*Differential Equations, Vol.30, No.9, 1994, pp. 1365–1375*

# MINIMIZATION OF CONVEX FUNCTIONS ON CONVEX SETS BY MEANS OF DIFFERENTIAL EQUATIONS[1]

**A.S. Antipin**

UDC 519.82

(Revised version 14 March 2003)

## 1. INTRODUCTION

Let us consider the problem on minimizing a function on a simple set, namely

$$x^* \in \text{Arg min}\,\{f(x) : x \in Q\}, \tag{1.1}$$

where $f(x)$ is a differentiable scalar function, $x \in Q \subseteq \mathbb{R}^n$, $\mathbb{R}^n$ is an Euclidean finite-dimension space, and $Q$ is a simple set, i.e., a set onto which one can easily project. Examples of such sets are the positive orthant, a parallelepiped, a ball, etc.

The gradient approach to solving problem (1.1) has long been a conventional method, but its continuous variants are not well investigated. First of all, this pertains to methods in which one should take into account constraints imposed on the variables. In this paper we consider the gradient methods of first and second order, where the constraints are taken into account by means of projection operators. The asymptotic and exponential stability of such processes is proved.

## 2. CONTINUOUS METHOD OF GRADIENT PROJECTION

The idea of the method can be presented in the following way. If $x^*$ is a minimum point of problem (1.1), then the necessary and sufficient conditions

$$x^* = \pi_Q(x^* - \alpha\nabla f(x^*)) \tag{2.1}$$

are satisfied, where $\pi_Q(\cdot)$ is the projection operator of a vector onto the set $Q$, $\alpha > 0$ is a parameter such as the step length, and $\nabla f(x)$ is the gradient of function $f(x)$ at the point $x$. Condition (2.1) has a simple geometric meaning: moving from the point $x^*$ along the antigradient, we return to the point after the projection operator, i.e., $x^*$ is a fixed point or an equilibrium point. The discrepancy $\pi_Q(x - \alpha\nabla f(x)) - x$ can be considered as a transformation of space $\mathbb{R}^n$ into $\mathbb{R}^n$. This transformation determines a vector field.

Let us formulate the problem on drawing a trajectory such that its tangent line coincides with the field vector at the given point. The problem is described by the system of differential equations

$$\frac{dx}{dt} + x = \pi_Q(x - \alpha\nabla f(x)), \qquad x(t_0) = x^0. \tag{2.2}$$

[1]Center of Program Research, Russian Academy of Sciences.
Translated from *Differentsial'nye Uravneniya*, **30**, No.9, pp. 1475–1486, September, 1994.
Original article submitted January 20, 1994.

The "dynamical" definition of the fixed point $x^*$ follows from (2.2), namely, $x^*$ is the trajectory point at which the velocity is zero. It follows from general theorems that the continuous right-hand side of system (2.2) ensures the existence of a solution on a finite interval. If the Lipschitz condition is satisfied for the right-hand side (it does so in our case), then the trajectory exists and is unique on the infinite interval, i.e., for all $t \geq t_0$.

If $\pi_Q(\cdot) = I$ is the unit matrix (i.e., $Q = \mathbb{R}^n$), then Eq. (2.2) becomes

$$\frac{dx}{dt} = \alpha \nabla f(x), \qquad x(t_0) = x^0. \tag{2.3}$$

The continuous gradient method without the projection operator has been considered in many papers, (e.g., see [1]–[5]). Regularized gradient equations have been investigated by Vasil'ev [6]. The paper [7] includes the review of papers written by non-Russian authors and devoted to equations like (2.3). Differential equations (2.2) with the projection operator have been studied in detail in [8]. Asymptotic and exponential stability of these systems is proved there.

Here we dwell on the case in which the projection operator $\pi_Q(\cdot)$ is linear. This variant of the gradient projection method has been studied in detail by Evtushenko and Zhadan [9, 10] and by Tanabe [11]. It is generated by the problem on minimizing the goal function $f(x)$ under the equality-type constraints $Q = \{x : g(x) = 0, \ \ x \in \mathbb{R}^n\}$, where $g(x)$ is a differentiable vector function. One should construct a gradient trajectory belonging to the manifold $Q$. To this end, at each point $x \in Q$ one constructs the tangent subspace $K$, which uniquely generates the operator of projection of the space $\mathbb{R}^n$ onto $K$ such that $\pi_k(K) = K$. Thus, in particular, the equality $\pi_k(x) = x$ is satisfied for the point $x$ being the tangency point of $K$ to the manifold $Q$. In order to construct the vector field at each point $x$, one projects the gradient $\nabla f(x)$ onto the tangent space. Hence, this field is described by the transformation $\pi_k(\nabla f(x))$.

Consider the following: construct a trajectory $x(t)$ belonging to the manifold $Q$ such that its tangent coincides with the field vector $\pi_k(\nabla f(x))$ at each point $x$. The problem is described by the system of differential equations

$$\frac{dx}{dt} = -\alpha \pi_k(\nabla f(x)), \qquad x(t_0) = x^0.$$

The latter can be obtained from (2.2) in a quite formal way if we take into account the fact that in this case the projection operator is linear and the condition $\pi_k(x) = x$ is satisfied for it.

The projection operator is determined by the analytical formula

$$\pi_k(\cdot) = I - \nabla g^\top(x) \left[\nabla g(x) \nabla g^\top(x)\right]^{-1} \nabla g(x),$$

where $\nabla g(x)$ is the gradient of function $g(x)$.

Differential equations of internal and external linearization for convex programming problems with equality-type constraints have been described in [12, 13]. Saddle methods have been investigated in [14, 15]. We also want to mention the approach presented in [16]. The properties of iterative analog to the continuous gradient projection method, namely,

$$x^{n+1} = \pi_Q(x^n - \alpha \nabla f(x^n)), \tag{2.4}$$

have been well studied in [17].

In what follows we formulate a series of statements on the asymptotic stability steadiness of the gradient projection method (2.2), successively imposing stricter conditions on the goal function $f(x)$ and the parameter $\alpha$.

First, we recall that the projection operator $\pi_Q(b)$ of vector $b$ onto the set $Q$ can be determined by solving the quadratic problem

$$\pi_Q(b) = \arg\min \left\{ \frac{1}{2} |z - b|^2 : z \in Q \right\},$$

which, in turn, is equivalent to solving the variational inequality

$$\langle \pi_Q(b) - b, z - \pi_Q(b) \rangle \geq 0 \tag{2.5}$$

for all $z \in Q$.

Let us rewrite equations (2.1) and (2.2) in the form (2.5). The former is equivalent to the variational inequality

$$\langle \nabla f(x^*), z - x^* \rangle \geq 0 \tag{2.6}$$

for all $z \in Q$, whereas the latter is equivalent to the variational inequality

$$\langle \dot{x} + x - (x - \alpha \nabla f(x)), z - \dot{x} - x \rangle \geq 0 \tag{2.7}$$

for all $z \in Q$, where $\dot{x} = dx/dt$.

In what follows we present a theorem on the method convergence, assuming that $f(x)$ is an arbitrary differentiable function.

**Theorem 1.** *If the goal function $f(x)$ is differentiable, the gradient $\nabla f(x)$ satisfies the Lipschitz condition, $Q$ is a convex closed bounded set, the initial condition is $x^0 \in Q$, and the parameter $\alpha$ is positive, then the trajectory of process* (2.2) *has a nonempty set of limit points, each being a stationary point of problem* (1.1).

**Proof**. Before proving the theorem, we note that the trajectory of the differential equation (2.2) can always be approximated on any finite time interval by a sequence of trajectories of iterative or discrete gradient projection method depending on the discretization step of time axis. But if the initial point $x^0$ belongs to the set $Q$, then so do all iterative trajectories. Hence, the continuous trajectory (their limit) also belongs to the set $Q$.

Let us show that the trajectory $x(t)$ converges (in the sense of subsequences) to the set of stationary points of the initial problem, i.e., $\rho(x(t), X^*) \to 0$ as $t \to \infty$.

Set $z = x$ in the inequality (2.7); then

$$|\dot{x}|^2 + \alpha \langle \nabla f(x), \dot{x} \rangle \leq 0, \tag{2.8}$$

or

$$\frac{df(x)}{dt} + \frac{1}{\alpha} |\dot{x}|^2 \leq 0. \tag{2.9}$$

Let us integrate the obtained inequality over $[t_0, t]$:

$$f(x) + \frac{1}{\alpha} \int_{t_0}^{t} |\dot{x}|^2 d\tau \leq f(x^0). \tag{2.10}$$

It follows that the integral $\int_{t_0}^{t} |\dot{x}|^2 d\tau < \infty$ converges as $t \to \infty$.

If we assume that there exists an $\varepsilon > 0$ such that $|\dot{x}| \geq \varepsilon$ for all $t \geq t_0$, then we obtain a contradiction. Therefore, there exists a subsequence of time moments $t_i \to \infty$ such that $|\dot{x}| \to 0$. Since $x(t)$ is bounded, we again can choose a subsequence of time moments, which we also denote by $t_i$ such that $x(t_i) \to x'$ and $\dot{x}(t_i) \to 0$.

Let us consider inequality (2.7) for all time moments $t_i \to \infty$. Passing to the limit, we write out the limit inequality

$$\langle \nabla f(x'), z - x' \rangle \geq 0$$

for all $z \in Q$. This inequality is sufficient but not necessary for minimum. In this case $x'$ is a stationary point. Thus, the theorem is proved. □

Note that in proving the theorem we have not used the Lipschitz condition imposed on the gradient $\nabla f(x)$, but it is necessary for the trajectory $x(t)$ to exist.

Next, let us cite the theorem on convergence and estimates of the rate of convergence with respect to a functional assuming that $f(x)$ is a convex differentiable function. Here we do not assume that the initial condition belongs to the set $Q$; moreover, $x^0 \in \mathbb{R}^n$.

**Theorem 2.** *Assume that*

- *(i) the set of solutions to problem* (1.1) *is not empty, i.e.,* $X^* \neq \emptyset$*;*
- *(ii) the goal function* $f(x)$ *is convex and differentiable, and its gradient satisfies the Lipschitz condition with constant* $L$*;*
- *(iii)* $Q$ *is a convex closed set.*

*Then the following statements hold:*

- *(a) if the parameter* $\alpha$ *in the process* (2.2) *is chosen from the condition* $\alpha < 2/L$*, then the trajectory* $x(t)$ *converges monotonically* ($|x(t + \Delta t) - x^*| \leq |x(t) - x^*|$) *with respect to norm to one of the solutions to the problem for all* $x^0 \in \mathbb{R}^n$*;*
- *(b) if the parameter* $\alpha$ *is chosen from the condition* $\alpha > 0$*, then the trajectory* $x(t)$ *converges to a solution to the problem for all* $x^0 \in \mathbb{R}^n$ *with the estimate* $f(x) - f(x^*) \leq C/t$*.*

**Proof**. Let us prove statement $(a)$. Set $z = x^*$ in (2.7) and $z = x + \dot{x}$ in (2.6) and add these two inequalities. We obtain

$$\langle \dot{x} + \alpha(\nabla f(x) - \nabla f(x^*)), x^* - x - \dot{x} \rangle \geq 0. \qquad (2.11)$$

Using the inequality [18]

$$\langle \nabla f(x_1) - \nabla f(x_3), x_3 - x_2 \rangle \leq \frac{L}{4}|x_1 - x_2|^2, \qquad (2.12)$$

which holds for all $x_1, x_2, x_3 \in Q$, where $L$ is the Lipschitz constant for $\nabla f(x)$ on $Q$, we estimate

$$\langle \nabla f(x) - \nabla f(x^*), x^* - x - \dot{x} \rangle \leq \frac{L}{4}|\dot{x}|^2. \qquad (2.13)$$

Taking the latter into account, we represent (2.11) as

$$\langle \dot{x}, x - x^* + \dot{x} \rangle - \alpha \frac{L}{4}|\dot{x}|^2 \leq 0.$$

Hence,

$$\frac{1}{2}\frac{d}{dt}|x - x^*|^2 + \left(1 - \alpha\frac{L}{4}\right)|\dot{x}|^2 \leq 0. \qquad (2.14)$$

Since $\alpha < 2/L$, we see that $c = 1 - \alpha L/4 > 0$.

By integrating inequality (2.14) over the range $[t_0, t]$, we obtain

$$|x - x^*|^2 + c \int_{t_0}^{t} |\dot{x}|^2 d\tau \leq |x^0 - x^*|^2, \tag{2.15}$$

where $x^0 = x(t_0)$. It follows from (2.15) that the trajectory is bounded, namely $|x(t) - x^*|^2 \leq$ $\leq |x^0 - x^*|^2$, and the trajectory tends to $x^*$ monotonically with respect to norm.

Since $|x(t) - x^*|^2$ decreases monotonically and since each limit point is a solution to the problem, it follows that the limit point is unique, i.e., the trajectory converges monotonically to solution to the original problem.

Let us now prove statement $(b)$. Let us represent (2.11) as

$$\langle \dot{x}, x^* - x \rangle + \alpha \langle \nabla f(x) - \nabla f(x^*), x^* - x \rangle - |\dot{x}|^2 - \alpha \langle \nabla f(x) - \nabla f(x^*), \dot{x} \rangle \geq 0. \tag{2.16}$$

Taking into account the fact that the gradient $\nabla f(x)$ is monotone, we obtain

$$\frac{d}{dt}|x - x^*|^2 + |\dot{x}|^2 + \alpha \frac{d}{dt}(f(x) - f(x^*) - \langle \nabla f(x^*), x - x^* \rangle) \leq 0. \tag{2.17}$$

Next, integrating (2.17) from $t_0$ to $t$, we obtain

$$\begin{aligned} |x - x^*|^2 \quad &+ \int_{t_0}^{t} |\dot{x}|^2 d\tau + \alpha(f(x) - f(x^*) - \langle \nabla f(x^*), x - x^* \rangle) \leq \\ &\leq \quad |x^0 - x^*|^2 + \alpha(f(x^0) - f(x^*) - \langle \nabla f(x^*), x^0 - x^* \rangle). \end{aligned} \tag{2.18}$$

Since $f(x)$ is convex, i.e., $f(x) - f(x^*) - \langle \nabla f(x^*), x - x^* \rangle \geq 0$, it follows from (2.18) that the trajectory $x(t)$ is bounded, i.e., $|x(t) - x^*| \leq C$, and decreases monotonically in the sense of $f(x) - f(x^*) - \langle \nabla f(x^*), x - x^* \rangle$. These properties are sufficient for the trajectory to converge to a limit point, namely, $x(t) \to x^* \in X^*$ as $t \to \infty$ for all $x^0$.

Let us estimate the rate at which the goal function decreases along the trajectory. Set $z = x^*$ in inequality (2.7) and represent the latter as

$$\langle \dot{x}, x^* - x - \dot{x} \rangle \geq \alpha \langle \nabla f(x), x + \dot{x} - x^* \rangle \geq 0. \tag{2.19}$$

Taking into account the fact that $f(x)$ is convex, consider the system of inequalities

$$f(x) - f(x^*) \leq \langle \nabla f(x), x - x^* \rangle = \langle \nabla f(x), x + \dot{x} - x^* \rangle - \langle \nabla f(x), \dot{x} \rangle. \tag{2.20}$$

Comparing (2.19) and (2.20) and taking into account the fact that $x(t) \to x^*$, $\dot{x} \to 0$, and $\nabla f(x(t)) \to \nabla f(x^*)$, we obtain

$$\begin{aligned} f(x) - f(x^*) \quad &\leq \quad (1/\alpha)\langle \dot{x}, x^* - x - \dot{x} \rangle - \langle \nabla f(x), \dot{x} \rangle = \\ &= \quad \langle \dot{x}, (1/\alpha)(x^* - x - \dot{x}) - \nabla f(x) \rangle \leq C|\dot{x}|. \end{aligned}$$

Hence, denoting $f(x^*) = f^*$, we have

$$f(x) - f^* \leq C|\dot{x}|. \tag{2.21}$$

By comparing (2.9) and (2.11), we obtain

$$\frac{d(f(x) - f^*)}{dt} + \frac{1}{C^2 \alpha}(f(x) - f^*)^2 \leq 0. \tag{2.22}$$

Next, we separate the variables and integrate inequality (2.22), thus obtaining

$$-(f(x)-f^*)^{-1}+(C^2\alpha)^{-1}\le C_1.$$

Therefore, $f(x)-f^*\le (C_2t-C_1)^{-1}\approx C/t$, and the theorem is proved. □

The iterative method (2.4) of gradient projection converges under the same conditions.

If the goal function in problem (1.1) is strongly convex, then we can estimate the rate of convergence of the process (2.2).

**Theorem 3.** *If, in addition to the conditions of Theorem* 2, *the goal function is strongly convex, then the only equilibrium point of system* (2.2) *is exponentially stable, i.e.,*

$$|x(t)-x^*|^2\le C\exp(-2s(\alpha)t), \tag{2.23}$$

*where*

$$s(\alpha)=\begin{cases}\alpha\ell(1-\alpha\ell/4) & if\ \ \alpha<4/(L+\ell),\\ \alpha L(1-\alpha L/4) & if\ \ \alpha>4/(L+\ell).\end{cases} \tag{2.24}$$

**Proof**. Let us rewrite inequality (2.11) in the form

$$\langle\dot x,x-x^*\rangle+|\dot x|^2+\alpha\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle+\alpha\langle\nabla f(x)-\nabla f(x^*),\dot x\rangle\le 0.$$

Then

$$\frac{1}{2}\frac{d}{dt}|x-x^*|^2+|\dot x|^2+\alpha\langle\nabla f(x)-\nabla f(x^*),\dot x\rangle+\alpha\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle\le 0. \tag{2.25}$$

Let us single out perfect squares from the second and third terms, namely,

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}|x-x^*|^2 \ &+\ \left|\dot x+\frac{\alpha}{2}(\nabla f(x)-\nabla f(x^*))\right|^2-\frac{\alpha^2}{4}|\nabla f(x)-\nabla f(x^*)|^2+\\ &+\ \alpha\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle\le 0.\end{aligned} \tag{2.26}$$

Next, we omit the second term in (2.26) and estimate the third term by means of the following inequality [18]:

$$|\nabla f(x_1)-\nabla f(x_2)|^2+L\ell|x_1-x_2|^2\le(L+\ell)\langle\nabla f(x_1)-\nabla f(x_2),x_1-x_2\rangle \tag{2.27}$$

which holds for all $x_1$ and $x_2$ from $Q$, where $\ell$ is the constant of strong convexity for the goal function, and $L$ is the Lipschitz constant for the gradient $\nabla f(x)$. Then we have

$$\frac{1}{2}\frac{d}{dt}|x-x^*|^2+\alpha^2\frac{L\ell}{4}|x-x^*|^2+\alpha(1-\alpha(L+\ell)/4)\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle\le 0. \tag{2.28}$$

If $\alpha<4/(L+\ell)$, then $1-\alpha(L+\ell)/4>0$. In this case, in order to estimate the last term in (2.28), we use the inequality

$$\ell|x-x^*|^2\le\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle. \tag{2.29}$$

Otherwise, if $\alpha>4/(L+\ell)$, we have $1-\alpha(L+\ell)/4<0$, and, therefore, apply the inequality

$$\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle\le L|x-x^*|^2. \tag{2.30}$$

Then (2.28) can be rewritten as

$$\frac{1}{2}\frac{d}{dt}|x-x^*|^2+s(\alpha)|x-x^*|^2\le 0, \tag{2.31}$$

where $s(\alpha)$ is determined by formula (2.24).

By integrating (2.31), we obtain

$$|x(t) - x^*|^2 \leq C \exp(-2s(\alpha)t), \tag{2.32}$$

where $C = |x^0 - x^*|^2$. To ensure that the trajectory converges exponentially, it is necessary that $s(\alpha) > 0$, i.e., $\alpha < \min\{4/\ell, 4/L\} = \dfrac{4}{L}$. Thus, although the process (2.2) converges for any parameter $\alpha$, the exponential convergence is ensured only for $\alpha \in (0, 4/L)$.

Since $s(\alpha)$ consists of branches of two parabolas (an ascending and a descending one), it is easy to see that the optimal $\alpha$ is equal to $\alpha_{opt} = 4/(L+\ell)$ with $s(\alpha_{opt}) = 4L\ell/(L+\ell)^2$. Thus, the theorem is proved. □

## 3. SECOND-ORDER CONTINUOUS METHOD OF THE GRADIENT PROJECTION

The discrepancy $\pi_Q(x - \alpha\nabla f(x)) - x$ generated by the necessary condition (2.1) determines the transformation of space $\mathbb{R}^n$ into $\mathbb{R}^n$. This transformation can be considered as a vector field.

Let us formulate the problem on drawing a trajectory such that some linear combination of velocity and acceleration on this trajectory coincides with the field vector at each point. The problem is described by the system of differential equations [8]

$$\mu\frac{d^2x}{dt^2} + \beta\frac{dx}{dt} + x = \pi_Q(x - \alpha\nabla f(x)), \qquad x(t_0) = x^0, \quad \dot{x}(t_0) = \dot{x}^0. \tag{3.1}$$

Here $\mu > 0$ and $\beta > 0$ are parameters. If $\mu = 0$ and $\beta = 1$, then process (3.1) coincides with (2.2). If $Q = \mathbb{R}^n$, i.e., $\pi_Q(*) = I$ is the identity operator, then (3.1) becomes

$$\mu\frac{d^2x}{dt^2} + \beta\frac{dx}{dt} = -\alpha\nabla f(x), \qquad x(t_0) = x^0, \quad \dot{x}(t_0) = \dot{x}^0. \tag{3.2}$$

If the gradient $\nabla f(x)$ satisfies the Lipschitz condition, then the solution $x(t)$ to system (3.1), (3.2) exists and is unique for all $t \geq t_0$. Here we should mention the papers [8, 19], in which equations of the form (3.2) have been considered earlier.

There exist various iterative analogs of the continuous method of gradient projection, e.g.,

$$x^{n+1} = \pi_Q\{x^n - \alpha\nabla f(x^n) + \mu(x^n - x^{n-1})\}. \tag{3.3}$$

The properties of this process have been studied in [8].

Let us represent the process (3.1) as the variational inequality

$$\langle \mu\ddot{x} + \beta\dot{x} + x - (x - \alpha\nabla f(x)), z - \mu\ddot{x} - \beta\dot{x} - x\rangle \leq 0 \tag{3.4}$$

for all $z \in Q$, where $\dot{x} = dx/dt$ and $\ddot{x} = d^2x/dt^2$.

Equation (3.1) and the variational inequality (3.4) are equivalent.

Let us present a theorem on convergence of the method to the solution of the initial problem.

**Theorem 4.** *Assume that*

*(i) the set of solutions to problem* (1.1) *is not empty, i.e.,* $X^* \neq \emptyset$;

*(ii) the goal function* $f(x)$ *is convex and twice continuously differentiable, with its second derivative bounded on a convex closed set containing the trajectory* $x(t)$, *i.e.,* $|\nabla^2 f(x)| \leq N$;

*(iii) $Q$ is a convex closed set;*

*(iv) the parameters $\alpha > 0$ and $\mu > 0$ satisfy the condition $\mu < \beta^2/(1+\alpha N)$.*

*Then the trajectory $x(t)$ converges to a solution to the problem, i.e., $x(t) \to x^* \in X^*$ as $t \to \infty$ for all $x^0$.*

**Proof**. Set $z = x^*$ in (3.4) and $z = \mu\ddot{x} + \beta\dot{x} + x$ in (2.6), and add these inequalities. Then

$$\langle \mu\ddot{x} + \beta\dot{x} + \alpha(\nabla f(x) - \nabla f(x^*)), x^* - \mu\ddot{x} - \beta\dot{x} - x\rangle \le 0. \tag{3.5}$$

Hence, we have

$$\begin{aligned} |\mu\ddot{x} + \beta\dot{x}|^2 &+ \langle \mu\ddot{x} + \beta\dot{x}, x - x^*\rangle + \alpha(\nabla f(x) - \nabla f(x^*)), x - x^*\rangle + \\ &+ \alpha\langle \nabla f(x) - \nabla f(x^*), \mu\ddot{x} + \beta\dot{x}\rangle \le 0. \end{aligned} \tag{3.6}$$

Next, we omit the nonnegative third term in (3.6) and then transform the other terms as follows:

$$\begin{aligned} \mu^2|\ddot{x}|^2 &+ \mu\beta\frac{d}{dt}|\dot{x}|^2 + \beta^2|\dot{x}|^2 + \mu\langle \ddot{x}, x - x^*\rangle + \frac{1}{2}\beta\frac{d}{dt}|x - x^*|^2 + \\ &+ \alpha\mu\langle \nabla f(x) - \nabla f(x^*), \ddot{x}\rangle + \alpha\beta\langle \nabla f(x) - \nabla f(x^*), \dot{x}\rangle \le 0. \end{aligned} \tag{3.7}$$

Using the identities

$$\begin{aligned} \frac{1}{2}\frac{d^2}{dt^2}|x - x^*|^2 &= |\dot{x}|^2 + \langle x - x^*, \ddot{x}\rangle, \\ \frac{d}{dt}\langle \nabla f(x) - \nabla f(x^*), \dot{x}\rangle &= \langle \nabla^2 f(x)\dot{x}, \dot{x}\rangle + \langle \nabla f(x) - \nabla f(x^*), \ddot{x}\rangle, \end{aligned} \tag{3.8}$$

we transform (3.7) so that

$$\begin{aligned} \mu^2|\ddot{x}|^2 &+ \mu\beta\frac{d}{dt}|\dot{x}|^2 + (\beta^2 - \mu)|\dot{x}|^2 + \mu\frac{1}{2}\frac{d^2}{dt^2}|x - x^*|^2 + \\ &+ \frac{1}{2}\beta\frac{d}{dt}|x - x^*|^2 - \alpha\mu\langle \nabla^2 f(x)\dot{x}, \dot{x}\rangle + \\ &+ \alpha\mu\frac{d}{dt}\langle \nabla f(x) - \nabla f(x^*), \dot{x}\rangle + \alpha\beta\langle \nabla f(x) - \nabla f(x^*), \dot{x}\rangle \le 0. \end{aligned} \tag{3.9}$$

According to the conditions of the theorem, the second derivative is bounded by a constant $N$ on some convex set containing the trajectory $x(t)$, i.e., $\langle \nabla^2 f(x)\dot{x}, \dot{x}\rangle \le |\nabla^2 f(x)|\,|x|^2 \le N|x|^2$. Taking into account this fact and the relation

$$\frac{d}{dt}\{f(x) - f(x^*) - \langle \nabla f(x^*), x - x^*\rangle\} = \langle \nabla f(x) - \nabla f(x^*), \dot{x}\rangle \tag{3.10}$$

and denoting

$$\varphi(x) = \frac{1}{2}|x - x^*|^2 + \alpha(f(x) - f(x^*) - \langle \nabla f(x^*), x - x^*\rangle),$$

we transform inequality (3.9) into

$$\mu\frac{d^2}{dt^2}\varphi(x) + \beta\frac{d}{dt}\varphi(x) + \mu\beta\frac{d}{dt}|\dot{x}|^2 + (\beta^2 - \mu(1 + \alpha N))|\dot{x}|^2 + \mu^2|\ddot{x}|^2 \le 0. \tag{3.11}$$

Since $\mu < \beta^2/(1+\alpha N)$, we see that $\beta^2 - \mu(1+\alpha N) > 0$, and, therefore, inequality (3.11) can be integrated from $t_0$ to $t$, namely,

$$\mu\frac{d}{dt}\varphi(x) + \beta\varphi(x) + \mu\beta|\dot{x}|^2 + (\beta^2 - \mu(1 + \alpha N))\int_{t_0}^{t}|\dot{x}|^2 d\tau + \mu^2\int_{t_0}^{t}|\ddot{x}|^2 d\tau \le C. \tag{3.12}$$

Let us show that the trajectory $x(t)$ is bounded. To this end, we write inequality (3.12) as

$$\mu\frac{d}{dt}\varphi(x)+\beta\varphi(x)\le C. \tag{3.13}$$

Hence,

$$\mu\exp\left(-\frac{\beta}{\mu}t\right)\frac{d}{dt}\left(\exp\left(\frac{\beta}{\mu}t\right)\varphi(x)\right)\le C.$$

Next,

$$\frac{d}{dt}\left(\exp\left(\frac{\beta}{\mu}t\right)\varphi(x)\right)\le C\frac{1}{\mu}\exp\left(\frac{\beta}{\mu}t\right). \tag{3.14}$$

By integrating the latter, we obtain

$$\exp\left(\frac{\beta}{\mu}t\right)\varphi(x)\le C_1\exp\left(\frac{\beta}{\mu}t\right)+C_2,$$

and, therefore,

$$\varphi(x)\le C_1+C_2\exp\left(-\frac{\beta}{\mu}t\right)\le C_3. \tag{3.15}$$

The function $\varphi(x)$ is strongly convex, and consequently, each of its Lebesgue sets is bounded. Thus, in particular, the Lebesgue set satisfying (3.15) is bounded. The trajectory $x(t)$ belongs to it, and, therefore, it is also bounded, i.e.,

$$|x(t)-x^*|\le C. \tag{3.16}$$

Let us show that the first term in (3.12) is bounded below. To this end, we first show that $|\dot{x}|^2$ is bounded. By taking into account the fact that all terms but the first are nonnegative (recall that $f(x)-f(x^*)-\langle\nabla f(x^*),x-x^*\rangle\ge 0$ 0 for all $x\in\mathbb{R}^n$ since $f(x)$ is convex) and by using identity (3.10), we can rewrite (3.12) in the form

$$\langle x-x^*+\alpha\langle\nabla f(x)-\nabla f(x^*),\dot{x}\rangle\rangle+\beta|\dot{x}|^2\le C/\mu,$$

or

$$\begin{aligned}&\left|(2\sqrt{\beta})^{-1}(x-x^*+\alpha(\nabla f(x)-\nabla f(x^*)))+\sqrt{\beta}\dot{x}\right|^2\le\\ \le\;&\frac{1}{\mu}C+(4\beta)^{-1}|x-x^*+(\nabla f(x)-\nabla f(x^*))|^2\le\text{Const}.\end{aligned}$$

Since $|x-x^*+(\nabla f(x)-\nabla f(x^*))|^2$ is bounded, it follows from the last inequality that the derivative is bounded, i.e.,

$$|\dot{x}|^2\le\text{Const}. \tag{3.17}$$

By taking into account the identity $0\le|a+b|^2=|a|^2+2\langle a,b\rangle+|b|^2$, we estimate the first term in (3.12) as

$$\frac{d}{dt}\varphi(x)=\langle x-x^*+\alpha\langle\nabla f(x)-\nabla f(x^*),\dot{x}\rangle\rangle\ge-\frac{C_4+C_5}{2}.$$

Now we again rewrite (3.12) as

$$(\beta^2-\mu(1+\alpha N))\int_{t_0}^{t}|\dot{x}|^2d\tau+\mu\int_{t_0}^{t}|\ddot{x}|^2d\tau\le C+\frac{C_4+C_5}{2}\le\text{Const}. \tag{3.18}$$

It follows from this estimate that the integrals

$$\int_{t_0}^{t} |\dot{x}|^2 d\tau < \infty, \qquad \int_{t_0}^{t} |\ddot{x}|^2 d\tau < \infty \tag{3.19}$$

converge as $t \to \infty$.

Let us prove that the set of equilibrium points is asymptotically stable. Assuming that there exists an $\varepsilon > 0$ such that $|\ddot{x}(t)| \geq \varepsilon$ and $|\dot{x}(t)| \geq \varepsilon$ for all $t \geq t_0$, we obtain a contradiction to the convergence of integrals. Hence, there exists a subsequence of time moments $t_i \to \infty$ such that $|\ddot{x}(t_i)| \to 0$ and $|\dot{x}(t_i)| \to 0$. Since $x(t)$ is bounded, we again choose a subsequence (which we also denote by $t_i$) such that $x(t_i) \to x'$, $|\ddot{x}(t_i)| \to 0$, and $|\dot{x}(t_i)| \to 0$.

Let us consider inequality (3.4) for all time moments $t_i \to \infty$, and write out the limit inequality

$$\langle \nabla f(x'), z - x' \rangle \geq 0$$

for all $z' \in Q$. Since $f(x)$ is convex, we have $x' = x^* \in Q$. Thus, any limit point of the trajectory $x(t)$ is a solution to the problem.

Omitting the last terms in inequality (3.11), we integrate it from $t_0$ to $t$ and obtain

$$\Phi(x) \equiv \mu \frac{d}{dt}\varphi(x) + \beta\varphi(x) + \mu\beta|\dot{x}|^2 \leq \mu\frac{d}{dt}\varphi(x(t_0)) + \beta\varphi(x(t_0)) + \mu\beta|\dot{x}(t_0)|^2.$$

It follows that $\Phi(x)$ decreases monotonically along any trajectory $x(t)$. If we assume that the trajectory $x(t)$ has two limit points, $x'$ and $x''$, then two subsequences converging to $x'$ and $x''$, respectively, result in $\Phi(x')$ and $\Phi(x'')$, respectively (it has been proved earlier that $d\varphi(x(t))/dt \to$ $\to 0$ and $|\dot{x}(t)|^2 \to 0$). Since $\Phi(x)$ is a strongly convex function, we see that $\Phi(x') \neq \Phi(x'')$. If we take two neighborhoods of the points $x'$ and $x''$ so that the corresponding intervals having the centers at the points $\Phi(x')$ and $\Phi(x'')$ do not intersect, then we see that terms of a monotonic numerical sequence lie in two nonintersecting intervals, which contradicts the monotonicity. Thus, the theorem is proved. □

Next, we cite the theorem on estimating the rate of convergence for the second-order method of gradient projection.

**Theorem 5.** *If $f(x)$ is a strongly convex differentiable function, $Q$ is a convex closed set, and the parameters $\mu$, $\beta$, and $\alpha$ satisfy the conditions $0 < \mu < \min\{(\sqrt{1+\beta^3} - 1)/2\beta, \sqrt{\beta/2}\}$, where $\beta < 3\sqrt[3]{s(\alpha)^2}$, and*

$$s(\alpha) = \begin{cases} \alpha\ell(1 - 3\alpha\ell/4) & if \quad \alpha < 4/3(L+\ell), \\ \alpha L(1 - 3\alpha L/4) & if \quad \alpha > 4/3(L+\ell), \end{cases} \tag{3.20}$$

*then system (3.1) has a unique equilibrium point, and this point is exponentially stable, i.e., $|x - x^*|^2 \leq C_3 \exp(-(\mu - \varepsilon)t)$, where $\varepsilon < \mu$.*

**Proof**. We rewrite inequality (3.6) in the form

$$\mu^2|\ddot{x}|^2 + 2\mu\beta\langle \dot{x}, \ddot{x}\rangle + \frac{1}{2}\beta^2|\dot{x}|^2 + \frac{1}{2}\beta^2|\dot{x}|^2 + \mu\langle x - x^*, \ddot{x}\rangle + \alpha\mu\langle \nabla f(x) - \nabla f(x^*), \ddot{x}\rangle +$$
$$+ \frac{1}{2}\beta\frac{d}{dt}|x - x^*|^2 + \alpha\beta\langle \nabla f(x) - \nabla f(x^*), \dot{x}\rangle + \alpha\langle \nabla f(x) - \nabla f(x^*), x - x^*\rangle \leq 0.$$

Next, we combine the first term with the sixth term and the fourth term with the eighth term and separate perfect squares from these pairs. We obtain

$$\begin{aligned}\frac{1}{2}\mu\frac{d^2}{dt^2}|x-x^*|^2 &+ \frac{1}{2}\beta\frac{d}{dt}|x-x^*|^2+2\mu\beta\langle\dot{x},\ddot{x}\rangle+\left(\frac{1}{2}\beta^2-\mu\right)|\dot{x}|^2+\\ &+ \left|\mu\ddot{x}+\frac{\alpha}{2}(\nabla f(x)-\nabla f(x^*))\right|^2-\frac{\alpha^2}{4}|\nabla f(x)-\nabla f(x^*)|^2+\\ &+ \left|\frac{\beta}{\sqrt{2}}\dot{x}+\frac{\alpha}{\sqrt{2}}(\nabla f(x)-\nabla f(x^*))\right|^2-\\ &- \frac{\alpha}{2}|\nabla f(x)-\nabla f(x^*)|^2+\alpha\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle\le 0.\end{aligned}$$

Omitting some positive terms, we rewrite the last inequality as

$$\begin{aligned}\frac{1}{2}\mu\frac{d^2}{dt^2}|x-x^*|^2+\frac{1}{2}\beta\frac{d}{dt}|x-x^*|^2+2\mu\beta\langle\dot{x},\ddot{x}\rangle+\left(\frac{1}{2}\beta^2-\mu\right)|\dot{x}|^2-\\ -\frac{3\alpha^2}{4}|\nabla f(x)-\nabla f(x^*)|^2+\alpha\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle\le 0.\end{aligned} \tag{3.21}$$

Next, we estimate the fifth term in (3.21) by means of inequality (2.27) and rewrite (3.21) as

$$\begin{aligned}&\frac{1}{2}\mu\frac{d^2}{dt^2}|x-x^*|^2+\frac{1}{2}\beta\frac{d}{dt}|x-x^*|^2+\frac{3}{4}\alpha^2L\ell|x-x^*|^2+\\ +\;&\alpha\left(1-\frac{3}{4}\alpha(L+\ell)\right)\langle\nabla f(x)-\nabla f(x^*),x-x^*\rangle\le 0.\end{aligned} \tag{3.22}$$

If the condition $\alpha<(4/3)(L+\ell)^{-1}$ is satisfied, then $1-(3/4)\alpha(L+\ell)>0$, and we use estimate (2.29); otherwise ($\alpha>(4/3)(L+\ell)^{-1}$) we use (2.30). Then, taking into account the previous considerations, we represent (3.21) in the form

$$\frac{1}{2}\mu\frac{d^2}{dt^2}|x-x^*|^2+\frac{1}{2}\beta\frac{d}{dt}|x-x^*|^2+s(\alpha)|x-x^*|^2+\mu\beta\frac{d}{dt}|\dot{x}|^2\left(\frac{1}{2}\beta^2-\mu\right)|\dot{x}|^2\le 0, \tag{3.23}$$

where the function $s(\alpha)$ is determined by (3.20).

The second-order differential inequality (3.23) is nonhomogeneous and includes the term $\mu\beta(d/dt)|\dot{x}|^2+(\beta^2/2-\mu)|\dot{x}|^2$ having indefinite sign. This makes it difficult to analyze the inequality. Therefore, following the idea proposed by A. Nedich, we multiply (3.23) by $\exp(\mu t)$ and, using the formula

$$\frac{d}{dt}(\exp(\mu t)y)=\mu\exp(\mu t)y+\exp(\mu t)\frac{d}{dt}y, \tag{3.24}$$

reduce it to

$$\begin{aligned}&\frac{1}{2}\mu\frac{d}{dt}\left(\exp(\mu t)\frac{d}{dt}|x-x^*|^2\right)+\frac{1}{2}(\beta-\mu^2)\frac{d}{dt}\left(\exp(\mu t)|x-x^*|^2\right)+\\ +\;&\left(s(\alpha)-\frac{1}{2}(\beta-\mu^2)\mu\right)\exp(\mu t)|x-x^*|^2+\\ +\;&\mu\beta\frac{d}{dt}(\exp(\mu t)|\dot{x}|^2)+\left(\left(\frac{1}{2}\beta^2-\mu\right)-\mu^2\beta\right)\exp(\mu t)|\dot{x}|^2\le 0.\end{aligned} \tag{3.25}$$

Putting

$$\frac{1}{2}(\beta-\mu^2)>0,\qquad s(\alpha)-\frac{1}{2}(\beta-\mu^2)\mu>0,\qquad \left(\frac{1}{2}\beta^2-\mu\right)-\mu^2\beta>0 \tag{3.26}$$

and integrating (3.25) from $t_0$ to $t$, after simple transformations we obtain

$$\mu\frac{d}{dt}|x-x^*|^2+(\beta-\mu^2)|x-x^*|^2\le 2C\exp(-\mu t). \tag{3.27}$$

Multiplying the latter by $\exp((\mu-\varepsilon)t)$, where $\mu-\varepsilon>0$, and applying formula (3.24), we rewrite (3.27) as

$$\frac{1}{2}\mu\frac{d}{dt}\psi(x)+\left(\beta-\frac{3}{2}\mu^2\right)\psi(x)\leq C\exp(-\varepsilon t), \tag{3.28}$$

where $\psi(x)=\exp((\mu-\varepsilon)t)|x-x^*|^2$.

Integrating (3.28) from $t_0$ to $t$ and then reasoning as in integrating inequality (3.13), we obtain

$$\exp((\mu-\varepsilon)t)|x-x^*|^2\leq\left(C_1\varepsilon^{-1}(\exp(-\varepsilon t_0)-\exp(-\varepsilon t))+C_2\right).$$

Hence,

$$|x-x^*|^2\leq C_3\exp(-(\mu-\varepsilon)t), \tag{3.29}$$

where $\varepsilon<\mu$.

Let us find the constraints on $\mu$ from conditions (3.26). The first condition results in $\mu<\sqrt{\beta/2}$. Another constraint can be obtained by solving the quadratic equation $\mu^2+(1/\beta)\mu-\beta/2=0$. The latter has the roots $\mu_{1,2}=(-1\pm\sqrt{1+2\beta^3})/2\beta$. Therefore, if $0<\mu<(\sqrt{1+2/\beta^3}-1)/2\beta$, then $(\beta^2/2-\mu)-\mu^2\beta>0$. Combining the two constraints for $\mu$, we have

$$0<\mu<\min\left\{\frac{\sqrt{1+2\beta^3}-1}{2\beta},\sqrt{\frac{\beta}{2}}\right\}. \tag{3.30}$$

Finally, the last constraint for $\mu$ can be found from the last condition in (3.26) by solving the cubic equation $\mu^3-\beta\mu+2s(\alpha)=0$. According to the Cardano formula, its roots are $\mu=v-\beta/3v$, where

$$v=\sqrt[3]{-s(\alpha)+\sqrt{s^2(\alpha)-\left(\frac{\beta}{3}\right)^2}}.$$

If $s(\alpha)^2-(\beta/3)^3>0$, then the equation has one negative root and two complex ones, i.e., the condition $s(\alpha)-(1/2)(\beta-\mu^2)\mu>0$ is satisfied for all $\mu\geq0$ if

$$\beta<3\sqrt[3]{s^2(\alpha)}. \tag{3.31}$$

Hence, if the parameters $\mu$, $\beta$, and $\alpha$ are related by (3.29) and (3.30), then the trajectory $x(t)$ converges to $x^*$ and the rate of convergence is exponential. Thus, the theorem is proved. □

Both the first- and second-order methods of gradient projection have exponential degree of convergence, but the convergence exponent in the second-order method can always be made larger than that in the first-order method by choosing the parameters $\alpha$, $\beta$, and $\mu$. In this sense the second-order method converges faster than the first-order one.

## LITERATURE CITED


1. S.I. Al'ber and Ya.I. Al'ber, *Zh. Vychislit. Mat. i Mat. Fiz.*, **7**, No. 1, 14–32 (1967).

2. M.V. Rybashev, *Avtomat. Telemekh.*, No. 9, 12–18 (1974).

3. V.I. Venets, *Avtomat. Telemekh.*, No. 9, 5–14 (1979).

4. Yu.G. Evtushenko and V.G. Zhadan, *Zh. Vychislit. Mat. i Mat. Fiz.*, **15**, No. 1, 101–112 (1975).

5. F.P. Vasil'ev, *Numerical Methods for Solving Extremum Problems* [in Russian], Moscow (1988).

6. F.P. Vasil'ev, *Methods for Solving Extremum Problems* [in Russian], Moscow (1981).

7. A.A. Brown and M.C. Bartholomew-Biggs, *J. Optim. Theory and Appl.*, **62**, No. 2, 211–224 (1989).

8. A.S. Antipin, Continuous and Iterative Processes with Projection Operators and Projection-Like Operators. In: *Voprosy Kibernetiki. Vychishtel'nye Voprosy Analiza Bol'shikh Sistem* [in Russian], AN SSSR, Scientific Counsel on the Complex Problem "Cybernetics", Moscow, 5–43 (1989).

9. Yu.G. Evtushenko and V.G. Zhadan, *Zh. Vychislit. Mat. i Mat. Fiz.*, **17**, No. 4, 73–87 (1977).

10. Yu.G. Evtushenko and V.G. Zhadan, *Threshold-Proyective and Threshold-Newton Numerical Methods of Optimization (Nonlinear Programming)* [in Russian], Preprint AN SSSR, Moscow (1991).

11. K.A. Tanabe, *J. Optim. Theory and Appl.*, **30**, No. 2, 181–210 (1980).

12. A.S. Antipin, *Problems of Nonlinear Dynamics: Optimization and Control* [in Russian], Sbornik Trudov ISA RAN, issue 3, Moscow, 5–14 (1993).

13. A.S. Antipin, *Zh. Vychislit. Mat. i Mat. Fiz.*, **33**, No. 12, 1776–1791 (1993).

14. A.S. Antipin, *Differents. Uravn.*, **28**, No. 11, 1846–1861 (1992).

15. A.S. Antipin, *Dokl. Akad. Nauk SSSR*, **333**, No. 6, 709–711 (1993).

16. S.D. Flam, *Mathematics of Operations Research*, **17**, No. 2, 290–302 (1992).

17. B.T. Polyak, *Introduction to Optimization* [in Russian], Moscow (1983).

18. A.S. Antipin, *Methods of Nonlinear Programming Based on the Direct and Dual Modification of the Lagrange Function* [in Russian], Preprint VNIISI, Moscow (1979).

19. F. Zirilli, S. Incerti, and F. Aluffi, *Systems of Equations and A-Stable Integration of Second-Order ODEs.* In: L. C. W. Dixon and G. P. Szego (eds.), *Numerical Optimization and Dynamic Systems*, Amsterdam (1980).