RR-75-13

GENERALIZED LAGRANGE MULTIPLIER TECHNIQUE FOR NONLINEAR PROGRAMMING

Yuri Evtushenko

June 1975

Generalized Lagrange Multiplier Technique For Nonlinear Programming

Yuri Evtushenko*

Abstract

Our aim here is to present numerical methods for solving a general nonlinear programming problem. These methods are based on transformation of a given constrained minimization problem into an unconstrained maximin problem. This transformation is done by using a generalized Lagrange multiplier technique. Such an approach permits us to use Newton and gradient methods for nonlinear programming. Convergence proofs are provided and some numerical results are given.

1. Statement of Problem and Description of Numerical Methods

We consider the following general nonlinear programming problem:

$$\text{minimize } F(x) \tag{1}$$

subject to constraints $x \in X = \{x | g(x) = 0,\ h(x) \leq 0,\ x \in E_n\}$, where F, g, h are real-valued, twice continuously differentiable functions defined on E_n, Euclidean n-space; $x = (x^1, x^2, \ldots, x^n)$ is a point in E_n: and vector functions $g(x)$, $h(x)$ define the mapping $g(x)$: $E_n \to E_e$, $h(x)$: $E_n \to E_c$.

We define the modified Lagrangian function $H(x,p,w)$

*International Institute for Applied Systems Analysis, Schloss Laxenburg, Laxenburg, Austria.

associated with problem (1) as:

$$H(x,p,w) = F(x) + \sum_{i=1}^{e} p^i g^i(x) + \sum_{i=1}^{c} (w^i)^2 h^i(x)$$

where

$$p = (p^1, p^2, \ldots, p^e) \in E_e \quad , \quad w = (w^1, w^2, \ldots, w^c) \in E_c \quad .$$

Consider an unconstrained maximin problem

$$\max_{p \in E_e} \max_{w \in E_c} \min_{x \in E_n} H(x,p,w) \quad . \qquad (2)$$

We shall solve this problem instead of (1). Under certain conditions, which we shall formulate later in 2, the solution x to problem (2) coincides with that to primal nonlinear programming problem (1). As a rule, the Lagrangian is defined as

$$C(x,p,w) = F(x) + \sum_{i=1}^{e} p^i g^i(x) + \sum_{i=1}^{c} w^i h^i(x) \quad ,$$

and the following problem is solved:

$$\max_{p \in E_e} \max_{w \in T} \min_{x \in E_n} C(x,p,w) \quad , \qquad (3)$$

where $T = \{w | w \geq 0\}$. Problem (3) is a constrained maximin problem and this circumstance complicates its solution.

When we use the modified Lagrangian $H(x,p,w)$ we have no such difficulties because (2) can be solved with the well-known numerical methods for unconstrained maximin and saddle point problems. For example, using the simplest gradient method yields the following method:

$$\dot{x} = -H_x, \ \dot{p} = H_p, \ \dot{w} = H_w, \tag{4}$$

$$x(0) = x_0, \ p(0) = p_0, \ w(0) = w_0,$$

where H_x, H_p, H_w are $n \times 1$, $e \times 1$, $c \times 1$ vectors, whose ith elements are

$$\delta H(x,p,w)/\delta x^i, \ \delta H(x,p,w)/\delta p^i, \ \delta H(x,p,w)/\delta w^i \quad ,$$

respectively.

From equation (4) on, a super dot denotes differentiation with respect to time variable t, i.e. $(\cdot) = d/dt$.

In 2 we shall prove that the solution $x(t)$, $p(t)$, $w(t)$ of system (4) locally converges to that of (2) as $t \to \infty$. The author presented (Refs. 1-2) a number of iterative methods for finding local solutions of an unconstrained maximin problem. Using three of these yields

$$\dot{x} = -H_x,\ \dot{p} = g - g_x^T H_{xx}^{-1} H_x,\ \dot{w} = 2D(w)\ [h - h_x^T H_{xx}^{-1} H_x] \tag{5}$$

$$\dot{x} = -H_x - H_{xx}^{-1}(g_x g + 4h_x D(w)\ D(w)\ h),\ \dot{p} = \dot{g},\ \dot{w} = 2D(w)h \tag{6}$$

$$\dot{x} = -H_{xx}^{-1} H_x,\ \dot{p} = g - g_x^T H_{xx}^{-1} H_x,\ \dot{w} = 2D(w)\ [h - h_x^T H_{xx}^{-1} H_x] \tag{7}$$

where g_x, h_x, H_{xx}, are $n \times e$, $n \times c$, $n \times n$, Jacobian matrices, respectively, whose ijth elements are

$$\delta g^j(x)/\delta x^i,\ \delta h^j(x)/\delta x^i,\ \delta^2 H(x,p,w)/\delta x^i \delta x^j \quad ,$$

respectively; $D(w)$ is the diagonal matrix whose ith diagonal element is w^i; superscript -1 denotes the inverse of a matrix; superscript T denotes the transpose of a matrix.

For simplicity we shall denote

$$z = (x,p,w) \in E_{n+e+c} \quad ,$$

$$z_* = (x_*,p_*,w_*) \in E_{n+e+c} \quad , \ ,$$

$$H(z) = H(x,p,w),\ H(z_*) = H(x_*,p_*,w_*) \quad .$$

Definition. The point z_* is a local maximin of function $H(z)$ in problem (2) if there exist neighborhoods A, Q, G about the points x_2, p_2, w_*, respectively, such that for all $x \in A$, $p \in Q$, $w \in G$, the following inequalities hold:

$$H(x(p,w), p, w) \leq H(x_*, p_*, w_*) \leq H(x, p_* w_*) \quad , \tag{8}$$

where

$$H(x(p,w), p, w) = \min_{x \in A} H(x,p,w) \quad .$$

The necessary conditions that z_2 be a local maximin of problem (2) are (see Ref. 1)

$$H_x(z_*) = 0, \quad H_p(z_*) = 0, \quad H_w(z_*) = 0 \quad . \tag{9}$$

All the points satisfying these conditions we will call stationary points. Now we apply the Newton method for computation of stationary points. We obtain the following continuous version of the method:

$$\begin{gathered} H_{xx}\,\dot{x} + H_{xp}\,\dot{p} + H_{xw}\,\dot{w} = -H_x \\ H_{px}\,\dot{x} = -H_p, \quad H_{wx}\,\dot{x} + H_{ww}\,\dot{w} = -H_w \quad , \end{gathered} \tag{10}$$

where H_{xp}, H_{xw}, H_{ww} are the matrices whose ijth elements are $\delta^2 H(x,p,w)/\delta x^i \delta p^j$, $\delta^2 H(x,p,w)/\delta x^i \delta w^j$, $\delta^2 H(x,p,w)/\delta w^i \delta w^j$, respectively; $H_{xp} = H_{px}^T, H_{xw} = H_{wx}^T$. Using abbreviated notations yields the following continuous and discrete version of system (10):

$$H_{zz}(z)\,\dot{z} = -H_z(z) \quad , \qquad z(0) = z_0 \tag{11}$$

$$z_{s+1} = z_s - H_{zz}^{-1}(z_s)\,H_z(z_s) \quad , \tag{12}$$

where z_0 is given, $s = 0,1,2,\ldots,$.

In the case when constraints are absent, these methods coincide with the Newton method. They are well known and are investigated when problem (1) has no inequality constraints (see Ref. 3).

On the basis of continuous methods (4) - (7), we can construct a number of discrete methods for finding saddle points. But we shall use only the simplest finite difference approximation to (4) - (7). For example, method (4) yields

$$x_{s+1} = x_s - \alpha H_x(z_s) \quad , \quad p_{s+1} = p_s + \alpha H_p(z_s) \quad ,$$

$$w_{s+1} = w_s + \alpha H_w(z_s) \quad , \tag{13}$$

where $0 < \alpha$ is the step length. The discrete version of other methods can be written in an analogous way, except in (12), where it is possible to use $\alpha = 1$.

2. Convergence Proofs

In this section we shall give rigorous convergence proofs of the methods suggested above. We shall first state some preliminary results.

Define the following set of integers:

$$B(x) = \{i | h^i(x) = 0, \ 1 \leq i \leq c\} \quad .$$

Definition. The constraint qualification holds at a point x if all gradients $\{g_x^i(x)\}$, $1 \leq i \leq c$ and all gradients $h_x^j(x)$, $j \in B(x)$ are linearly independent.

Definition. The strict complementarity holds at a point z_* if from $h^i(x_*) = 0$ it follows that $w_*^i \neq 0$, $i \leq i \leq c$.

Lemma 1. If $\bar{z} = (\bar{x}, \bar{p}, \bar{w})$ is a saddle point of function $H(z)$ in problem (2), then $\bar{x}$ solves problem (1), and $F(\bar{x}) = H(\bar{x}, \bar{p}, \bar{w})$.

Lemma 2. Let A be a neighborhood of $\bar{x}$ and let the following inequalities hold:

$$H(\bar{x}, p, w) \leq H(\bar{x}, \bar{p}, \bar{w}) < H(x, \bar{p}, \bar{w}) \qquad (14)$$

for any $p \in E_e$, $w \in E_c$, $x \in A$, $x \neq x_*$; then $\bar{x}$ is a local, isolated solution to problem (1).

Lemma 3. If $\bar{x} \in X$ then

$$F(\bar{x}) = \sup_{p \in E_e} \ \sup_{w \in E_c} \ H(\bar{x}, p, w) \quad .$$

The proof of these lemmas is quite similar to that of analogous results for problem (3)(see for example Ref. 4), and is therefore not given here.

Consider the following auxiliary problem

$$\max_{u \in E_k} \min_{x \in E_n} P(x,u) , \tag{15}$$

where P(x,u) is a continuous function of x and u. Use will be made of the following lemma, which is stated here without proof (for proof see Ref. 1).

Lemma 4. Suppose that function P(x,u) is twice continuously differentiable on $E_n \times E_k$, and a solution to problem (15) exists. Sufficient conditions for $y_* = (x_*,u_*)$ to be an isolated (unique locally) maximin point of problem (15) are that

1) y_* is a stationary point, i.e.

   $P_x(y_*) = 0, \quad P_u(y_*) = 0,$

2) $P_{xx}(y_*)$ and $M(y_*) =$

   $P_{ux}(y_*)\ P_{xx}^{-1}(y_*)\ P_{xu}(y_*) - P_{uu}(y_*)$ are positive definite matrices.

If matrices $P_{xx}(x,u)$ and $M(x,u)$ are positive definite for arbitary $x \in E_n$, $u \in E_k$, then the stationary point y_* is a global maximin point of $P(x,u)$. However, y_* may not be a saddle point of $P(x,u)$ (see also Ref. 1).

Lemma 5. Suppose that the constraint qualification and strict complementarity hold at a stationary point z_*, the Hessian $H_{xx}(z_*)$ is positive definite, and $h(x_*) \leq 0$. Then the Hessian $H_{zz}(z_*)$ is nonsingular, the symmetric block matrix

$$N = [H_{xp} \vdots H_{xw}]^T H_{xx}^{-1} [H_{xp} \vdots H_{xw}] - \begin{bmatrix} H_{pp} & H_{pw} \\ H_{wp} & H_{ww} \end{bmatrix}$$

is positive definite, z_* is a local, isolated solution of problem (1).

For shorthand in the formula for N, we omit the argument, which is z_*. We shall use the same abbreviations later.

Proof. Stationary conditions (9) and inequality $h(x_*) \leq 0$ imply that $x_* \in X$, i.e. x_* is a feasible point for problem (1).

To prove nonsingularity $H_{zz}(z_*)$ we need only show that there is no non-zero solution of the following system of linear equations:

$$H_{xx}(z_*)\bar{x} + g_x(x_*)\bar{p} + 2h_x(x_*)\ D(w_*)\bar{w} = 0 \tag{16}$$

$$g_x^T(x_*)\bar{x} = 0, \quad D(w_*)h_x^T(x_*)\bar{x} + D(h(x_*))\bar{w} = 0 \tag{17}$$

From the last system and strict complementarity, it follows that for all i such that $i \in B(x_*)$,

$$h_x^i(x_*)\bar{x} = 0, \quad h^i(x_*) = 0, \quad w_*^i \neq 0 \quad ,$$

and that for all i such that $i \bar{\in} B(x_*)$,

$$h^i(x_*) < 0, \quad w_*^i = 0, \quad \bar{w}^i = 0 \quad .$$

In both cases $h^i(x_*)\bar{w}^i = 0$ and $D(w_*)h_x^T(x_*)\bar{x} = 0$. Let $\bar{x} \neq 0$; then premultiplying (16) on the left by $\bar{x}^T$ and taking into account (17) yields

$$\bar{x}^T\ H_{xx}(z_*)\bar{x} = 0 \quad .$$

This holds only if $\bar{x} = 0$. Consider this case. From (16) and (17) we find

$$g_x(x_*)\bar{p} + 2h_x(x_*)\ D(w_*)\bar{w} = 0 \quad , \quad D(h(x_*))\bar{w} = 0 \quad .$$

The first system can be written in the form

$$g_x(x_*)\bar{p} + 2 \sum_{i\in B(x_*)} h_x^i(x_*) w_*^i \bar{w}^i = 0 \tag{18}$$

All $w_*^i > 0$ for $i \in B(x_*)$; with the assumed constraint qualification all the gradients in (18) are linearly independent; and (18) holds if $\bar{p} = 0$ and $\bar{w}^i = 0$ for all $i \in B(x_*)$. But we found above that $\bar{w}^i = 0$ for $i \bar{\in} B(x_*)$; thus $\bar{x} = 0$, $\bar{p} = 0$, $\bar{w} = 0$ for all solutions. This contradiction proves that the matrix $H_{zz}(x_*)$ is nonsingular. We can assume without loss of generality that $h^i(x_*) = 0$ for $1 \leq i \leq s$ and $h^i(x_*) < 0$ for $1 + s \leq i \leq c$. Introduce the vectors

$$v = \left[p^1, p^2, \ldots, p^e,\ w^1, w^2, \ldots, w^s\right] \in E_k \quad , \quad k = e + s$$

and

$$\tilde{h} = \left[h^{s+1}, h^{s+2}, \ldots, h^c\right] \in E_r \quad , \quad r = c - s \quad .$$

Making use of strict complementarity, we obtain $w_*^i = 0$ for all $1 + s \leq i \leq c$. Therefore, omitting arguments we can rewrite N as follows:

$$N = \left[H_{xv} \;\vdots\; 0_{nr} \right]^T H_{xx}^{-1} \left[H_{xv} \;\vdots\; 0_{nr} \right] - 2 \left[\begin{array}{c|c} 0_{ee} & 0_{ec} \\ \hline 0_{ce} & D(h) \end{array} \right]$$

where O_{ij} is the $i \times j$ matrix whose elements are all equal to zero, and $D(h)$ is the diagonal matrix whose ith diagonal element is h^i. The matrix N can be written in the four blocks form

$$N = \left[\begin{array}{c|c} H_{vx} \quad H_{xx}^{-1} \quad H_{xv} & O_{kr} \\ \hline O_{rk} & -2D(\tilde{h}) \end{array}\right]$$

where

$$H_{xv} = \left[g_x^1(x_*) \,\vdots\, g_x^2(x_*) \,\vdots\, \ldots \,\vdots\, g_x^e(x_*) \,\vdots\, 2w_*^1 h_x^1(x_*) \,\vdots\, \ldots \,\vdots\, 2w_*^s h_*^s(x_*)\right]$$

is the $n \times k$ matrix. Assuming strict complementarity, $w_*^i \neq 0$ for all $1 \leq i \leq s$. Since the constraint qualification holds, all gradients $g_x^i(x_*)$, $1 \leq i \leq e$ and $w_*^i h_x^i(x_*)$, $1 \leq i \leq s$ are linearly independent columns; that is, H_{xv} has maximum rank k. Since $H_{xx}^{-1}(z_*)$ is a nonsingular matrix, there exists a symmetric, nonsingular matrix W such that $H_{xx}^{-1}(z_*) = W \cdot W$. It is well known (Ref. 5) that if a matrix is multiplied on the left or on the right by a nonsingular matrix, the rank of the original matrix remains unchanged. Thus matrices $H_{xv}^T W$ and $W H_{xv}$ have maximum rank k. Their product $H_{xv}^T W W H_{xv} = H_{xv}^T H_{xx}^{-1} H_{xv}$ is a nonsingular symmetric matrix. Because of assumption $\tilde{h} < 0$, matrix $-D(\tilde{h})$ is positive definite and consequently N is also positive definite.

According to the sufficient conditions formulated in lemma 4, the stationary point z_* is the local, isolated maximin point of problem (2); hence, taking into account that x_* is a feasible point for problem (1), we obtain from lemma 3 that

$$F(x_*) = H(z_*) = \max_{p\in Q} \max_{w\in G} \min_{x\in A} H(x,p,w)$$

$$= \sup_{p\in E_e} \sup_{w\in E_c} H(x_*,p,w) \quad . \qquad (19)$$

where Q, G, A are neighborhoods about points p_*, w_*, x_*, respectively. From (8) and (19) the inequalities (14) follow. Therefore z_* is a local, isolated solution of (1).

We shall show now that z_* is an isolated saddle point of H(z) in problem (2). If it is not true, then for any neighborhood of point z_* there would exist a saddle point z_1 of H(z). The point would be stationary. Applying the Taylor formula for first-order expansions, we obtain

$$H_z(z_1) = H_z(z_*) + H_{zz}(z_* + t(z_1 - z_*))(z_1 - z_*) = 0 \quad , \qquad (20)$$

where $0 \leq t \leq 1$. The Hessian $H_{zz}(z_*)$ is nonsingular. As the Hessian is continuous, we may select z_1 so close to z_* that the Hessian $H_{zz}(z_* + t(z_1 - z_*))$ is also nonsingular for arbitary $0 \leq t \leq 1$. Hence, the system (20) has only trivial solution $z_1 = z_*$. The contradiction is evident. Local uniqueness of the saddle point is proved.

Theorem 1. Suppose that the assumptions of lemma 5 are satisfied. Then the solutions of systems (4) - (7) and (10) locally, exponentially converge to z_* as $t \to \infty$ (i.e. positive numbers ε, μ exist such that $||z(t) - z_*|| \leq \phi(\varepsilon)e^{-\mu t}$ if $||z_0 - z_*|| \leq \varepsilon$). There exists a number $\bar{\alpha}$ such that for any $0 < \alpha < \bar{\alpha}$ the solutions of finite difference approximations to (4) - (7), similar to (13), converges locally and linearly to z_* (i.e. $0 < \varepsilon$, $0 \leq q \leq 1$ exist such that $||z_s - z_*|| \leq \phi(\varepsilon)q^s$ if $||z_* - z_0|| < \varepsilon$).

Proof. All the systems suggested above have two common properties. They are autonomous, and for all these systems any stationary point z_* is an equilibrium position. This permits us to use for proof the linearization principle first proved by Liapunov (Ref. 6) and often called "the first method of Liapunov". With this technique we shall prove the asymtotic stability of solution $z(t) \equiv z_*$ of systems (4) - (7) and (10). This result implies local convergence of the solutions z(t) to a stationary point z_*.

Set $\delta x = x(t) - x_*$, $\delta p = p(t) - p_*$, $\delta w = w(t) - w_*$, $\delta z = (\delta x, \delta p, \delta w)$. Using the Taylor formula for first-order expansions using stationary condition $H_z(z_*) = 0$, we obtain

$$H_z(z_* + \delta z) = H_z(z_*) + H_{zz}(z_*)\delta z + O(||\delta z||^2)$$

$$H_{xx}^{-1}(z_* + \delta z)\, H_x(z_* + \delta z) = \delta z + O(||\delta z||^2)$$

where $O(||y||)$ is a quantity such that

$$\lim \frac{O(||y||)}{||y||} < \infty \qquad \text{as} \quad ||y|| \mapsto 0 \quad .$$

The equation of the first approximation of system (4) about the equilibrium point z_* is

$$\delta\dot{z}(t) = M\,\delta z(t) \text{ where } M = \left[\begin{array}{c|c|c} -H_{xx} & -g_x & -2h_x\, D(w) \\ \hline g_x^T & 0_{ee} & 0_{ec} \\ \hline 2D(w)h_x^T & 0_{ce} & 2D(h) \end{array}\right] \tag{21}$$

All elements of matrix M are computed at the point $z = z_*$. The convergence of method (4) will be proved if we show that all eigenvalues λ of matrix M have negative real parts. Let $\delta z = (\delta x, \delta p, \delta w)$ be a characteristic vector of M, i.e. $M\,\delta z = \lambda\,\delta z$. Let $\overline{\delta z} = (\overline{\delta x}, \overline{\delta p}, \overline{\delta w})$ be a complex conjugate to vector δz; Re b denotes the real part of complex number b. From (21) we obtain

$$\mathrm{Re}\ \overline{\delta z}^T M \delta z = \mathrm{Re}\ \lambda ||\delta z||^2 = \mathrm{Re}\left[-\delta\bar{x}^T H_{xx}(z_*)\delta x + 2\delta\bar{w}^T\ D(h(x_*))\delta w\right] \le 0 \quad .$$

Here we take into account that $H_{xx}(z_*)$ is positive definite and x_* is a feasible point. Consider the case when Re $\lambda = 0$. Then Re $\left[-\overline{\delta x}^T H_{xx}\delta x + 2\delta\bar{w}D(h(x_*))\delta w\right] = 0$ if and only if $\delta x = 0$, $\delta w^i \neq 0$ for all i such that $i \bar{\in} B(x_*)$. From the characteristic equation we have

$$g_x(x_*)\delta p + 2 \sum_{i \in B(x_*)} h_x^i(x_*) w_*^i \delta w^i = 0 \quad .$$

From the constraint qualification it follows that $\delta w^i = 0$ for any $i \in B(x_*)$. Hence $||\delta z|| = 0$; the case Re $\lambda = 0$ is thus impossible and strict inequality Re $\lambda < 0$ holds.

The convergence of methods (4) - (7) can be proved by similar analysis of their characteristic equations. Their eigenvalues proved to be real and this circumstance simplifies investigation. For example, the linearized system of equation (5) about the stationary point z_* is

$$\delta\dot{x} = -H_{xx}\delta x - g_x\delta p - 2h_x\ D(w)\delta w$$

$$\delta\dot{p} = -g_x^T H_{xx}^{-1}\ [g_x\ \delta p + 2h_x\ D(w)\delta w]$$

$$\delta\dot{w} = H_{ww}\delta w - 2D(w)h_x^T H_{xx}^{-1}\ [g_x\delta p + 2h_x D(w)\delta w]\ .$$

The condition for asymptotic stability can be expressed by means of the characteristic roots of the following secular equation

$$\left|\begin{array}{c|c|c} H_{xx} + \lambda I_n & g_x & 2h_x D(w) \\ \hline 0_{en} & g_x^T H_{xx}^{-1} g_x + \lambda I_e & 2g_x^T H_{xx}^{-1} h_x D(w) \\ \hline 0_{cn} & 2D(w)h_x^T H_{xx}^{-1} g_x & 4D(w)h_x^T H_{xx}^{-1} h_x D(w) - H_{ww} + \lambda I_c \end{array}\right| = 0 \tag{22}$$

where I_j is the $j \times j$ unit matrix.

It is easy to see that determinant (22) is equal to the product of the determinants of the diagonal cells:

$$|H_{xx} + \lambda I_n| \cdot |N + \lambda I_{e+c}| = 0 \tag{23}$$

According to lemma 5 the matrices H_{xx} and N are symmetrical and positive definite; hence, the characteristic roots of equation (23) are real and strictly negative.

After some transformation it can be shown that secular equations for systems (6) and (7) also have real, strictly negative roots. From the integration of (10) along a solution, we have

$$H_z(z(t)) = H_z(z(0))e^{-t} \quad , \qquad z(0) = z_0 \quad .$$

This shows that if for any initial state z_0 there exists the solution z(t) of system (10) for all $t \geq 0$, then this solution converges to a stationary point, which may not be feasible for problem (1), nor be a saddle point in problem (2). But if z_0 is chosen sufficiently close to a saddle point z_* at which all assumptions of lemma 5 hold, then the solution z(t) of (10) exists for all $t \geq 0$, and z(t) converges to the saddle point z_* as $t \to \infty$.

The principle of determining the stability from the equation of the first approximation about an equilibrium state is also valid for discrete systems. Denote $\Delta x_s = x_s - x_{s*}$, $\Delta p_s = p_s - p_*$, $\Delta w_s = w_s - w_*$, $\Delta z_s = (\Delta x_s, \Delta p_s, \Delta w_s)$. The linearized system of (13) about equilibrium point z_* is

$$\Delta z_{s+1} = \phi \Delta z_s \tag{24}$$

where $\phi = I_{n+e+c} + \alpha M$ (M is defined by (21)).

The solution $z_s \equiv z_*$ of the autonomous discrete system (24) is asymptotically stable if all eigenvalues of the matrix ϕ have magnitudes smaller than 1.

Let u and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{n+e+c})$ be eigenvalues of matrices ϕ and M respectively, i.e.

$$|\phi - uI_{n+e+c}| = 0 \qquad |M - \lambda I_{n+e+c}| = 0$$

Consequently, we have relationship $u = 1 + \alpha\lambda$.

Denote

$$|\lambda_j|^2 = \max \left[|\lambda_1|^2, |\lambda_2|^2, \ldots, |\lambda_{n+e+c}|^2 \right]$$

$$\mathrm{Re}\ \lambda_s = \max \left[\mathrm{Re}\ \lambda_1, \ \mathrm{Re}\ \lambda_2, \ldots, \mathrm{Re}\ \lambda_{n+e+c} \right]$$

$$\bar{\alpha} = -2\ \mathrm{Re}\ \lambda_s / |\lambda_j|^2$$

We proved that all λ have negative real parts, hence $\bar{\alpha} > 0$. The magnitudes of all u are smaller than I (in modulus) if α is sufficiently small, $0 < \alpha < \bar{\alpha}$. This follows from inequalities:

$$|u|^2 = |1+\alpha\lambda|^2 = I+\alpha|\lambda|^2[\alpha + \frac{2 \text{ Re } \lambda}{|\lambda|^2}] \le 1+\alpha|\lambda|^2[\alpha-\bar{\alpha}] < 1$$

For computation it is desirable to take step length α as large as possible. But in the case of large α values we may lose convergence. The maximum admissible α value depends on function F, g, h, point z_* and the computational method. In all other discrete versions of systems (5) - (7), the proof of convergence follows from that of the respective continuous system, as was shown above.

Theorem 2. Suppose that the assumptions of lemma 5 are satisfied and the function $H_{zz}(z)$ satisfies the Lipschitz condition in a neighborhood of point z_*. Then the solution z_s of (11) locally quadratically converges to the saddle point z_*; i.e. q,ε exist such that

$$||z_s - z_*|| \le c(\varepsilon)q^{2^n} \text{ if } ||z_0 - z_*|| \le \varepsilon) \quad .$$

The proof is analogous to that of the Newton method of convergence theorem (Ref. 7), and is therefore omitted. To hasten convergence to the solution of problem (1) we can, in methods (4) - (7), (10) and (11) use the following function instead of H:

$$\Gamma = H + a \sum_{i=1}^{e} [g^i(x)]^2 + b \sum_{i=1}^{c} (w^i)^4 (h^i(x))^2$$

where a, b are some positive coefficients. From (4), for example, we obtain

$$\dot{x} = - \Gamma_x, \quad \dot{p} = \Gamma_p, \quad \dot{w} = \Gamma_w \tag{25}$$

All other methods can be modified in a similar way. It is easy to prove that if the assumptions of theorem 1 hold, then the solution of (25) locally converges to z_* for any $0 \leq a$, $0 \leq b$.

3. Numerical Examples

We shall give an example that was solved using the three methods presented to illustrate their convergence properties. The function to be minimized is

$$F(x) = [x^1 + 3x^2 + x^3]^2 + 4 (x^1 - x^2)^2$$

The constraints are

$$g(x) = 1 - x^1 - x^2 - x^3 = 0 \quad , \quad h^1 = - x^1, h^2 = - x^2 \quad ,$$

$$h^3 = - x^3 \quad , \quad h^4 = 3 - 4x^3 - 6x^2 + [x^1]^3 \quad .$$

The starting point is assumed to be

$$x^1 = 0.1 \quad , \quad x^2 = 0.7 \quad , \quad x^3 = 0.2 \quad , \quad p^1 = -0.1 \quad ,$$

$$w^4 = 1 \quad , \quad w^1 = w^2 = w^3 = 0.1 \quad .$$

The step length was $\alpha = 0.02$.

The approximate solution of this problem is $F_* = 1.8311$. The iterations were terminated if the difference between the current value of $F(x_s)$ and the following one remained less than 10^{-5}. If the number of iterations was more than 100, then the process was also manually terminated.

Denote the maximum number of steps by N. Let δ be a difference between $F(x_N)$ and F_* and T be the time of computations. For the discrete version of (4) $N = 100$, $\delta = 0.0064$. $T = 11$ sec were obtained; for the discrete version of (5) $N = 100$, $\delta = 0.0056$, $T = 16$ sec; for method (11), $N = 4$, $\delta = 0.0001$, $T = 3$ sec.

The modified Newton method converges after 4 iterations. While this method has the best rate of convergence, it also requires more time per iteration than the others, and the size of the region of convergence was also less. The simplest method (4) has the largest region of convergence.

It is not possible to state without ambiguity that one numerical method is superior to another. It is also doubtful whether a universally best method exists. For computation the combination of different methods seems to be most expedient. For finding a rough solution, the simplest methods, such as (4), may be used; a more accurate solution would be found by a more complicated method such as (11).

The difference $\delta(s) = F(x_s) - F_*$ as a function of step number s is shown in Fig. 1 for method (13). Various values $\alpha = 0.05$, $\alpha = 0.04$, $\alpha = 0.02$ were used. For $\alpha = 0.2$, method (13) does not converge. Increasing the step length α hastens the rate of convergence, but the solution becomes less stable.

The influence of coefficient a on the rate of convergence of method (25) is shown in Fig. 2. For computation, a discrete approximation similar to (13) was used with $\alpha = 0.02$, $b = 0$. Using a small value of a ($a = 1$, $a = 2$) hastens convergence, but for a larger value ($a = 5$) the convergence rate decreases.

4. Some Generalizations

Consider the following minimax problem. Find

$$\min_{x \in X} \max_{y \in Y} K(x,y) \tag{26}$$

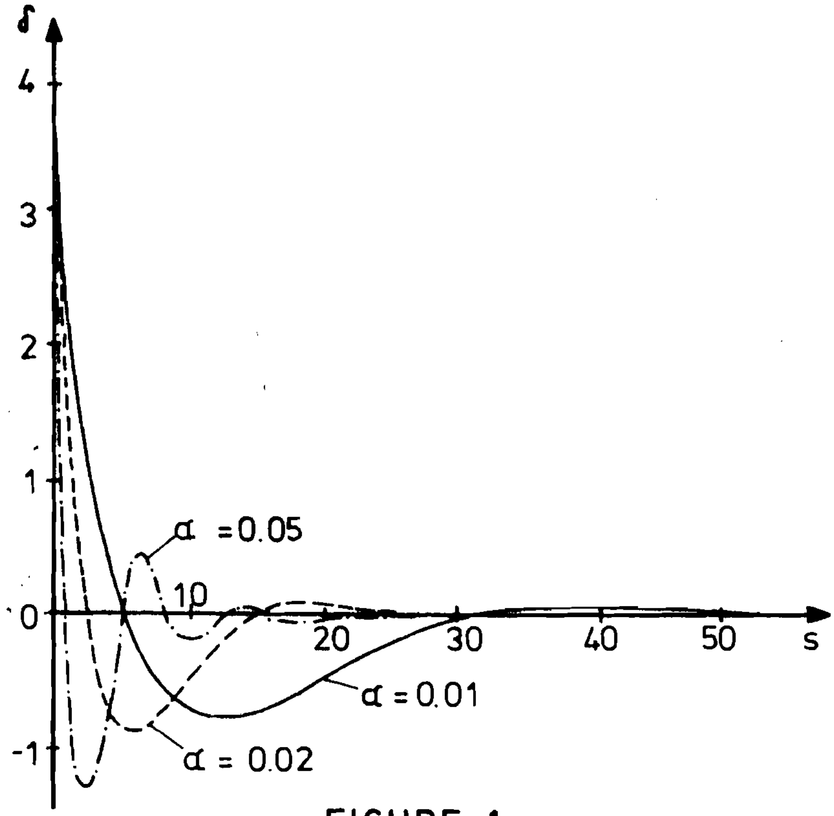

FIGURE 1

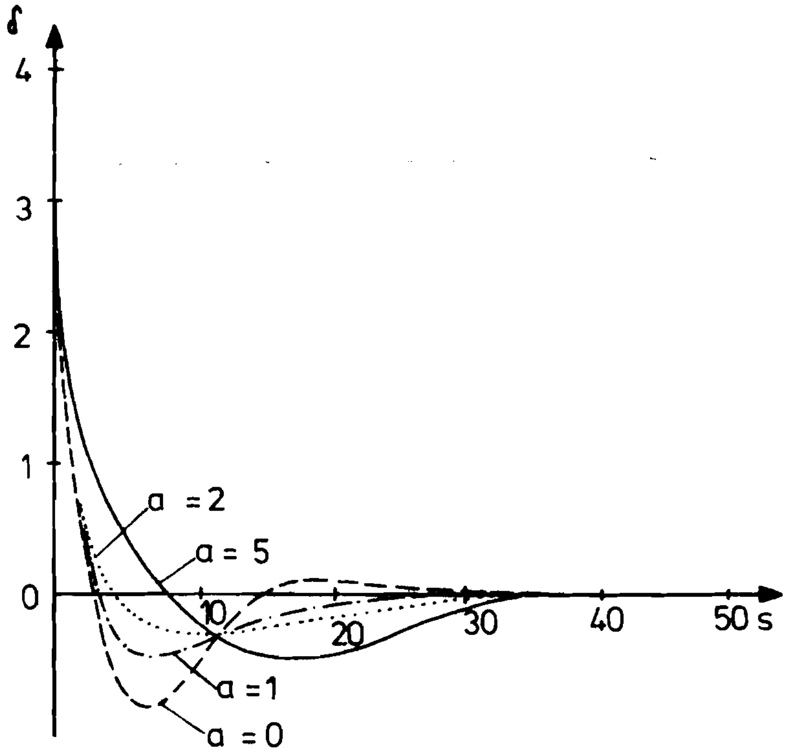

FIGURE 2

where $X = \{x \in E_n | g(x) = 0,\ h(x) \leq 0\}$, $Y = \{y \in E_n | G(x) = 0,$ $H(x) = 0\}$, $x \in E_n$, $y \in E_m$, $g \in E_e$, $h \in E_c$, $G \in E_k$, $H \in E_s$. Functions K, g, h, G, H are continuously differentiable.

Introduce the Lagrangian as follows:

$$\Phi(x,y,p,w,P,W) = K(x,y) + \sum_{i=1}^{e} p^i g^i(x)$$

$$+ \sum_{i=1}^{c} (w^i)^2 h^i(x) - \sum_{i=1}^{k} P^i G^i(y) - \sum_{i=1}^{s} (W^i)^2 H^i(y)$$

where $P \in E_k$, $W \in E_s$, $p \in E_e$, $w \in E_c$.

Consider an unconstrained maximin problem

$$\max_{y \in E_m} \max_{p \in E_e} \max_{w \in E_c} \min_{x \in E_n} \min_{P \in E_k} \min_{W \in E_s} \Phi(x,y,p,w,P,W) \tag{27}$$

Lemma 5. If $\bar{z} = (\bar{x},\bar{y},\bar{p},\bar{w},\bar{P},\bar{W})$ is a saddle point of function $L(z)$ in problem (27), then $(\bar{x},\bar{y})$ is a saddle point of function $K(x,y)$ in problem (26).

For solving problem (27) any of the above methods can be used. For example, the simplest method, (4), yields

$$\begin{aligned} \dot{x} &= -\Phi_x, & \dot{p} &= \Phi_p, & \dot{w} &= \Phi_w \\ \dot{y} &= \Phi_y, & \dot{P} &= -\Phi_P, & \dot{W} &= -\Phi_W \end{aligned} \tag{28}$$

We shall call stationary points those points z_* where the right-hand sides of equation (28) are equal to zero.

Theorem 3. Suppose the constraint qualifications (for constraints g,h and G,H) and strict complementarity hold at a point z_* which is feasible for problem (26), and matrices $\Phi_{xx}(z_*)$ and $-\Phi_{yy}(z_*)$ are positive definite. Then the solution of system (28) locally, exponentially converges to z_* as $t \to \infty$.

The proof is similar to that of theorem 1 and therefore is omitted. Analogously to (28), all other methods can be generalized.

References

[1] Evtushenko, Yu. G., *Some Local Properties of Minimax Problems*, (in Russian), Zhurnal vych. matematiki i matematicheskoy fiziki, Vol.14 , 3, 1974.

[2] Evtuschenko, Yu. G., *Iterative Computational Methods for Solving Minimax Problems*, (in Russian), Zhurnal vych. matematiki i matematicheskoy fiziki, Vol.14,5, 1974.

[3] Polyak, B.T., *Iterative Methods using the Lagrange Multiplier Technique for Solving Extreme Problems with Equality Constraints*, (in Russian), Zhurnal vych. matematiki i matematicheskoy fizicki, Vol.10, 5, 1970.

[4] Zangwill, W.T., *Nonlinear Programming. A Unified Approach*, Prentice-Hall, Englewood Cliffs, N.J., 1969.

[5] Gantmacher, F.R., *The Theory of Matrices*, Vol.1, Chelsea Publishing Co., New York, 1974.

[6] Liapunov, M.A., *Probleme General de la Stabilite du Mouvement*, Reprinted in Annals of Mathematical Studies, 17, Princeton University Press, Princeton, N.J., 1949.

[7] Kantarovich, L.V., and G.P. Akilov, *Functional Analysis in Normed Space*, Ch.15, Macmillan, New York, 1964.