased towards infinity but held at a constant value. In this way the ill-conditioning, which is characteristic of the penalty function methods, is avoided. Hestenes showed that $F(x,\mu^*)$ has a local minimum at x^* if c is chosen large enough and if the problem is nonsingular. This result can be viewed as a special case of the theorems given later in 3.5.

To use this result μ^* must be known. Hestenes suggested the following iterative scheme.

1) put $k = 0$; select μ^0.

2) minimize $F(x,\mu^k)$; call the minimum x^k.

3) update μ according to

$$\mu^{k+1} = \mu^k + 2ch(x^k)$$

4) if $\|h(x^k)\|$ is sufficiently small stop, otherwise increment k to k+1 and go to 2).

From the fact that $F(x,\mu^k)$ is minimized it follows that

$$f_x(x^k) + \mu^{k^T} h_x(x^k) + 2ch^T(x^k)h_x(x^k) = 0$$

which means that

$$f_x(x^k) + \mu^{k+1^T} h_x(x^k) = 0$$

Since at the minimum μ^* satisfies

$$f_x(x^*) + \mu^{*T} h_x(x^*) = 0$$

this is a natural way of updating μ which turns out to work quite well in practice.

3.4. Powell's Method.

A method working along similar lines has been developed by Powell [4]. The difference is that in Powell's method the function to be minimized is

$$F(x,\mu) = f(x) + \mu^T h(x) + \sum_{i=1}^{m} c_i h_i(x)^2$$

The different components of h are given different positive weights c_i. The updating of μ is done in the same way as in the previous section. The parameters c_i are adjusted in the following way:

If $\|h(x^k)\|_\infty < \frac{1}{4}\|h(x^{k-1})\|_\infty$ then the parameters c_i remain the same. If this condition is not satisfied the c_i corresponding to components of h, that are too large, are increased by a factor 10. Powell shows that his strategy ensures convergence for quite general functions.

3.5. Methods That Use Lagrange Multiplier Functions.

In these methods the Lagrange multipliers are replaced by functions $\tilde{\mu}(x)$. Fletcher [5] and Mårtensson [7] have studied the following function:

$$\phi(x) = f(x) + \tilde{\mu}(x)^T h(x) + c h^T(x) h(x)$$

where

$$\tilde{\mu}(x) = -\left(h_x(x) h_x^T(x)\right)^{-1} h_x(x) f_x^T(x).$$

With this definition $\tilde{\mu}(x^*) = \mu^*$ and $\phi_x(x^*) = L_x(x^*,\mu^*) = 0$.

In [5] and [7] it is shown that, if the problem is nonsingular and c great enough, $\phi_{xx}(x^*)$ is positive definite. ϕ then has a local minimum at x^*. Fletcher has also studied the function

$$\phi_2(x) = f(x) + \tilde{\mu}^T(x)h(x) + ch^T(x)\left(h_x(x)h_x^T(x)\right)^{-1}h(x)$$

A problem with the choice of $\tilde{\mu}(x)$ is that the computation of ϕ itself requires derivatives of f and h, while for ϕ_x second derivatives of f and h would be needed. In [6] Fletcher discusses this problem. He suggests that in the formula

$$\phi_x = f_x + \tilde{\mu}^T h_x + h^T\tilde{\mu}_x + 2ch^T h_x$$

one should replace $\tilde{\mu}_x$ with an approximation which is updated so that $\tilde{\mu}_x \Delta x = \Delta\mu$. In this way a minimization method using the gradient can be used even if f_{xx} and h_{xx} are not available.

A generalization of these ideas is given by Mårtensson in [7]. He studies the function

$$F(x) = f(x) + \tilde{\mu}^T(x)h(x) + ch^T(x)h(x)$$

where $\tilde{\mu}(x)$ is an unspecified function, and formulates the criteria $\tilde{\mu}(x)$ has to satisfy. These are

(i) $\tilde{\mu}(x)$ exists and is twice differentiable in a neighbourhood of x^*.

(ii) $\tilde{\mu}(x^*) = \mu^*$.

Then it can be shown that

Theorem 3.1. If the sufficient conditions of Theorem 2 are satisfied and c is great enough F(x) has a local minimum at $x = x^*$ with $F_{xx}(x^*) > 0$.

If the problem is singular a third condition on $\tilde{\mu}(x)$ is needed.

(iii) for every $y \neq 0$ such that $h_x(x^*)y = 0$ and $y^T L_{xx}(x^*,\mu^*) = 0$ $\tilde{\mu}(x)$ satisfies

$$\left\{h_x(x^*)L_{xx}(x^*,\mu^*) + h_x(x^*)h_x^T(x^*)\tilde{\mu}_x(x^*)\right\}y = 0$$

Then the following theorem can be shown.

Theorem 3.2. Assume that $h_x(x^*)$ has full rank and that $\tilde{\mu}(x)$ satisfies conditions (i) - (iii). Then there exists a c_0 such that $F_x(x^*) = 0$ and $F_{xx}(x^*) \geq 0$ for $c > c_0$.

It can be shown that the choice $\tilde{\mu}(x) = -\left(h_x(x)h_x^T(x)\right)^{-1} \cdot h_x(x)f_x^T(x)$ satisfies the conditions (i) - (iii). Obviously $\tilde{\mu}(x) = \mu^*$ satisfies (i) and (ii), but this choice does not satisfy (iii) for arbitrary f and h.

Mårtensson has also shown the connection between the multiplier function method and Hestenes' method. Suppose the following algorithm is used.

a) set $k = 0$; select an x^0.

b) compute $\mu^k = -\left(h_x(x^k)h_x^T(x^k)\right)^{-1}h_x(x^k)f_x^T(x^k)$.

c) minimize $F(x) = f(x) + \mu^{k^T}h(x) + ch^T(x)h(x)$. Call the minimum x^{k+1}.

d) if $\|h(x^{k+1})\|$ is sufficiently small then stop else increment k to k+1 and go to b).

Then this algorithm is equivalent to Hestenes's method. That follows because x^{k+1} satisfies $f_x(x^{k+1}) + \mu^{k^T} h_x(x^{k+1}) + 2ch^T(x^{k+1})h_x(x^{k+1}) = 0$. The updating formula used by Hestenes would now give a μ^{k+1} that satisfies

$$f_x(x^{k+1}) + \mu^{k+1^T} h_x(x^{k+1}) = 0$$

or

$$\mu^{k+1} = -\left(h_x(x^{k+1})h_x^T(x^{k+1})\right)^{-1} h_x(x^{k+1})f_x^T(x^{k+1})$$

which is the expression used in step b).

4. SOME THEORETICAL RESULTS FOR LAGRANGE MULTIPLIER FUNCTIONS:

In the previous section it was shown that the function

$$F(x) = f(x) + \tilde{\mu}^T(x)h(x) + ch^T(x)h(x)$$

has a local minimum at x* if the problem is nonsingular and c great enough , if $\tilde{\mu}(x) = -(h_x h_x^T)^{-1} h_x f_x^T$ or $\tilde{\mu}(x) = \mu^*$. For $\tilde{\mu}(x) = \mu^*$ this result cannot be extended to singular problems which is shown by the following example.

$$f = x_2^4 + x_1 x_2$$

$$h = x_1$$

Here $x^* = (0,0)$ and $\mu^* = 0$ and

$$F(x) = x_2^4 + x_1 x_2 + cx_1^2$$

$$F_{xx}(0) = \begin{pmatrix} 2c & 1 \\ 1 & 0 \end{pmatrix}$$

$F_{xx}(0)$ is indefinite for all c which means that x* is not a minimum of F(x).

In the previous section it was shown that for multiplier functions satisfying the extra third condition the second derivative of F(x) could be made positive semidefinite. This is, however, not a sufficient condition for a local minimum. The following theorem shows that for $\tilde{\mu}(x) = -(h_x h_x^T)^{-1} h_x f_x^T$ it is indeed possible to choose c so that F(x) has a local minimum.

Theorem 4.1. Suppose that f and h are twice continuously differentiable and that $h_x(x^*)$ has full rank. Then there is a c_0 such that $F(x) = f(x) + \tilde{\mu}^T(x)h(x) + ch^T(x)h(x)$, where $\tilde{\mu}(x) = -(h_x h_x^T)^{-1} h_x f_x^T$, has a local minimum at x^* for $c > c_0$.

Proof: Since $h_x(x^*)$ has full rank and h_x is continuous it follows that there is a $\delta > 0$ such that $h_x(x)$ has full rank in $M = \{x \mid \|x-x^*\| \leq \delta\}$. Then there exists a constant K such that all elements of $h_x^T(h_x h_x^T)^{-1}$ are less than $K/\sqrt{n}$ for all $x \in M$.

From the facts that $h(x^*) = 0$ and $h(x)$ is continuous it follows that there exists a δ' such that $\|h(x)\| \leq \delta/2K$ if $\|x-x^*\| \leq \delta'$. Let $\delta'' = \min(\delta/2, \delta')$ and define $M' = \{x \mid \|x-x^*\| \leq \delta''\}$. Then $\|h(x)\| < \delta/2K$ for all $x \in M'$ and $M' \subset M$.

Let x_0 be any point in M' with $h(x_0) \neq 0$ and study the differential equation

$$\begin{cases} x'(t) = -h_x^T(h_x h_x^T)^{-1} h & (*) \\ x(0) = x_0 \end{cases}$$

The right hand side of (*) is a continuously differentiable function on M and therefore satisfies a Lipschitz condition which means that a solution exists in a neighbourhood of x_0. Study the function

$P(t) = h^T(x(t))h(x(t))$. It satisfies $P'(t) = -2P(t)$ which gives $P(t) = P(0)e^{-2t}$. Along the curve $x(t)$ it is consequently true that

$$\|h(x(t))\| = \|h(x_0)\| e^{-t}$$

Since

$$x(t) = x_0 - \int_0^t h_x^T(h_x h_x^T)^{-1} h\,dt$$

it follows that

$$\| x(t)-x^* \| \leq \| x_0-x^* \| + K\| h(x_0) \| \int_0^t e^{-t}dt \leq \delta'' + \delta/2 \leq \delta$$

The solution $x(t)$ therefore never leaves M and can be continued indefinitely. From $\| x'(t) \| \leq K\| h(x_0) \| e^{-t}$ it follows that

$$\lim_{t\to\infty} x(t) = \bar{x}$$

exists and from $\| h(x(t))\| = \| h(x_0) \| e^{-t}$ it follows that $h(\bar{x}) = 0$.

Now study $F(x_0)$.

$$F(x_0) - F(\bar{x}) = - \int_0^\infty F_x x'(t)dt =$$

$$= + \int_0^\infty [h^T \tilde{\mu}_x h_x^T(h_x h_x^T)^{-1}h + 2ch^T h]dt$$

The function $\tilde{\mu}_x h_x^T(h_x h_x^T)^{-1}$ is continuous on M and therefore there exists a constant C such that $|h^T\tilde{\mu}_x h_x^T(h_x h_x^T)^{-1}h| \leq$ $\leq C\| h \|^2$ for all x in M. If c is chosen such that $c > C/2$ then

$$\int_0^\infty [h^T\tilde{\mu}_x h_x^T(h_x h_x^T)^{-1}h + 2ch^T h]dt \geq (2c-C) \int_0^\infty \| h(x) \|^2 dt =$$

$$= (2c-C)\| h(x_0) \|^2 > 0$$

Therefore $F(x_0) > F(\bar{x}) \geq F(x^*)$ where the last inequality follows from the fact that $\bar{x}$ satisfies $h(\bar{x}) = 0$. Consequently $F(x_0) > F(x^*)$ for all x with $\| x-x^* \| \leq \delta$ and $h(x) \neq 0$. Since $F(x) \geq F(x^*)$ for all x with $h(x) = 0$ it follows that x^* is a local minimum.

Corollary. If $f(x) > f(x^*)$ for all $x \neq x^*$ with $h(x) = 0$, then $F(x) > F(x^*)$ for all $x \neq x^*$ such that $\| x-x^* \| \leq \delta$.

So far it has only been proved that in certain cases x^* is a local minimum of $F(x)$. This means that there exists a $\delta > 0$ such that $F(x) \geq F(x^*)$ for all x satisfying $\| x-x^* \| < \delta$. In the following theorems it is proved that for a large class of problems δ can be made arbitrarily large by choosing c great enough.

Theorem 4.2. Assume that f and h are such that $f(x) > f(x^*)$ for all $x \neq x^*$ such that $h(x) = 0$. Let $F(x) = f(x) + \tilde{\mu}^T(x) \cdot h(x) + ch^T(x)h(x)$ where $\tilde{\mu}(x)$ is a continuous function with $\tilde{\mu}(x^*) = \mu^*$. Assume that for $c \geq c_0$ $F(x)$ has a proper local minimum at x^*. (From the preceding results it follows that this is true for $\tilde{\mu}(x) = -(h_x h_x^T)^{-1} h_x f_x^T$ and, if the problem is nonsingular, for $\tilde{\mu}(x) = \mu^*$.) Then for every $A > 0$ there exists a constant $\delta > 0$ such that, for $c \geq c_0$, it holds that $F(x) > F(x^*)$ for all $x \neq x^*$ with $d(x,M_h) \leq \delta$,*) where $M_h = \{x | h(x) = 0 \text{ and } \| x-x^* \| \leq A\}$.

Proof: Assume to begin with that $c = c_0$. Since $F(x)$ has a proper local minimum at x^* there is a constant $d > 0$ such that $F(x) > F(x^*)$ for all $x \neq x^*$ with $\| x-x^* \| \leq d$. Define $M_1 = \{x | \frac{d}{2} \leq \| x-x^* \| \leq d\}$ and

$$\varepsilon_1 = \min_{x \in M_1} \left(F(x) - F(x^*)\right) > 0.$$

*) $d(x,M_h)$ is defined as $\inf_{y \in M_h} \{ ||x-y|| \}$

Also define

$$M = \{x | h(x) = 0, \frac{d}{2} \leq \| x-x^* \| \leq A\}$$

and

$$\varepsilon_M = \min_{x \in M}\left(F(x) - F(x^*)\right) > 0$$

$F(x)$ is uniformly continuous in $M_A = \{x | \| x-x^* \| \leq A\}$ so there is a $\delta' > 0$ such that $\| F(x)-F(y) \| \leq \varepsilon_M/2$ for all $x,y \in M_A$ with $\| x-y \| \leq \delta'$. Put $\delta = \min(\delta', d/2)$. Study $M' = \{x | d(x,M) \leq \delta\}$. If $x \in M'$ there exists a $y \in M$ with $- F(x) + F(y) \leq \varepsilon_M/2$ and $F(y) \geq F(x^*) + \varepsilon_M$ so it follows that $F(x) \geq F(x^*) + \varepsilon_M/2$. Now take an arbitrary $x \neq x^*$ with $d(x,M_h) \leq \delta$. If $x \in M'$ then $F(x) \geq F(x^*) + \varepsilon_M/2 > F(x^*)$. Suppose that $x \notin M'$. Then there exists $y \in M_h$ such that $\| x-y \| \leq \delta$ and $y \notin M$. Therefore $\| y-x^* \| < d/2$ and $\| x-x^* \| \leq$ $\leq \| x-y \| + \| y-x^* \| < \delta + d/2 \leq d$ which means that $F(x) >$ $> F(x^*)$. Consequently $F(x) > F(x^*)$ for all $x \neq x^*$ with $d(x,M_h) \leq \delta$. Since $F(x^*) = f(x^*)$ and, for all other values of x, $F(x)$ increases with increasing c, the theorem is also true for all $c \geq c_0$.

<u>Theorem 4.3</u>. Assume that the conditions of Theorem 4.2 hold. Then for every $A > 0$ there is a c_1 such that $F(x) > F(x^*)$ for all $x \neq x^*$ with $\| x-x^* \| \leq A$ if $c \geq c_1$.

Proof: According to Theorem 4.2 there is a $\delta > 0$ such that $F(x) > F(x^*)$ for $x \neq x^*$ with $d(x,M_h) \leq \delta$ if $c \geq c_0$. Define $M' = \{x | \| x-x^* \| \leq A \text{ and } d(x,M_h) \geq \delta\}$,

$$m_1 = \min_{x \in M'} \| h(x) \|^2 > 0 \quad \text{and} \quad m_2 = \min_{x \in M'} \left[f(x) + \tilde{\mu}^T(x)h(x)\right]$$

Then for $x \in M'$ $F(x) \geq m_2 + cm_1 > F(x^*)$ if

$$c > \max\left(c_0,\ \frac{F(x^*) - m_2}{m_1}\right)$$

Since it is possible to make x^* the minimum of an arbitrarily large part of $\mathbb{R}^n$ one could ask if it is possible to make x^* a global minimum by choosing c large enough. That this is not possible is shown by the following example.

Take $f(x) = x_2^2 - x_1^3 x_2^2$, $h(x) = x_1$.

Then $x^* = 0$ and $\mu^* = 0$.

For $\tilde{\mu}(x) = \mu^*$ one gets

$$F_1(x) = x_2^2 - x_1^3 x_2^2 + c x_1^2$$

Let x_1 be constant, $x_1 = \alpha > 1$. Then $F_1 = c\alpha^2 - (\alpha^3 - 1)x_2^2$ which is monotonically decreasing towards $-\infty$ when x_2 varies from 0 to ∞. This reveals a fundamental difficulty when using numerical methods to find x^*. Suppose the method is started at $x = (x_1, x_2)$ with $x_1 > 1$. $F(x)$ can be made greater than $F(x^*)$ by choosing c great enough, but this is no guarantee that the algorithm will find x^*. It might just as well move towards infinity. If instead $\tilde{\mu}(x) = - (h_x h_x^T)^{-1} h_x f_x^T$ one gets $F_2(x) = x_2^2 + 2x_1^3 x_2^2 + c x_1^2$ which has the same behaviour.

To be sure that these difficulties can be avoided one has to make more assumptions about $f(x)$.

Theorem 4.4. Assume that

the minimum is nonsingular
$f(x) > f(x^*)$ for all $x \neq x^*$ with $h(x) = 0$
$h_x(x^*)$ has full rank
$f(x) \to \infty$ when $\| x \| \to \infty$.

Then there exists a c_0 such that for $c > c_0$ $F(x) = f(x) + \mu^{*T}h(x) + ch^T(x)h(x)$ has a unique global minimum at x^* and $F(x) \to \infty$ when $\| x \| \to \infty$.

Proof: Since

$$F(x) = f(x) + \mu^{*T}h(x) + ch^T(x)h(x) =$$

$$= f(x) + c\left(h(x) + \frac{\mu^*}{2c}\right)^T\left(h(x) + \frac{\mu^*}{2c}\right) - \frac{\mu^{*T}\mu^*}{4c} \geq$$

$$\geq f(x) - \frac{\mu^{*T}\mu^*}{4c} \geq f(x) - \frac{\mu^{*T}\mu^*}{4c_1} \quad \text{for all } c \geq c_1$$

it follows that $F(x) \to \infty$ when $\| x \| \to \infty$. Consequently there is an $A > 0$ such that $F(x) > F(x^*)$ for $\| x-x^* \| > A$ and $c \geq c_1$. But according to Theorem 4.3 there is a c_2 such that $F(x) > F(x^*)$ for $x \neq x^*$ and $\| x-x^* \| \leq A$ if $c \geq c_2$. Finally choose $c_0 = \max(c_1, c_2)$.

The theorem is not true when applied to $F(x) = f(x) + \tilde{\mu}^T(x)h(x) + ch^T(x)h(x)$ with $\tilde{\mu}(x) = -(h_x h_x^T)^{-1}h_x f_x^T$. This is shown by the following example.

Take $f(x) = x_2^2 + (1+x_2^4)x_1^2$, $h = x_1$.

Then $\tilde{\mu}(x) = -2x_1(1+x_2^4)$ and $F(x) = x_2^2 - x_1^2(1+x_2^4) + cx_1^2$ which is not bounded from below.

5. THE ITERATION WITH RESPECT TO μ.

5.1. Basic Theory.

To use the results from the previous section about the multiplier function $\tilde{\mu}(x) = \mu^*$ the value of μ^* has to be known. It is then necessary to have some iterative method of determining μ^*. Therefore it is natural to study the function of two variables

$$F(x,\mu) = f(x) + \mu^T h(x) + ch^T(x)h(x)$$

and the function which is a result of the minimization with respect to x

$$G(\mu) = \min_x F(x,\mu).$$

The basis of the theory for $G(\mu)$ is the following theorem.

Theorem 5.1. Suppose that x^* satisfies the sufficient conditions of Theorem 2.2. Let the constant c be chosen such that $F_{xx}(x^*,\mu^*) > 0$. Then there exists an $\varepsilon > 0$ and a $\delta >$ > 0 such that the function

$$G(\mu)_S = \min_{x \in S} F(x,\mu)$$

exists for all $\mu \in D$, where $S = \{x \mid \| x-x^* \| < \varepsilon\}$ and $D =$ $= \{\mu \mid \| \mu - \mu^* \| < \delta\}$. The unique minimum of $F(x,\mu)$ in S is given by $x = \varphi(\mu)$ where φ is continuously differentiable with $\varphi(\mu^*) = x^*$.

Proof: Since F_{xx} is continuous it is possible to find a $\delta_1 > 0$ and an $\varepsilon > 0$ such that $F_{xx}(x,\mu) > 0$ when $x \in S = \{x \mid$ $\| x-x^* \| < \varepsilon\}$ and $\mu \in D_1 = \{\mu \mid \| \mu-\mu^* \| < \delta_1\}$.

Study the equation $F_x(x,\mu) = 0$ for $x \in S$, $\mu \in D_1$. Since $F_x(x^*,\mu^*) = 0$ and $F_{xx}(x^*,\mu^*) > 0$ it follows from the implicit function theorem (see for instance [8]) that there exists a solution of $F_x(x,\mu) = 0$ for $\mu \in D_2 = \{\mu \mid \|\mu-\mu^*\| < \delta_2\}$ if δ_2 is chosen small enough. It also follows that $x = \varphi(\mu)$ with φ continuously differentiable. From the strict convexity of F in S it also follows that $x = \varphi(\mu)$ is the unique minimum of F in S. With $\delta = \min(\delta_1,\delta_2)$ the theorem is proved.

The following theorem shows that μ^* can be computed as the maximum of $G(\mu)_S$.

<u>Theorem 5.2.</u>

$$\max_{\mu \in D} G(\mu)_S = G(\mu^*)_S = f(x^*)$$

Proof: For all $\mu \in D$ it holds that

$$G(\mu)_S = \min_{x \in S} F(x,\mu) \leq F(x^*,\mu) = f(x^*) = F(x^*,\mu^*) =$$

$$= \min_{x \in S} F(x,\mu^*) = G(\mu^*)_S$$

Theorem 5.3. $G(\mu)_S$ is concave in D.

Proof: Let θ be a real number satisfying $0 \leq \theta \leq 1$. Then

$$G\big((1-\theta)\mu_1 + \theta\mu_2\big) = \min_{x\in S}\Big\{f(x) + (1-\theta)\mu_1^T h(x) + \theta\mu_2^T h(x) +$$

$$+ ch^T(x)h(x)\Big\} =$$

$$= \min_{x\in S}\big((1-\theta)F(x,\mu_1) + \theta F(x,\mu_2)\big) \geq$$

$$\geq \min_{x\in S}\Big\{(1-\theta)F(x,\mu_1)\Big\} + \min_{x\in S}\Big\{\theta F(x,\mu_2)\Big\} =$$

$$= (1-\theta)G(\mu_1)_S + \theta G(\mu_2)_S$$

For numerical calculations it is essential to know the derivatives of G which are given by

Theorem 5.4. For $\mu\in D$ $G(\mu)_S$ is twice continuously differentiable with

$$G_\mu(\mu)_S = h\big(\varphi(\mu)\big)$$

$$G_{\mu\mu}(\mu)_S = -\,h_x\big(\varphi(\mu)\big)F_{xx}^{-1}\big(\varphi(\mu),\mu\big)h_x^T\big(\varphi(\mu)\big)$$

where $\varphi(\mu)$ is defined in Theorem 5.1.

Proof:

$$G(\mu)_S = \min_{x\in S} F(x,\mu) = F\big(\varphi(\mu),\mu\big)$$

$$G_\mu(\mu)_S = F_x\big(\varphi(\mu),\mu\big)\varphi_\mu(\mu) + F_\mu\big(\varphi(\mu),\mu\big) = h\big(\varphi(\mu)\big)$$

since $F_x(\varphi(\mu),\mu) = 0$.

$$G_{\mu\mu}(\mu)_S = \frac{d}{d\mu} h(\varphi(\mu)) = h_x(\varphi(\mu))\varphi\ (\mu)$$

φ_μ is computed in the following way. Differentiate the identity $F_x(\varphi(\mu),\mu) = 0$ with respect to μ

$$F_{xx}(\varphi(\mu),\mu)\varphi_\mu(\mu) + F_{x\mu}(\varphi(\mu),\mu) = 0$$

This gives

$$\varphi_\mu(\mu) = - F_{xx}^{-1}(\varphi(\mu),\mu)h_x^T(\varphi(\mu))$$

and the theorem is proved.

If f(x) tends to infinity when $\| x \| \to \infty$ a stronger version of the theorems can be proved. It is then possible to take D and S as the whole of $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively.

Theorem 5.5. Suppose that the conditions of theorem 4.4 are satisfied. Then for $c > c_0$ the function

$$G(\mu) = \min_{x\in\mathbb{R}^n} F(x,\mu)$$

exists for all $\mu\in\mathbb{R}^m$. It is concave and attains its maximum for $\mu = \mu^*$.

Proof:

$$F(x,\mu) = f(x) + c\left(h + \frac{\mu}{2c}\right)^T\left(h + \frac{\mu}{2c}\right) - \frac{\mu^T\mu}{4c}$$

and consequently $F(x) \to \infty$ when $\| x \| \to \infty$. This means that

$F(x,\mu)$ has a global minimum in x so that $G(\mu)$ is well defined. The concavity of $G(\mu)$ is proved in the same way as in Theorem 5.3.

$$G(\mu) = \min_x F(x,\mu) \leq F(x^*,\mu) = f(x^*)$$

From Theorem 4.4 it follows that

$$G(\mu^*) = \min_x F(x,\mu^*) = F(x^*,\mu^*) = f(x^*)$$

so that $G(\mu) \leq G(\mu^*)$.

<u>Theorem 5.6</u>. Suppose that the conditions of Theorem 4.4 are satisfied. Then there exists a $\delta > 0$ such that for $\| \mu-\mu^* \| < \delta$ the minimum of $F(x,\mu)$ is unique and given by $x = \varphi(\mu)$ where φ is continuously differentiable.

Proof: From the continuity of F_{xx} it follows that there are constants $\delta_1 > 0$ and $\varepsilon_1 > 0$ such that $F_{xx}(x,\mu) > 0$ if $\| x-x^* \| < \varepsilon_1$ and $\| \mu-\mu^* \| < \delta_1$.

$$F(x,\mu) = f(x) + c\left(h + \frac{\mu}{2c}\right)^T\left(h + \frac{\mu}{2c}\right) - \frac{\mu^T\mu}{4c} \geq f(x) - \frac{\| \mu \|^2}{4c}$$

Since $f(x) \to \infty$ when $\| x \| \to \infty$ it follows that there exists an $A > 0$ such that for a given $\alpha > 0$

$$f(x) > \frac{(\| \mu^* \| + \delta_1)^2}{4c} + f(x^*) + \alpha \quad \text{if} \quad \| x \| > A$$

If $\| \mu-\mu^* \| < \delta_1$ then

$$F(x,\mu) \geq f(x) - \frac{\| \mu \|^2}{4c} \geq f(x) - \frac{(\| \mu^* \| + \delta_1)^2}{4c} > f(x^*) + \alpha$$

if $\| x \| > A$.

Now study $F(x,\mu)$ on the compact set $M = \{x \mid \varepsilon_1 \leq \| x \| \leq A\}$.
Define

$$\alpha_1 = \min_{x \in M} \left(F(x,\mu^*) - f(x^*)\right) > 0$$

and

$$m = \max_{x \in M} \| h(x) \|$$

Then $F(x,\mu) = F(x,\mu^*) + h^T(x)(\mu-\mu^*) \geq \alpha_1 - m\delta_2 + f(x^*) = \alpha_1/2 + f(x^*)$ if $x \in M$ and $\| \mu-\mu^* \| < \delta_2 = \alpha_1/2m$. Now choose $\delta_3 = \min(\delta_1, \delta_2)$. For $\| \mu-\mu^* \| < \delta_3$

$$\inf_{\| x-x^* \| \geq \varepsilon_1} F(x,\mu) > f(x^*)$$

But $\min F(x,\mu) \leq f(x^*)$.

The global minimum of $F(x,\mu)$ therefore lies in the region $\| x-x^* \| < \varepsilon_1$ where $F_{xx} > 0$. That $x = \varphi(\mu)$ with φ continuously differentiable then follows with the same reasoning as in Theorem 5.1.

<u>Theorem 5.7</u>. For $\| \mu-\mu^* \| < \delta$ (with δ the same as in Theorem 5.6) $G(\mu)$ is twice differentiable with

$$G_\mu(\mu) = h^T(\varphi(\mu))$$

$$G_{\mu\mu}(\mu) = -h_x(\varphi(\mu))F_{xx}(\varphi(\mu),\mu)^{-1}h_x^T(\varphi(\mu))$$

Proof: Same as for Theorem 5.4.

5.2. Computational Consequences.

In the light of these theorems the method of Hestenes can be interpreted in the following way. Since the gradient is given by $G_\mu = h(\varphi(\mu))$ the updating formula $\mu^{k+1} = \mu^k + 2ch$ can be regarded as a step in the direction of the gradient in the μ-space. The method tries to locate the maximum by a "steepest ascent" method. It is known that this method can be rather ineffective. Powell [4] has shown that when $c \to \infty$ the updating of μ comes arbitrarily close to a Newton-Raphson step, which would be very effective. Since it is desirable not to make c larger than necessary to prevent ill-conditioning of $F(x,\mu)$, it would be interesting to use an updating formula for μ which is a Newton-Raphson step for any value of c. This is dealt with in the following section.

5.3. A Second Order Algorithm.

This algorithm makes use of $G_{\mu\mu}$ to take a Newton-Raphson step in μ-space. At first sight this seems to require second derivatives of f an h, since F_{xx} is used. However, many algorithms give an estimate of the second derivative (or its inverse) at the minimum and this is utilized in the algorithm making it possible to use only function values and gradients of f and h. When μ has been updated it is desirable to estimate where the new minimum of $F(x,\mu)$ lies to get a good starting point for the algorithm. This is done by differentiating $F_x(x,\mu) = 0$ which gives $F_{xx}(x,\mu) \cdot \delta x + h_x^T \delta\mu = 0$ if higher order terms are neglected.

The second order algorithm uses the following iterative scheme:

1) Set $k = 0$, select a μ^0 and $\tilde{x}^0$.

2). Compute the minimum point x^k of $F(x,\mu^k)$, using $\tilde{x}^0$ as starting point for the minimization algorithm.

Assume that this algorithm also gives an estimate of $F_{xx}(x^k,\mu^k)$ (or $F_{xx}(x^k,\mu^k)^{-1}$).

3) Calculate $G_\mu(\mu^k) = h^T(x^k)$ and $G_{\mu\mu}(\mu^k) = - h_x(x^k) \cdot$ $\cdot\ F_{xx}(x^k,\mu^k)^{-1}h_x^T(x^k)$.

4) Compute $\delta\mu^k$ from $G_{\mu\mu}\delta\mu^k = - G_\mu$ and δx^k from $F_{xx}(x^k,\mu^k) \cdot$ $\cdot\ \delta x^k = - h_x^T\delta\mu^k$

5) Put $\mu^{k+1} = \mu^k + \delta\mu^k$, $\tilde{x}^{k+1} = x^k + \delta x^k$.
Put $k = k+1$ and go to 2).

6. SUMMARY AND DISCUSSION OF THE THEORETICAL RESULTS.

From the preceding sections it follows that the methods for minimization using Lagrange multipliers can be divided into two groups, (I) methods that use the function
$F_1(x)=f(x)+\tilde{\mu}^T(x)h(x)+ch^T(x)$ with $\tilde{\mu}(x)=-(h_x h_x^T)^{-1}h_x f_x^T$
and (II) methods that use
$F_2(x,\mu)=f(x)+\mu^T h(x)+ch^T(x)h(x)$
and iterate in μ in such a way that $\mu\to\mu^*$.

For the methods of group I the interesting theorem is 4.1, which states that for all functions f and h which are twice continously differentiable and satisfy the constraint qualification, the solution x^* of the problem (P) can be made into a local minimum of $F_1(x)$ by choosing the parameter c large enough. This means that any unconstrained minimization method should be able to find x^* if the starting point is near enough. On the other hand the example on page 17 shows that it might be impossible to find the minimum if no good initial approximation can be found.

The possibilities of the methods of group II are indicated by the theorems 5.1-5.4 . In these theorems it is assumed that the function $F_2(x,\mu)$ is minimized with respect to x in the set S,which is a sufficiently small neighbourhood of x^*. This means that the minimization algorithm must be started near enough to x^*. The function $G(\mu)$ can then be computed by the minimization algorithm.The multipliers μ has to be updated in a way that leads to the maximization of $G(\mu)$, using the expressions for $G_\mu(\mu)$ and possibly $G_{\mu\mu}(\mu)$. Since these results are only valid for values of μ which are near μ^*, a sufficiently good initial approximation of μ^* has to be provided. The results of these theorems form the theoretical basis of the method of Hestenes and the method of section 5.3 . As shown by the example on page 12 the methods of type II do not necessarily work on singular problems, in contrast to the methods of type I .

So far the results only deal with local properties indicating that algorithms will work if the initial approximation of the minimum is good enough. It would be nice to have global results, making it possible to guarantee convergence for algorithms starting from an arbitrary point. A result of this type has only been obtained for a restricted class of functions f and h. These are the functions satisfying the conditions of theorem 4.4 : In addition to the constraint qualification and the condition that the minimum should be nonsingular, f(x) must tend to infinity when $||x||$ tends to infinity and f(x) must have a proper minimum under the constraint h(x)=0. Under these conditions, with the help of theorem 4.4,it is possible to prove theorem 5.5, which states that, for any μ, $G(\mu)$ is defined as the global minimum of $F(x,\mu)$. The maximum of the concave function $G(\mu)$ is obtained for $\mu=\mu^*$. If algorithms for global unconstrained minimization or maximization were available the solution could be obtained from any initial value of x and μ by a minimization in an inner loop and a maximization in an outer loop. When conventional minimization algorithms are used the result is not clear as there might be local minima as well as the global minimum of $F(x,\mu)$. Some global optimization algorithms have been suggested,[14],[15],[16], but it is not yet clear how well they work in practise. Also the theorems 5.6 and 5.7 only give G_μ and $G_{\mu\mu}$ in a neighbourhood of μ^*.Note that even when the stronger conditions of theorem 4.4 apply, the methods of type I can not be used directly for arbitrary initial values of x, as shown by the example on page 18.

7. NUMERICAL RESULTS.

In order to get an idea of the usefulness of some of the methods presented in the previous sections, a comparison has been made using a number of test-problems found in the literature. The following methods were tested.

1. The ordinary penalty function method (OPF).
This is the method presented in 3.1. Since it is the classical method to solve the problem it was chosen as a comparison for the more recent Lagrange multiplier methods. In the tests the method was used together with the extrapolation technique of [1] .

2. The method of Hestenes and Powell (HEPO).
Since these methods are very similar they have been represented by one method in the test, that of Powell [4] .

3. The method of section 5.3 (MYNEW).

4. The method of Fletcher and Mårtensson (FLE).
This is the method of section 3.5 with $\tilde{\mu}(x)=-(h_x h_x^T)^{-1} h_x f_x^T$.
In order to make it possible to use an unconstrained minimization method requiring the gradient, without using second derivatives of f and h, $\tilde{\mu}_x(x)$ is approximated as suggested by Fletcher, [6]. An initial approximation of $\tilde{\mu}_x(x)$ is obtained by difference approximations and this approximation is then updated according to the formula
$\tilde{\mu}_x := \tilde{\mu}_x + (\Delta\mu - \tilde{\mu}_x \Delta x)\Delta x^T / \Delta x^T \Delta x$.

In all cases the unconstrained minimization which is required is performed by an algorithm given by Fletcher in [9] . This algorithm uses function values and gradients of the function to be minimized. This means that the four methods which are compared all use the same information: function values and derivatives of f and h.

The

The following test problems are used:

POW, see [4]

minimize $x_1x_2x_3x_4x_5$

when $x_1^2+x_2^2+x_3^2+x_4^2+x_5^2-10=0$

$x_2x_3-5x_4x_5=0$

$x_1^3+x_2^3+1=0$

starting point (-2,2,2,-1,-1)

solution (-1.7171,1.5957,1.8272,-0.7636,-0.7636)

PAV, see [12]

minimize $1000-x_1^2-2x_2^2-x_3^2-x_1x_2-x_1x_3$

when $x_1^2+x_2^2+x_3^2-25=0$

$8x_1+14x_2+7x_3-56=0$

starting point (10,10,10)

solution (3.512,0.217,3.552)

EXP, see [12]

minimize $\sum_{i=1}^{10} \{\exp(x_i)(c_i+x_i-\ln \sum_{j=1}^{10} \exp(x_j))\}$

when $\exp(x_1)+2\exp(x_2)+2\exp(x_3)+\exp(x_6)+\exp(x_{10})-2=0$

$\exp(x_4)+2\exp(x_5)+\exp(x_6)+\exp(x_7)-1=0$

$\exp(x_3)+\exp(x_7)+\exp(x_8)+2\exp(x_9)+\exp(x_{10})-1=0$

where $c_1=-6.089$ $c_2=-17.164$ $c_3=-34.054$ $c_4=-5.914$ $c_5=-24.721$
$c_6=-14.986$ $c_7=-24.100$ $c_8=-10.708$ $c_9=-26.662$ $c_{10}=22.179$

Starting point (-2.3,-2.3,-2.3,-2.3,-2.3,-2.3,-2.3,-2.3,-2.3,-2.3)

solution (-3.2,-1.9,-0.24,-∞,-0.72,-∞,-3.6,-4.0,-3.3,-2.3)

COL1, see [6] and [13]

minimize $\sum_{j=1}^{5} e_j x_j + \sum_{j=1}^{5} \sum_{i=1}^{5} c_{ij} x_i x_j + \sum_{j=1}^{5} d_j x_j^3$

5

when $-3.5x_1+2x_3= -0.25$

$-9x_2-2x_3+x_4-2.8x_5= -4$

$2x_1-4x_3= -1$

$x_1+2x_2+3x_3+4x_4+5x_5= 5$

where e_j, c_{ij} and d_j are given by

	j	1	2	3	4	5
e_j	i	-15	-27	-36	-18	-12
c_{ij}	1	30	-20	-10	32	-10
	2	-20	39	-6	-31	32
	3	-10	-6	10	-6	-10
	4	32	-31	-6	39	-20
	5	-10	32	-10	-20	30
d_j		4	8	10	6	2

starting point (0,0,0,0,1)

solution (0.3000,0.3335,0.4000,0.4283,0.2240)

TRIG

minimize $\sum_{i=1}^{n} (\theta_i E_i - f_i(x))^2$

when $E_i - f_i(x) = 0 \qquad i=1,..,m$

where $\theta_i = \begin{cases} \text{random number} & i=1,..,m \\ 1 & i=m+1,..,n \end{cases}$

$$f_i(x) = \sum_{j=1}^{n} (A_{ij}\sin(x_j) + B_{ij}\cos(x_j))$$

$E_i = f_i(x^*)$

x^* = point chosen to be the minimum

A_{ij} and B_{ij} are generated by a random number generator.

Result: Number of times that (f, f_x, h, h_x) has been evaluated to reach the given absolute accuracy in x.

problem		accuracy	HEPO	OPF	MYNEW	FLE
POW		10^{-4}	37	41	36	43
PAV		10^{-3}	54*	64	61	47
EXP		10^{-1}	192	211	163	47
COL1		10^{-4}	54	45	20	13
TRIG	n=2 m=1	10^{-3}	28	22	14	7
		10^{-5}	35	22	14	8
	n=4 m=2	10^{-3}	92	130	85	21
		10^{-5}	105	-	101	-
	n=6 m=3	10^{-3}	57	62	38	25
		10^{-5}	79	-	49	-
	n=8 m=4	10^{-2}	172	194	101	-
		10^{-4}	-	-	136	-

*) found a different minimum

Entries which are marked "-" indicate that the accuracy in question was never reached. FLE had a tendency to stop too early, before the desired accuracy was obtained. It is not clear if this depends on the approximation of μ_x that is used or on the way the algorithm is implemented. In the latter case it would probably be possible to avoid the problem. Apart from this FLE has very good results especially on TRIG with n=4,m=2 and EXP. The increase in speed obtained with the different updating of μ in MYNEW compared to HEPO is also shown although MYNEW is usually slower than FLE. The comparison also indicates that use of the Lagrange multiplier technique makes it possible to design algorithms that are faster than the ordinary penalty function algorithm.

REFERENCES.

[1] A.V.Fiacco and G.P.McCormick:"Nonlinear Programming: Sequential Unconstrained Minimization Techniques",J.Wiley and Sons, New York,1968.

[2] J.D.Schoeffler:"Static Multilevel Systems" in D.A.Wismer,ed.: "Optimization Methods for Large Scale Systems", McGraw Hill 1971.

[3] M.R.Hestenes:"Multiplier and Gradient Methods". J.Opt.Theory Appl. Vol.4,1969,303-320.

[4] M.J.D.Powell:"A Method for Nonlinear Constraints in Minimization Problems", in R.Fletcher, ed.:"Optimization", Academic Press, London,1969.

[5] R.Fletcher:"Methods for Nonlinear Programming",in Abadie, ed.: "Integer and Nonlinear Programming",North Holland Publishing Company,1970.

[6] R.Fletcher and S.A.Lill:"A Class of Methods for Nonlinear Programming II. Computational Experience", in J.B.Rosen, O.L.Mangasarian and K.Ritter, ed.:"Nonlinear Programming", Academic Press, London, 1970.

[7] K.Mårtensson:"New Approaches to the Numerical Solution of Optimal Control Problems", Report 7206, March 1972, Lund Institute of Technology, Division of Automatic Control.

[8] J.M.Ortega and W.C.Rheinboldt:"Iterative Solution of Nonlinear Equations in Several Variables", Academic Press, New York,1970.

[9] R.Fletcher:"Fortran Subroutines for Minimization by Quasi-Newton Methods", Harwell Report AERE-R7125, 1972

[10] H.Ramsin:"Kvasinewtonmetoder för minimering av en reellvärd funktion av flera variabler utan bivillkor", Report NA 71.44 Department of Information Processing and Computer Science, The Royal Institute of Technology, Stockholm.

[11] W.Murray ed.:"Numerical Methods for Unconstrained Optimization", Academic Press, London, 1972.

[12] D.M.Himmelblau:"Applied Nonlinear Programming",McGraw Hill, New York, 1972.

[13] A.R.Colville:"A Comparative Study of Nonlinear Programming Codes",IBM New York Scientific Center Report nr 320-2949, June 1968.

[14] R.P.Brent:"Algorithms for Minimization without Derivatives" Prentice Hall,1973.

[15] F.H.Branin, S.K.Hoo :"A Method for Finding Multiple Extrema of a Function of n Variables", in F.A. Lootsma, ed.:"Numerical Methods for Non-Linear Optimization",Academic Press, London, 1972.

[16] J.Opacic:"A Heuristic Method for Finding Most Extrema of a Nonlinear Function", IEEE Trans. Syst. Man Cyb., vol.SMC-3,nr 1, Jan.1973.