G. M. Korpelevich (USSR)

THE EXTRAGRADIENT METHOD FOR FINDING SADDLE POINTS AND OTHER PROBLEMS*

In this paper we study a certain modification of the gradient method that uses the idea of extrapolation.

The gradient method [1, 2] is one of the simplest and most natural general methods for finding saddle points and minimizing. However, while the gradient method always converges for the most important classes of minimization problems, it converges for saddle point problems only under extremely rigid assumptions, such as strong concavity-convexity of the function (the results of convergence under such assumptions are given, for example, in [3]). A weaker sufficient condition for convergence is the property of "stability" (see [4]). Even it, however, does not hold for saddle point functions, which is what Lagrange functions for convex programming problems are. It can be shown that for a bilinear saddle point function (the Lagrange function of the linear programming problem or a matrix game payoff function) the gradient method never converges for any length of step (except for certain degenerate cases). Therefore, the problem of using gradient methods for saddle point functions in general remains open.

*Russian text © 1976 by "Nauka" Publishers. "Ekstragradientnyi metod dlia otyskaniia sedlovykh tochek i drugikh zadach," Ekonomika i matematicheskie metody, 1976, Vol. 12, No. 4, pp. 747-756.

One possible way of solving this problem is to replace the initial saddle point function by a modified version that has the same saddle point but has a stability property. (1) In this paper we use a somewhat different solution that consists of modifying the gradient method itself by using extrapolation. It is precisely by sequential approximation that we make a trial step along a gradient, and the value of the gradient in a new "extrapolated" point is used as the direction for actual movement for the next approximation. It turns out that, generally speaking, this method (which we will call an "extragradient" method) converges for a convex-concave function (sufficiently smooth).

For the case of bilinear functions, it can be shown that the method converges with the speed of a geometric progression.

The extragradient method also converges for minimization, solutions of the operator equations, and other problems, but generally speaking under the same assumptions as in the gradient method.

We note that the basic idea of using extrapolated "prices" to give "stability" to the gradient process has already been expressed (for example, [2, chap. II]).

1. Convergence Theorems for the Extragradient Method

1. The extragradient method may be applied to the same class of problem as the gradient method. However, a more interesting application of the extragradient method is the problem of finding saddle points (we will treat this case in particular here).

This problem consists of the following. Suppose that $Q\subset R^n$, $S\subset R^m$ are subsets of Euclidean space; $\varphi(x, y)$ is a numerical function given on $R^n\times R^m$. We must find the point $[x^*, y^*]\in Q\times S$ (the so-called saddle point) such that

$$\varphi(x^*, y)\leqslant\varphi(x^*, y^*)\leqslant\varphi(x, y^*). \qquad (1)$$
$$\forall y\in S \qquad \forall x\in Q$$

We will also assume in problem (1):

a) the sets Q and S are closed and convex;

b) the function $\varphi(x, y)$ is convex on x, concave on y, and differentiable, and its partial derivatives satisfy a Lipshitz condition on $Q\times S$, i.e.,

$$\|\varphi_x(x, y)-\varphi_x(x', y')\|\leqslant L(\|x-x'\|^2+\|y-y'\|^2)^{1/2},$$
$$\|\varphi_y(x, y)-\varphi_y(x', y')\|\leqslant L(\|x-x'\|^2+\|y-y'\|^2)^{1/2};$$

c) the set $X^*\times Y^*$ of saddle points of the function $\varphi(x, y)$ on $Q\times S$ is nonempty.

The extragradient method for finding saddle points of the function $\varphi(x, y)$ is determined by the following recurrence relations:

$$\begin{aligned} \bar{x}^k&=P_Q(x^k-\alpha\varphi_x(x^k, y^k)), \\ \bar{y}^k&=P_S(y^k+\alpha\varphi_y(x^k, y^k)), \\ x^{k+1}&=P_Q(x^k-\alpha\varphi_x(\bar{x}^k, \bar{y}^k)), \\ y^{k+1}&=P_S(y^k+\alpha\varphi_y(\bar{x}^k, \bar{y}^k)), \end{aligned} \tag{2}$$

where $\alpha>0$ is a numerical parameter and P_Q, P_S are operators of the projection on the corresponding sets.

Each iteration of the process (2) under consideration consists of a "trial" gradient step at the point $[\bar{x}^k, \bar{y}^k]$, calculation of the gradients and the resulting "extrapolated" point $\varphi_x'(\bar{x}^k, \bar{y}^k)$, $\varphi_y'(\bar{x}^k, \bar{y}^k)$, and use of these "extrapolated" gradients as the actual direction for movement from the point $[x^k, y^k]$ to get the next approximation $[x^{k+1}, y^{k+1}]$.

Under certain assumptions, the sequence of points $\{[x^k, y^k]\}$, determined by the relations (2) converges to one of the saddle points $[\hat{x}, \hat{y}]\in X^*\times Y^*$.

Theorem 1. If assumptions a)-c) hold and in addition

d) $0<\alpha<1/L$,

there exists a saddle point $[\hat{x}, \hat{y}]\in X^*\times Y^*$ such that $[x^k, y^k]\to[\hat{x}, \hat{y}]$ as $k\to\infty$.

Proof. We will first convert (1) to a slightly different form. It is well known (for example, [3]) that, with the assumptions of the theorem, the point $(x^*, y^*)\in Q\times S$ is a saddle point if and only if the inequalities

$$\begin{aligned} (\varphi_x(x^*, y^*), \quad & x-x^*)\geqslant 0, \quad \forall x\in Q, \\ (\varphi_y(x^*, y^*), \quad & y-y^*)\leqslant 0, \quad \forall y\in S. \end{aligned}$$

hold.

If we denote $u=[x, y]$, $T(u)=[\varphi_x(x, y), -\varphi_y(x, y)]$, $\theta=Q\times S$, the necessary and sufficient conditions may be written as

$$(T(u^*), u-u^*)\geqslant 0, \quad \forall u\in\theta, \tag{3}$$

where $u^*=[x^*, y^*]\in U^*=X^*\times Y^*$.

We further note that conditions b) mean that the operator $T(u)$ is single-valued, definite, and monotonic, i.e.,

$$(T(u)-T(v), u-v)\geqslant 0, \quad u, v\in\theta, \tag{4}$$

and, in addition, satisfies the Lipshitz condition with constant L

$$\|T(u)-T(v)\|\leqslant L\|u-v\|. \tag{5}$$

The iterative problems (2) may now be written as

$$\begin{aligned} \bar{u}^k&=P_\theta(u^k-\alpha T(u^k)),\\ u^{k+1}&=P_\theta(u^k-\alpha T(\bar{u}^k)). \end{aligned} \tag{6}$$

We should note that the sequence $\{u^k\}$, defined by the relations (6), converges to some point $\hat{u}\in U^*$. For arbitrary $u^*\in U^*$ we estimate $\|u^{k+1}-u^*\|^2$. Using for this the well-known property of a projection on a convex set $\forall u$

$$(u-P_\theta(u), v-P_\theta(u))\leqslant 0, \quad \forall v\in\theta, \tag{7}$$

from which it follows that $\|u-v\|^2\geqslant\|u-P_\theta(u)\|^2+\|v-P_\theta(u)\|^2$, $\forall v\in\theta$, $\forall u$, for $v=u^*$, $u=u^k-\alpha T(\bar{u}^k)$, in accord with conditions (6), gives us

$$\begin{aligned} \|u^{k+1}-u^*\|^2&\leqslant\|u^k-\alpha T(\bar{u}^k)-u^*\|^2-\|u^k-\alpha T(\bar{u}^k)-u^{k+1}\|^2\\ &=\|u^k-u^*\|^2-\|u^k-u^{k+1}\|^2+2\alpha(T(\bar{u}^k), u^*-u^{k+1}). \end{aligned} \tag{8}$$

We also note that from the monotonicity of operator $T(u)$ and inequality (3) follows the relation

$$\begin{aligned} 0&\leqslant(T(u)-T(u^*), u-u^*)=(T(u), u-u^*)\\ &-(T(u^*), u-u^*)\leqslant(T(u), u-u^*), \quad \forall u\in\theta. \end{aligned} \tag{9}$$

In particular, for $u=\bar{u}^k$ inequality (9) gives $(T(\bar{u}^k), u^*-\bar{u}^k)\leqslant 0$, whence the estimate

$$(T(\bar{u}^k),\ u^*-u^{k+1})=(T(\bar{u}^k),\ u^*-\bar{u}^k)+(T(\bar{u}^k),\ \bar{u}^k-u^{k+1})\leqslant(T(\bar{u}^k),\ \bar{u}^k-u^{k+1}) \tag{10}$$

follows immediately.

We use (10) in the basic chain of inequalities (8)

$$\begin{aligned}\|u^{k+1}-u^*\|^2\leqslant\|u^k-u^*\|^2-\|u^k-u^{k+1}\|^2+2\alpha(T(\bar{u}^k),\ \bar{u}^k-u^{k+1})\\=\|u^k-u^*\|^2-\|u^k-\bar{u}^k\|^2-\|\bar{u}^k-u^{k+1}\|^2-\\-2(u^k-\bar{u}^k,\ \bar{u}^k-u^{k+1})+2\alpha(T(\bar{u}^k),\ \bar{u}^k-u^{k+1})=\\=\|u^k-u^*\|^2-\|u^k-\bar{u}^k\|^2-\|\bar{u}^k-u^{k+1}\|^2+\\+2(u^k-\alpha T(\bar{u}^k)-\bar{u}^k,\ u^{k+1}-\bar{u}^k).\end{aligned} \tag{11}$$

We now estimate the scalar product in the last expression (11), breaking it into the sum of two other scalar products; the nonpositiveness of one follows from (7) for $u=u^k-\alpha T(u^k)$, $v=u^{k+1}$; and the second we estimate using the Cauchy-Buniakovskii inequality

$$\begin{aligned}(u^k-\alpha T(\bar{u}^k)-\bar{u}^k,\ u^{k+1}-\bar{u}^k)=(u^k-\alpha T(u^k)-\bar{u}^k,\ u^{k+1}-\bar{u}^k)\\+(\alpha T(u^k)-\alpha T(\bar{u}^k),\ u^{k+1}-\bar{u}^k)\leqslant\alpha(T(u^k)-T(\bar{u}^k),\ u^{k+1}-\bar{u}^k)\\\leqslant\alpha\|T(u^k)-T(\bar{u}^k)\|\,\|u^{k+1}-\bar{u}^k\|.\end{aligned} \tag{12}$$

In addition,

$$\|u^k-\bar{u}^k\|^2+\|\bar{u}^k-u^{k+1}\|^2\geqslant2\|u^k-\bar{u}^k\|\,\|\bar{u}^k-u^{k+1}\|. \tag{13}$$

We now continue the basic chain of equalities (11), using (12) and (13) and also proceeding in such a way that the operator $T(u)$ satisfies the Lipshitz condition (5)

$$\begin{aligned}\|u^{k+1}-u^*\|^2\leqslant\|u^k-u^*\|^2-\|u^k-\bar{u}^k\|^2-\|\bar{u}^k-u^{k+1}\|^2\\+2\alpha L\|u^k-\bar{u}^k\|\,\|u^{k+1}-\bar{u}^k\|\leqslant\|u^k-u^*\|^2-\|u^k-\bar{u}^k\|^2\\-\|\bar{u}^k-u^{k+1}\|^2+\alpha^2L^2\|u^k-\bar{u}^k\|^2+\|u^{k+1}-\bar{u}^k\|^2.\end{aligned}$$

Thus, we finally reach the estimate

$$\|u^{k+1}-u^*\|\leqslant\|u^k-u^*\|^2-(1-\alpha^2L^2)\,\|u^k-\bar{u}^k\|^2. \tag{14}$$

According to the conditions of theorem 1, $\alpha^2L^2>0$, so that from (14) it follows that

$$\|u^k-\bar{u}^k\|\to0\quad\text{as}\quad k\to\infty, \tag{15}$$

and the sequence $\|u^k-u^*\|$ is nonincreasing and therefore $\{u^k\}$ is bounded. The finiteness of the space examined and the closedness of the quantity θ confirm the existence of the similar subsequence $\{u^{k_i}\}$

$$u^{k_i}\to\hat{u} \quad \text{as} \quad k_i\to\infty, \quad \hat{u}\in\theta. \tag{16}$$

We consider the function $\Phi(u)=\|u-\bar{u}\|^2$, where $\bar{u}=P_\theta(u-\alpha T(u))$. We will show that, if for $u=\tilde{u}\in\theta$ the function $\Phi(\tilde{u})=0$, then $\tilde{u}\in U^*$. Actually, $\Phi(\tilde{u})=0$, i.e., $\tilde{u}=P_\theta(\tilde{u}-\alpha T(\tilde{u}))$. Then by the property of projections (7) for $\forall v\in\theta$, we get $(\tilde{u}-\alpha T(\tilde{u})-\tilde{u},\ v-\tilde{u})\leqslant 0$, whence we get $(T(\tilde{u}),\ v-\tilde{u})\geqslant 0,\ \ \forall v\in\theta$. According to (3) this means that $\tilde{u}\in U^*$.

From the continuity of $\Phi(u)$ and (15) and (16) it follows that $\Phi(\hat{u})=0$, as soon as it is shown that $\hat{u}\in U^*$.

Inequality (14) holds for an arbitrary point $u^*\in U^*$; thus it follows that $\{\|u^k-\hat{u}\|\}$ is monotonic for $\hat{u}$, as well. Therefore, not only the sequence $\{u^{k_i}\}$, but also the sequence $\{u^k\}$ converges to $\hat{u}$. The theorem is proved.

<u>Observation</u>. To reach inequality (14) we nowhere used the finite dimensionality of the space. Therefore, if the initial problem consists of finding a saddle point in Hilbert space, it follows from (14) that $\|u^k-\bar{u}^k\|\to 0$ as $k\to\infty$. Thus, in this case we can say that for the sequence $\{u^k\}$ in the limit the condition $u=\bar{u}$, which characterizes saddle points, holds.

2. We now consider the application of the extragradient method to other problems. We note, first of all, that in the proof of the theorem of convergence, the necessary and sufficient conditions for the saddle point in problem (1) were written in the form of a variational inequality [7]

$$(h(z),\ \ v-z)\geqslant 0, \tag{17}$$

where $z,\ v\in R^l$; $h(z)$ is an operator carrying R^l into R^l. A solution of this type of inequality on the set $\Omega\subset R^l$ is any point $z^*\in\Omega$ such that $(h(z^*),\ v-z^*)\geqslant 0,\ \forall v\in\Omega$. Thus, in the proof of the theorem, the solution of problem (1) on saddle points was reduced to a solution of a variational inequality (3).

The extragradient method for solving the variational inequality (17) has the form

$$\bar{z}^k = P_\Omega(z^k - \alpha h(z^k)),$$
$$z^{k+1} = P_\Omega(z^k - \alpha h(\bar{z}^k)). \qquad (18)$$

With regard to the convergence of the sequence $\{z^k\}$, defined by (18), in the proof of theorem 1 we nowhere used the fact that the operator $T(u)$ is connected in some special way to the saddle point problem (1); it was recognized only that the operator $T(u)$ is monotonic (4) and satisfies a Lipshitz condition (5), and that the set θ, on which the solution of the variational inequality (3) was being sought, was convex and closed. Thus, in fact we proved a more general statement.

Theorem 2. Suppose that the variational inequality (17) on the set $\Omega \subset R^l$, is solved and:

a') Ω is a convex, closed set; $(h(z)-h(v), z-v) \geqslant 0$, $\forall z, v \in \Omega$,

b') $h(z)$ is a monotonic operator satisfying a Lipshitz condition: $\|h(z)-h(v)\| \leqslant L\|z-v\|$, $\forall z, v \in \Omega$;

c') the set of solutions Z^* of inequalities (17) on Ω is nonempty; and

d') $0<\alpha<1/L$. Then the sequence $\{z^k\}$, defined by the recurrence relations (18), converges to some $\hat{z} \in Z^*$, the solution of the variational inequality being considered.

We note that the gradient process $z^{k+1}=P_\Omega(z^k-\alpha h(z^k))$ for inequality (17) converges, generally speaking, only under stronger assumptions, such as in [7], where strong monotonicity of the operator $h(z)$

$$(h(z)-h(v), z-v) \geqslant m\|z-v\|^2, \quad m>0.$$

is required.

Certain important classes of problems — such as minimization problems, the solution of operator equations, the search for equilibrium points in n-person games, and others, as well as the saddle point problem — converge to the solution of variational inequalities [7] and consequently for all these problems, given certain assumptions, the extragradient method corresponding to (18) converges. Below we will also give examples of the use of these.

We consider the minimization problem

$$\min_{z\in\Omega} f(z). \tag{19}$$

Under the condition that $\Omega\subset R^l$ is closed and convex, and $f(z)$ is a convex, differentiable function with a derivative satisfying a Lipshitz condition $\|f'(z)-f'(z')\|\leqslant L\|z-z'\|$, the necessary and sufficient conditions for $z^*\in\Omega$ to be a minimum in (19) may be written as: $(f'(z^*),\ v-z^*)\geqslant 0,\ \forall v\in\Omega$. Thus, the minimization problem reduces to solution of a variational inequality (17) in which the operator $h(z)=f'(z)$. It is easy to see that, in assuming the existence of a solution of problem (19), all the conditions of theorem 2 hold, and it can be claimed that the extragradient process for the minimization problem

$$\bar{z}^k=P_\Omega(z^k-\alpha f'(z^k)),$$
$$z^{k+1}=P_\Omega(z^k-\alpha f'(\bar{z}^k))$$

for $\alpha<1/L$ converges to some solution of problem (19). However, we observe that for the same assumptions (with $\alpha<2/L$) the usual gradient method also converges:

$$z^{k+1}=P_\Omega(z^k-\alpha f'(z^k)).$$

To solve the operator equation

$$z=S(z),\quad z\in R^l, \tag{20}$$

with a nonexpanding operator $S(z)$: $\|S(z)-S(z')\|\leqslant\|z-z'\|$, the necessary and sufficient conditions for z^* to be a solution may be written as $(z^*-S(z^*),\ v-z^*)\geqslant 0,\ \forall v\in R^l$. The problem converges to solution of a variational inequality (17) with $h(z)=z-S(z)$, $\Omega=R^l$. In assuming the existence of a solution of this operator equation, the conditions in theorem 2 hold with $0<\alpha<{}^1/_2$ and the extragradient method

$$\bar{z}^k=z^k-\alpha(z^k-S(z^k)),$$
$$z^{k+1}=z^k-\alpha(\bar{z}^k-S(\bar{z}^k))$$

converges to solution of equation (20). The method of sequential approximation $z^{k+1}=S(z^k)$ for this problem generally speaking diverges, although the "gradient" method $z^{k+1}=z^k-\alpha(z^k-S(z^k))$ with $0<\alpha<1$ converges [8].

Thus, for the minimization problem (19) and the solution of the operator equations (20), the extragradient method has no advantages over the gradient method. The situation is different with regard to the problem of finding equilibrium points in n-person games, which is just like the saddle point problem we have already treated – that is, in contrast to the gradient method, the extragradient method generally converges for an arbitrary game.

2. The Extragradient Method for Linear Programming and the Solution of Matrix Games

In the present section we consider in greater detail the extragradient method for finding saddle points of a bilinear functional. Linear programming and the solution of matrix games reduce to a problem of this type. If we consider the pair of dual linear problems

$$\begin{array}{ll} (c, x)\to\min, & (b, y)\to\max, \\ A^T x\geqslant b, & Ay\leqslant c, \\ x\geqslant 0, & y\geqslant 0, \end{array}$$

where x, $c\in R^n$; y, $b\in R^m$; A is a matrix of dimension $n\times m$, and A^T is the transpose, the solution of these problems reduces to the search for the saddle points of the corresponding Lagrange function on the set $x\geqslant 0$, $y\geqslant 0$

$$\min_{x\geqslant 0}\max_{y\geqslant 0} L(x, y),\ L(x, y)=(c, x)+(b, y)-(Ay, x). \tag{21}$$

The problem of finding optimal strategies of matrix games with a payoff matrix A also reduces to the search for saddle points of a bilinear functional

$$M(x, y)=(e, x)+(e, y)-(Ay, x) \tag{22}$$

on the set $x\geqslant 0$, $y\geqslant 0$. Here $e=(1, 1,\dots, 1)$ is the vector of similar dimension. (Problem (22) is a special case of (21).)

The extragradient method for finding saddle points of the bilinear functional (21) on the set $x \geqslant 0$, $y \geqslant 0$ may be written in the following way:

$$\begin{aligned} \bar{x}^k &= [x^k + \alpha(Ay^k - c)]^+, \\ \bar{y}^k &= [y^k - \alpha(A^T x^k - b)]^+, \\ x^{k+1} &= [x^k + \alpha(A\bar{y}^k - c)]^+, \\ y^{k+1} &= [y^k - \alpha(A^T \bar{x}^k - b)]^+, \end{aligned} \tag{23}$$

where $[p]^+ = \max\{0, p\}$ for the scalar p and $[p]^+ = ([p_1]^+, [p_2]^+, \ldots, [p_l]^+)$ for the vector $p = (p_1, p_2, \ldots, p_l)$.

We will prove the following theorem concerning the convergence of the process (23).

Theorem 3. If: a) the bilinear functional $(c, x) + (b, y) - (Ay, x)$ on the set $x \geqslant 0$, $y \geqslant 0$ has a unique saddle point $[x^*, y^*]$; and b) $0 < \alpha < 1/\|A\|$, then $\{[x^k, y^k]\}$ converges geometrically; $\|x^{k+1} - x^*\|^2 + \|y^{k+1} - y^*\|^2 \leqslant q(\|x^k - x^*\|^2 + \|y^k - y^*\|^2)$, where $0 < q < 1$ (the implicit form of q is used in the proof of the theorem).

Proof. We will denote by a^i the i-th row of matrix A; by $\bar{a}^j$, the j-th column of matrix A; and by z_i, the i-th component of the vector $z = (z_1, z_2, \ldots, z_l)$. Suppose that for the saddle point $[x^*, y^*]$ of the functional (21) which we are considering

$$\begin{aligned} I^* &= \{i : (a^i, y^*) = c_i\}, \\ J^* &= \{j : (\bar{a}^j, x^*) = b_j\}. \end{aligned} \tag{24}$$

Then, as is known, it follows from the assumption of uniqueness of the saddle point that

$$\begin{aligned} &(a^i, y^*) < c_i, \quad x_i^* = 0, \quad i \bar{\in} I^*, \\ &(\bar{a}^j, x^*) > b_j, \quad y_j^* = 0, \quad j \bar{\in} J^*, \\ &x_i^* > 0, \quad i \in I^*, \\ &y_j^* > 0, \quad j \in J^*. \end{aligned} \tag{25}$$

We will now choose an arbitrarily small $\varepsilon > 0$, such that for all $[x, y]$ in an ε-neighborhood $[x^*, y^*]$: $\|x - x^*\|^2 + \|y - y^*\|^2 \leqslant \varepsilon$ the inequalities

$$
\begin{aligned}
&(\bar{a}^j, x) - b_j \geqslant \gamma > 0, \quad j \bar{\in} J^*,\\
&(a^i, y) - c_i \leqslant -\gamma, \quad i \bar{\in} I^*,\\
&x_i + \alpha(a^i y - c_i) \geqslant 0, \quad i \in I^*,\\
&y_j - \alpha(\bar{a}^j x - b_j) \geqslant 0, \quad j \in J^*.
\end{aligned} \tag{26}
$$

hold.

According to (24) and (25) and the continuity of the functions in (26), there exist such $\gamma > 0$ and $\varepsilon > 0$.

It is easy to see that for method (23) all the assumptions of theorem 1 hold (in particular, for L we may take $\|A\|$). Therefore, the sequence $\{[x^k, y^k]\}$ converges and for sufficiently large $k > K$ all points $\{[x^k, y^k]\}$ and $\{[\bar{x}^k, \bar{y}^k]\}$ will be contained within the ε-neighborhood of $[x^*, y^*]$ that we are considering, and the inequalities of (26) will hold for them. Therefore, for $k > K$

$$
x_i^{k+1} = [x_i^k + \alpha((a^i, \bar{y}^k) - c_i)]^+ \leqslant [x_i^k - \alpha\gamma]^+, \quad i \bar{\in} I^*,
$$

$$
y_j^{k+1} = [y_j^k - \alpha((\bar{a}^j, \bar{x}^k) - b_j)]^+ \leqslant [y_j^k - \alpha\gamma]^+, \quad j \bar{\in} J^*.
$$

From this it follows that after a finite number of steps it turns out that

$$
\begin{aligned}
&x_i^k = 0, \quad i \bar{\in} I^*,\\
&y_j^k = 0, \quad j \bar{\in} J^*, \quad \forall k > K_1 > K.
\end{aligned} \tag{27}
$$

Thus, for $k > K_1$ the method of (23) takes the form

$$
\begin{aligned}
&\bar{x}_i^k = x_i^k + \alpha((a^i, y^k) - c_i), \quad i \in I^*,\\
&\bar{x}_i^k = 0, \quad i \bar{\in} I^*,\\
&\bar{y}_j^k = y_j^k - \alpha((\bar{a}^j, x^k) - b_j), \quad j \in J^*,\\
&\bar{y}_j^k = 0, \quad j \bar{\in} J^*,
\end{aligned} \tag{28}
$$

$$
x_i^{k+1} = x_i^k + \alpha((a^i, \bar{y}^k) - c_i), \quad i \in I^*,
$$

$$
x_i^{k+1} = 0, \quad i \bar{\in} I^*,
$$

$$
y_j^{k+1} = y_j^k - \alpha((\bar{a}^j, \bar{x}^k) - b_j), \quad j \in J^*,
$$

$$
y_j^{k+1} = 0, \quad j \bar{\in} J^*.
$$

If we denote by v^k, w^k the nonzero components corresponding to x^k and y^k, and by B the matrix derived from A by deleting the rows with subscripts $i\overline{\in}I^*$ and columns with subscripts $j\overline{\in}J^*$, (28) may be written as

$$\begin{aligned} \bar{v}^k &= v^k + \alpha(Bw^k - c), \\ \bar{w}^k &= w^k - \alpha(B^T v^k - b), \\ v^{k+1} &= v^k + \alpha(B\bar{w}^k - c), \\ w^{k+1} &= w^k - \alpha(B^T \bar{v}^k - b). \end{aligned} \tag{29}$$

Through the assumption of uniqueness of the saddle point, the matrix B is square, nonsingular, and of dimension $l \leqslant \min\{n, m\}$, and, evidently,

$$\|Bw\| \geqslant \frac{1}{\|B^{-1}\|}\|w\|, \quad \forall w \in R^l. \tag{30}$$

Using the equalities

$$B^T v^* = b, \quad Bw^* = c, \tag{31}$$

as well as (30), we get the following estimates from (29):

$$\begin{aligned} \|\bar{v}^k - v^k\|^2 = \alpha^2\|B(w^k - w^*)\|^2 &\geqslant \frac{\alpha^2}{\|B^{-1}\|^2}\|w^k - w^*\|^2, \\ \|\bar{w}^k - w^k\|^2 &\geqslant \frac{\alpha^2}{\|B^{-1}\|^2}\|v^k - v^*\|^2. \end{aligned} \tag{32}$$

For our problem the basic estimate (14) derived in the proof of theorem 1 has the form

$$\|v^{k+1} - v^*\|^2 + \|w^{k+1} - w^*\|^2 \leqslant \|v^k - v^*\|^2 + \|w^k - w^*\|^2 - (1 - \alpha^2\|A\|^2)(\|\bar{v}^k - v^k\|^2 + \|\bar{w}^k - w^k\|^2).$$

Since for the sequence under investigation the estimates (32) are appropriate, we get from this inequality

$$\|v^{k+1} - v^*\|^2 + \|w^{k+1} - w^*\|^2 \leqslant \left(1 - \frac{\alpha^2}{\|B^{-1}\|^2}(1 - \alpha^2\|A\|^2)\right)(\|v^k - v^*\|^2 + \|w^k - w^*\|^2).$$

Since $1-\alpha^2\|A\|^2>0$ by assumption

$$q=1-\frac{\alpha^2}{\|B^{-1}\|^2}(1-\alpha^2\|A\|^2)<1. \tag{33}$$

The theorem is proved.

We will consider q and (33) as a function of α. It is easy to see that $\min_\alpha q(\alpha)=1-\frac{1}{4\|B^{-1}\|^2\|A\|^2}$ and is achieved for $\alpha=\frac{1}{\sqrt{2}\|A\|}$. Thus, if we choose $\alpha=\frac{1}{\sqrt{2}\|A\|}$, we can guarantee that there will be convergence at the rate of a geometric progression with ratio $q=\left(1-\frac{1}{4\|B^{-1}\|^2\|A\|^2}\right)$.

Observation 1. According to the proof of theorem 3, after a finite number of iterations of the extragradient method (23), we find an optimal basis (27) after which we can solve the pair of systems of linear equations (31) with quadratic nonsingular matrices B and B^T. This enables us to combine the extragradient method with any finite method for solving systems of linear equations to get a finite method for solving the initial problem (21). The method of accompanied gradients of [9] — which is close to the extragradient method in the operations it uses, and in addition makes it possible to use the properties of the initial matrix of the system (in this case B) — is entirely appropriate for this purpose. We will apply the method of accompanied gradients to the system with a nonsingular symmetric matrix [9, chap. 6] and therefore to the pair of systems of interest to us as well, represented in the form

$$\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}\begin{pmatrix} v \\ w \end{pmatrix}=\begin{pmatrix} c \\ b \end{pmatrix},$$

and it has the form

$$v^{k+1}=v^k+\gamma^k p^{k+1},$$
$$w^{k+1}=w^k+\gamma^k q^{k+1},$$
$$\beta^k=\frac{\|c-Bw^k\|^2+\|b-B^Tv^k\|^2}{\|c-Bw^{k-1}\|^2+\|b-B^Tv^{k-1}\|^2}, \tag{34}$$

$$p^{k+1}=(c-Bw^k)+\beta^k p^k,$$
$$q^{k+1}=(b-B^T v^k)+\beta^k q^k,$$
$$\gamma^k=\frac{\|c-Bw^k\|^2+\|b-B^T v^k\|^2}{(p^{k+1}, Bq^{k+1})+(q^{k+1}, B^T p^{k+1})}.$$

Method (34) converges after a number of steps, equal to the doubled order of matrix B (except for degenerate cases [9]).

Observation 2. Suppose that we are considering a block linear programming problem of the form

$$\min \sum_{i=1}^{N} (c_i, z_i), \quad c_i, z_i \in R^{n_i},$$
$$A_i z_i \leqslant b_i, \quad b_i \in R^{m_i}, \tag{35}$$
$$\sum_{i=1}^{N} B_i z_i = d, \quad d \in R^l, \quad i=1, 2, \ldots, N,$$

where A_i is a matrix of size $m_i \times n_i$; B_i is a matrix of size $l \times n_i$; and the set $Q_i=\{z_i : A_i z_i \leqslant b_i\}$, $i=1, 2, \ldots, N$. Then the solution of problem (35) reduces to the search for saddle points of a Lagrange function

$$\min_{z\in Q} \max_{y\in R^l} L(z, y), \quad L(z, y)=\sum_{i=1}^{N} (c_i, z_i)+\left(y, \sum_{i=1}^{N} B_i z_i - d\right),$$
$$z=[z_1, z_2, \ldots, z_N], \quad Q=Q_1\times Q_2\times \ldots \times Q_N. \tag{36}$$

The extragradient method for problem (36) has the form

$$\bar{z}_i^k=P_{Q_i}[z_i^k-\alpha(c_i+B_i^T y^k)],$$
$$\bar{y}^k = y^k + \alpha\left(\sum_{i=1}^{N} B_i z_i^k - d\right),$$
$$z_i^{k+1} = P_{Q_i}[z_i^k-\alpha(c_i+B_i^T \bar{y}^k)],$$
$$y^{k+1} = y^k + \alpha\left(\sum_{i=1}^{N} B_i \bar{z}_i^k - d\right), \quad i=1, \ldots, N.$$

Thus, the extragradient method is a simple convergent (for an appropriate choice of α) method for solving block linear programming problems.

Note

1) For a Lagrange function such approaches are described in [4-6].

References

1. K. J. Arrow, L. Hurwicz, and H. Uzawa, Studies in Linear and Nonlinear Programming, Moscow, Foreign Languages Press, 1962.

2. B. T. Poliak, "Gradient Methods for Minimizing Functionals," Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki, 1963, Vol. 3, No. 4.

3. V. F. Demianov and A. B. Pevnyi, "Numerical Methods for Finding Saddle Points," Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki, 1972, Vol. 12, No. 5.

4. E. G. Gol'shtein, "A Generalized Gradient Method for Finding Saddle Points," Ekonomika i matematicheskie metody, 1972, Vol. 8, No. 4.

5. B. T. Poliak and N. V. Tretiakov, "An Iteration Method of Linear Programming and Its Economic Interpretation," Ekonomika i matematicheskie metody, 1972, Vol. 8, No. 5.

6. E. G. Gol'shtein and N. V. Tretiakov, "Modified Lagrange Functions," Ekonomika i matematicheskie metody, 1974, Vol. 10, No. 3.

7. A. B. Bakushinskii and B. T. Poliak, "The Solution of Variational Inequalities," Doklady AN SSSR, 1974, Vol. 219, No. 5.

8. M. A. Krasnosel'skii, "Two Observations on the Method of Sequential Approximations," Uspekhi matematicheskoi nauki, 1955, Vol. 10, No. 1.

9. D. K. Fadeev and V. N. Fadeeva, Computing Methods of Linear Algebra, Moscow, Fizmatgiz, 1960.

Received April 14, 1975