Pergamon

Comput. Maths. Math. Phys., Vol.34, No.5, pp. 579–590, 1994

0965-5425(94)00086-7

BARRIER-PROJECTIVE METHODS FOR NON-LINEAR PROGRAMMING[1]

YU.G. EVTUSHENKO and V.G. ZHADAN

Moscow

(Revised 6 March 2002)

The class of barrier-projective methods for solving non-linear programming problems is considered. A general approach to their construction, based on a transformation of spaces, is developed. The main attention is devoted to asymptotically stable versions of the methods. Convergence of the continuous and discrete versions of the methods is proved and estimates of the rate of convergence are given.

INTRODUCTION

The gradient projection method was one of the first numerical methods to be used to solve linear and non-linear programming problems [1] – [5]. The ideas of gradient projection and the barrier functions method were subsequently combined, and the approach was used to solve problems of linear and quadratic programming in [6, 7]. It was then applied to general problems of non-linear programming and operations research in [8] – [12]. Interest in this subject intensified considerably after the publication of a paper by Karmarkar in 1984 [13]. This was followed by further papers [14, 15, 16], in which different versions of interior point methods for solving linear programming problems were proposed.

This paper develops the results obtained in [8] – [12]. We describe a unified approach to the construction of barrier-projective methods which involves a change to new spaces in which the structure of the admissible set is considerably simpler than in the initial space. The gradient projection method in its pure form can then be used to find solutions in the transformed space. Returning to the original space, we obtain different versions of the barrier-projective methods described in [12], concentrating on stable versions for which the initial approximation does not have to belong to the admissible set. If the initial approximations are taken from the admissible set, these methods are similar to interior point relaxation, in that the trajectories that they generate do not go out of the admissible set and the minimized function decreases along trajectories. For admissible sets of quite general form it is shown that, if the selected transformation of spaces satisfies certain conditions, the solution of the non-linear programming problem is an asymptotically stable position of equilibrium for all the continuous versions of barrier-projective methods, and their discrete analogues converge locally to this solution at the rate of a geometric progression.

[1] *Zh. Vȳchisl. Mat. Mat Fiz.*, Vol.34, No.5, pp. 669-684, 1994.

Financial support for this research was provided by the Russian Fund for Fundamental Research (Project Code 94-01-01379).

1. A STABLE VERSION OF THE GRADIENT PROJECTION METHOD FOR SOLVING PROBLEMS WITH EQUALITY-TYPE CONSTRAINTS

Let x be a vector of n-dimensional Euclidean space $\mathbb{E}^n$. The scalar function $f(x)$ and vector-function $g(x) : \mathbb{E}^n \to \mathbb{E}^m$ are defined everywhere in $\mathbb{E}^n$. Suppose that f and g and their first derivatives are continuous everywhere. We will assume that all the optimization problems formulated below have solutions. Suppose that we are looking for

$$f_* = \min_{x \in X} f(x), \qquad X = \{x \in \mathbb{E}^n : g(x) = 0_m\}. \tag{1.1}$$

Here and below 0_{ij} is an $i \times j$ null matrix, 0_i is the zero i-dimensional vector, and I_i is the unit matrix of order i. The subscripts in the null and unit matrices will be omitted whenever this does not lead to confusion.

We will introduce the vector of binary variables $u \in \mathbb{E}^m$ and, forming the Lagrange function, compute its gradient:

$$L(x, u) = f(x) + u^\top g(x), \qquad L_x(x, u) = f_x(x) + g_x^\top(x) u.$$

Here g_x is the $m \times n$ matrix of the first derivatives.

We will consider the following system of n ordinary differential equations:

$$dx/dt = -L_x(x, u(x)), \qquad x(x_0, 0) = x_0. \tag{1.2}$$

The function $u(x)$ on the right-hand side of (1.2) is chosen so that all trajectories of the system approach the admissible set X as $t \to +\infty$. To do this, we require that

$$dg/dt = -\tau g(x), \qquad \tau > 0. \tag{1.3}$$

Differentiating the function $g(x)$, from (1.2) and (1.3) we have

$$dg/dt = -g_x(x)[f_x(x) + g_x^\top(x) u(x)] = -\tau g(x). \tag{1.4}$$

If the Gram matrix $\Gamma(x) = g_x(x) g_x^\top(x)$ is non-degenerate, we can find the function $u(x)$ from this condition:

$$u(x) = \Gamma^{-1}(x)[\tau g(x) - g_x(x) f_x(x)].$$

Substituting this function into the right-hand side of (1.2), we have

$$dx/dt = -\{f_x(x) + g_x^\top(x) \Gamma^{-1}(x)[\tau g(x) - g_x(x) f_x(x)]\}. \tag{1.5}$$

By virtue of this system, the derivative of the function $f(x)$ is equal to

$$df/dt = -\|L_x(x, u(x))\|^2 + \tau u^\top(x) g(x). \tag{1.6}$$

Here and below $\|\cdot\|$ is the Euclidean norm in $\mathbb{E}^n$, and $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product.

If the point $x_0 \in X$ or $\tau = 0$, system (1.2), (1.4) represents the gradient projection method in which

$$dx/dt = -[f_x(x) + g_x^\top(x) u(x)], \qquad x_0 \in X, \tag{1.7}$$

$$g_x(x) g_x^\top(x) u(x) + g_x(x) f_x(x) = 0. \tag{1.8}$$

According to (1.6), the function $f(x(x_0, t))$ is monotone decreasing in this case. This function might be non-monotone in the general case, and only monotone decreasing in the set X or near it, where the norm $\|g(x)\|$ is small. The system (1.7), (1.8) is neutrally stable with respect to the admissible set, since if $g(x_0) = c$, $||c|| \neq 0$, we have $g(x(x_0, t)) \equiv c$. By (1.4), the system (1.5) is asymptotically stable relative to the constraints. In fact, if the solutions of system (1.5) are continuable as $t \to \infty$, we have

$$g(x(x_0, t)) = g(x_0)e^{-\tau t}, \qquad \lim_{t\to\infty} g(x(x_0, t)) = 0.$$

Hence, as $t \to \infty$ the trajectories of the system approach points of the admissible set X.

Definition 1. *The constraint regularity condition (CRC) for problem* (1.1) *is satisfied at the point $x \in \mathbb{E}^n$ if the columns of the matrix $g_x^\top(x)$ are linearly independent.*

Let $K^\perp(x)$ denote the vector space generated by vectors $g_x^1(x), \ldots, g_x^m(x)$. If the CRC holds at the point x, the subspace $K^\perp$ is of dimensions m, and its orthogonal complement will be the subspace $K(x) = \{\bar{x} \in \mathbb{E}^n : g_x(x)\bar{x} = 0_m\}$ of dimensions $d = n - m$.

Let W be an $m \times n$ matrix of maximum rank m. The $n \times m$ matrix W^+ denotes the right pseudo-inverse of W, that is, $W^+ = W^\top(WW^\top)^{-1}$, $WW^+ = I_m$. Let $\pi(W)$ denote the matrix $\pi(W) = I_n - W^+W$. System (1.5) can be written in projective form:

$$dx/dt = -\pi(g_x(x))f_x(x) - \tau(g_x(x))^+g(x). \tag{1.9}$$

The first vector on the right-hand side is the projection of the antigradient of $f(x)$ on to the tangent subspace $K(x)$ of the manifold $g(x) = const$, and the second vector lies in the orthogonal subspace $K^\perp(x)$.

Points $x \in \mathbb{E}^n$, at which the right-hand sides (1.9) vanish are called stationary.

Lemma 1. *A point x_* at which the CRC is satisfied is stationary if and only if the pair $[x_*, u_*]$, where $u_* = u(x_*)$, is a Kuhn–Tucker point, that is,*

$$L_x(x_*, u_*) = 0_n, \qquad g(x_*) = 0_m. \tag{1.10}$$

Proof. Sufficiency is obvious, and so we need only prove necessity. Let x_* be a stationary point. Then, in view of the fact that the vectors on the right-hand side of (1.9) are orthogonal, we have

$$\pi(g_x(x_*))f_x(x_*) = 0_n, \qquad (g_x(x_*))^+g(x_*) = 0_n.$$

The first n relations here are the same as the first n relations of (1.10). By virtue of the CRC, the next n relations can only hold if $g(x_*) = 0$. Thus $[x_*, u_*]$ is a Kuhn–Tucker point, which it was required to prove. □

If the sufficient conditions of second order of an isolated local minimum for problem (1.1) given in [17] are satisfied at the stationary point x_*, then the trajectories of system (1.9) locally converge exponentially to x_*, and the trajectories of the system (1.7), (1.8) locally converge to X_* on the admissible set X. This follows from the more general result obtained in Section 3.

2. ALLOWANCE FOR ADDITIONAL CONSTRAINTS OF SIMPLE STRUCTURE

The approach explained in the previous section can be used for problems with more complicated constraints. Suppose we have the problem of finding

$$f_* = \min_{x\in X} f(x), \qquad X = \{x \in \mathbb{E}^n : g(x) = 0_m, \quad x \in \Pi\}, \tag{2.1}$$

where $\Pi \subset \mathbb{E}^n$ is a closed convex set with a non-empty interior.

We introduce the new n-dimensional Euclidean space $\mathbb{E}^n$ with coordinates $[y^1, \ldots, y^n]$. We use the transformation $x = \xi(y)$ to change this space to the initial space. This transformation is constructed so that it is a surjection from $\mathbb{E}^n$ into Π or at least from $\mathbb{E}^n$ into $\operatorname{int}\Pi$. Then each element of Π (or int Π) is the image of not fewer than one element of $\mathbb{E}^n$ and the closure of the image of the set $\mathbb{E}^n$ coincides with Π. Suppose also that the mapping $\xi(y)$ is continuously differentiable everywhere on $\mathbb{E}^n$. We replace the initial problem (2.1) by the following: it is required to find

$$\tilde{f}_* = \inf_{y\in Y} \tilde{f}(y), \qquad Y = \{y \in \mathbb{E}^n : \tilde{g}(y) = 0_m\}. \tag{2.2}$$

Here $\tilde{f}(y) = f(\xi(y))$, $\tilde{g}(y) = g(\xi(y))$, with $\tilde{f}_* = f_*$.

Suppose that the point y_*, is a solution of problem (2.2). Then the point $x_* = \xi(y_*)$ will be a solution of problem (2.1). Thus problem (2.2) can be solved by the method described in the previous section, after which one can revert to coordinates $[x^1, \ldots, x^n]$, giving a method of solving problem (2.1) in the initial space. In a space $y \in \mathbb{E}^n$, the method (1.2), (1.4) takes the form:

$$dy/dt = -[\tilde{f}_y(y) + \tilde{g}_y^\top(y)\tilde{u}(y)], \qquad y(y_0, 0) = y_0 \in \mathbb{E}^n, \tag{2.3}$$

$$\tilde{g}_y(y)\tilde{g}_y^\top(y)\tilde{u}(y) + \tilde{g}_y(y)\tilde{f}_y(y) = \tau\tilde{g}(y). \tag{2.4}$$

The gradients of the functions $\tilde{f}$, $\tilde{g}$ and f, g are connected by the obvious relations $\tilde{f}_y(y) = \tilde{J}^\top(y)f_x(\xi(y))$, $\tilde{g}_y^\top(y) = \tilde{J}^\top(y)g_x^\top(\xi(y))$. Here $\tilde{J}(y) = d\xi(y)/dy$ is a Jacobian matrix. At non-singular points of the transformation $x = \xi(y)$, where the Jacobian is non-zero, the inverse transformation $y = \xi^{-1}(x)$ exists. If this transformation is used and the vector x is taken as the argument of the Jacobian, we will obtain the matrix $J(x) = \tilde{J}(\xi^{-1}(x))$ which now depends on x. Using the relations

$$\frac{dx}{dt} = \frac{d\xi(y)}{dy}\frac{dy}{dt} = \tilde{J}(y)\frac{dy}{dt} = J(x)\frac{dy}{dt},$$

we can find differential equations for the trajectories corresponding to (2.3), (2.4) in the space of x. It must be borne in mind here that if the condition $x_0 = \xi(y_0) \in \Pi$ holds automatically for (2.3), (2.4), we must have $x_0 \in \Pi$ in the space of x. From (2.3), (2.4) we have

$$dx/dt = -G(x)L_x(x, u(x)), \qquad x(x_0, 0) = x_0 \in \Pi, \tag{2.5}$$

$$\Gamma(x)u(x) + g_x(x)G(x)f_x(x) = \tau g(x). \tag{2.6}$$

The two Gram matrices $\Gamma(x) = g_x(x)G(x)g_x^\top(x)$ and $G(x) = J(x)J^\top(x)$ have been introduced here. The analogue of formula (1.6) will be

$$df/dt = -||J^\top(x)L_x(x, u(x))||^2 + \tau u^\top(x)g(x). \tag{2.7}$$

When $\tau = 0$, the method becomes:

$$dx/dt = -G(x)L_x(x, u(x)), \qquad x(x_0, 0) = x_0 \in X, \tag{2.8}$$

$$\Gamma(x)u(x) + g_x(x)G(x)f_x(x) = 0. \tag{2.9}$$

Methods of the type (2.5), (2.6), (2.8), (2.9) are called barrier-projective methods.

Let $K(x|\Pi)$ and $K^*(x|\Pi)$, respectively, denote the cone of admissible directions at the point x relative to the set Π and its dual:

$$\begin{aligned} K(x|\Pi) &= \{z \in \mathbb{E}^n : \exists\lambda(z) > 0 \text{ s.t. } x + \lambda z \in \Pi \quad \forall\, 0 < \lambda < \lambda(z)\}, \\ K^*(x|\Pi) &= \{z \in \mathbb{E}^n : \langle z, y\rangle \geq 0 \quad \forall y \in K(x|\Pi)\}. \end{aligned}$$

Let $S(x|\Pi)$ be a linear hull of the cone $K^*(x|\Pi)$.

Definition 2. *The CRC for problem* (2.1) *holds at a point* $x \in \Pi$ *if all the vectors* $g_x^i(x)$, $1 \leq i \leq m$, *and any non-zero vector* $p \in S(x|\Pi)$ *are linearly independent.*

We will impose the following condition on the transformation $\xi(y)$.

Condition 1. *At each point* $x \in \Pi$ *the matrix* $J(x)$ *is defined and the kernel* $\ker J^\top(x)$ *is the same as* $S(x|\Pi)$.

In particular, it follows from this condition that at all interior points $x \in \operatorname{int}\Pi$ the matrix $J(x)$ is non-degenerate, only becoming singular on the boundary of the set Π.

Note also that, according to Condition 1, the subspace $S^\perp(x|\Pi)$ orthogonal to $S(x|\Pi)$ coincides with the space of the columns of the matrix $J(x)$, and since the vector $-J(x)J^\top(x)$ $L_x(x, u(x))$ belongs to that space, the velocity vector $\dot{x}$ always lies in the orthogonal subspace $S^\perp(x|\Pi)$. Thus, if x is a boundary point of Π, owing to the degeneracy of the matrix $J^\top(x)$ the vectors $\dot{x}$ will belong to the characteristic subspace of the space $\mathbb{E}^n$, which coincides with the space $M(x) - x$, where $M(x)$ is the intersection of all support planes of the set Π at the point x. If the cone $K^*(x|\Pi)$ has a non-empty interior, this subspace degenerates to a single point (the origin of coordinates).

In (2.5), (2.6) the matrix $G(x)$ acts as a barrier to the trajectories in the set Π. In fact, for the trajectory $x(x_0, t)$ starting inside Π to leave Π at a time $t_1 > 0$, there must be a vector $p \in K^*(\mathrm{e}(x_0, t_1)|\Pi)$ such that $\langle \dot{x}(x_0, t_1), p\rangle < 0$. But, as we have noted, the vector $\dot{x}$ always lies in the orthogonal subspace of $S(x|\Pi)$, to which the vector p belongs.

Lemma 2. *Let the transformation* $\xi(y)$ *satisfy Condition* 1. *Then if the CRC is satisfied at the point* x, *the matrix* $\Gamma(x)$ *is positive definite.*

Proof. We will show that the rank of the matrix $B(x) = J^\top(x)g_x^\top(x)$ is equal to m. Then, from the relation $\Gamma(x) = B^\top(x)B(x)$, it will follow that the matrix $\Gamma(x)$ is non-singular and positive semi-definite. If $x \in \mathrm{int}\Pi$, this is obvious, since the matrix $J(x)$ is non-singular, and by the regularity condition the rank of the matrix $g_x^\top(x)$ is equal to m.

Now let $x \in \mathrm{fr}\Pi$. If the rank of the matrix $B(x)$ is less than m, there is a non-zero vector $z \in \mathbb{E}^n$ such that $B(x)z = J^\top(x)g_x^\top(x)z = 0$. But then, according to Condition 1, the vector $p = g_x^\top(x)z$, which is non-zero, will belong to the space $S(x|\Pi)$. Thus the vectors $g_x^i(x)$, $1 \leq i \leq m$, and the vector p will be linearly dependent, which contradicts the CRC. Thus, the matrix $B(x)$ has total rank m. $\square$

On the basis of the statement of Lemma 2, we know that the matrix $\Gamma(x)$ will be non-singular at all points $x \in \Pi$ if the CRC holds. Thus, the unique dependence $u(x)$ can be found by solving (2.6):

$$u(x) = \Gamma^{-1}(x)[\tau g(x) - g_x(x)G(x)f_x(x)].$$

After substituting $u(x)$ into the right-hand side of (2.5), we can rewrite method (2.5), (2.6) in projective form, similar to (1.9):

$$dx/dt = -J(x)[\pi(g_x(x)J(x))J^\top(x)f_x(x) + \tau(g_x(x)J(x))^+g(x)]. \tag{2.10}$$

Definition 3. *The point* $[x_*, u_*] \in \Pi \times \mathbb{E}^n$ *is called a Kuhn–Tucker point for problem* (2.1) *if*

$$L_x(x_*, u_*) \in K^*(x_*|\Pi), \qquad g(x_*) = 0. \tag{2.11}$$

If $L_x(x_*, u_*) \in S(\mathrm{e}_*|\Pi)$ *and the second equation of* (2.11) *holds, then* $[\mathrm{e}_*, u_*]$ *is called a weak Kuhn–Tucker point.*

Lemma 3. *Let the transformation* $\xi(y)$ *satisfy Condition* 1. *Then the point* $x_* \in \Pi$ *at which the CRC holds is stationary for system* (2.10) *if and only if the pair* $[x_*, u_*]$, *where* $u_* = u(x_*)$, *is a weak Kuhn–Tucker point for problem* (2.1).

Proof. At a stationary point

$$G(x_*)L_x(x_*,u_*)=0,$$

that is, $L_x(x_*,u_*)\in \ker G(x_*)$. But the matrices $J^\top(\mathrm{e}_*)$ and $G(x_*)=J(x_*)J^\top(x_*)$ both have rank m, and their null spaces coincide. Thus $L_x(x_*,u_*)\in \ker J^\top(x_*)$ and, therefore, the first inclusion of (2.11) holds. The equation $g(x_*)=0$ follows from (2.6). This proves the lemma. □

The strict supplementary stiffness condition (SSSC) holds at a Kuhn–Tucker point if

$$L_x(x_*,u_*)\in \mathrm{ri}K^*(x_*|\Pi), \tag{2.12}$$

where $\mathrm{ri}\,B$ is the relative interior of the set B.

Now suppose that all the functions $f(x)$, $g^i(x)$, $1\le i\le m$, are twice continuously differentiable. Also, let $N(x)=\{h\in\mathbb{E}^n : g_x(x)J(x)h=0\}$. The sufficient second-order conditions of [17] for problem (2.1) can be reformulated as follows.

Theorem 1. *Let the function $\xi(y)$ satisfy Condition* 1. *Furthermore, let the SSSC be satisfied at the Kuhn–Tucker point $[x_*,u_*]$ and*

$$\langle h, J^\top(x_*)L_{xx}(x_*,u_*)J(x_*)h\rangle>0 \tag{2.13}$$

for any $h\in N(x_*)$ such that $J(x_*)h\neq 0_n$. Then x_* is the point of an isolated local minimum in problem (2.1).

Proof. If x_* is not the point of an isolated local minimum in problem (2.1), there is a sequence of admissible points $\{x_k\}$ which converges to x_* such that $f(x_k)\le f(x_*)$. We will represent x_k in the form $x_k=x_*+\lambda_k s_k$, where $||s_k||=1$, $\lambda_k>0$, $\lambda_k\to 0$. Without loss of generality, we can assume that $s_k\to s_*$, $\|s_*\|=1$. Since $s_k\in K(x_*|\Pi)$ for all $k>0$, $s_*\in \mathrm{cl}K(x_*|\Pi)$, where бl$B$ is the closure of the set B. We have

$$f(x_k)-f(x_*)=\lambda_k\langle f_x(x_*+\lambda_k\theta_k^0 s_k),\ s_k\rangle\le 0, \tag{2.14}$$

$$g^i(x_k)=\lambda_k\langle g_x^i(x_*+\lambda_k\theta_k^i s_k),\ s_k\rangle=0, \qquad 1\le i\le m. \tag{2.15}$$

Here $0\le\theta_k^i\le 1$, $0\le i\le m$. Multiplying (2.15) by u_*^i and adding them to (2.14), after dividing by λ_k and taking the limit we have $\langle L_x(x_*,u_*),s_*\rangle\le 0$. But, according to (2.12), $\langle L_x(x_*,u_*),s_*\rangle\ge 0$, and $L_x(x_*,u_*)\neq 0_n$. Comparing these two inequalities, we conclude that $\langle L_x(x_*,u_*),s_*\rangle=0$, that is, the vector s_* is orthogonal to the vector $L_x(x_*,u_*)\in S(x_*|\Pi)$.

We will show that the vector s_* belongs to the orthogonal subspace $S^\perp(x_*|\Pi)$. In fact, if it did not, the vector s_* could be represented in the form $s_*=a+b$, where $a\in S(x_*|\Pi)$, $b\in S^\perp(x_*|\Pi)$, $a\neq 0$. We have $\langle L_x(x_*,u_*),s_*\rangle=\langle L_x(x_*,u_*),a\rangle=0$. Thus, these two non-zero vectors lie in the same subspace and are orthogonal to one another. But the linear hull $\mathrm{ri}\,K^*(x_*|\Pi)$ actually coincides with the linear hull of the cone $K^*(x_*|\Pi)$, equal to $S(x_*|\Pi)$. Moreover, if $p\in \mathrm{ri}\,K^*(x_*|\Pi)$, then any vector of the subspace $S(x_*|\Pi)$ which lies in some neighborhood of the vector p will also belong to $\mathrm{ri}\,K^*(x_*|\Pi)$. Thus, by virtue of (2.12), we can find a vector $q\in \mathrm{ri}\,K^*(x_*|\Pi)$ for which $\langle s_*,q\rangle<0$, contrary to the inclusion $s_*\in \mathrm{cl}\,K(x_*|\Pi)$. Thus, $s_*\in S^\perp(x_*|\Pi)$. It follows from Condition 1 that $S^\perp(x_*|\Pi)$ is the same as the space of columns of the matrix $J(x_*)$. Thus $s_*=J(x_*)h_*$ for some non-zero vector $h_*\in\mathbb{E}^n$.

Now expanding the functions $f(x)$ and $g^i(x)$ in a Taylor series up to the second term inclusive, we have

$$f(x_k)-f(x_*) \;=\; \lambda_k\langle f_x(x_*),s_k\rangle+\frac{\lambda_k^2}{2}\langle s_k, f_{xx}(x_*+\lambda_k\theta_k^0 s_k)s_k\rangle\le 0, \tag{2.16}$$

$$g^i(x_k) \;=\; \lambda_k\langle g_x^i(x_*),s_k\rangle+\frac{\lambda_k^2}{2}\langle s_k, g_{xx}^i(x_*+\lambda_k\theta_k^i s_k)s_k\rangle=0, \tag{2.17}$$

$0 \le \theta_k^i \le 1,\ 0 \le i \le m$. Again multiplying (2.17) by u_*^i and adding them to (2.16), we have

$$\langle L_x(x_*, u_*), s_k\rangle + \frac{\lambda_k}{2}\left[\langle s_k, f_{xx}(x_* + \lambda_k\theta_k^0 s_k)s_k\rangle + \sum_{i=1}^{m} u_*^i\langle s_k, g_{xx}^i(x_* + \lambda_k\theta_k^i s_k)s_k\rangle\right] \le 0. \quad (2.18)$$

It follows from $s_k \in K(x_*|\Pi)$, $L_x(x_*, u_*) \in K^*(x_*|\Pi)$ that $\langle L_x(x_*, u_*), s_k\rangle \ge 0$. Thus, apart from (2.18), we have

$$\langle s_k, f_{xx}(x_* + \lambda_k\theta_k^0 s_k)s_k\rangle + \sum_{i=1}^{m} u_*^i\langle s_k, g_{xx}^i(x_* + \lambda_k\theta_k^i s_k)s_k\rangle \le 0.$$

Taking the limit in these inequalities we obtain $\langle s_*, L_{xx}(x_*, u_*)s_*\rangle \le 0$, or

$$\langle h_*, J^\top(x_*)L_{xx}(x_*, u_*)J(x_*)h_*\rangle \le 0, \quad (2.19)$$

with $h_* \in N(x_*)$ and $||J(x_*)h_*|| \ne 0$. The inequality (2.19) contradicts (2.13). This proves the theorem. □

In the special case when the set Π coincides with the whole space $\mathbb{E}^n$, taking $\xi(y)$ as the identity transformation $x = y$ we find that the statement of the theorem reduces to the sufficient conditions for an isolated local minimum for problem (1.1) given in [17].

3. THE CONVERGENCE OF BARRIER-PROJECTIVE METHODS

We will investigate the local behaviour of trajectories of system (2.10) in the neighborhood of the point x_*. Suppose that the function $\xi(y)$ is such that the matrix $G(x)$ is continuously differentiable. Suppose $p \in \mathbb{E}^n$. Let $G_x(x;p)$ denote a square matrix of order n whose element (i,j) is equal to

$$G_x^{ij}(x;p) = \sum_{k=1}^{n} \frac{\partial G^{ik}(x)}{\partial x^j} p^k.$$

We impose two additional conditions on the transformation $\xi(y)$:

Condition 2. *At each point* e $\in \Pi$ *for any vector* $p \in \mathrm{ri}\, K^*(x|\Pi)$ *the matrix* $G_x(x;p)$ *is symmetric and its null space coincides with* $S^\perp(x|\Pi)$.

Condition 3. *If* $x \in \Pi$, *then* $h^\top G_x(x;p)h > 0$ *for any non-zero vector* $h \in S(x|\Pi)$ *and for any vector* $p \in \mathrm{ri}K^*(x|\Pi)$.

Under Conditions 1 and 2, the matrices $G(x)$ and $G_x(x;p)$ at each point e $\in \Pi$ commute with one another. For since $J(x)a \in S^\perp(x|\Pi)$ for any vector $a \in \mathbb{E}^n$, we have $G_x(x;p)J(x)a \equiv \equiv 0_n$. But this means that the matrix $G_x(x;p)J(x)$ is a zero matrix and so $G_x(x;p)G(x) = = G_x(x;p)J(x)J^\top(x) = 0_{nn}$. On the other hand, since $G_x(x;p)$ is symmetric,

$$G(x)G_x(x;p) = J(x)J^\top(x)G_x(x;p) = J(x)[G_x(x;p)J(x)]^\top = 0_{nn}.$$

Notice that, owing to the fact that $K^*(x|\Pi) = \{0\}$, at points $x \in \mathrm{int}\,\Pi$ the matrix $G_x(x;p)$ itself will always be zero.

Theorem 2. *Let the function* $\xi(y)$ *satisfy Conditions* $1-3$. *Also let the CRC and the sufficient conditions of second order of Theorem* 1 *hold at a point* x_* *which is a solution of problem* (2.1). *Then for any* $\tau > 0$ *the point* x_* *is an exponentially stable position of equilibrium for system* (2.10).

Proof. The equation in variations for system (2.10) has the form

$$\delta\dot{x} = -Q(x_*, u_*)\delta x, \quad (3.1)$$

where

$$Q(x,u) = \tilde{M}(x)[G(x)L_{xx}(x,u) + G_x(x;L_x(x,u))] + \tau G(x)\tilde{P}(x), \tag{3.2}$$
$$\tilde{M}(x) = I_n - G(x)\tilde{P}(x), \qquad \tilde{P}(x) = g_x^\top(x)[g_x(x)G(x)g_x^\top(x)]^{-1}g_x(x).$$

Suppose, to fix our ideas, that the point x_* is such that the rank of the matrix $G(x_*)$ is equal to s, where $s < n$. Since $G(x_*)$ is a symmetric matrix, we can find an orthogonal matrix U such that $G(x_*) = UHU^\top$ and the matrix H has the form

$$H = \begin{bmatrix} H^B & 0_{s,n-s} \\ 0_{n-s,s} & 0_{n-s,n-s} \end{bmatrix}.$$

Here H^B is a diagonal matrix of order s, whose diagonal elements are the non-zero eigenvalues of the matrix $G(x_*)$. Since $G(x_*)$ is a non-negative definite matrix, they are all strictly positive. Moreover, since the matrix $G_x(x_*;L_x(x_*,u_*))$ is symmetric and the matrices $G(x_*)$ and $G_x(x_*;L_x(x_*,u_*))$ commute, the matrix U can be chosen in such a way that the matrix $Y = U^\top G_x U$ will also be diagonal. Thus $Q(x_*,u_*)$ can be represented in the form

$$Q(x_*,u_*) = URU^\top, \qquad R = (I_n - HU^\top\tilde{P}U)(HU^\top L_{xx}U + Y) + \tau HU^\top\tilde{P}U$$

and, therefore, its eigenvalues are equal to the eigenvalues of the matrix R.

Let V and $V^\perp$ be the null space and orthogonal complement of the matrix H, respectively:

$$\begin{aligned} V &= \ker H = \{y \in \mathbb{E}^n : y^1 = \ldots = y^s = 0\}, \\ V^\perp &= \{y \in \mathbb{E}^n : y^{s+1} = \ldots = y^n = 0\}. \end{aligned}$$

The relation which holds between the subspaces $S(x_*|\Pi)$, $S^\perp(x_*|\Pi)$ and V, $V^\perp$ is: $S(x_*|\Pi) = UV$, $S^\perp(x_*|\Pi) = UV^\perp$.

According to Condition 2, $G_x z = 0$ if $z \in S^\perp(x_*|\Pi)$. Thus $G_x U y = 0$ for all $y \in V^\perp$. It follows that the matrix Y has the form

$$Y = [0_{n,s}, B], \qquad B^\top = [0_{n-s,s}, C],$$

where C is a diagonal non-degenerate matrix of order $n-s$.

Let U^B and U^N denote submatrices of the matrix U consisting, respectively, of the first s and last $n-s$ of its rows. Also, let H^B denote the left-hand upper square submatrix of order s of the matrix H. Let

$$P^B = (U^B)^\top g_{x_B}^\top [g_{x_B} U^B H^B (U^B)^\top g_{x_B}^\top]^{-1} g_{x_B} U^B, \qquad L_{xx}^B = (U^B)^\top L_{xx} U^B.$$

Then the matrix R can be written in the following block-diagonal form:

$$R = \begin{bmatrix} R_1 & R_3 \\ 0_{n-s,s} & R_2 \end{bmatrix}.$$

Here

$$R_1 = (I_s - H^B P^B)H^B L_{xx}^B + \tau H^B P^B, \qquad R_2 = (U^N)^\top G_x U^N.$$

The characteristic equation for the matrix R splits into two equations:

$$|R_1 - \lambda_i I_s| = 0, \qquad |R_2 - \lambda_j I_{n-s}| = 0, \qquad 1 \le i \le s, \qquad s+1 \le j \le n.$$

We will first find the solution of the second equation. If λ_j is the eigenvalue corresponding to the eigenvector $z_j \in \mathbb{E}^{n-s}$, we have

$$(U^N)^\top G_x U^N z_j = \lambda_j z_j, \qquad s+1 \le j \le n,$$

or, multiplying on the right by $z_j^\top$,

$$z_j^\top (U^N)^\top G_x U^N z_j = \lambda_j ||z_j||^2, \qquad s+1 \le j \le n.$$

Since the vectors $h_j = U^N z_j \in S(x_*|\Pi)$, according to Condition 3 the numbers

$$\lambda_j = \frac{z_j^\top (U^N)^\top G_x U^N z_j}{||z_j||^2} = \frac{h_j^\top G_x h_j}{||z_j||^2} \tag{3.3}$$

will be real and strictly positive. Put

$$\hat{\lambda}_1 = \min_{s+1 \le j \le n} \lambda_j > 0.$$

Let Λ be the square root of the matrix H, and Λ^B the upper left-hand square submatrix of order s. Instead of finding the roots of the first equation, we can find the eigenvalues of the matrix $W_1 = (\Lambda^B)^{-1} R_1 \Lambda^B$, similar to the matrix R_1. After elementary algebra we have

$$|W_1 - \lambda_i I_s| = |\hat{M}\hat{L}_{xx}^B + \tau \hat{P} - \lambda_i I_s| = 0, \qquad 1 \le i \le s,$$

where $\hat{M} = I_s - \hat{P}$, $\hat{P} = \Lambda^B P^B \Lambda^B$, $\hat{L}_{xx}^B = \Lambda^B L_{xx}^B \Lambda^B$.

Thus, the matrices $\hat{M}$ and $\hat{P}$ are idempotent, and $\hat{M} \times \hat{M} = \hat{M}$, $\hat{P} \times \hat{P} = \hat{P}$, $\hat{M} \times \hat{P} = 0$. The matrix $\hat{M}$ projects any s-dimensional vector on to the tangent manifold:

$$\hat{K}(x_*) = \{\bar{x} \in \mathbb{E}^s : g_{x_B}(x_*) U^B \Lambda^B \bar{x} = 0_m\}.$$

The matrix $\hat{P}$ projects s-dimensional vectors on to the orthogonal complement $\hat{K}^\perp(x_*)$ of this space.

Let z_i be an eigenvector and λ_i the corresponding eigenvalue of the matrix W_1; then

$$(\hat{M}\hat{L}_{xx}^B + \tau \hat{P}) z_i = \lambda_i z_i, \qquad z_i \in \mathbb{E}^s. \tag{3.4}$$

If the non-zero eigenvector z_i is such that $||\hat{P} z_i|| \neq 0$, multiplying (3.4) on the left by the matrix $\hat{P}$ we obtain $\lambda_i = \tau$. Now if $||\hat{P} z_i|| = 0$, that is, $z_i \in \hat{K}(x_*)$, multiplying (3.4) on the left by $z_i^\top$, we find

$$\lambda_i = \frac{z_i^\top \Lambda^B L_{xx}^B \Lambda^B z_i}{||z_i||^2}. \tag{3.5}$$

We will now take into account that $U^B \Lambda^B z_i = U \Lambda h_i = J h_i$ for some vector $h_i \in \mathbb{E}^n$, whose first s components are equal to the corresponding components of the vector z_i. Then (3.5) can be rewritten in the form

$$\lambda_i = \frac{h_i^\top J^\top L_{xx} J h_i}{||z_i||^2}. \tag{3.6}$$

It follows from $z_i \in \hat{K}(x_*)$, $z_i \neq 0_s$, that $h_i \in N(x_*)$, $Jh_i \neq 0_n$. Thus we conclude on the basis of inequality (2.13) that any eigenvector z_i of the matrix W_1 of the tangent manifold $\hat{K}(x_*)$ corresponds to a positive eigenvalue λ_i and

$$\hat{\lambda}_2 = \min_{i \in \Delta(x_*)} \lambda_i > 0, \qquad \Delta(x_*) = \{i : z_i \in \hat{K}(x_*)\}.$$

Thus, the eigenvalues of the matrix Q split into three groups:
(1) the $n-s$ roots of (3.3),
(2) the k roots $\lambda = \tau$, and
(3) the $s-k$ roots of (3.6).

If $\tau > 0$, then all the eigenvalues of the matrix Q are strictly positive and, according to Lyapunov's theorem about asymptotic stability with respect to a first approximation, the position of equilibrium $x = x_*$ is locally exponentially stable. $\square$

The proof of Theorem 2 implies the following estimate for the rate of convergence of solutions of system (2.10):

$$\lim_{t\to\infty} \frac{\ln|x^i(x_0,t) - x^i_*|}{t} \le -\lambda_*, \qquad \lambda_* = \min[\tau, \hat{\lambda}_1, \hat{\lambda}_2], \qquad 1 \le i \le n. \tag{3.7}$$

Notice that if $x_0 \in X$, then the trajectories of system (2.5), (2.6) coincide with those of system (2.8), (2.9). Thus, on the basis of Theorem 2 it can be stated that the method (2.8), (2.9) is locally exponentially convergent to the point x_* on the admissible set X. The estimate (3.7) still applies, but with λ_* replaced by $\lambda_* = \min[\hat{\lambda}_1, \hat{\lambda}_2]$. Notice also that if the set Π is the whole space $\mathbb{E}^n$, taking $x = y$ as $\xi(y)$, we go from (2.5), (2.6) to the method (1.2), for which the convergence conditions given in Theorem 2 now hold.

Consider the discrete version of method (2.10), which can be written in the form

$$\begin{aligned} x_{k+1} &= x_k - \alpha_k G(x_k) L_x(x_k, u_k), &(3.8)\\ u_k &= \Gamma^{-1}(x_k)[\tau g(x_k) - g_x(x_k) G(x_k) f_x(x_k)], &(3.9)\end{aligned}$$

where $\alpha_k > 0$ is the integration step of system (2.10) according to Euler's scheme.

Theorem 3. *In problem* (2.1) *let the conditions of Theorem* 2 *hold. Then the method* (3.8), (3.9) *converges locally to the point x_* at the rate of a geometric progression if the step α_k is constant and equal to α, where*

$$0 < \alpha < 2/\lambda^*, \tag{3.10}$$

and λ^* is the maximum eigenvalue of the matrix Q given by (3.2).

Proof. We will represent (3.8), (3.9) as a simple iteration:

$$x_{k+1} = \Phi(x_k), \qquad \Phi(x) = x - \alpha G(x) L_x(x, u(x)). \tag{3.11}$$

The point $x = x_*$ is a fixed point of the operator $\Phi(x)$. According to Ostrowski's theorem (see [18]) a sufficient condition for linear local convergence of method (3.11) is that the spectral radius ρ of the matrix $\Phi_x(x_*)$ shall be less than one. Consider the characteristic equation $|\Phi_x(x_*) - \chi I_n| = 0$. It is easy to see that $\Phi_x(x_*) = I_n - \alpha Q$, where the matrix Q is defined by formula (3.2). Let λ be any eigenvalue of the matrix Q; the corresponding value of χ is equal to $1 - \alpha\lambda$. In the proof of Theorem 2 it was established that all the eigenvalues of the matrix Q are real positive numbers. Thus if the step a satisfies Condition (3.10), then $|\chi| < 1$ and hence $\rho < 1$. $\square$

Theorem 3 gives sufficient conditions for the local convergence of method (3.8), (3.9). The rate of convergence is linear, that is, for any $\varepsilon > 0$ and sufficiently large k,

$$||x_k - x_*|| \le C(\rho + \varepsilon)^k,$$

where $\rho = \max[|1 - \alpha\lambda_*|, |1 - \alpha\lambda^*|]$, λ_* is the smallest eigenvalue of the matrix Q, and ' is a positive constant.

We shall consider the important special case of problem (2.1) where the set Π is the positive orthant $\mathbb{E}^n_+$ of the space $\mathbb{E}^n$. Then a transformation of coordinates can be constructed in the following separable form: $\mathrm{e}^i = \xi^i(y^i)$, $1 \le i \le n$. The matrices $J(x)$ and $G(x)$ for such transformations will be diagonal:

$$\begin{aligned} J(x) &= D(\gamma(x)) = \mathrm{diag}(\gamma^1(x^1), \ldots, \gamma^n(x^n)), & \gamma^i(t) &= \xi^i((\xi^i)^{-1}(t)),\\ G(x) &= D(\theta(x)) = \mathrm{diag}(\theta^1(x^1), \ldots, \theta^n(x^n)), & \theta^i(t) &= [\gamma^i(t)]^2.\end{aligned}$$

Let $\sigma(x) = \{i : x^i = 0\}$ be the set of active indices at the point $x \in \Pi$. The cone $K^*(x|\Pi)$ and the subspace $S(x|\Pi)$ for $\Pi = \mathbb{E}^n_+$ have the form

$$\begin{aligned} K^*(x|\Pi) &= \{z \in \mathbb{E}^n_+ : z^i = 0, \quad i \notin \sigma(x)\}, \\ S(x|\Pi) &= \{z \in \mathbb{E}^n : z^i = 0, \quad i \notin \sigma(x)\}. \end{aligned}$$

Thus in this case Condition 1 reduces to the form $\gamma^i(0) = 0$ and $\gamma^i(t) > 0$ if $t > 0$. For Conditions 2 and 3 to hold, we only need the functions $\theta^i(t)$, $1 \le i \le n$, to be differentiable and

$$\dot\theta^i(0) > 0. \tag{3.12}$$

The simplest examples of transformations of this form are:

$$x^i = (y^i)^2/4, \qquad x^i = e^{-y^i}, \qquad 1 \le i \le n.$$

For these we have, respectively, $G(x) = D(x)$ and $G(x) = D^2(x)$. Condition (3.12) holds for the first but is violated for the second.

We shall now consider the case of problem (2.1) where the set Π is a “parallelepiped”:

$$\Pi = \{x \in \mathbb{E}^n : a \le x \le b\}, \qquad a \in \mathbb{E}^n, \qquad b \in \mathbb{E}^n.$$

Here too it is better to construct the transformation $\xi(y)$ in separable form, leading to diagonal matrices $G(x) = D(\theta(x))$. For example, if we use the transformations

$$x = \frac{1}{2}[a + b + (b-a)\sin y], \qquad x = \frac{1}{2}\left[a + b + \frac{2(b-a)}{\pi}\operatorname{arctg} y\right], \tag{3.13}$$

we obtain, respectively,

$$\theta(x) = (b-x)(x-a), \qquad \theta(x) = \frac{(b-a)^2}{\pi^2}\cos^4\frac{\pi(2x-a-b)}{2(b-a)}.$$

Condition 3 reduces to the conditions $\dot\theta(a^i) > 0$, $\dot\theta(b^i) < 0$. It is satisfied only by the first transformation of (3.13).

4. PROBLEMS WITH INEQUALITY-TYPE CONSTRAINTS

We will now consider problem (2.1) with the equality replaced by inequality-type constraints:

$$\min_{x \in X} f(x), \qquad X = \{x \in \mathbb{E}^n : g(x) \le 0_m, \quad x \in \Pi\}. \tag{4.1}$$

The introduction of additional non-negative variables enables the problem to be reduced to the form (2.1) and enables method (2.5), (2.6) to be used to solve it, avoiding equations for the additional artificial variables in the final numerical schemes. This method was investigated for the case when $\Pi = \mathbb{E}^n_+$ in [10, 12].

There is another approach to the construction of barrier-projective methods for solving problem (4.1) similar to that considered in [19]. We form the modified Lagrange function:

$$M(x,u,\tau) = L(x,u) - \frac{1}{2\tau}\langle L_x(x,u), G(x)L_x(x,u)\rangle.$$

The dependence $u(x)$, found from (2.6), is the solution of the unconditional parametric maximization

$$\max_{u \in \mathbb{E}^m} M(x,u,\tau),$$

and method (3.8), (3.9) can be interpreted as simple iteration for solving the system of equations

$$G(x)L_x(x,u(x)) = 0_n. \tag{4.2}$$

It can be extended to case (4.1) by the following iteration:

$$u_k = \arg\max_{u\in\mathbb{E}^m_+} M(x_k,u,\tau), \qquad x_{k+1} = x_k - \alpha G(x_k)L_x(x_k,u_k), \tag{4.3}$$

where $\alpha > 0$ is the integration step of the system.

We will put $\sigma_0(\mathrm{e}) = \{1 \le i \le m : g^i(x) = 0\}$, $N_1(x) = \{h \in \mathbb{E}^n : g^i_x(x)J(x)h = 0, i \in \sigma_0(x)\}$. The CRC and SSSC for problem (4.1) can be reformulated as follows.

Definition 4. *The CRC for problem* (4.1) *holds at the point* $x \in \Pi$ *if all the vectors* $g^i_x(x)$, $i \in \sigma_0(x)$, *and any non-zero vector* $p \in S(x|\Pi)$ *are linearly independent.*

Definition 5. *The point* $[x_*, u_*] \in \Pi \times \mathbb{E}^m_+$ *is a Kuhn–Tucker point for problem* (4.1) *and the SSSC holds there if*

$$L_x(x_*,u_*) \in \mathrm{ri}K^*(x_*|\Pi), \qquad g(x_*) = L_u(x_*,u_*) \in -\mathrm{ri}K^*(u_*|\mathbb{E}^m_+).$$

The following statements are the analogues of Theorems 1 and 2.

Theorem 4. *Let the function* $\xi(y)$ *satisfy Condition* 1. *Also, let the SSSC for problem* (4.1) *hold at the Kuhn–Tucker point* $[x_*, u_*]$, *and for any* $h \in N_1(x_*)$ *such that* $J(x_*)h \neq 0_n$ *let* (2.13) *be satisfied. Then* x_* *is the point of an isolated local minimum in problem* (4.1).

Theorem 5. *Let the function* $\xi(y)$ *satisfy Conditions* $1-3$. *Also, let the CRC and the sufficient conditions of Theorem* 4 *of second order be satisfied at the point* x_*, *which is a solution of problem* (4.1). *Then for any* $\tau > 0$ *and any* α *satisfying inequality* (3.10), *the iterative process* (4.3) *converges locally to* x_* *at a linear rate.*

The proof of Theorem 5 is based on the use of Ostrowsky's theorem and repeats the proof of Theorem 2 almost word-for-word, while the matrix $\tilde{P}$, which occurs in the definition of the matrix (3.2), comprises only the gradients of the active constraints $g^i_x(x_*)$, $i \in \sigma_0(x_*)$.

REFERENCES

1. ROZEN J.B., The gradient projection method for non-linear programming. Part I: Linear constraints. *SIAM J. Appl. Math.*, 8, 1, 181-217; Part II: Non-linear constraints, 9, 4, 514–532, 1961.

2. LEVITIN Ye.S. and POLYAK B.T., Minimization with constraints. *Zh. Vȳchisl. Mat. Mat. Fiz.*, 6, 787–823, 1966.

3. DEM'YANOV V.P. and RUBINOV A.M., *The Approximate Solution of Extremal Problems.* Izd. LGU, Leningrad, 1968.

4. ANTIPIN A.S., Continuous and iterative processes with projection operators. In: *Problems of Cybernetics. Computational Problems of the Analysis of Large Systems*, 1–43. Nauka, Moscow, 1989.

5. TANABE K., A geometric method in non-linear programming. *J. Optimization Theory and Applications*, 30, 2, 181–210, 1980.

6. DIKIN I.I., Iterative solution of problems of linear and quadratic programming. *Dokl. Akad. Nauk SSSR*, 174, 745–747, 1967.

7. DIKIN I.I., On the convergence of an iterative process. *Control Systems*, No.12, 54–60. Novosibirsk, 1974.

8. YEVTUSHENKO Yu.G. and ZHADAN V.G., Numerical methods in operations research. *Zh. Vȳchisl. Mat. Mat. Fiz.*, 13, 583–597, 1973.

9. YEVTUSHENKO Yu.G., Two numerical methods for non-linear programming. *Dokl. Akad. Nauk SSSR*, 215, 38–40, 1974.

10. YEVTUSHENKO Yu.G. and ZHADAN V.G., A relaxation method for solving problems of non-linear programming. *Zh. Vȳchisl. Mat. Mat. Fiz.*, 17, 890–904, 1977.

11. YEVTUSHENKO Yu.G., *Methods of Solving Extremal Problems and their Application to Optimization.* Nauka, Moscow, 1982.

12. YEVTUSHENKO Yu.G. and ZHADAN V.G., Barrier-projective and barrier-Newtonian optimization (for non-linear programming). In: *Communications in Computational Mathematics.* VTs Akad. Nauk S.S.S.R., Moscow, 1991.

13. KARMARKAR N., A new polynomial-time algorithm for linear programming. *Combinatorica*, 4, 373–395, 1984.

14. BARNES E.R., A variation on Karmarkar's algorithm for solving linear programming problems. *Math. Prog.*, 36, 174–182, 1986.

15. VANDERBEI R.J., MECETON M.S. and FREEMAN B.A, On a modification of Karmarkar's linear programming algorithm. *Algorithmica*, 1, 395–407, 1986.

16. WEI ZI-LUAN, An interior point method for linear programming. *J. '®ma. Math.*, Oct., 342–350, 1987.

17. FIACCO A. and McCORMICK G.P., *Non-linear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, New York, 1968.

18. ORTEGA J.M. and RHEINBOLDT W.C., *Iterative Solution of Non-linear Equations in Several Variables.* Academic Press, New York, 1970.

19. ZHADAN V.G., Modified Lagrange functions in non-linear programming. *Zh. Vȳchisl. Mat. Mat. Fiz.*, 22, 296–308, 1982.