Pergamon

Comput. Maths. Math. Phys., Vol.33, No.12, pp. 1555–1568, 1993

0965-5425(94)E0014-A

AN INTERIOR LINEARIZATION METHOD[1]

A.S. ANTIPIN
Moscow
(Revised version 25 December 2002)

An interior linearization method is described. Its convergence to the solution of the convex programming problem is proved. Bounds are obtained for the rate of convergence. A relation with internal modified Lagrange functions methods is established.

1. STATEMENT OF THE PROBLEM

Consider the convex programming problem

$$x^* \in \text{Argmin}\,\{f(x) : g(x) \le 0,\ x \in Q\}, \tag{1.1}$$

where $f(x)$ and each component of the vector function $g(x)$ is a convex scalar function, and $Q \subseteq \mathbb{R}^n$ is a convex closed set in $\mathbb{R}^n$.

There are several approaches to the solution of this problem. One of these, based on the gradient method, has been used as the basis of very many modifications intended for the solution of problem (1.1). We will be discussing modifications associated with the idea of linearization, arrived at as follows: if x^* is a minimum point of problem (1.1), then we have the necessary and sufficient conditions

$$x^* = \pi_\Omega[x^* - \alpha \nabla f(x^*)], \tag{1.2}$$

where $\pi_\Omega(\cdot)$ is the projection of a certain vector on the admissible set $\Omega = \{x : g(x) \le 0,\ x \in Q\}$, $\alpha > 0$ is a parameter of the step length type, and $\nabla f(x)$ is the gradient of the function $f(x)$ at the point x. The geometric meaning of condition (1.2) is simple: a step along the antigradient from the point x^* after the projection operation again reaches the point x^*, that is, x^* is a fixed point, or a point of equilibrium. The discrepancy $\pi_\Omega[x - \alpha \nabla f(x)] - x$ can be regarded as a transformation of space $\mathbb{R}^n$ into $\mathbb{R}^n$. This transformation defines a vector field. The problem is to find a trajectory whose tangent coincides with the field vector at that point. Formally, the problem can be described by the system of differential equations

$$dx/dt + x = \pi_\Omega[x - \alpha \nabla f(x)], \quad x(t_0) = x^0. \tag{1.3}$$

It follows from the general theorems that the continuous right-hand side of system (1.3) guarantees the existence of a solution in a finite interval. But if the Lipschitz condition holds for the right-hand side (as it does in our case), existence and uniqueness of a trajectory in an infinite interval, that is, for all $t \ge t_0$, is guaranteed.

[1] *Zh. Vychisl. Mat. Mat. Fiz.*, Vol.33, No.12, pp. 1776–1791, 1993.

The behaviour of the system as $t \to \infty$ has been studied in detail in [1], where it was established that, for convex $f(x)$ and with certain constraints on the values of the parameter α, the set of points of equilibrium of system (1.3) is asymptotically stable. Special cases of the process (1.3), when the set Ω coincides with the space $\mathbb{R}^n$, that is, $\pi_\Omega(\cdot) = I$, the identity operator, have been investigated in many publications [2]–[5].

Computational experience has shown that application of the projection operation is justified if Ω is a simple set, a positive orthant, parallelepiped or sphere. But if the admissible set has a complicated structure of the type $\Omega = \{x : g(x) \leq 0,\ \ x \in Q\}$, projection becomes too complex an operation, in which case it is better to approximate the set Ω by a family of simpler sets. It seems natural to take approximating families of the admissible set as the family of polygons for an exterior approximation, and the family of spheres for an interior approximation.

2. CONTINUOUS METHOD OF INTERIOR LINEARIZATION

Iterative linearization methods have long been used in non-linear programming and have been investigated in detail. The numerous versions differ mainly in the choice of the step length [6]–[8].

Linearization methods generate trajectories which converge to the solution of the convex programming problem from outside the admissible domain. In many cases, practical necessity demands the construction of a trajectory in the admissible domain which converges to the solution of the problem within that domain. In that case, the approximation obtained will always be admissible and will be meaningful in some situation or other. For instance, admissible components might be interpreted as the presence of necessary ingredients in a certain mixture.

To construct interior trajectories, we use the idea of the approximation of an admissible set from inside by means of a family of intersecting spheres. We will explain the idea of the approach [9, 10] using the example of problem (1.1) with inequality-type constraints. Assuming that the initial condition x^0 is a strictly admissible point, that is, $g(x^0) < 0$, we determine the trajectory $x(t)$ by means of the process

$$\begin{gathered} dx/dt + x = \pi_{S(t)}(x - \alpha \nabla f(x)), \qquad x(t_0) = x^0, \\ S(t) = \{z : 0.5|z - x|^2 e + \alpha[\nabla g(x)(z - x) + g(x)] \leq 0,\ \ z \in Q\}. \end{gathered} \tag{2.1}$$

Here $e = (1, \ldots, 1)$ is a vector of dimensions m, all the components of which are one, the parameter $\alpha > 0$, $\nabla g(x)$ is a matrix, each row of which is the vector gradient of the corresponding functional constraint. The set $S(t)$ is non-empty (it contains the point x) and is the intersection of m spheres and the simple set Q. The set $S(t)$ is the intersection of spheres $\{z : |z - [x - \alpha \nabla g_i(x)]|^2 \leq \alpha[\alpha|\nabla g_i(x)|^2 - 2g_i(x)],\ i = 1, 2, \ldots, m\}$ with centres $x - \alpha \nabla g_i(x)$ and squares of radius $\alpha[\alpha|\nabla g_i(x)|^2 - 2g_i(x)]$. A family of intersecting spheres is constructed at each point of the trajectory $x(t)$ and the gradient step $x - \alpha \nabla f(x)$ is projected onto it. As $t \to \infty$, the family of intersecting spheres approaches closer and closer to the admissible set of problem (1.1) at the minimum point.

It will be shown below that if, at time t, the point $x(t)$ is strictly admissible, that is, $g_i(x(t)) < 0$ for all $i = 1, 2, \ldots, m$, then when the point $x(t)$ is shifted by the vector $\dot{x}$, the new point $x + \dot{x}$, where $\dot{x} = dx/dt$, will also be strictly admissible.

We will now derive the dual of the process (2.1). We first recall that the projection operator $\pi_S(b)$ of the vector b onto the set S is realized by the solution of the quadratic problem

$$\pi_S(b) = \operatorname{argmin}\{0.5|z - b|^2 : z \in S\} \tag{2.2}$$

or the variational inequality

$$\langle \pi_S(b) - b, z - \pi_S(b)\rangle \geq 0 \qquad \forall z \in S. \tag{2.3}$$

We represent process (2.1) in the form (2.2):

$$x+\dot{x} \in \mathrm{Argmin}\,\{0.5|z-[x-\alpha\nabla f(x)]|^2 : 0.5|z-x|^2 e+\alpha[\nabla g(x)(z-x)+g(x)] \leq 0,\ \ z \in Q\}. \tag{2.4}$$

The projection of the vector $x - \alpha\nabla f(x)$ onto the set $S(t)$ at the current time t reduces to the minimization of a strongly convex quadratic function for strongly convex quadratic constraints. The Lagrange function for this auxiliary problem has the form

$$L(z, y) = 0.5|z - [x - \alpha\nabla f(x)]|^2 + \langle y, 0.5|z - x|^2 e + \alpha[\nabla g(x)(z - x) + g(x)]\rangle \ \ \forall\ z \in Q \text{ and } y \geq 0.$$

The saddle point $x + \dot{x}, p$ of this function satisfies the system of inequalities

$$L(x + \dot{x}, y) \leq L(x + \dot{x}, p) \leq L(z, p), \tag{2.5}$$

which hold for all $z \in Q$ and $y \geq 0$. Here $x + \dot{x}$ is the minimum of the quadratic programming problem, and p is its dual solution (Lagrange multipliers).

We will make the simplifying assumption that $Q = \mathbb{R}^n$. In that case, the minimum point $x + \dot{x}$ of the Lagrange function, according to the right-hand inequality of (2.5), is the solution of the equation obtained by equating the gradient of the function to zero:

$$(1 + \langle e, y\rangle)(z - x) + \alpha[\nabla f(x) + \nabla g^\top(x)y] = 0, \tag{2.6}$$

or

$$z - x + \alpha\frac{\nabla f(x) + \nabla g^\top(x)y}{1 + \langle e, y\rangle} = 0. \tag{2.7}$$

We eliminate the variable z from $L(z, y)$ with the help of the last equation, thereby obtaining a function with respect to the dual variable y. Technically this can be done as follows. We find the scalar product of both sides of Eq. (2.6) by the vector $z - x$; then

$$(1 + \langle e, y\rangle)|z - x|^2 = -\alpha\langle\nabla f(x), z - x\rangle - \alpha\langle y, \nabla g(x)(z - x)\rangle. \tag{2.8}$$

We use this equation to transform the Lagrange function:

$$\begin{aligned} L(z,y) &= \frac{1}{2}|z-x|^2 + \alpha\langle\nabla f(x), z-x\rangle + \frac{\alpha^2}{2}|\nabla f(x)|^2 + \\ &+ \langle y, \frac{1}{2}|z-x|^2 e + \alpha[\nabla g(x)(z-x) + g(x)]\rangle = \\ &= \frac{1}{2}|z-x|^2 + \frac{1}{2}|z-x|^2\langle y, e\rangle + \alpha\langle y, g(x)\rangle - (1+\langle e,y\rangle)|z-x|^2 + \frac{\alpha^2}{2}|\nabla f(x)|^2 = \\ &= -\frac{1}{2}(1+\langle e,y\rangle)|z-x|^2 + \alpha\langle y, g(x)\rangle + \frac{\alpha^2}{2}|\nabla f(x)|^2. \end{aligned}$$

Using Eq. (2.8), we eliminate the variable $|z - x|^2$ from the last equation, and then

$$\Psi(y) = -\frac{\alpha}{2}\frac{|\nabla f(x) + \nabla g^\top(x)y|^2}{1 + \langle e, y\rangle} + \langle y, g(x)\rangle.$$

This function is fractional quadratic-linear with respect to the variable y. Maximizing it with respect to y in the positive orthant, we can find the dual solution p (Lagrange multipliers) of the direct problem. It follows from (2.7) that, in terms of that function, the process dual to (2.1) for $Q = \mathbb{R}^n$ takes the form

$$p = \operatorname{Argmax}\left\{-\frac{\alpha}{2}\frac{|\nabla f(x) + \nabla g^\top(x)y|^2}{1 + \langle e, y\rangle} + \langle y, g(x)\rangle : y \geq 0\right\}, \tag{2.9}$$

$$\frac{dx}{dt} = -\alpha\frac{\nabla f(x) + \nabla g^\top(x)p}{1 + \langle e, p\rangle}, \qquad x(t_0) = x^0. \tag{2.10}$$

Unlike the direct process (2.1), which describes the direct trajectory $x(t)$ in explicit form, the dual process (2.9), (2.10) gives an explicit description of the direct trajectory $x(t)$, as well as the dual trajectory $p(t)$, the trajectory of Lagrange multipliers.

The fractional quadratic-linear objective function $\Psi(y)$ can be transformed to a quadratic function by substitution. With this aim, we will interpret the function

$$1 + \langle e, y\rangle = 1 + \sum_{i=1}^{m} y_i = v^2$$

as the square of the norm of the vector y, and $(1/v)y$ as the normalized vector. In that notation, the fractional function $\Psi(y)$ takes the form

$$\Psi(y) = -\frac{\alpha}{2}\left|\nabla f(x)\frac{1}{v} + \nabla g^\top(x)\frac{1}{v}y\right|^2 + \langle y, g(x)\rangle.$$

Introducing the new variables $1/v = u$ and $(1/v)y = w$, we can represent the function $\Psi(y)$ as:

$$\Psi(v, u, w) = -\frac{\alpha}{2}|\nabla f(x)u + \nabla g^\top(x)w|^2 + v\langle w, g(x)\rangle.$$

Thus, the problem of maximizing the fractional quadratic-linear objective function $\Psi(y)$ in the positive orthant $y \geq 0$ reduces to the simplest quadratic programming problem of the form

$$v^*, u^*, w^* \in \operatorname{Argmax}\{-\frac{\alpha}{2}|\nabla f(x)u + \nabla g^\top(x)w|^2 + v\langle w, g(x)\rangle : u + \langle e, w\rangle = w, v \geq 0, u \geq 0, w \geq 0\}.$$

My attention was drawn to this substitution by A.I. Golikov.

If the dimensions of the vector x in the original convex programming problem (1.1) is low and there are sufficiently many constraints, it is preferable to use the direct process (2.1), while conversely, if x is large and the number of constraints is small, the dual process (2.9), (2.10) is more efficient.

Processes (2.1) and (2.9), (2.10) can be examined independently of one another, without regard to the intrinsic relation between them. In particular, their convergence will be proved separately. For this purpose, we derive some relations which hold for both.

We consider the auxiliary optimization problem in (2.9), (2.10) and represent it in the equivalent form of a variational inequality:

$$\left\langle \frac{-\alpha\nabla g(x)[\nabla f(x) + \nabla g^\top(x)p]}{1 + \langle e, p\rangle} + \frac{\alpha|\nabla f(x) + \nabla g^\top(x)y|^2}{2(1 + \langle e, y\rangle)^2}e + g(x), p - y\right\rangle \geq 0 \ \ \forall\ y \geq 0. \tag{2.11}$$

We convert (2.11) with the help of Eq. (2.10) to the form

$$\langle \nabla g(x)\dot{x} + \frac{1}{2\alpha}|\dot{x}|^2 e + g(x), p - y\rangle \geq 0 \ \ \forall\ y \geq 0. \tag{2.12}$$

Hence, inequality (2.11) is equivalent to the relations

$$\alpha\nabla g(x)\dot{x} + 0.5|\dot{x}|^2 e + \alpha g(x) \leq 0, \tag{2.13}$$

$$\langle \alpha\nabla g(x)\dot{x} + 0.5|\dot{x}|^2 e + \alpha g(x), p\rangle = 0. \tag{2.14}$$

The differential equation of (2.9), (2.10) can also be represented in the form of a variational inequality:

$$\langle (1 + \langle e, p\rangle)\dot{x} + \alpha[\nabla f(x) + \nabla g^\top(x)p], z - x - \dot{x}\rangle \geq 0 \quad \forall\, z \in Q. \tag{2.15}$$

Thus, the system of inequalities (2.12) and (2.15) (or, what amounts to the same thing, (2.13), (2.14) and (2.15)) fully characterizes the process (2.9), (2.10).

We will show that the procedure (2.1) is subject to the same system of inequalities. All that is needed is to represent the right-hand and left-hand inequalities of (2.5) in variational form:

$$\langle (1 + \langle e, p\rangle)\dot{x} + \alpha[\nabla f(x) + \nabla g^\top(x)p], z - x - \dot{x}\rangle \geq 0 \quad \forall\, z \in Q, \tag{2.16}$$

$$\langle y - p, 0.5|\dot{x}|^2 e + \alpha[\nabla g(x)\dot{x} + g(x)]\rangle \leq 0 \qquad \forall\, y \geq 0. \tag{2.17}$$

It is obvious that inequalities (2.16) and (2.17) are identical to (2.12) and (2.15).

We will show that both processes, (2.1) and (2.9), (2.10) are admissible, or interior processes, by showing that if $x(t)$ is a strictly interior point at time t, then it remains strictly interior when that point is displaced by the length of the vector $\dot{x}$. For this purpose, we use the Lipschitz inequality

$$g(z) - g(x) - \nabla g(x)(z - x) \leq \frac{L}{2}|z - x|^2 \tag{2.18}$$

for all x and z of Q, where L is a vector constant with components L_i, $i = 1, 2, \ldots, m$. Each ith component is the Lipschitz constant for the ith functional constraint. From (2.13) and (2.18) we have

$$g(x) \leq -\nabla g(x)(x + \dot{x} - x) - \frac{1}{2\alpha}|\dot{x}|^2 e \leq g(x) - g(x + \dot{x}) + \frac{L}{2}|\dot{x}|^2 - \frac{1}{2\alpha}|\dot{x}|^2 e.$$

This yields

$$g(x + \dot{x}) \leq 0.5\left(L - \frac{1}{\alpha}e\right)|\dot{x}|^2 < 0. \tag{2.19}$$

Thus, if $\alpha < 1/L_i$, $i = 1, 2, \ldots, m$, then $g(x + \dot{x}) < 0$, that is, after a small displacement of the point x along the trajectory $x(t)$ in the direction $\dot{x}$, the point $x + \dot{x}$ remains strictly inside for any $t \geq t_0$.

As the process (2.1) evolves over time, generally speaking, the family of spheres $S(t) = \{z : 0.5|z-x|^2 e + \alpha[\nabla g(x)(z-x) + g(x)] \leq 0,\ z \in Q\}$ approaches the boundary of the admissible set (if the optimum lies on the boundary) and approximates this set from inside. It is important to note that the spheres $\{z : |z - [x - \alpha\nabla g_i(x)]|^2 \leq \alpha[\alpha|\nabla g_i(x)|^2 - 2g_i(x)],\ i = 1, 2, \ldots, m\}$ do not degenerate over time and do not contract to a point, since their radii do not tend to zero. In fact, in order for the square of the radius to be equal to zero at the limit point $x' : \alpha|\nabla g_i(x')|^2 - 2g_i(x') = 0$, we must have $\nabla g_i(x') = 0$ and $g(x') = 0$. If the point x' lies on the boundary of the ith constraint, $g(x') = 0$, but then $\nabla g_i(x') \neq 0$, since Slater's condition, which the initial problem satisfies by assumption, would be violated if the gradient $\nabla g_i(x')$ were equal to zero. If x' is a strictly interior point, we have $g(x') < 0$.

Since the diameter of spheres of the family $S(t)$ does not contract to zero over time, the spheres do not degenerate, and generally speaking the Lagrange multipliers $p(t)$ corresponding to those spheres as if to constraints do not tend to infinity, and so the assumption that they are bounded can be regarded as completely justified.

We will now give a theorem on the convergence of methods (2.1) and (2.9), (2.10).

Theorem 1. *If problem* (2.1) *satisfies Slater's regularity condition,* $f(x)$, $g(x) = (g_1(x), \ldots, \ldots, g_m(x))$ *are convex differentiable functions, the gradients of all of which satisfy the Lipschitz condition with constants* L_0, $L = (L_1, \ldots, L_m)$, Q *is a convex closed set (in* (2.9), (2.10) *the set* $Q = \mathbb{R}^n$*), the parameter* α *in* (2.1) *and* (2.9), (2.10) *is chosen from the condition*

$$\alpha < \min\{2/L_0, 1/L_1, \ldots, 1/L_m\},$$

the trajectory $p(t)$ *is bounded by a constant:* $|p(t) \leq C$*, then the set of equilibrium points* X^* *of systems* (2.1) *and* (2.9), (2.10) *is asymptotically stable, that is,* $x(t) \to x^* \in X^*$ *as* $t \to \infty$ *for all* x^0 *and the trajectory is strictly interior for all* $t \geq t_0$.

The proof of Theorem 1 is given in Appendix 1. If the objective function of problem (2.1) is strongly convex, bounds can be found for the rate of convergence of processes (2.1) and (2.9), (2.10).

Theorem 2. *If, in addition to the conditions of Theorem* 1*, the objective function is strongly convex, then the only point of equilibrium of systems* (2.1) *and* (2.9), (2.10) *is exponentially stable, that is,* $|x(t) - x^*|^2 \leq C \exp[-2a(\alpha)t]$*, where*

$$a(\alpha) = \begin{cases} \dfrac{\alpha\ell(2 - \alpha\ell/2)}{1 + \langle e, C\rangle}, & \text{if } \alpha < \dfrac{4}{L_0 + \ell}, \\ \dfrac{\alpha L_0(2 - \alpha L_0/2)}{1 + \langle e, C\rangle}, & \text{if } \alpha > \dfrac{4}{L_0 + \ell}, \end{cases}$$

and C *is the vector constant bounding the trajectory:* $0 \leq p(t) \leq C$.

The proof of Theorem 2 is given in Appendix 1.

3. AN ITERATIVE METHOD OF INTERIOR LINEARIZATION

By approximating the derivative $dx(t)/dt$ by the finite difference $(x^{n+1} - x^n)/\Delta t$, we can obtain a discrete trajectory which is a good approximation to the continuous trajectory for sufficiently small Δt. But since it is not the trajectory itself, but its limit point as $t \to \infty$ that is the required solution, it is reasonable to use large time steps to reach the goal ($\Delta t = 1$). However, the convergence of the iteration must be examined separately in that case.

Thus, consider the iterative analogues of the direct process (2.1):

$$x^{n+1} = \pi_{S(n)}[x^n - \alpha\nabla f(x^n)], \tag{3.1}$$

$$S(n) = \{z : 0.5|z - x^n|^2 e + \alpha[\nabla g(x^n)(z - x^n) + g(x^n)] \leq 0 \;\; x \in 0\}, \tag{3.2}$$

where $e = (1, \ldots, 1)$ is an m-dimensional vector, all the elements of which are equal to one, the parameter $\alpha > 0$, and of the dual process (2.9), (2.10):

$$p^n = \operatorname{Argmax}\left\{-\frac{\alpha}{2}\frac{|\nabla f(x^n) + \nabla g^\top(x^n)y|^2}{1 + \langle e, y\rangle} + \langle y, g(x^n)\rangle : y \geq 0\right\}, \tag{3.3}$$

$$x^{n+1} = x^n - \alpha\frac{\nabla f(x^n) + \nabla g^\top(x^n)p^n}{1 + \langle e, p^n\rangle}. \tag{3.4}$$

In the process (3.1), (3.2) the gradient step is projected onto a set which is the intersection of spheres. This is a quadratic programming problem with quadratic constraints. The justification for posing this auxiliary problem was given in the last section, where it was also shown that it reduces to a quadratic programming problem.

The process (3.1), (3.2) is efficient if the dimensions of x are low and there are many constraints. If the dimensions of x are sufficiently high and there is a relatively small number of constraints, that is, the dimensions of the dual vector y are small enough, it is sensible to use the dual process (3.3), (3.4).

We will now write out the iterative analogues of relations (2.12) – (2.15) characterizing processes (3.1), (3.2) and (3.3), (3.4):

$$\langle (1 + \langle e, p^n \rangle)(x^{n+1} - x^n) + \alpha[\nabla f(x^n) + \nabla g^\top(x^n)p^n], z - x^{n+1} \rangle \geq 0 \quad \forall\, z \in Q, \quad (3.5)$$
$$\langle 0.5|x^{n+1} - x^n|^2 e + \alpha[\nabla g(x^n)(x^{n+1} - x^n) + g(x^n)], p^n - y \rangle \geq 0 \qquad \forall\, y \geq 0. \quad (3.6)$$

This inequality, in turn, is equivalent to the relations

$$0.5|x^{n+1} - x^n|^2 e + \alpha[\nabla g(x^n)(x^{n+1} - x^n) + \alpha g(x^n)] \leq 0, \tag{3.7}$$
$$\langle 0.5|x^{n+1} - x^n|^2 e + \alpha[\nabla g(x^n)(x^{n+1} - x^n) + \alpha g(x^n)], p^n \rangle = 0. \tag{3.8}$$

Inequalities (3.5) and (3.6) are equivalent to processes (3.1), (3.2) and (3.3), (3.4) . As in the case of continuous processes, this fact can be proved using the Lagrange function $L(z, y) = 0.5|z - [x^n - \alpha\nabla f(x^n)]|^2 + \langle y, 0.5|z - x^n|^2 e + \alpha[\nabla g(x^n)(z - x^n) + g(x^n)]\rangle$ for all $z \in Q$ and $y \geq 0$, associated with the system of inequalities

$$L(x^{n+1}, y) \leq L(x^{n+1}, p^n) \leq L(z, p^n),$$

valid for all $z \in Q$ and $y \geq 0$.

We will show that processes (3.1), (3.2) and (3.3), (3.4) are admissible (interior processes). From (3.7) and the Lipschitz condition (2.18) we have

$$\begin{aligned} g(x^n) &\leq \nabla g(x^n)(x^n - x^{n+1}) - \frac{1}{2\alpha}|x^{n+1} - x^n|^2 e \leq \\ &\leq g(x^n) - g(x^{n+1}) + \frac{L}{2}|x^{n+1} - x^n|^2 - \frac{1}{2\alpha}|x^{n+1} - x^n|^2 e. \end{aligned}$$

Thus

$$g(x^{n+1}) \leq \frac{1}{2}\left(L - \frac{1}{\alpha}e\right)|x^{n+1} - x^n|^2. \tag{3.9}$$

Since $\alpha < 1/L_i$, $i = 1, 2, \ldots, m$, we have $g(x^{n+1}) \leq 0$, that is, if x^n is an admissible point, the following approximation will also be admissible.

Theorem 3. *If problem* (2.1) *satisfies Slater's regularity condition, $f(x)$, $g(x) = (g_1(x), \ldots, \ldots, g_m(x))$ are convex differentiable functions, the gradients of all the functions satisfy the Lipschitz condition with constants L_0, $L = (L_1, \ldots, L_m)$, Q is a convex closed set (in* (3.3), (3.4) *the set $Q = \mathbb{R}^n$), and the parameter α in processes* (3.1), (3.2) *and* (3.3), (3.4) *can be chosen from the condition*

$$\alpha < \min\{2/L_0, 1/L_1, \ldots, 1/L_m\},$$

the trajectory $p(t)$ being bounded by the constant $|p(t)| \leq C$, then $x(t) \to x^* \in X^*$ as $t \to \infty$ for all x^0. The trajectory $x(t)$ is an interior one.

The proof of Theorem 3 is given in Appendix 2. If the objective function of problem (1.1) is strongly convex, bounds can be found for the rate of convergence of processes (3.1), (3.2) and (3.3), (3.4).

Theorem 4. *If, in addition to the conditions of Theorem 3, the objective function is strongly convex, processes* (3.1), (3.2) *and* (3.3), (3.4) *converge to a unique solution at the rate* *of a geometric progression (at a linear rate), that is,* $|x^{n+1} - x^*|^2 \le q(\alpha)|x^n - x^*|^2$, *where*

$$q(\alpha) = \begin{cases} 1 - \dfrac{\alpha\ell(2-\alpha\ell)}{1+\langle e, C\rangle}, & \textit{if } \alpha < \dfrac{2}{L_0+\ell}, \\ 1 - \dfrac{\alpha L_0(2-\alpha L_0)}{1+\langle e, C\rangle}, & \textit{if } \alpha > \dfrac{2}{L_0+\ell}, \end{cases}$$

and $p^n \le C$ *is a vector constant.*

The proof of Theorem 4 is given in Appendix 2.

4. OTHER POINTS OF VIEW

The method under discussion has been interpreted as a method of interior linearization, but there are alternative ways of looking at it. In the process (2.9), (2.10) at each time t, the function

$$\Phi(x,y) = f(x) + \langle y, g(x)\rangle - \frac{\alpha}{2}\frac{|\nabla f(x) + \nabla g^\top(x)y|^2}{1+\langle e, y\rangle} \tag{4.1}$$

is maximized in the positive orthant of the dual variables y for fixed x. The addition of the constant $f(x)$ to the expression $\Phi(x,y)$ has no effect on the optimization, since the latter is carried out with respect to the dual variables. It is now clear, however, that the optimized function has a precise structure: it consists of the Lagrange function $L(x,y) = f(x) + \langle y, g(x)\rangle$ plus (or minus, in this case) the square of the discrepancy of the linear equation (necessary conditions of the original problem (1.1)) $\nabla f(x) + \nabla g^\top(x)y$, multiplied by the reciprocal of the square of the "norm" of the dual variable $\alpha/(1+\langle e, y\rangle)$. The fact that the function $\Phi(x,y)$ contains the square of the discrepancy of a certain equation makes it akin to the Lagrange functions modified with respect to dual variables studied in [8, 11]; the fact that the expression $\Phi(x,y)$ contains the inverse of the norm of the dual variable makes the function similar to barrier penalty functions of the form $g^{-1}(y)$, which have been investigated in detail in [12]. All this justifies us in regarding the function $\Phi(x,y)$ as an interior modified Lagrange function. Accordingly, the process (2.9), (2.10) can be regarded as the method of an interior modified Lagrange function.

There is a considerable difference between interior penalty function methods and interior modified Lagrange function methods. In the former, the conditionality of the auxiliary problems deteriorates as the penalty parameter increases, while it remains constant in the latter.

In [13]–[15], in which interior modified Lagrange functions are considered, the modification is made with respect to the direct variables x, unlike the modification with respect to the dual variable y considered here.

A different standpoint is offered by the theory of control of computational methods [16, 17]. On this view, process (2.9), (2.10) can be regarded as a controlled interior gradient method

$$dx/dt = -\beta[\nabla f(x) + \nabla g^\top(x)p], \qquad x(t_0) = x^0, \tag{4.2}$$

where the dual and parametric controls are taken, respectively, by the dual variable p and parametric variable β. Control with respect to the dual variable is realized by feedback of the form

$$p = \operatorname{Argmax}\left\{-\frac{\alpha}{2}\frac{|\nabla f(x) + \nabla g^\top(x)y|^2}{1+\langle e, y\rangle} + \langle y, g(x)\rangle : y \ge 0\right\}, \tag{4.3}$$

and the parametric control by feedback of the type

$$\beta = \frac{\alpha}{1+\langle e, p\rangle}. \tag{4.4}$$

Feedbacks (4.3) and (4.4) in combination guarantee that the trajectory is admissible and the set of equilibrium points of the initial problem (1.1) is asymptotically stable.

On this interpretation, method (4.2) and its direct form (2.1), even more so, belongs to the class of methods of feasible directions [18].

Appendix 1

Theorem 1. For convenience, we will repeat relations (2.13) – (2.15), which completely characterize processes (2.1) and (2.9), (2.10):

$$\alpha\nabla g(x)\dot{x} + 0.5|\dot{x}|^2 e + \alpha g(x) \le 0, \tag{A1.1}$$
$$\langle \alpha\nabla g(x)\dot{x} + 0.5|\dot{x}|^2 e + \alpha g(x), p\rangle = 0, \tag{A1.2}$$
$$\langle (1 + \langle e, p\rangle)\dot{x} + \alpha[\nabla f(x) + \nabla g^\top(x)p], z - x - \dot{x}\rangle \ge 0 \ \ \forall\, z \in Q. \tag{A1.3}$$

In addition, we give the necessary and sufficient conditions for a solution of problem (2.1) to exist:

$$\langle \nabla f(x^*) + \nabla g^\top(x^*)p^*, z - x^*\rangle \ge 0 \qquad \forall\, z \in Q, \tag{A1.4}$$
$$\langle y - p^*, g(x^*)\rangle \le 0 \qquad \forall\, y \ge 0. \tag{A1.5}$$

We put $z = x^*$ in (A1.3) and $z = x + \dot{x}$ in (A1.4); we add the two inequalities, obtaining

$$\langle (1 + \langle e, p\rangle)\dot{x} + \alpha[\nabla f(x) - \nabla f(x^*) + \nabla g^\top(x)p - \nabla g^\top(x^*)p^*], x^* - x - \dot{x}\rangle \ge 0. \tag{A1.6}$$

We transform the separate components of (A1.6), using the convexity of $g(x)$ and Eq. (A1.2):

$$\begin{aligned} & \langle \nabla g^\top(x)p, x^* - x - \dot{x}\rangle = \langle p, \nabla g(x)(x^* - x - \dot{x})\rangle = \\ = \ & \langle p, \nabla g(x)(x^* - x)\rangle - \langle p, \nabla g(x)\dot{x}\rangle \le \langle p, g(x^*) - g(x)\rangle + \left\langle p, g(x) + \frac{1}{2\alpha}|\dot{x}|^2 e\right\rangle = \\ = \ & \langle p, g(x^*)\rangle + \frac{1}{2\alpha}\langle e, p\rangle|\dot{x}|^2 \le \frac{1}{2\alpha}\langle e, p\rangle|\dot{x}|^2. \end{aligned} \tag{A1.7}$$

Then, using the convexity of $g(x)$ and the inequality (2.19), together with (A1.5), we obtain

$$\begin{aligned} & \langle \nabla g(x^*)p^*, x^* - x - \dot{x}\rangle \ge \langle p^*, g(x^*) - g(x + \dot{x})\rangle \ge \\ \ge - \ & \langle p^*, g(x + \dot{x})\rangle \ge -\frac{1}{2}\left\langle p^*, \left(L - \frac{1}{\alpha}e\right)\right\rangle |\dot{x}|^2 > 0. \end{aligned} \tag{A1.8}$$

In this chain of inequalities, the last value is positive, and so, by the condition,

$$\alpha < \min\{1/L_1, \dots, 1/L_m\} \tag{A1.9}$$

A bound can be found for the second term of (A1.6) using the inequality (see [8])

$$\langle \nabla f(x_1) - \nabla f(x_3), x_3 - x_2\rangle \ge \frac{L_0}{4}|x_1 - x_2|^2, \tag{A1.10}$$

which is true for all x_1, x_2, x_3 of Q, where L_0 is the Lipschitz constant for $\nabla f(x)$ on Q. Then

$$\langle \nabla f(x) - \nabla f(x^*), x^* - x - \dot{x}\rangle \le \frac{L_0}{4}|\dot{x}|^2. \tag{A1.11}$$

From the resulting inequalities we obtain (A1.6) in the form

$$\langle (1 + \langle e, p\rangle)\dot{x}, x - x^* + \dot{x}\rangle - \alpha\frac{L}{4}|\dot{x}|^2 - \frac{1}{2}\langle e, p\rangle|\dot{x}|^2 \le 0.$$

Thus

$$\frac{1}{2}(1+\langle e,p\rangle)\frac{d}{dt}|x-x^*|^2+\left(1+\frac{1}{2}\langle e,p\rangle\right)|\dot{x}|^2-\alpha\frac{L_0}{4}|\dot{x}|^2\le 0, \tag{A1.12}$$

or

$$\frac{1}{2}(1+\langle e,p\rangle)\frac{d}{dt}|x-x^*|^2+\left(1+\frac{1}{2}\langle e,p\rangle-\frac{\alpha}{4}L_0\right)|\dot{x}|^2\le 0. \tag{A1.13}$$

Then from (A1.13) we have

$$\frac{d}{dt}|x-x^*|^2+\left(1+\frac{1-\alpha L_0/2}{1+\langle e,p\rangle}\right)|\dot{x}|^2\le 0. \tag{A1.14}$$

If $\alpha<1/L_0$, then $1-\alpha L_0/2>0$, and thus

$$1+\frac{1-\alpha L_0/2}{1+\langle e,p\rangle}\ge 1.$$

From this inequality it follows that (A1.13) can be put in the form

$$\frac{d}{dt}|x-x^*|^2+|\dot{x}|^2\le 0. \tag{A1.15}$$

Inequality (A1.15) holds, provided that $\alpha<2/L_0$, to which we add the condition $\alpha<$ $<\min\{1/L_1,\dots,1/L_m\}$ previously obtained. Combining the two, we have $\alpha<\min\{2/L_0,1/L_1,$ $\dots,1/L_m\}$.

We integrate inequality (A1.15) from t_0 to t:

$$|x-x^*|^2+\int_{t_0}^{t}|\dot{x}|^2d\tau\le|x^0-x^*|^2, \tag{A1.16}$$

where $x^0=x(t_0)$. It follows from (A1.16) that the trajectory is bounded: $|x(t)-x^*|^2\le|x^0-x^*|^2$, that is, the set of equilibrium points of the system is Lyapunov stable and the integral is convergent:

$$\int_{t_0}^{t}|\dot{x}|^2d\tau<\infty \quad\text{as}\quad t\to\infty.$$

Assuming $\varepsilon>0$ exists such that $|\dot{x}|\ge\varepsilon$ for all $t\ge t_0$, we obtain a contradiction with convergence of the integral. Thus there is a subsequence of times $t_i\to\infty$ such that $|\dot{x}|\to 0$. Since $x(t)$ and $p(t)$ are bounded, we once more select a subsequence of times, which we also denote by t_i, such that $x(t_i)\to x'$, $p(t_i)\to p'$ and $\dot{x}(t_i)\to 0$.

Consider relations (A1.1) – (A1.3) for all times $t_i\to\infty$ (remember that the first two relations are equivalent to the variational inequality (A1.14)). Taking the limit and writing out the limiting relations:

$$\langle\nabla f(x')+\nabla g^\top(x')p',z-x'\rangle\ge 0,\qquad \langle y-p',g(x')\rangle\le 0$$

for all $z\in Q$ and $y\ge 0$. These inequalities are obviously the same as (A1.4) and (A1.5), which means that x', p' are, respectively, the direct and dual solutions of the initial problem (1.1).

From the fact that the value $|x(t)-x^*|^2$ is monotone decreasing and any limit point is a solution of the problem, it follows that the limit point is unique, that is, the trajectory converges monotonically to a certain solution of the initial problem. This proves the theorem.

Proof of Theorem 2. Allowing for the inequalities (A1.7) and (A1.8), we represent inequality (A1.6) in the form

$$\begin{aligned}\langle(1+\langle e,y\rangle)\dot{x}, x-x^*\rangle \quad &+ \quad (1+\langle e,p\rangle)|\dot{x}|^2+\alpha\langle\nabla f(x)-\nabla f(x^*), x-x^*\rangle + \\ &+ \quad \alpha\langle\nabla f(x)-\nabla f(x^*), \dot{x}\rangle-\frac{1}{2}\langle e,p\rangle|\dot{x}|\le 0,\end{aligned}$$

or

$$(1+\langle e,p\rangle)\frac{d}{dt}|x-x^*|^2+c|\dot{x}|^2+2\alpha\langle\nabla f(x)-\nabla f(x^*),\dot{x}\rangle+2\alpha\langle\nabla f(x)-\nabla f(x^*), x-x^*\rangle\le 0,$$

where $c=2+\langle e,p\rangle$.

We complete the square in the second and third terms:

$$\begin{aligned}&(1+\langle e,p\rangle)\frac{d}{dt}|x-x^*|^2+|\sqrt{c}\dot{x}+(\alpha/\sqrt{c})[\nabla f(x)-\nabla f(x^*)]|^2- \\ -\quad &(\alpha^2/c)|\nabla f(x)-\nabla f(x^*)|^2+2\alpha\langle\nabla f(x)-\nabla f(x^*), x-x^*\rangle\le 0.\end{aligned}$$

We now drop the second term, and for the third use the inequality [8]

$$|\nabla f(x_1)-\nabla f(x_2)|^2+L_0\ell|x_1-x_2|^2\le(L_0+\ell)\langle\nabla f(x_1)-\nabla f(x_2), x_1-x_2\rangle, \tag{A1.17}$$

which is true for all x_1 and x_2 of Q, where ℓ is the strong convexity constant of the objective function, and L_0 is the Lipschitz constant for the gradient $\nabla f(x)$,

$$(1+\langle e,p\rangle)\frac{d}{dt}|x-x^*|^2+\alpha^2\frac{L_0\ell}{c}|x-x^*|^2+\alpha\left(2-\alpha\frac{L_0+\ell}{c}\right)\langle\nabla f(x)-\nabla f(x^*), x-x^*\rangle\le 0. \tag{A1.18}$$

If $\alpha<2c/(L_0+\ell)$, then $2-\alpha(L_0+\ell)/c>0$; the last term of (A1.18) satisfies the inequality

$$\ell|x-x^*|^2\le\langle\nabla f(x)-\nabla f(x^*), x-x^*\rangle.$$

Otherwise, for $\alpha>2c/(L_0+\ell)$, $2-\alpha(L_0+\ell)/c<0$, and so we use the inequality

$$\langle\nabla f(x)-\nabla f(x^*), x-x^*\rangle\le L_0|x-x^*|^2.$$

Then (A1.18) takes the form

$$\frac{d}{dt}|x-x^*|^2+a_1(\alpha)|x-x^*|^2\le 0, \tag{A1.19}$$

where

$$a_1(\alpha)=\begin{cases}\dfrac{\alpha\ell(2-\alpha\ell/c)}{1+\langle e,C\rangle}, & \text{if } \alpha<\dfrac{2c}{L_0+\ell}, \\ \dfrac{\alpha L_0(2-\alpha L_0/c)}{1+\langle e,C\rangle}, & \text{if } \alpha>\dfrac{2c}{L_0+\ell},\end{cases}$$

and $c=2+\langle e,p\rangle$, and $p(t)\le C$ is the vector constant bounding the trajectory $0\le p(t)\le C$.

Since $1/c=1/(2+\langle e,p\rangle)\le 1/2$, $\alpha_1(t)\ge\alpha(t)$; then

$$a(\alpha)=\begin{cases}\dfrac{\alpha\ell(2-\alpha\ell/c)}{1+\langle e,C\rangle}, & \text{if } \alpha<\dfrac{4}{L_0+\ell}, \\ \dfrac{\alpha L_0(2-\alpha L_0/c)}{1+\langle e,C\rangle}, & \text{if } \alpha>\dfrac{4}{L_0+\ell}.\end{cases}$$

Thus, (A1.19) can be rewritten once more as:

$$\frac{d}{dt}|x - x^*|^2 + a(\alpha)|x - x^*|^2 \leq 0. \tag{A1.20}$$

Integrating (A1.20), we have

$$|x(t) - x^*|^2 \leq C \exp[-2a(\alpha)t],$$

where $C = |x^0 - x^*|^2$. A necessary condition for the trajectory to be exponentially convergent is that $a(\alpha) > 0$, that is, $\alpha < \min\{4/\ell, 4/L_0\} = 4/L_0$. The optimum value of α is equal to $\alpha_{opt} = 4/(L_0 + \ell)$, when $a(\alpha_{opt}) = 8L_0\ell/(L_0 + \ell)^2$. This proves the theorem.

Appendix 2

Proof of Theorem 3. We put $z = x^*$ in (3.3) and $z = x^{n+1}$ in (3.4); then

$$\langle (1+\langle e, p^n\rangle)(x^{n+1}-x^n)+\alpha[\nabla f(x^n)-\nabla f(x^*)+\nabla g^\top(x^n)p^n-\nabla g^\top(x^*)p^*], x^*-x^{n+1}\rangle \geq 0. \tag{A2.1}$$

We transform the terms of (A2.1) separately. We use the convexity of $g(x)$ and Eq. (3.4). Then

$$\begin{aligned} & \langle \nabla g^\top(x^n)p^n, x^* - x^{n+1}\rangle = \langle p^n, \nabla g(x^n)(x^* - x^{n+1})\rangle = \\ = \; & \langle p^n, \nabla g(x^n)(x^* - x^n)\rangle - \langle p^n, \nabla g(x^n)(x^{n+1} - x^n)\rangle \leq \\ \leq \; & \langle p^n, g(x^*) - g(x^n)\rangle + \langle p^n, \nabla g(x^n) + \frac{1}{2\alpha}|x^{n+1} - x^n|^2 e\rangle = \\ = \; & \langle p^n, g(x^*)\rangle + \frac{1}{2\alpha}\langle e, p^n\rangle|x^{n+1} - x^n|^2 \leq \frac{1}{2\alpha}\langle e, p^n\rangle|x^{n+1} - x^n|^2. \end{aligned} \tag{A2.2}$$

Then, once again using the convexity of $g(x)$ and inequality (3.7), we have

$$\begin{aligned} & \langle \nabla g^\top(x^*)p^*, x^* - x^{n+1}\rangle \geq \langle p^*, g(x^*) - g(x^{n+1})\rangle \geq \\ \geq \; & -\langle p^*, g(x^{n+1})\rangle \geq -\frac{1}{2}\left\langle p^*, \left(L - \frac{1}{\alpha}e\right)\right\rangle |x^{n+1} - x^n|^2 > 0. \end{aligned} \tag{A2.3}$$

The last expression in (A2.3) is positive since, by the condition of the theorem, $\alpha < 1/L_i$, $i = 1, 2, \ldots, m$. The second term of (A2.1), by (A1.10), satisfies the inequality

$$\langle \nabla f(x^n) - \nabla f(x^*), x^* - x^{n+1}\rangle \leq \frac{L_0}{4}|x^{n+1} - x^n|^2,$$

where L_0 is the Lipschitz constant for $\nabla f(x)$ on Q.

Using the inequalities thus obtained, we represent (A2.1) in the form

$$\langle (1 + \langle e, p^n\rangle)(x^{n+1} - x^n), x^* - x^{n+1}\rangle - \alpha\frac{L_0}{4}|x^{n+1} - x^n|^2 - \frac{1}{2}\langle e, p^n\rangle|x^{n+1} - x^n|^2 \leq 0.$$

A bound for the first term can be found with the help of the identity

$$|x_1 - x_2|^2 = |x_1 - x_3|^2 + 2\langle x_1 - x_3, x_3 - x_2\rangle + |x_3 - x_2|^2, \tag{A2.4}$$

then

$$\begin{aligned} & \frac{1}{2}(1 + \langle e, p^n\rangle)|x^{n+1} - x^*|^2 + \frac{1}{2}(1 + \langle e, p^n\rangle)|x^{n+1} - x^n|^2 - \\ - \; & \frac{1}{2}\langle e, p^n\rangle|x^{n+1} - x^n|^2 - \alpha\frac{L_0}{4}|x^{n+1} - x^n|^2 \leq \frac{1}{2}(1 + \langle e, p^n\rangle)|x^n - x^*|^2. \end{aligned}$$

or

$$(1+\langle e,p^n\rangle)|x^{n+1}-x^*|^2+(1/\alpha L_0/2)|x^{n+1}-x^n|^2\le(1+\langle e,p^n\rangle)|x^n-x^*|^2.$$

Thus,

$$|x^{n+1}-x^*|^2+\frac{1}{2}\frac{2-\alpha L_0}{1+\langle e,p^n\rangle}|x^{n+1}-x^n|^2\le|x^n-x^*|^2.$$

Since $|p^n|\le C$ and the fixed parameter $\alpha<2/L_0$, we have

$$\frac{2-\alpha L_0}{1+\langle e,p^n\rangle}\ge\frac{2-\alpha L_0}{1+\langle e,C\rangle}=c.$$

Finally,

$$|x^{n+1}-x^*|^2+c|x^{n+1}-x^n|^2\le|x^n-x^*|^2. \tag{A2.5}$$

We sum (A2.5) from $n=0$ to $n=N$, and then

$$|x^{N+1}-x^*|^2+c\sum_{n=0}^{n=N}|x^{n+1}-x^n|^2\le|x^0-x^*|^2.$$

It follows from (A2.5) that the series $\sum\limits_{n=0}^{\infty}|x^{n+1}-x^n|^2<\infty$ and the value $|x^{n+1}-x^n|^2$ tends to zero as $n\to\infty$.

Let $x^n\to x'$, $p^n\to p'$; then the limit inequalities for (3.3) and (3.4) will have the form

$$\langle\nabla f(x')+\nabla g^\top(x')p',z-x'\rangle\ge 0,\qquad\langle y-p',g(x')\rangle\le 0$$

for all $z\in Q$ and $y\ge 0$. It is obvious that this system of inequalities is identical with (A1.4) and (A1.5), and so $x'=x^*\in Q$, $p'=p^*\ge 0$. Thus, on the one hand, all the limit points of the trajectory are solutions of problem (1.1), and on the other, by the monotonicity of $|x^{n+1}-x^*|^2\le|x^n-x^*|^2$, there can only be one limit point. It follows that the trajectory converges to a solution of the initial problem. This proves the theorem.

Proof of Theorem 4. Using inequalities (A2.2) and (A2.3), we write (A2.1) in the form

$$\begin{aligned}&\langle(1+\langle e,p^n\rangle)(x^{n+1}-x^n),x^{n+1}-x^*\rangle+\alpha\langle\nabla f(x^n)-\nabla f(x^*),x^n-x^*\rangle+\\+\ &\alpha\langle\nabla f(x^n)-\nabla f(x^*),x^{n+1}-x^n\rangle-0.5\langle e,p^n\rangle|x^{n+1}-x^n|^2\le 0.\end{aligned} \tag{A2.6}$$

Using identity (A2.1), we expand the first term of (A2.6) in the sum of squares

$$\begin{aligned}&(1+\langle e,p^n\rangle)|x^{n+1}-x^*|^2+|x^{n+1}-x^n|^2+2\alpha\langle\nabla f(x^n)-\nabla f(x^*),x^{n+1}-x^n\rangle+\\+\ &2\alpha\langle\nabla f(x^n)-\nabla f(x^*),x^n-x^*\rangle\le(1+\langle e,p^n\rangle)|x^n-x^*|^2,\end{aligned}$$

completing the square in the second and third terms:

$$\begin{aligned}&(1+\langle e,p^n\rangle)|x^{n+1}-x^*|^2+|x^{n+1}-x^n+\alpha[\nabla f(x^n)-\nabla f(x^*)]|^2-\\-\ &\alpha^2|\nabla f(x^n)-\nabla f(x^*)|^2+2\alpha\langle\nabla f(x^n)-\nabla f(x^*),x^n-x^*\rangle\le(1+\langle e,p^n\rangle)|x^n-x^*|^2.\end{aligned}$$

We now omit the second term, and use inequality (A1.17) to obtain a bound for the third; then

$$\begin{aligned}&(1+\langle e,p^n\rangle)|x^{n+1}-x^*|^2+\alpha^2L_0\ell|x^n-x^*|^2+\\+\ &\alpha[2-\alpha(L_0+\ell)]\langle\nabla f(x^n)-\nabla f(x^*),x^n-x^*\rangle\le(1+\langle e,p^n\rangle)|x^n-x^*|^2.\end{aligned} \tag{A2.7}$$

If $\alpha<2/(L_0+\ell)$, then $2-\alpha(L_0+\ell)>0$; for the last term of (A2.7), we use the inequality

$$\ell|x^n-x^*|^2\le\langle\nabla f(x^n)-\nabla f(x^*),x^n-x^*\rangle.$$

Otherwise, for $\alpha > 2/(L_0 + \ell)$, we have $2 - \alpha(L_0 + \ell) < 0$.

We apply the inequality

$$\langle \nabla f(x^n) - \nabla f(x^*), x^n - x^* \rangle \le L_0 |x^n - x^*|^2,$$

then (A2.7) takes the form

$$|x^{n+1} - x^*|^2 \le q(\alpha)|x^n - x^*|^2, \tag{A2.8}$$

where

$$q(\alpha) = \begin{cases} 1 - \dfrac{\alpha\ell(2 - \alpha\ell)}{1 + \langle e, C\rangle}, & \text{if } \alpha < \dfrac{2}{L_0 + \ell}, \\ 1 - \dfrac{\alpha L_0(2 - \alpha L_0)}{1 + \langle e, C\rangle}, & \text{if } \alpha > \dfrac{2}{L_0 + \ell}, \end{cases}$$

and $p^n \le C$ is a vector constant.

Summing (A2.8), we have

$$|x^{n+1} - x^*|^2 \le C q(\alpha)^{n+1},$$

where $C = |x^0 - x^*|^2$.

The optimum value of the parameter is $\alpha_{opt} = 2/(L_0 + \ell)$, the best estimate for the common ratio of the progression being equal to

$$q(\alpha_{opt}) = 1 - \frac{1}{1 + \langle e, C\rangle} \frac{4L_0\ell}{(L_0 + \ell)^2}.$$

This proves the theorem.

I wish to thank A. I. Golikov for helpful discussion of the results.

REFERENCES

1. ANTIPIN A.S., Continuous and iterative processes with projection and projection-type operators. In *Problems of Cybernetics. Computational Problems of the Analysis of Large Systems*, 5–43. Akad. Nauk SSSR, Research Council on Computational Problems of "Cybernetics", Moscow, 1989.

2. YEVTUSHENKO Yu.G. and ZHADAN V.G., The use of the Lyapunov's functions method to investigate the convergence of numerical methods. *Zh. Vychisl. Mat. Mat. Fiz.* **15**, 1, 101–112, 1975.

3. VENETS V.I. and RYBASHOV M.V., The Lyapunov's functions method in the investigation of continuous mathematical programming algorithms. *Zh. Vychisl. Mat. Mat. Fiz.* **17**, 3, 622–633, 1977.

4. BROWN A.A. and BARTHOLOMEW-BIGGS M.C., Some effective methods for unconstrained optimization based on the solution of systems of ordinary differential equations. *J. Optimizat. Theory and Applic.* **62**, 2, 211–224, 1989.

5. VASIL'YEV F.P., *Numerical Methods of Solving Extremum Problems.* Nauka, Moscow, 1988.

6. PSHENICHNYI B.N. and DANILIN Yu.M., *Numerical Methods in Extremum Problems.* Nauka, Moscow, 1975.

7. FEDORENKO R.P., *The Approximate Solution of Problems of Optimal Control.* Nauka, Moscow, 1978.

8. ANTIPIN A.S., Non-linear programming methods based on direct and dual modification of the Lagrange function. Preprint, Izd. VNIISI, Moscow, 1979.

9. ANTIPIN A.S., On feasible convex programming methods. In *Mathematical Methods of Optimization and Their Application to Large Economic and Technical Systems*, 26–29. TsEMI, Moscow, 1980.

10. ANTIPIN A.S., A feasible method akin to gradient projection method for solution of convex programming. In *Methods Math. Programming*, 7–11. PWN-Polish Scien. Publs, Warsaw, 1981.

11. GOLIKOV A.I., *Modified Lagrange Functions in Non-linear Programming.* VTs Akad. Nauk S.S.S.R., Moscow, 1988.

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

13. CHARALAMBOUS C., A method to overcome the ill-conditioning problem of differentiable penalty functions. *Operat. Res.* **28**, 3, 650–667, 1980.

14. KALININ I.N. and STERLIN A.M., On a version of the modified Lagrange function. *Dokl. Akad. Nauk SSSR* **267**, 4, 787–789, 1982.

15. POLYAK R., Modified barrier functions (theory and methods). *Math. Program.* **54**, 177–222, 1992.

16. ANTIPIN A.S., Controlled proximal differential systems for solving saddle problems. *Differents. Uravneniya* **28**, 11, 1846–1861, 1992.

17. ANTIPIN A.S., Proximal differential systems, controlled by negative feedback. *Dokl. Ross. Akad. Nauk* **329**, 2, 1–3, 1993.

18. ZOUTENDIJK G., *Methods of Feasible Directions.* Elsevier, Amsterdam, 1960.