GRADIENT METHODS FOR FINDING SADDLE POINTS*

1. In 1958, H. Uzawa proposed the method of Lagrange [1] to solve concave programming problems, the basic questions of whose convergence — in distinction from the differential prototype investigated in [2-5] — remained unanswered (see, for example, [6]). The objective of the present paper is to study this method, which is a gradient method for finding saddle points of a Lagrange function. Since the details of the latter do not as a rule play an important role, the results of the work have been formulated for general concave-convex functions.

Under certain constraints on the problem, which are natural for this particular method, Uzawa formulated a proposition about the \( \varepsilon \)-convergence of the method to the set \( X^* \times Y^* \) of saddle points on the assumption that the length of the step of the method was sufficiently small in comparison with \( \varepsilon \). The proof of this statement in [1], however, contained an error (see [7]). Later, U. Zangwill [8] proved a weaker statement; for any \( \varepsilon > 0 \) and step that is sufficiently small with respect to \( \varepsilon \), there is the sequence \( \{(x^k, y^k)\}_{k=0}^\infty \) generated by the method, at least one vector \( x^k \), lying in an \( \varepsilon \)-neighborhood of this set \( X^* \). The behavior of the sequence \( \{x^k\}_{k=0}^\infty \) for \( k > k_0 \) and the sequence

\{y^n\}_{n=1}^\infty \text{ in } [8] \text{ remained unclear. In this paper we show that Uzawa's claim is completely justified.}

The existence of an \(\varepsilon\)-neighborhood, of course, does not guarantee the convergence of the method, although it suggests a way of constructing a convergent method. Instead of a constant step \(\tau\) we can choose a sequence of steps \(\{\tau^k\}_{k=0}^\infty\) satisfying the conditions

\[ \tau^k > 0, \quad \tau^k \to 0, \quad \sum_{k=0}^\infty \tau^k = \infty. \tag{1} \]

The gradient method with such a choice of step applied to the general concave-convex function was proposed by E. G. Gol'shtein [9], who showed its convergence to the set \(X^* \times Y^*\) for a certain class of functions that does not, however, include the Lagrange function. Convergence for some subsequence follows from the last of the results of [9]. A similar result for one method, which is close to that considered, was derived earlier by Iu. M. Ermol'ev [10]. Below, it is proved that the method of Lagrange, with a sequence of steps satisfying conditions (1), converges to the set \(X^* \times Y^*\).

The method with a length of step that tends to zero will not automatically converge linearly (i.e., with the speed of a geometric progression). In this connection the question arises of the conditions under which convergence can be guaranteed and under which it is linear for a method with a constant step \(\tau\). It turns out that a sufficient condition for global linear convergence of this method for small \(\tau\) is linear independence of the gradients of the active constraints at the point of minimum and strong concavity of the Lagrange function with respect to \(x\). In the case of strict regularity of the problem, the last condition may be weakened to require only strong concavity of the constraints in the Lagrange function for certain subspaces. The weakened condition is necessary in the sense that without it the method converges with zero probability.

2. Let \(X\) and \(Y\) be convex closed solid subsets of Euclidean spaces \(\mathbb{R}^n\) and \(\mathbb{R}^m\), respectively, and \(Z = X \times Y\), \(\psi\) be a concave
function of class $C^r$ on $Z$. Here a derivative at the boundary point $z$ of the set $Z$ is understood to mean a vector $\psi'(z)$, that satisfies the condition

$$\psi(\tilde{z}) - \psi(z) = (\psi'(z), \tilde{z} - z) + o(\|\tilde{z} - z\|), \quad \tilde{z} \in Z.$$

Since $Z$ is a convex solid set, this vector is defined uniquely. We note that the assumption of solid sets $X$ and $Y$ is not restrictive, since, when this is not the case, we could consider affine hulls of the sets $X$ and $Y$ instead of the sets $\mathbb{R}^n$ and $\mathbb{R}^m$. We will assume that the set $Z^* = X^* \times Y^*$ of saddle points of the function $\psi$ is not empty and is bounded.

We will denote the operator of projection in $\mathbb{R}^n$ on the set $X$ by $\pi_1$, and that on $X^*$, by $\pi_1^*$; we will denote the projection operator in $\mathbb{R}^m$ on the set $Y$ by $\pi_2$, and that on the set $Y^*$, by $\pi_2^*$.

We will define for any $\tau > 0$ the transformation $\Gamma_\tau$ on the set $Z$ by the equation

$$\Gamma_\tau(z) = (\pi_1(x + \tau \psi_x(z)), \pi_2(y - \tau \psi_y(z))),$$

$$z = (x, y), \quad \psi'(z) = (\psi_x(z), \psi_y(z)).$$

The gradient process with constant step is defined by the recurrence formula

$$z^{k+1} = \Gamma_\tau(z^k), \quad (2)$$

and that with variable step, by the formula

$$z^{k+1} = \Gamma_{\tau^k}(z^k). \quad (3)$$

Here $\{\tau^k\},^\omega$ is any numerical sequence satisfying conditions (1).

If $x^* \in X^*, y^* \in Y^*, x \in X, \psi(x, y^*) \leq \psi(x^*, y^*)$.

We will assume that the function $\psi$ under consideration satisfies the strict inequality

$$\psi(x, y^*) < \psi(x^*, y^*), \quad x \in X \setminus X^*. \quad (4)$$

Using the terminology of [9], this property can be called stability of the set of saddle points with respect to $x$. The stability property plays an important role in what follows. When it is absent, methods (2) and (3), generally speaking, do...
not converge to the neighborhood of the set \( Z^* \). Thus, for example, we are dealing with a function

\[
\psi(x, y) = (1 - y)x, \quad X = R^1, \quad Y = R^1,
\]

which is the Lagrangian of the problem \( \{ x \mid -x \geq 0 \} \).

We note that the assumptions hold for a Lagrange function of the concave programming problem

\[
\max \{ \varphi(x) \mid f_i(x) \geq 0, \quad i = 1, 2, \ldots, m \}
\]

with differentiable functions \( \varphi \) and \( f_i \), if the function \( \varphi \) is strictly convex and the Slater condition is met. We also note in this connection the error of the assertion in [8] to the effect that condition (4) for the Lagrangian of problem (6), for which the set \( X^* \) consists of one point, holds automatically. The function in (5) is an example.

3. We will take arbitrary positive numbers \( \rho_1 \) and \( \rho_2 \), \( \rho_1 < \rho_2 \), and consider the set

\[
Z(\rho_1, \rho_2) = \{ z \mid z \in Z, \quad \rho_1 \leq \rho(z, Z^*) \leq \rho_2 \},
\]

where \( \rho(z, Z^*) \) is the distance from the point \( z \) to the set \( Z^* \). At this point we will study the behavior of the mapping \( \Gamma \), on the set \( Z(\rho_1, \rho_2) \) for sufficiently small \( \tau \). In earlier studies of gradient methods for finding saddle points [1, 8-10] the closeness of an arbitrary point \( z \) to the set \( Z^* \) was measured by the function \( \rho^2(z, Z^*) \), which is a Liapunov function for the operator \( \Gamma \), outside a certain neighborhood of the set

\[
T = \{ z = (x, y) \mid z \in Z, \quad (x - \pi_1^*(x), \psi_1(z)) - (y - \pi_2^*(y), \psi_2(z)) = 0 \},
\]

but is not such within a small neighborhood of the set \( T \). Below it will be shown that the function \( \psi \) is a Liapunov function for the operator \( \Gamma \), within a small neighborhood of the set \( T \cap Z(\rho_1, \rho_2) \) and, consequently, for sufficiently small positive \( \gamma \), the function \( G_\gamma(z) = \rho^2(z, Z^*) - \gamma \psi(z) \) is a Liapunov function for the operator \( \Gamma \), on the whole set \( Z(\rho_1, \rho_2) \).

We will denote by \( C_1, C_2, C_3, \ldots \), the positive numbers that are encountered in the estimates and that depend for a fixed function \( \psi \) only on \( \rho_1 \) and \( \rho_2 \).
Lemma 1. There exist $C_1$, $C_2$, and $C_3$ such that for all $z \in \mathbb{Z}(\rho_1, \rho_2)$, all $\gamma \leq C_1$, and all $\tau \leq C_2 \gamma$, the inequality
\[ G_\gamma(\Gamma_\gamma(z)) - G_\gamma(z) \leq -C_3 \gamma \tau. \] (7)
holds.

Proof. For a concave-convex function, we have the inequality \[ (x-x^*, \psi_\tau(x, y)) - (y-y^*, \psi_\tau(x, y)) \leq \psi(x, y^*) - \psi(x^*, y), (x, y), (x^*, y^*) \in \mathbb{Z}. \] (8)
Therefore, for $(x^*, y^*) \in \mathbb{Z}$,
\[ (x-x^*, \psi(x, y)) - (y-y^*, \psi(x, y)) \leq 0, \] (9)
in which, if there is equality in (8),
\[ \psi(x, y^*) = \psi(x^*, y^*) = \psi(x^*, y). \] (10)
The second of these equalities means that
\[ \psi(x^*, y) \leq \psi(x^*, \bar{y}), \forall \bar{y} \in Y, \] (11)
and because of condition (4) it follows from the first that
\[ x \in X^*. \] (12)
Let $(x, y) \in T$, $x^* = \pi^*(x)$, $y^* = \pi^*(y)$. In this special case (12) gives $x^* = x$ and from (9) we conclude that
\[ \psi(x, y) \leq \psi(x, \bar{y}), \forall \bar{y} \in Y, \] (13)
i.e., $y$ is a minimum point of function $g(\bar{y}) = \psi(x, \bar{y})$, $\bar{y} \in Y$. But one of the necessary conditions for a minimum is that for any $\tau \geq 0$, $\pi_2(y - \tau g'(y)) = y$.
Thus, if $(x, y) \in T$, $(x', y') = \Gamma_\tau(x, y)$,
\[ y' = y. \] (14)
We will now show that for all $(x, y) \in \mathbb{Z}(\rho_1, \rho_2) \cap T$
\[ \psi(x, y) \leq \max \{\psi(\bar{x}, y) | \bar{x} \in X, \rho(\bar{x}, X^*) \leq \rho_2\} - C_1. \] (15)
The function $\psi$ is a continuously differentiable function.
Therefore, the set $T$ is closed and, thus the set $\mathbb{Z}(\rho_1, \rho_2) \cap T$ is
compact. Since \( \psi \) is a continuous function, to prove \((13)\) it is sufficient for any point \((x, y) \in Z(p_1, p_2) \cap T\) to prove the inequality

\[
\psi(x, y) \leq \max \{ \psi(\bar{x}, y) \mid \bar{x} \in X, \rho(\bar{x}, X^*) \leq \rho_2 \}.
\]

(14)

We note that there exists a point \( \bar{x} \in X \) such that

\[
\psi(x, y) \leq \psi(\bar{x}, y),
\]

(15)

for, otherwise, it would follow from \((11)\) that \((x, y) \in Z^*\), and this would contradict the condition \((x, y) \in Z(p_1, p_2)\). If \( \rho(\bar{x}, X^*) \leq \rho_2 \), \((14)\) is already proved. Let \( \rho(\bar{x}, X^*) > \rho_2 \). We will set \( \bar{x}_0 = t \bar{x} + (1-t)x \), where \( t = \rho_2 \| \bar{x} - x \|^{-1} \). Since \( x \in X^* \) and the set \( X \) is convex, \( \bar{x}_0 \in X, \rho(\bar{x}_0, X^*) \leq \rho_2 \). But the function \( \psi \) is concave with respect to \( x \) and, therefore, it follows from \((15)\) that \( \psi(x, y) < \psi(\bar{x}_0, y) \).

Thus, inequality \((14)\) and, consequently, \((13)\) is proved.

We will now show that there exist numbers \( C_s \) and \( C_0 \) such that for \( \tau \leq C_s \)

\[
\psi(x', y) - \psi(x, y) \geq C_s \tau, \ (x, y) \in Z(p_1, p_2) \cap T, \ (x', y') = f_t(x, y).
\]

(16)

For this we will consider for fixed \( y \) the concave function

\[
g(\bar{x}) = \psi(\bar{x}, y), \ \ \bar{x} \in X.
\]

Moving from point \( x \) to point \( x' \) is a step in the method of gradient projection applied to the maximization of this function. We will use two inequalities that have been established in the investigation of this method ([11, theorem 5.1]),

\[
\begin{align*}
g(\bar{x}) - g(x) \leq (\tau^{-1} \| x' - \bar{x} \| + \| g'(x) \|) \| x' - x \|, & \quad \forall \bar{x} \in X, \\
g(x') - g(x) \geq (\tau^{-1} \| x' - x \| - s(\| x' - x \|)) \| x' - x \|, & \quad (17)
\end{align*}
\]

where in our situation

\[
s(t) = \max \{ \| \psi_z(x + h, y) - \psi_z(x, y) \| \mid (x, y) \in Z(p_1, p_2), \| h \| \leq t \}.
\]

Setting \( \bar{x} \in \text{Arg max} \{ \psi(\bar{x}, y) \mid \bar{x} \in X, \rho(\bar{x}, X^*) \leq \rho_2 \} \), we get by means of \((13)\)

\[
\| x' - x \| \geq C_1 \tau.
\]

(19)

Besides this, it is obvious that
\[ \| x' - x \| \leq C_s \tau. \]  
(20)

Combining inequalities (18)-(20), we get
\[ g(x') - g(x) \geq C_s \tau - C_s \tau s(C_s \tau). \]

But the function \( \psi \) is continuously differentiable, so that \( s(t) \to 0 \) as \( t \to 0 \). Thus, the latter inequality follows from (16). Putting together (12) and (16), we get
\[ \psi(\Gamma, (z)) - \psi(z) \geq C_s \tau, \quad z \in Z(\rho_1, \rho_2) \cap T, \quad \tau \leq C_s. \]  
(21)

From inequality (21) it follows by continuity that for \( \tau \leq C_s \) for all \( z \) from a certain neighborhood \( U \) of the set \( Z(\rho_1, \rho_2) \cap T \)
\[ \psi(\Gamma, (z) - \psi(z) \geq \frac{1}{2} C_s \tau. \]  
(22)

On the other hand, for all \( z = (x, y) \in \mathcal{Z} \) and \( \tau \geq 0 \)
\[ \rho^2(\Gamma, (z), Z^*) \leq \rho^2(z, Z^*) + 2\tau [(x - \pi_z(x), \psi_z(z)) - (y - \pi_z(y), \psi_z(z))] + \tau \| \psi'(z) \|^2. \]  
(23)

Hence, using (8), we get for \( z \in \mathcal{Z}(\rho_1, \rho_2) \)
\[ \rho^2(\Gamma, (z), Z^*) \leq \rho^2(z, Z^*) + C_s \tau. \]

Combining this inequality with (22), we reach the relation
\[ \psi(\Gamma, (z)) \leq \frac{1}{2} C_s \tau + C_s \tau^2, \quad \tau \leq C_s, \quad z \in U. \]  
(24)

According to (8) and the definition of the sets \( T \) and \( U \), there exists a number \( C_{10} \) such that
\[ (x - \pi_z(x), \psi_z(z)) - (y - \pi_z(y), \psi_z(z)) \leq -C_{10}, \quad z \in \mathcal{Z}(\rho_1, \rho_2) \setminus U. \]

Therefore, from (23) it follows that for \( z \in \mathcal{Z}(\rho_1, \rho_2) \setminus U \)
\[ \rho^2(\Gamma, (z), Z^*) \leq \rho^2(z, Z^*) - 2C_{10} \tau + C_s \tau^2. \]

Since the function \( \psi \) on the compact set satisfies a Lipschitz condition and \( \| \Gamma, (z) - z \| \leq \tau \| \psi'(z) \| \) we have
\[ |\psi(\Gamma, (z)) - \psi(z)| \leq C_{11} \tau, \quad \tau \leq C_s, \quad z \in \mathcal{Z}(\rho_1, \rho_2). \]
From the two last inequalities, it follows that
\[ G_{v}(\Gamma_{v}(z)) \leq G_{v}(z) - 2C_{10} \tau + C_{v} \tau^2 + C_{11} \gamma \tau, \ \tau \leq C_{6}, \ z \in \mathbb{Z}(\rho_{1}, \rho_{2}) \backslash U. \ (25) \]

We will set

\[ C_{1} = 4C_{10}[C_{5} + 2C_{11}]^{-1}, \ \ C_{2} = \frac{1}{4} C_{3} C_{9}^{-1}, \ \ C_{3} = \frac{1}{4} C_{5}. \]

Inequality (7) follows from (24) and (25).

Lemma 1 is proved.

4. Let \( \{z^{k}\}_{\infty} \) be a sequence generated by process (2) or (3). We are interested in the behavior of the sequence \( \alpha_{k} = \rho^{k}(z^{k}, Z^{k}) \).

Lemma 1 guarantees the decreasing nature of the sequence of the \( \beta^{k} = G_{v}(z^{k}) \), for subscript \( k \), for which \( \rho_{k}^{2} \leq \alpha_{k}^{2} \leq \rho_{2}^{2} \).

Lemma 2. Let \( \rho_{1} \) and \( \rho_{2} \) be positive numbers,

\[ \rho_{1}^{2} \leq \frac{1}{8} \rho_{2}^{2}; \]

\( \{\alpha^{k}\}_{\infty}, \ \{\beta^{k}\}_{\infty}, \ \{t^{k}\}_{\infty} \)

be sequences, and

\[ K_{1} = \{k | \rho_{k}^{2} \leq \alpha^{k}\}, \ \ K_{2} = \{k | \alpha^{k} \leq \rho_{2}^{2}\}. \]

We will assume that the relations

\[ \alpha^{k} \geq 0, \ \ k = 0, 1, 2, \ldots, \ (27) \]

\[ \alpha^{0} \leq \frac{1}{2} \rho_{2}^{2}, \ (28) \]

\[ |\alpha^{k} - \beta^{k}| \leq \frac{1}{2} \rho_{1}^{2}, \ \ k \in K_{1}, \ (29) \]

\[ |\alpha^{k+1} - \alpha^{k}| \leq \frac{1}{2} \rho_{1}^{2}, \ \ k \in K_{2}, \ (30) \]

\[ \beta^{k+1} \leq \beta^{k} - t^{k}, \ \ k \in K_{1} \cap K_{2}, \ (31) \]

hold.

\[ t^{k} \geq 0, \ \ k = 0, 1, 2, \ldots, \ \sum_{k=0}^{\infty} t^{k} = \infty. \ (32) \]
We will define $k_0$ as the smallest number for which

$$
\sum_{i=0}^{k_0} i \geq \rho_2^2.
$$

(33)

Then for all $k \geq k_0$

$$
\alpha^3 \leq 3 \rho_1^2.
$$

(34)

**Proof.** We will first justify by induction the inequalities $j=0, 1, 2, \ldots$

$$
\alpha^j \leq \frac{3}{4} \rho_2^2, \quad \beta^j \leq \frac{5}{8} \rho_2^2.
$$

(35)

(36)

For $j=0$ inequality (35) follows from (28), and (36), from (29), (26), and (28). Let (35)-(36) hold for some $j$.

We will first consider the case $\alpha^j < \rho_1^2$. Then from (30) and (26),

$$
\alpha^{j+1} \leq \frac{3}{16} \rho_2^2,
$$

(37)

and from (29), (37), and (26), it follows that

$$
\beta^{j+1} \leq \frac{1}{4} \rho_2^2.
$$

Now let $\alpha^j \geq \rho_1^2$, i.e., $j \notin K_1$. This permits us to use inequality (31) for $k=j$. We note that from (30), (35), and (26) it follows that $j+1 \notin K_2$. Therefore, we are justified in using expression (29) with $k=j+1$. The inequality

$$
\beta^{j+1} \leq \frac{5}{8} \rho_2^2
$$

(38)

follows from (31) and (36), and

$$
\alpha^{j+1} \leq \frac{3}{4} \rho_2^2
$$
from (29), (38), and (26). Thus, inequalities (35) and (36) have been proved. From (35) it follows that all \( k \in K_2 \) and we can use inequalities (29) and (30), and, if \( \alpha^k \geq \rho_1^2 \), (31) as well.

Let \( j \) be some number such that \( \alpha^k \geq \rho_1^2, k=0, 1, 2, \ldots, j \).

We will successively use inequalities (27), (29), and (31)

\[-\frac{1}{2} \rho_1^2 \leq \alpha^{j+1} - \frac{1}{2} \rho_1^2 \leq \beta^{j+1} \leq \beta^0 - \sum_{i=0}^{j} t_i.\]

Hence, using (29), (28), (26), and (33), we get

\[\sum_{i=0}^{j} t_i \leq \beta^0 + \frac{1}{2} \rho_1^2 \leq \alpha^0 + \rho_1^2 \leq \frac{5}{8} \rho_2^2 < \sum_{i=0}^{k_0} t_i.\]

Therefore, we have found at least one number \( j < k_0 \) such that

\[\alpha^j \leq \rho_1^2.\] (39)

Now let \( k \) be an arbitrary number greater than \( j \). If \( \alpha^k \leq \rho_1^2 \), inequality (34) holds. Let

\[\alpha^k > \rho_1^2.\] (40)

From (39) and (40) follows the existence of some number \( k_1 < k \), such that

\[\alpha^{k_1} \leq \rho_1^2 \text{ and } \alpha^s > \rho_1^2, \quad s = k_1 + 1, \ldots, k.\] (41)

Using successively inequalities (29), (31), (29), (30), and (41), we get the chain

\[\alpha^k \leq \beta^k + \frac{1}{2} \rho_1^2 \leq \beta^{k+1} + \frac{1}{2} \rho_1^2 \leq \alpha^{k+1} + \rho_1^2 \leq \alpha^k + \frac{3}{2} \rho_1^2 \leq \frac{5}{2} \rho_1^2.\]

Lemma 2 is proved.

For any number \( \rho > 0 \) we will set

\[Z(\rho) = \{z \mid z \in Z, \rho(z, Z^*) < \rho\}.\]

**Theorem 1.** For any numbers \( \varepsilon \) and \( \rho \) that satisfy inequality \( 0 < \varepsilon < \rho \), there exist positive numbers \( \tau(\varepsilon, \rho) \) and \( l(\varepsilon, \rho) \), such that, for any \( \tau \leq \tau(\varepsilon, \rho) \) and any \( z \in Z(\rho) \),
for process (2) for all $k \geq 1, \rho \tau^{-1}$.

Proof. We set $\rho_2 = \rho - 2$ and take an arbitrary $\rho_1 < \frac{1}{\sqrt{8}}$. Let $C_i = C_i(\rho_1, \rho_2)$, $i = 1, 2, 3$ be the numbers that figure in lemma 1.

$$C_{12} = \max \{ |\psi(z)|, \, z \in \mathbb{Z}(\rho_2) \}, \quad (42)$$

$$\gamma = \min \left\{ C_1, \frac{1}{2} \rho_1^2 C_{12}^{-1} \right\}. \quad (43)$$

We now observe that for any $z$ and $z'$, $|\rho^2(z', Z^*) - \rho^2(z, Z^*)| \leq (\rho(z', Z^*) + \rho(z, Z^*)) \|z' - z\|$. In particular, for $z \in \mathbb{Z}(\rho_2)$, $z' = \Gamma_1(z)$; from this, and from inequality $\|z' - z\| \leq \tau \|\psi(z)\|$ it follows that there exists $C_{13} > 0$ such that for $\tau \leq C_{13} \gamma$

$$|\rho^2(\Gamma_1(z), Z^*) - \rho^2(z, Z^*)| \leq C_{13} \tau. \quad (44)$$

We will set

$$\tilde{\tau} = \gamma \min \{ C_2, C_{12}C_{13}^{-1} \}. \quad (45)$$

We will set an arbitrary $z \in \mathbb{Z}(\rho)$ and an arbitrary $\tau \leq \tilde{\tau}$ and consider process (2) for them. Let

$$\alpha^k = \rho^2(z^k, Z^*), \quad \beta^k = G_{\gamma}(z^k), \quad t^k = C_3 \gamma \tau, \quad k = 0, 1, 2, \ldots.$$ 

For the sequences $\{\alpha^k\}_{k=0}^\infty$, $\{\beta^k\}_{k=0}^\infty$, $\{t^k\}_{k=0}^\infty$ the conditions of lemma 2 hold. Actually, it is obvious that (26)-(28) hold. Inequality (29) follows from inequalities $|\alpha^k - \beta^k| \leq \gamma |\psi(z^k)|$ and (42)-(43), and inequality (30) follows from (44), (45), (43), and the condition $\tau \leq \tilde{\tau}$; inequality (31) follows from lemma 1. From lemma 2 it follows that $\alpha^k \leq 3 \rho^2$ for $k \geq \rho^2(\Gamma_1 \rho)$. It remains to set $\rho_1 = \epsilon/2$, $\tau(\epsilon, \rho) = \tilde{\tau}$, $l(\epsilon, \rho) = \rho^2(\Gamma_1 \rho)^{-1}$.

Theorem 1 is proved.

Theorem 2. For any positive number $\rho$ there exists a positive number $\tau(\rho)$, such that for any $z \in \mathbb{Z}(\rho)$ and any sequence $\{t_k\}_{k=0}^\infty$ that satisfies condition (1) and is such that $\tau^k \leq \tau(\rho)$, process (3) converges to the set $Z^*$.

Proof. We will set $\rho_1 = \frac{1}{2} \rho$, $\rho_2 = \rho/2$, determine the value of $\gamma$
by equation (43), and determine the value $\tau(\rho)$ by equation (45). From lemma 1 it follows that, for the sequence

$$a^k = \rho^k(z^k, Z^*)$$,  \quad $b^k = G_\gamma(z^k)$,  \quad $t^k = C_\gamma t^k$,  \quad $k=0, 1, 2, \ldots$,

the conditions of lemma 2 hold. Therefore, there exists a number $k_0$, such that $z^k \in Z(\rho)$, $k \geq k_0$.

We will now set an arbitrarily small positive $\varepsilon$ and set $\rho_1 = \frac{1}{2} \varepsilon$, $\rho_2 = \rho \sqrt{2}$. For a new value $\rho_1$, we will determine the value of $\gamma$ and $\bar{\tau}$ by equations (43) and (45). Since $\tau^k \to 0$ as $k \to \infty$, there exists a number $k_1$, such that $\tau^k \leq \bar{\tau}$, $k \geq k_1$.

We will set $k_2 = \max \{k_0, k_1\}$ and consider the sequences $a^k = a^k_1 + a^k_2$,  \quad $\bar{\tau} = \bar{\tau} + \bar{\tau}_2$,  \quad $k=0, 1, 2, \ldots$, that satisfy the conditions of lemma 2 and, therefore, $a^k \leq \varepsilon^2$ for all sufficiently small $k$. Theorem 2 is proved.

5. We will now consider the conditions of convergence of process (2). Here, besides the conditions set upon the function $\psi$ above, we will set additional conditions. In particular, we will assume that the set $Z^*$ consists of one point $z^* = (z^*, y^*)$ and that the function $\psi$ at this point has a second derivative.

We will first consider the case in which $z^*$ is an interior point of the set of $Z$. We will set

$$A = -\psi_{xx}(z^*, y^*), \quad B = \psi_{xy}(z^*, y^*), \quad C = \psi_{yy}(z^*, y^*)$$.

The operator conjugate to operator $B$, we will denote by $B^*$. Theorem 3. Let $z^*$ be an interior point of the set $Z$ and suppose that there exists no number $\lambda$ nor nonzero vector $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$Au = 0, \quad Cv = 0, \quad Bu = \lambda v, \quad B^* v = \lambda u.$$ (46)

Then for any $\rho > 0$ there is a number $\tau(\rho) > 0$ such that, for all $\tau \leq \tau(\rho)$ and all $z^0 \in Z(\rho)$, process (2) converges linearly to the point $z^*$.

Proof. We will consider the transformation $T: Z \to \mathbb{R}^n \times \mathbb{R}^m$, defined by the equation $Tz = (\psi_x(z), -\psi_y(z))$.

Since $z^* \in \text{Int} Z$, we have $\psi'(z^*) = 0$, and consequently $Tz^* = 0$. Let $D = T'(z^*)$. We will assume that the real part of all eigenvalues ($^1$)
of the operator $D$ are strictly negative. Then, according to the theorem of Lyapunov, there exists a $\tau, >0$, such that for any $\tau \in (0, \tau_1]$, the iteration process

$$z^{t+1} = (I+\tau T)z^t$$  \hspace{1cm} (47)

converges linearly to the point $z^*$, if for any number $k$, $\|z^t - z^k\| \leq \varepsilon_t$, where $\varepsilon_t = \varepsilon_1(\tau, \tau_1)$ is some positive number. By looking in detail at the proof of the theorem, we can show that the number $\varepsilon_t$ can be chosen independently of $\tau$. We will take it so that $\varepsilon_t \leq \rho$. We combine this fact with theorem 1. Let $\varepsilon_2$ be an arbitrary number in the interval $(0, \varepsilon_1)$; $\tau_2 = \tau(\varepsilon_2, \rho)$, where $\tau(\varepsilon, \rho)$ is the variable figuring in the formulation of theorem 1; $\tau \leq \tau_2$; $z^\varepsilon \varepsilon Z(\rho)$; $\{z^t\}_t$ is the corresponding process 1. Then $z^\varepsilon \varepsilon Z(\varepsilon_2)$ for all sufficiently large $k$. Since $z^\varepsilon \varepsilon \text{Int} Z$, there exists an $\varepsilon_2$ such that, if $z^\varepsilon \varepsilon Z(\varepsilon_2)$ and $\tau \leq \tau_2$, then $(I+\tau T)z^\varepsilon \varepsilon \text{Int} Z$, and consequently $\Gamma(z) = (I+\tau T)z$. We will assume that $\varepsilon_2$ satisfies this condition. Therefore, for all sufficiently large $k$, process 2 is determined by the recurrence relation (47), and therefore the theorem is applicable to it. Thus, theorem 3 will hold if we set $\tau(\rho) = \min \{\tau_1, \tau_2\}$.

Thus, it remains only to check that the spectrum of operator $D$ lies in an open left-hand halfplane. Let $\mu + i\lambda$ be an arbitrary eigenvalue of operator $D$; $(u+ir, s+iv)$ be the corresponding nonzero eigenvector. Then $\|u\|^2 + \|v\|^2 \neq 0$ or $\|r\|^2 + \|s\|^2 \neq 0$. For the sake of definiteness, we will assume that $\|u\|^2 + \|v\|^2 \neq 0$. We have

$$-Au + B's = \mu u - \lambda r, \quad -Bu - Cs = \mu s - \lambda v,$$
$$-Ar + B'v = \lambda u + \mu r, \quad -Br - C'v = \lambda s + \mu v. \hspace{1cm} (48)$$

By scalar multiplication on each of these equations and the vectors $u, s, r, \text{and} v$, respectively, and adding, we get

$$-[(Au, u) + (Ar, r) + (Cv, v) + (Cs, s)]$$
$$= \mu [\|u\|^2 + \|v\|^2 + \|r\|^2 + \|s\|^2]. \hspace{1cm} (49)$$

Since $\psi$ is a concave-convex function, $A \geq 0, \ C \geq 0$. Therefore,
from (49) it follows that \( \mu \leq 0 \). If \( \mu = 0 \), (49) gives \( Au = Ar = 0 \), \( Cv = Cs = 0 \). From this and from (48) we get (46), which contradicts the assumption of the theorem. Thus, \( \mu < 0 \). This proves theorem 3.

We will apply this theorem to the study of the method of Lagrange multipliers for problem (6). Let \( x^* \) be a stationary point of the problem. For the sake of definiteness we will assume that

\[
\begin{align*}
f_1(x^*) &= 0, \ldots, f_p(x^*) = 0, \\
&f_{p+1}(x^*) > 0, \ldots, f_m(x^*) > 0.
\end{align*}
\]

We will consider together with the \( \psi \)-function of Lagrange for problem (6), a Lagrange function considering only the constraints active at the point \( x^* \), i.e.,

\[
\psi(x, \tilde{y}) = \varphi(x) + \sum_{i=1}^{p} y_i f_i(x), \quad \tilde{y} = (y_1, \ldots, y_p) \in \mathbb{R}_+^p.
\]

Since \( x^* \) is a stationary point of problem (6), there exists a vector \( \tilde{y}^* = (y_1^*, \ldots, y_p^*) \) such that \( \psi(x^*, \tilde{y}^*) = 0 \).

We will set \( A = -\tilde{\psi}_x(x^*, \tilde{y}^*), B = \tilde{\psi}_x(x^*, \tilde{y}^*) \).

We recall that problem (6) is called strictly regular if: 1) there exists no nonzero vector \( u \in \mathbb{R}^n \) such that \( Au = 0, Bu = 0 \); 2) there exists no nonzero vector \( v \in \mathbb{R}^p \) such that \( B^*v = 0 \) (i.e., the vectors \( f_1'(x^*), \ldots, f_p'(x^*) \) are linearly independent); and 3) the condition of strict supplementary rigidity is observed: \( \tilde{y}^* \in \text{Int} \mathbb{R}_+^p \), i.e., \( y_i^* > 0, i = 1, 2, \ldots, p \).

Strict regularity of the problem is a necessary and sufficient condition for its Lagrange function to have a unique saddle point \( z^* \) and for this point to be stable with respect to disturbances of the problem in the metric \( C^2 \) (see \([12, 13]\)).

Theorem 3'. Suppose that problem (6) is strictly regular and there exists no number \( \lambda \neq 0 \) nor nonzero vector \( u \in \mathbb{R}^n \) such that

\[
Au = 0, \quad B^*Bu = \lambda^2 u.
\]

Then for any \( \rho > 0 \) there is a number in \( \tau(\rho) > 0 \) such that, for all \( \tau \leq \tau(\rho) \) and all \( z^* \in \mathbb{Z}(\rho) \), the method of Lagrange multipliers converges linearly to the point \( z^* \).
Proof. By theorem 1 we can assume that the process for all sufficiently large $k$ is in a small neighborhood of the point $(x^*, y^*)$. Therefore, from (50) we conclude that there exists a constant $C_{13} > 0$ such that

$$f_i(x^*) \geq C_{13}, \quad i = p+1, \ldots, m. \tag{52}$$

Since in this situation $\pi_2$ is a projection operation on the positive octant $R_+^m$, equation $y^{k+1} = \pi_2(y^k - \tau \psi_v(x^k, y^k))$ takes the form

$$y^{k+1}_i = [y^k_i - \tau f_i(x^k)]_+, \quad i = 1, 2, \ldots, m,$$

where $[\alpha]_+$ is the positive part of the number $\alpha$. Therefore it follows from (52) that there exists a number $k_0$, such that

$$y^k_i = 0, \quad i = p+1, \ldots, m, \quad k \geq k_0.$$

Thus, the sequence $\{(x^k, y^k)\}_{k \geq k_0}$ is determined completely by the sequence $\{(x^k, \tilde{y}^k)\}_{k \geq k_0}$, where $\tilde{y}^k$ is the projection of the vector $y^k$ on the space $R^p$ corresponding to the first $p$ coordinates. The sequence $\{(x^k, \tilde{y}^k)\}_{k \geq k_0}$ is given by the recurrence relations

$$x^{k+1} = \pi_1(x^k + \tau \tilde{\psi}_v(x^k, \tilde{y}^k)), \quad \tilde{y}^{k+1} = \pi_2(y^k - \tau \psi_v(x^k, \tilde{y}^k)),$$

where $\pi_1$ is a unit transformation in $R^n$ and $\pi_2$ is the projection operation in $R^p$ on the set $Y = R_+^p$.

Since $\tilde{x}^* = (x^*, \tilde{y}^*) \in \text{Int}(X \times Y)$, theorem 3 may be applied to the function $\tilde{\psi}$. It is necessary only to show that there exists no number $\lambda$ nor nonzero vector $(u, v)$ that satisfy (46). For $\lambda = 0$ the equations (46) cannot hold — this follows from conditions 1) and 2) of strict regularity. Suppose equations (46) hold for some $\lambda \neq 0$. Then the corresponding vector $u$ is different from zero, since otherwise $v = \lambda^{-1} Bu = 0$. Therefore, (51) follows from (46) with $\lambda \neq 0$ and $u \neq 0$, which contradicts the assumption of theorem 3'. Theorem 3' is proved.

Observation. The assumption of theorem 3' is not only sufficient, but in some sense is necessary for convergence of the Lagrange method in the strictly regular problem. If this condition does not hold, then for any $\tau > 0$ the spectral radius of an
arbitrary iterative transformation at the point is greater than unity. Therefore, from the theorem of Hadamard and Perron, it follows that the zone of attraction of the point $\bar{z}$ locally is not a manifold of less than full dimension. Thus, on the set of strictly regular problems for which the assumption of theorem 3' does not hold, the method of Lagrange multipliers converges with zero probability.

We note that the method of Lagrange multipliers for conditions of strict regularity of problem (6) is locally determined by the same formulas as the method of Lagrange multipliers for the problem with equality constraints. The local behavior of the method for the latter was studied by B. T. Poliak [14]. Our discussion regarding proof of the local part of theorem 3 is close to that in [14].

We will now turn to the question of the convergence of process (2) without assuming that $z^* \in \text{Int } Z$. We can perform the corresponding investigation only in the case where $X$ and $Y$ are special sets, to wit $X=R^n$, $Y=R_+^m$. If $\psi$ is a Lagrange function of problem (6), this assumption obviously holds.

Let $\{e_i\}_{1}^{m}$ be a canonical basis in $R^m$. We will set $v_i = (\psi(x^*, y^*), e_i)$. Evidently, $v_i \geq 0$. We will assume, for the sake of definiteness, that $v_1 = 0, \ldots , v_p = 0, v_{p+1} > 0, \ldots , v_m > 0$.

For any point $y = (y_1, \ldots , y_m) \in R^m$ we will denote by $\bar{y}$ the projection of the point $y$ on the subspace $R^p$, corresponding to the first $p$ coordinates. Let $\bar{\psi}$ be the constraint of the function $\psi$ on the set $R^n \times R_+^p$. In distinction from the situation of theorem 3', the point $\bar{z}^* = (x^*, \bar{y}^*)$ is not, generally speaking, an interior point of the set $R^n \times R_+^p$. This condition does not permit us to use the theorem of Liapunov. We investigate the condition of convergence of process (2), using as a Liapunov function, the function

$$G_\gamma(z) = \|z - \bar{z}^*\|^2 - \gamma(\bar{\psi}(z) - \bar{\psi}(\bar{z}^*)), \ z = (x, \bar{y}).$$

We will set

$$A = -\bar{\psi}_{xx}(x^*, \bar{y}^*), B = \bar{\psi}_{y}(x^*, \bar{y}^*), C = \bar{\psi}_{uv}(x^*, \bar{y}^*).$$

**Theorem 4.** Suppose there exists no vector $u \neq 0$ such that
\( A\mathbf{u} = 0 \), and no vector \( \mathbf{v} \neq 0 \) such that \( B\mathbf{v} = 0 \), \( C\mathbf{v} = 0 \). Then for any \( \rho > 0 \), there is a number \( \tau(\rho) > 0 \) such that for any \( \tau \leq \tau(\rho) \) and any \( z^* \in Z(\rho) \), process (2) converges linearly to the point \( z^* \).

**Proof.** Let \( \varepsilon \) be a sufficiently small positive number, and \( \tau(\varepsilon, \rho) \) be the variable figuring in the formulation of theorem 1, \( \tau \leq \tau(\varepsilon, \rho) \). Then for all sufficiently large \( k \) for process (2) we have

\[
\|z^k - z^*\| \leq \varepsilon. \tag{53}
\]

As in theorem 3', here it is sufficient to prove the linear convergence of the process

\[
z^{k+1} = \Gamma_\tau(z^k),
\]

where \( \Gamma_\tau : R^n \times R_+ \to R^n \times R_+ \) is a transformation defined by the equality

\[
\Gamma_\tau(z) = (x + \tau \bar{\Phi}_x(x, \bar{y}), \pi_2(\bar{y} - \tau \bar{\Phi}_y(x, \bar{y}))), \tag{54}
\]

\[
z = (x, \bar{y}).
\]

Let \( \|z - z^*\| \leq \varepsilon, z' = \Gamma_\tau(z) \). From (54) we get, by Taylor's formula,

\[
|\Phi(z') - \Phi(z) - \tau \|\Phi_z(z)\| \| \bar{z} - z^*\|^2 | \leq \tau \|\Phi_y(z)\|^2 + o(\tau\|z - z^*\|^2).
\]

From this and from (23) it follows that

\[
\mathcal{G}_\tau(z') - \mathcal{G}_\tau(z) \leq \tau \{2[(x - x^*, \Phi_x(z))]
\]

\[
- (\bar{y} - \bar{y}^*, \Phi_y(z))] - \gamma[\|\Phi_x(z)\|^2 - \|\Phi_y(z)\|^2]
\]

\[
+ o(\tau\|z - z^*\|^2).
\]

Setting \( u = x - x^*, v = \bar{y} - \bar{y}^*, w = (u, v) \), we rewrite this inequality in the form

\[
\mathcal{G}_\tau(z') - \mathcal{G}_\tau(z) \leq \tau \{-2[(Au, u) + (Cv, v)]
\]

\[
- \gamma[\|Au + B^*v\|^2 - \|Bu + Cv\|^2] + o(\tau\|w\|^2).
\]

Making use of the estimates that follow directly from the triangle inequality

\[
-\|Au + B^*v\|^2 \leq -\frac{1}{2} \|B^*v\|^2 + \|Au\|^2,
\]

\[
\|Bu + Cv\|^2 \leq 2\|Bu\|^2 + 2\|Cv\|^2.
\]
Then
\[
G_1(\bar{z}') - G_1(\bar{z}) \leq -2\tau \left\{ \left( (Au, u) - \frac{\gamma}{2} \|Au\|^2 - \gamma\|Bu\|^2 \right) + \left( (Cv, v) + \frac{\gamma}{4} \|B^*v\|^2 - \gamma\| Cv\|^2 \right) \right\} + o(\tau\|w\|^2).
\]

If the number \( \gamma \) is small enough, it follows from the assumptions of the theorem that the value in the braces of the last inequality has a lower estimate \( C_{1s}\gamma\|\bar{z} - \bar{z}^*\|^2 \), where \( C_{1s} \) is some positive constant. Thus,
\[
G_1(\bar{z}') - G_1(\bar{z}) \leq -2C_{1s}\gamma\|\bar{z} - \bar{z}^*\|^2 + o(\tau\|\bar{z} - \bar{z}^*\|^2) \leq -C_{1s}\gamma\|\bar{z} - \bar{z}^*\|^2,
\]
if the number \( \tau(\varepsilon, \rho)\varepsilon^2 \) is small enough. We will assume that this condition holds. We now know that if \( \gamma \) is small enough,
\[
\frac{1}{2}\|\bar{z} - \bar{z}^*\|^2 \leq G_1(\bar{z}) \leq \frac{3}{2}\|\bar{z} - \bar{z}^*\|^2. \tag{55}
\]

Therefore,
\[
G_1(\Gamma_1(\bar{z})) - G_1(\bar{z}) \leq -\frac{2}{3}C_{1s}\gamma\|\bar{z} - \bar{z}^*\| \leq \varepsilon, \quad \varepsilon \leq \tau(\varepsilon, \rho).
\]

From this and from (53), it follows that
\[
\lim_{k \to \infty} \frac{\psi}{\sqrt{\kappa}C_1(\bar{z}_k)} < 1.
\]

It remains to use the left-hand inequality in (55). Theorem 4 is proved.

In cases where \( \psi \) is the Lagrange function of problem (6), the conditions of theorem 4 imply linear independence of the gradients \( f'_i(x'), \ldots, f'_p(x') \) of the active constraints and strong convexity of the function \( \psi(x, y') \) in a neighborhood of the point \( x \)

**Note**

1) The eigenvalue and eigenvector of operator \( D \), which acts
in a real space, we understand to mean, respectively, the eigenvalue and eigenvector of the complex extension of operator $D$. The terms spectrum and spectral radius are used correspondingly.

References


matematicheskoi fiziki, 1966, Vol. 6, No. 5.


Received December 15, 1975