Generalized Lagrange Multiplier Technique for Nonlinear Programming

Evtushenko, Y.

IIASA Working Paper

WP-74-071

1974

GENERALIZED LAGRANGE MULTIPLIER TECHNIQUE FOR NONLINEAR PROGRAMMING

Yuri Evtushenko

December 1974 WP-74-71

Introduction

The numerical methods we present are based on transforming a given constrained minimization problem into an unconstrained maximin problem. This transformation is accomplished by utilizing generalized Lagrange multipler technique. Such an approach permits us to use Newton's 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 non-linear programming problem:

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

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

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

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

Consider an unconstrained maximin problem

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

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 solution to primal nonlinear programming problem (1). As a rule, the Lagrangian is defined as

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

and the following problem is solved

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

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 do not have such difficulties because (2) is an unconstrained maximin problem and for solving (2) we can use all well-known numerical methods for solving unconstrained maximin and saddle point problems. For example, using the simplest gradient method yields the following method

$$\dot{x} = -H_x, \ \dot{p} = H_p, \ \dot{w} = H_w, \ x(0) = x_0, \ p(0) = p_0, \ w(0) = w_0. \tag{4}$$

where H_x, H_p, H_w are n x 1, e x 1, c x 1 vectors, whose ith elements are

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

respectively.

In (4) and everywhere later 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 solution of (2) as $t \to \infty$. The author presented in [1, 2] a number of iterative methods for finding local solutions of a maximin unconstrained problem. Using three of them 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 x e, n x c, n x n Jacobian matrices respectively, whose ijth elements are

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

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

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

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

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

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_*, p_*, w_* respectively such that for all $x \varepsilon A$, $x \neq x_*$, $p \varepsilon Q$, $p \neq p_*$, $w \varepsilon G$, $w \neq w_*$ the following inequalities hold

$$H(x(p,\ w),\ p,\ w) \leq H(x_*,\ p_*,\ w_*) \leq H(x,\ p_*,\ w_*) \qquad (8)$$

where

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

The necessary conditions that z_* be a local maximin of problem (2) are (see [1])

$$H_x(z_*) = 0, \qquad H_p(z_*) = 0, \qquad H_w(z_*) = 0 \qquad (9)$$

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

$$H_{xx}\,\dot{x} + H_{xp}\,\dot{p} + H_{xw}\,\dot{w} = -H_x$$

$$H_{px}\,\dot{x} = -H_p, \qquad H_{wx}\,\dot{x} + H_{ww}\,\dot{w} = -H_w \tag{10}$$

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; $H_{xp} = H^T_{px}$, $H_{xw} = H^T_{wx}$. Utilizing abbreviated notations yields the following continuous and discrete version of (10).

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

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

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

In the case when constraints are absent, these methods coincide with Newton's methods. These methods are also well known and are studied when problem (1) has no inequality constraints (see [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

$$x_{s+1} = x_s - \alpha H_x(z_s),\ p_{s+1} = p_s + \alpha H_p(z_s),\ w_{s+1} = w_s + \alpha H_w(z_s) \quad (13)$$

where $0 < \alpha$ is the step length. The discrete version of other methods can be written in a completely similar 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 all the methods suggested above. Now we shall state some preliminary results.

Define the following set of integers

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

Definition: The constraint qualification holds at a point x if all gradients $\{g^i_x(x)\}$, $1 \le i \le c$ and all gradients $h^j_x(x)$, $j \in B(x)$ are linearly independent.

Definition: The strict complementarity holds at a point z_* if from $h^i(x_*) = 0$ follows that $w^i_* \ne 0$, $1 \le i \le c$.

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

$$H(\bar{x}, p, w) \le H(\bar{x}, \bar{p}, \bar{w}) < H(x, \bar{p}, \bar{w}) \quad (14)$$

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

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

We shall not give a proof of these lemmas, because it is quite similar to the proof of analogous results for problem (3) (see for example [4]).

Consider the following auxilliary problem

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

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 [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 that $y_* = (x_*, u_*)$ be an isolated (unique locally) maximin point of problem (15) are that

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

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

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 $x \varepsilon E_n$, $u \varepsilon E_k$ then the stationary point y_* is a global maximin point of $P(x, u)$. Though y_* may not be a saddle point of $P(x, u)$ (see also [1]).

Lemma 5: Suppose that 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 saddle point of $H(z)$, and x_* 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 $x_* \varepsilon X$, 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 equation

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

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

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

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

also

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

for all i such that $i \bar{\in} B(x_*)$. 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 x^{-T} and taking into account (17) yields

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

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

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

The first system can be rewritten in 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 $i \in B(x_*)$, by assumed constraint qualification all the gradients in (18) are linearly independent, (18) holds if $\bar{p} = 0$ and $\bar{w}^i = 0$ for all $i \in B(x_*)$. But we obtained above that $\bar{w}^i = 0$ for $i \in B(x_*)$. 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 $v = [p^1, p^2, \ldots, p^e, w^1, w^2, \ldots, w^s] \varepsilon E_k, k = e + s$ and $\tilde{h} = [h^{s+1}, h^{s+2}, \ldots, h^c] \varepsilon E_r$, $r = c-s$. Making use of strict complementarity, we obtain $w_*^i = 0$ for all $1 + s \leq i \leq c$. Therefore, omitting arguments we can rewrite matrix 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 0_{ij} is i x j matrix whose elements are all equal to zero; $D(h)$ is the diagonal matrix whose ith diagonal element is h^i. The matrix N can be written in the four blocks form

$$N = \left[\begin{array}{c|c} H_{vx} \quad H_{xx}^{-1} \quad H_{xv} & 0_{kr} \\ \hline 0_{rk} & -2D(\tilde{h}) \end{array}\right]$$

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 n x k matrix. Under assumption of strict complementarity, $w_*^i \neq 0$ for all $1 \leq i \leq s$. Since 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 [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 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 get from lemma 3 that

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

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

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). This point would be stationary. Applying the Taylor formula for first-order expansions, we obtain

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

where $0 \leq t \leq 1$. The Hessian $H_{zz}(z_*)$ is not singular. By continuity of the Hessian 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.

The proof of the lemma 5 is now complete.

Theorem 1: Suppose that the assumptions of lemma 5 are satisfied. Then the solutions of systems (4) - (7), (10) locally, exponentially converge to z_* as $t \rightarrow \infty$ (i.e. exist such positive numbers ε, μ 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), converge 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 $||z_* - z_*|| < \varepsilon$).

Proof: All the methods suggested above have two common properties. They are autonomous and any stationary point z_* is an equilibrium position for all these systems. This permits us to use for proof the linearization principle which was first proved by Liapunov [6] and often called "the first method of

Liapunov". We shall prove on the basis of this technique asymtotic stability of solution $z(t) \equiv z_*$ of systems (4) - (7), (10). This result implies local convergence of their solutions $z(t)$ to a stationary point z_*.

Denote $\delta x = x(t) - x_*$, $\delta p = p(t) - p_*$, $\delta w = w(t) - w_*$, $\delta z = (\delta x, \delta p, \delta w)$. By 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 vector such that

$$\lim \quad \frac{O(||y||)}{||y||} < \infty \quad \text{when } ||y|| \longmapsto 0 \quad .$$

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 complex conjugate to vector δz, Re b denotes real part of complex number b. From (21) we obtain

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

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

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

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

The convergence of methods (4) - (7) can be proved by the 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 expresses by means of the characteristic roots of the following secular equation

$$\begin{vmatrix} H_{xx} + \lambda I_n & g_x & 2h_x D(w) \\ 0_{en} & g_x^T H_{xx}^{-1} g_x + \lambda I_e & 2g_x^T H_{xx}^{-1} h_x D(w) \\ 0_{cn} & 2D(w) h_x^T H_{xx}^{-1} g_x & 4D(w) h_x^T H_{xx}^{-1} h_x D(w) - H_{ww} + \lambda I_c \end{vmatrix} = 0 \qquad (22)$$

where I_j is j x j unit matrix.

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

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

According to lemma 5 the matrices H_{xx} and N are symmetrical and positive definite; hence, 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 $\Delta x_s = x_s - x_*$, $\Delta p_s = p_s - p_*$, $\Delta w_s = w_s - w_*$, $\Delta z_s = (\Delta x_s, \Delta p_s, \Delta w_s)$. 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 asympototically 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 eigenvalue 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

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

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

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

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

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

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 proof of convergence of respective continuous system, as it was shown above.

Theorem 2: Suppose that the assumptions of lemma 5 are satisfied and the function $H_{zz}(z)$ satisfies Lipschitz condition in a vicinity of the 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 completely similar to the proof of Newton's method of convergence theorem [7], and is therefore omitted. To hasten convergence to solution of problem (1) we can use in methods (4) - (7), (10), (11) instead of H the function

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

where a, b - 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 assumptions of theorem 1 hold, then the solution of (25) locally converges to z_* for any $0 \le a$, $0 \le b$.

§3. Numerical examples.

We shall give an example that was solved using three presented methods 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

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

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

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 $\alpha = 0.02$.
Approximate solution of this problem is $F_* = 1.8311$. The iterations were terminated if the difference between current value of $F(x_s)$ and the following one remained less than 10^{-5}. If number of iterations was more than 100, then the process was also manually terminated.

Denote 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 sek were obtained. For the discrete version of (5) N = 100, δ = 0.0056, T = 16 sek, for method (11), N = 4, δ = 0.0001, T = 3 sek.

The modified Newton's method converges after 4 iterations. While this method has the best rate of convergence, it also requires more time per iteration than the other methods. The size of the region of convergence of this method was also less than for the other methods. 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 some other method. 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 like (4), may be used; a more accurate solution would be found by a more complicated method like (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 α = 0.05, α = 0.04, α = 0.02 were used . For α = 0.2 the method (13) does not converge. The increasing of 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 α = 0.02, b = 0.

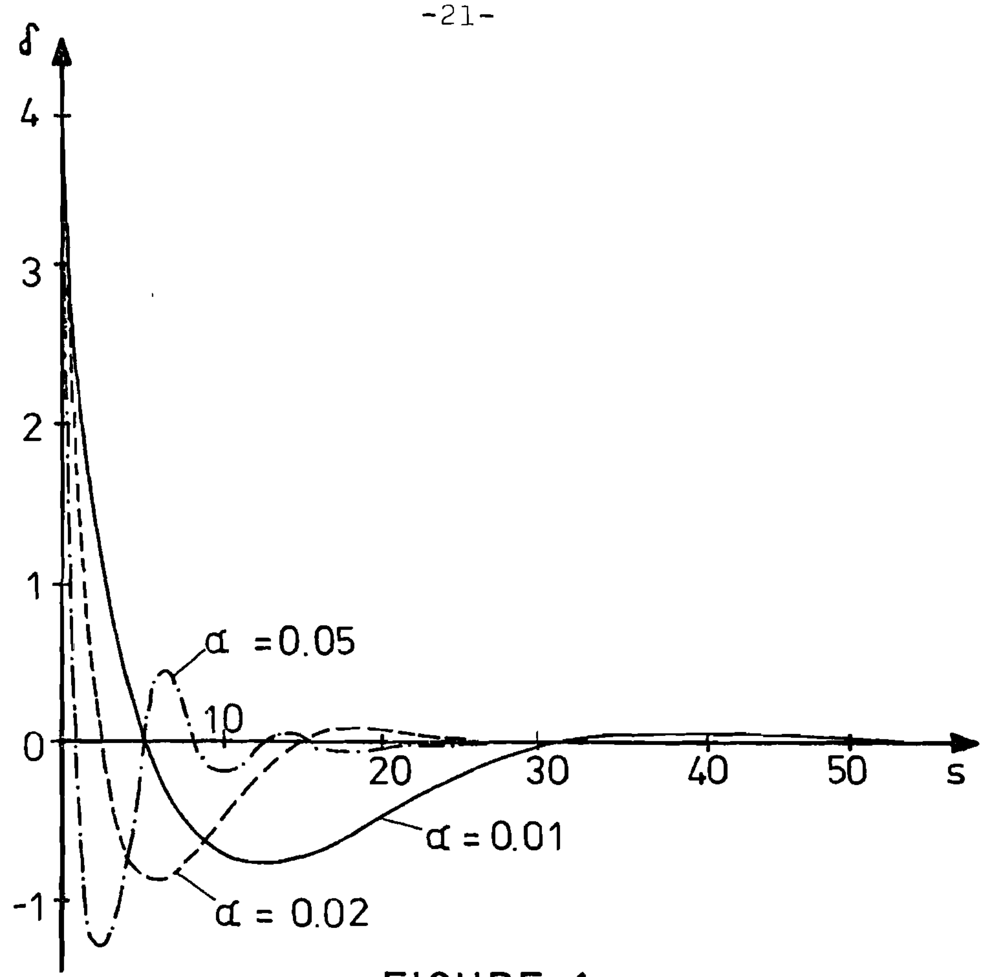

FIGURE 1

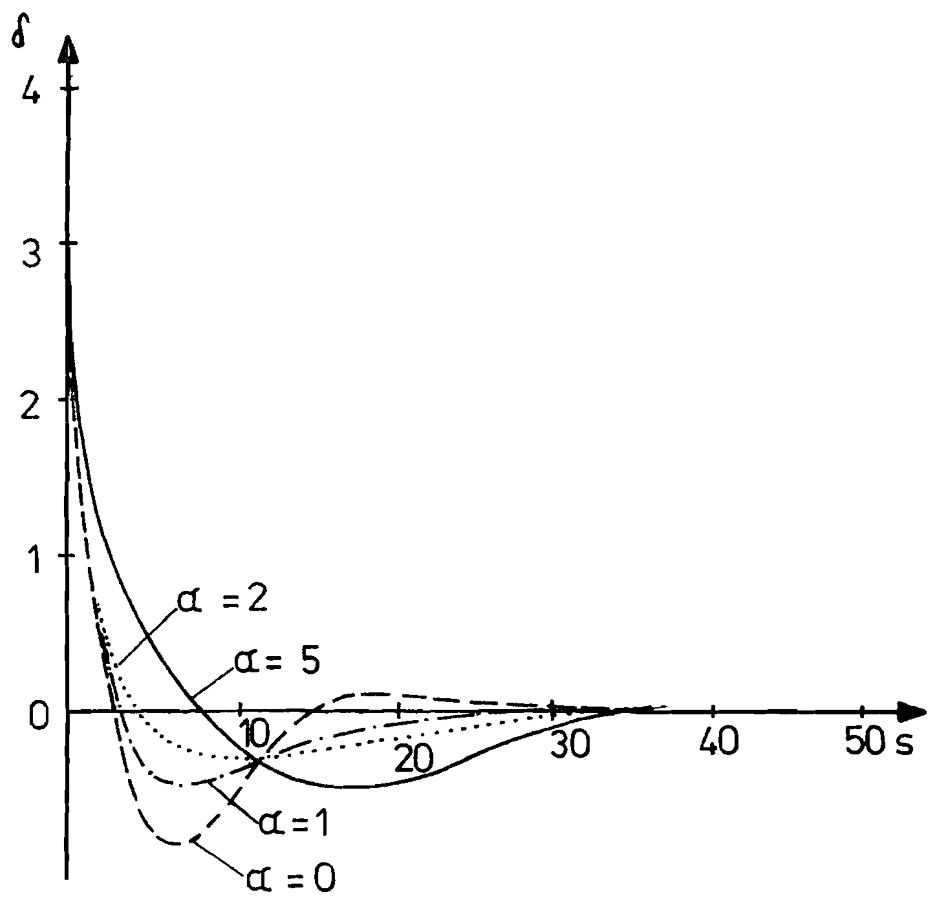

FIGURE 2

Utilization of 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

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

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

Introduce Lagrangian as follows

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

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

Where $P \varepsilon E_k$, $W \varepsilon E_s$, $p \varepsilon E_e$, $w \varepsilon E_c$.

Consider an unconstrained maximin problem

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

<u>Lemma 5</u>: 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. Utilizing, for example, the simplest method (4), yields

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

Those points z_*, where the right-hand sides of the equations of this system are equal to zero, we shall call stationary points.

Theorem 3: Suppose constraint qualifications (for constraints g, h, and G, H) and strict complementarity hold at a point z_* which is a 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 the proof of theorem 1 and therefore is omitted. Analogous to (28), all other methods can be generalized.

REFERENCES

[1] Evtushenko, Yu.G. Some Local Properties of Minimax Problems. Zhurnal vych. matematiki i matematicheskoy fiziki, 1974, Vol. 13, N3, 583-598.

[2] Evtuschenko, Yu.G. Iterative Computational Methods for Solving the Minimax Problems. Zhurnal vych. matematiki i matematicheskoy fiziki, 1974, Vol. 13, N5, 1136-1149.

[3] Polyak, B.T. Iterative Methods using Lagrange Multiplier Technique for Solving Extremium Problems with Equality Constraints. Zhurnal vych. matematiki i matematicheskoy fizicki. 1970. Vol. 10, N5, 1098-1108.

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

[5] Gantmacher, F.R. (1974) The Theory of Matrices, v.I. Chelsea Publishing Company, New York, N.Y.

[6] Liapunov, M.A. 1893, 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 , Chapter 15, MacMillan, New York, 1964.