Differential Equations, Vol. 30, No. 11, 1994, pp. 1703–1713

ON FINITE CONVERGENCE OF PROCESSES TO A SHARP MINIMUM AND TO A SMOOTH MINIMUM WITH A SHARP DERIVATIVE[1][2]

A.S. Antipin

UDC 517.97+519.82

(Revised version 14 February 2003)

1. INTRODUCTION

Let us consider an illustrating example. The problem is to minimize the Euclidean norm in a finite-dimensional space

$$x^* \in \text{Argmin}\{|x| : x \in \mathbb{R}^n\}, \tag{1.1}$$

where the function $f(x) = |x| = (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}$ is differentiable everywhere except for the origin and its gradient is $\nabla f(x) = x/|x|$ (this can easily be checked by direct differentiation of the Euclidean norm). At the point $x^* = 0$ the goal function has a sharp minimum. The gradient $\nabla f(x)$ generates a vector field at each point of which the unit vector is directed to the origin along the radius vector. The subdifferential at the point of minimum is the unit ball centered on the origin. The vector field has a finite jump at the point of minimum.

We pose a problem to find a trajectory such that at each time moment its tangent is parallel to the corresponding vector of the field. In terms of differential equations this is equivalent to the system

$$\frac{dx}{dt} = -\alpha \frac{x}{|x|}, \qquad \alpha > 0, \qquad x(t_0) = x^0. \tag{1.2}$$

The trajectory of the process $x(t)$ exists for all $t \geq t_0$. There is a question about its behavior as $t \to \infty$ and, in particular, about its convergence to the point of minimum.

First notice that as it follows from (1.2) all tangent vectors to the trajectory have the same length for all $t \geq t_0$, namely,

$$|dx/dt| = \alpha. \tag{1.3}$$

We take the inner product of Eq. (1.2) by dx/dt and find

$$\left|\frac{dx}{dt}\right|^2 = -\alpha \left\langle \frac{x}{|x|}, \frac{dx}{dt} \right\rangle = -\alpha \langle \nabla f(x), \dot{x} \rangle = -\alpha \frac{df(x)}{dt}, \tag{1.4}$$

where $\dot{x} = dx/dt$. Comparing (1.3) and (1.4) we get

$$\frac{d(f(x) - f(x^*))}{dt} + \alpha = 0. \tag{1.5}$$

[1]Computing Center of Russian Academy of Sciences.
Translated from *Differentsial'nye Uravneniya*, **30**, No.11, pp. 1843–1854, November, 1994.
Original article submitted June 6, 1994.
[2]This work was supported by the Russian Foundation for Basic Research (grant No. 94-01-00005).

We now integrate (1.5) from t_0 to t and derive

$$f(x) - f(x^*) + \alpha \int\limits_{t_0}^{t} d\tau = f(x^0) - f(x^*),$$

or

$$f(x) - f(x^*) + \alpha(t - t_0) = f(x^0) - f(x^*). \tag{1.6}$$

Since $f(x) - f(x^*) \geq 0$ as $t - t_0 \to \infty$ and the right-hand side of Eq.(1.6) is a constant, we arrive at a contradiction. The quantity $f(x) - f(x^*)$ vanishes at the point $t = t_0 + \alpha^{-1}(f(x^0) - f(x^*))$, therefore the trajectory $x(t)$ passes through the point of minimum x^*. Thus the optimal value on the trajectory is attained in a finite time but unlike smooth processes the velocity at the optimal point can take any value of the unit ball centered at the origin.

Since the velocity at the optimal point is nonzero, the trajectory does not terminate at the point of minimum (like in smooth process) but leaves it in any direction. In this case the point of minimum is unstable. Instability also appears when the time corresponding to the optimal point of the process is computed not exactly. A small error in this time may lead to the case in which process (1.2) passes through the optimal point. The divergence can be close to unity.

Hence, there are two characteristics of the problems with a sharp minimum. On one hand, these problems are of interest because gradient, proximal, and other methods can converge to the solution to such a problem in a finite time. On the other hand, these processes are extremely unstable, i.e., the trajectory of the solution in equilibrium can be chaotically deviated (we assume that the velocity at a point of equilibrium can be either zero or take any other value from the unit ball centered at the origin). To make the equilibrium state of process (1.2) rough it is useful to use regularization, which is developed in [1, 2].

Considered example shows that if the gradient is bounded from below by a positive constant, then the process comes to the equilibrium state in a finite time. Another situation is much more difficult.

Consider the problem

$$x^* \in \operatorname{Argmin}\{|x|^{1+\nu} : x \in \mathbb{R}^n\}, \tag{1.7}$$

where $0 < \nu < 1$. The goal function is differentiable everywhere including the origin, and its gradient is $\nabla f(x) = (1+\nu)|x|^\nu x/|x| = (1+\nu)x/|x|^{1-\nu}$. The goal function has a smooth minimum at the point $x^* = 0$, where the first derivative is sharp and the second derivative is discontinuous. It is important to emphasize that the vector field generated by the gradient of this function is smooth in a neighborhood of the minimum point although the second derivative of the field is discontinuous. The intuition suggests that when moving along the gradient trajectory in a neighborhood of the point of minimum the process infinitely slows down because the gradient is small and the trajectory comes to the point of equilibrium in an infinite time.

Consider from this point of view the gradient method for the solution of problem (1.7). The process has the form

$$\frac{dx}{dt} = -\alpha(1+\nu)\frac{x}{|x|^{1-\nu}}, \qquad \alpha > 0, \quad x(t_0) = x^0. \tag{1.8}$$

We take the inner product of Eq. (1.8) by x and obtain

$$\frac{1}{2}\frac{d}{dt}|x|^2 = \left\langle \frac{dx}{dt}, x \right\rangle = -\alpha(1+\nu)|x|^{1+\nu}. \tag{1.9}$$

We differentiate the latter equation and cancel $|x|$ in both its sides to get

$$\frac{d}{dt}|x| + \alpha(1+\nu)|x|^\nu = 0. \tag{1.10}$$

We separate the variables and single out the total derivative:

$$\frac{1}{1-\nu}\frac{d}{dt}|x|^{1-\nu} + \alpha(1+\nu) = 0. \tag{1.11}$$

Let us integrate Eq. (1.11) from t_0 to t and find

$$|x|^{1-\nu} + \alpha(1-\nu^2)(t-t_0) = c_0. \tag{1.12}$$

If one of the positive term in Eq. (1.12) is increasing, then the other must be decreasing. When $t_f = t_0 + [\alpha(1-\nu^2)]^{-1}c_0$, then $|x|^{1-\nu}$ vanishes. This implies that the trajectory comes to the equilibrium state $x(t_f) = x^*$ at $t = t_f$. This example shows that in vector fields with the smooth minimum but sharp derivative at the point of minimum the equilibrium state is attained in a finite time.

The equilibrium state in this problems has advantages of a smooth solution (the equilibrium point is a fixed point) and those of a sharp solution (the equilibrium is attained in a finite time).

If we set $\nu = 0$ in Eq. (1.10), then we obtain the gradient method (1.2) for the solution of problem (1.1) with a sharp minimum, namely,

$$\frac{d}{dt}|x| + \alpha = 0. \tag{1.13}$$

Comparing Eq. (1.10) with Eq. (1.13), we see that outside any neighborhood of the origin $\{x : |x| \geq \varepsilon,\ x \in \mathbb{R}^n\}$ the value of $|x|^\nu$ is close to unity for small ν. This implies that the trajectory of Eq. (1.10) is close to that of Eq. (1.13), and converges in a finite time. Moreover, the fast drop of $|x|^\nu$ to zero in the ε-neighborhood of the point of minimum cannot make the process worse and it remains finite.

2. A SHARP MINIMUM AND A SMOOTH MINIMUM WITH A SHARP DERIVATIVE

Consider the minimization problem for a function on a convex set

$$x^* \in \operatorname{Argmin}\{f(x) : x \in Q\}, \tag{2.1}$$

where $f(x)$ is a convex goal function, $Q \subset \mathbb{R}^n$, $\mathbb{R}^n$ is the finite-dimensional Euclidean space, and Q is a convex closed set. Different necessary conditions, depending on the smoothness of the goal function, are satisfied at the point of minimum. If the function is nondifferentiable, then we can formulate these conditions using a proximal operator

$$x^* \in \operatorname{argmin}\{2^{-1}|z-x^*|^2 + \alpha f(z) : z \in Q\}. \tag{2.2}$$

For a differentiable goal function necessary conditions of minimum can be formulated using a projection operator

$$x^* = \pi_Q\{x^* - \alpha\nabla f(x^*)\}, \tag{2.3}$$

where $\nabla f(x^*)$ is the gradient of $f(x)$ at the point x^*, $\pi_Q\{\ldots\}$ is the projection operator taking a vector onto the set Q, and $\alpha > 0$ is a parameter like the step length.

We introduce the concept of a sharp minimum and a smooth minimum with a sharp derivative. Following [3], we define the *sharp minimum* as a point where the inequality

$$f(x) - f(x^*) \geq \gamma|x - x^*| \tag{2.4}$$

is satisfied for all $x \in Q$ and for a positive parameter γ. For example, the function $f(x) = |x|$ has a sharp minimum on $\mathbb{R}^n$.

Following [4], we extend the concept of a sharp minimum to include the set of sharp minima. The set of solutions to minimization problem is a set of *sharp minima* if the inequality

$$f(x) - f(\pi_{X^*}(x)) \geq \gamma |x - \pi_{X^*}(x)| \tag{2.5}$$

is satisfied for all $x \in Q$, $\gamma > 0$, where X^* is the set of solutions to the original problem, $\pi_{X^*}(x)$ is the operator of projection of a vector x onto the set X^*.

Let us now show that if a system of linear equations $Ax = b$ is solvable (i.e., there exists an x^* such that $Ax^* = b$), then the function $|Ax - b|$ satisfies Eq. (2.5). First, we consider the case of a nonsingular matrix, in which

$$\begin{aligned} |Ax - b| = |A(x - x^*)| &= \sqrt{|A(x-x^*)|^2} = \sqrt{\langle A^\top A(x-x^*), x - x^*\rangle} \geq \\ &\geq \sqrt{\mu |x - x^*|^2} = \sqrt{\mu}|x - x^*|. \end{aligned} \tag{2.6}$$

Here μ is the least eigenvalue of the nonsingular matrix $A^\top A$. If the matrix $A^\top A$ is singular, we decompose the space $\mathbb{R}^n$ into the direct sum $\mathbb{R}^n = H_1 + H_2$, where H_1 is the kernel of the matrix $A^\top A$ and H_2 is the orthogonal complement of H_1. In this case each vector $x - x^* \in \mathbb{R}^n$ has the representation $x - x^* = h_1 + h_2$, where $h_1 = \pi_{H_1}(x - x^*)$ and $h_2 = \pi_{H_2}(x - x^*)$ and, in addition, $A^\top A h_1 = 0$, $A^\top A h_2 \in H_2$. Taking this into account, we derive the bound (2.6) for the singular situation:

$$\begin{aligned} |Ax - b| &= \sqrt{\langle A^\top A(x-x^*), x - x^*\rangle} = \sqrt{\langle (A^\top A)^{1/2}(x - x^*), (A^\top A)^{1/2}(x-x^*)\rangle} = \\ &= \sqrt{\langle (A^\top A)^{1/2}(h_1 + h_2), (A^\top A)^{1/2}(h_1+h_2)\rangle} = \\ &= \sqrt{\langle (A^\top A)^{1/2}h_2, (A^\top A)^{1/2}h_2\rangle} = \sqrt{\langle A^\top A h_2, h_2\rangle} \geq \sqrt{\mu |h_2|^2} = \\ &= \sqrt{\mu}|h_2| = \sqrt{\mu}|x - x^* - h_1| = \sqrt{\mu}|x - x^* - \pi_{H_1}(x - x^*)| = \\ &= \sqrt{\mu}|x - x^* - \pi_{H_1}(x) + \pi_{H_1}(x^*)| = \sqrt{\mu}|x - \pi_{H_1}(x)|. \end{aligned} \tag{2.7}$$

Here μ is the minimal nonzero eigenvalue of the singular matrix $A^\top A$. When deriving the bound (2.7), we used the existence of a square root of the symmetric nonnegative matrix $A^\top A$ and the linearity of the projector operator $\pi_{H_1}(x - x^*)$, moreover $\pi_{H_1}(x^*) = x^*$.

Sharp minima occur in a lot of problem classes, e.g., in linear programming problems. For example, condition (2.5) for the problem

$$x^* \in \text{Argmin}\,\{\langle c, x\rangle : Ax \leq b, \ \ x \geq 0\} \tag{2.8}$$

becomes

$$\langle c, x\rangle - \langle c, \pi_{X^*}(x)\rangle \geq \gamma |x - \pi_{X^*}(x)| \tag{2.9}$$

for all x of the admissible set. It is proved [3] that inequality (2.9) holds. Some smooth problems of convex programming also satisfy the sharpness condition [5].

Let us introduce a new concept of a smooth minimum with a sharp derivative. The set of solutions to a minimization problem is called a *set of smooth minima with a sharp derivative at a point* if the inequality

$$f(x) - f(\pi_{X^*}(x)) \geq \gamma |x - \pi_{X^*}(x)|^{1+\nu} \tag{2.10}$$

is satisfied for all $x \in Q$, $\gamma > 0$, and $0 < \nu < 1$, where X^* is the set of solutions to the original problem (2.1) and $\pi_{X^*}(x)$ is the operator of projection of a vector x onto the set X^*.

Repeating the reasoning of (2.7), we can show that function $|Ax-b|^{1+\nu}$ satisfies inequality (2.10), namely,

$$|Ax-b|^{1+\nu} \geq \sqrt{(\mu/(1+\nu))}\,|x-\pi_{H_1}(x)|^{1+\nu} \tag{2.11}$$

for all $x \in Q$, provided the solvability condition for the equation $Ax = b$. Here μ is the minimal nonzero eigenvalue of the singular matrix $A^\top A$ and H_1 is its kernel.

If the function $f(x)$ is differentiable, then a smooth minimum with a sharp derivative can be defined by the following inequalities

$$\begin{aligned} f(x) - f(\pi_{X^*}(x)) - \langle \nabla f(\pi_{X^*}(x)), x - \pi_{X^*}(x)\rangle &\geq \gamma|x-\pi_{X^*}(x)|^{1+\nu}, && (2.12)\\ \langle \nabla f(x) - \nabla f(\pi_{X^*}(x)), x - \pi_{X^*}(x)\rangle &\geq \gamma|x-\pi_{X^*}(x)|^{1+\nu} && (2.13)\end{aligned}$$

for all $x \in Q$, $\gamma > 0$, and $0 < \nu < 1$. Taking into account the convexity of $f(x)$, it is easy to show that (2.12) yields (2.10) and (2.13).

Now we verify, for example, that inequality (2.13) is satisfied for the function of the form $f(x) = |Ax-b|^{1+\nu}$. We write its gradient $\nabla f(x) = (1+\nu)|Ax-b|^{\nu}A^\top(Ax-b)/|Ax-b| =$ $= (1+\nu)A^\top(Ax-b)/|Ax-b|^{1-\nu}$, then we use (2.11) and get

$$\begin{aligned} &\langle \nabla f(x) - \nabla f(\pi_{X^*}(x)), x - \pi_{X^*}(x)\rangle = (1+\nu)\left\langle \frac{A^\top(Ax-b)}{|Ax-b|^{1-\nu}}, x - x^*\right\rangle = \\ &= \frac{1+\nu}{|Ax-b|^{1-\nu}}\langle Ax-b, Ax-Ax^*\rangle = \frac{1+\nu}{|Ax-b|^{1-\nu}}|Ax-b|^2 = \\ &= (1+\nu)|Ax-b|^{1+\nu} \geq \sqrt{\mu(1+\nu)}\,|x-\pi_{H_1}(x)|^{1+\nu}\end{aligned} \tag{2.14}$$

for all $x \in Q$. Here μ is the minimal nonzero eigenvalue of the singular matrix $A^\top A$ and H_1 is its kernel.

3. CONVERGENCE OF THE PROXIMAL CONTINUOUS METHOD IN A FINITE TIME

We can consider a residual argmin $\{2^{-1}|z-x|^2 + \alpha f(z) : z \in Q\} - x$ as a transformation of $\mathbb{R}^n$ into itself. This transformation generates a vector field in the space $\mathbb{R}^n$. We start at any point x^0 and draw a trajectory along the directions of this field, which is described by the system of differential equations

$$\frac{dx}{dt} + x = \operatorname{argmin}\{2^{-1}|z-x|^2 + \alpha f(z) : z \in Q\} \quad x(t_0) = x^0. \tag{3.1}$$

Since the proximal operator is non-expanding, the trajectory $x(t)$ exists for all $t \geq t_0$.

Before we discuss asymptotic stability of method (3.1) for the solution of problem (2.1), let us prove a very useful inequality, used below in the proof of convergence of the proximal method, namely [6],

$$2^{-1}|z_x - x|^2 + \alpha f(z_x) \leq 2^{-1}|z-x|^2 + \alpha f(z) - 2^{-1}|z-z_x|^2, \tag{3.2}$$

where $z \in Q$ and z_x is the point of minimum of the function $\varphi_x(x) = 2^{-1}|z-x|^2 + \alpha f(z)$ on Q for a fixed vector x. This inequality can simply be proved. Indeed, let z_x be the point of minimum of $\varphi_x(x)$ on Q. Then, by the necessary condition, the subdifferential $\partial\varphi_x(z)$ at the point of minimum z_x contains a positive subgradient

$$\langle z_x - x + \alpha\nabla f(z_x), z - z_x\rangle \geq 0 \tag{3.3}$$

for all $z \in Q$. We combine the inequality

$$f(z) \geq f(z_x) + \langle \partial\varphi(z_x), z - z_x \rangle, \tag{3.4}$$

which implies that $f(z)$ is a convex function, and the identity

$$2^{-1}|z-x|^2 = 2^{-1}|z-z_x|^2 + \langle z - z_x, z_x - x \rangle + 2^{-1}|z_x - x|^2. \tag{3.5}$$

According to (3.3), we get inequality (3.2).

Let us represent Eq. (3.1) as an inequality (3.2), namely,

$$2^{-1}|x + \dot{x} - x|^2 + \alpha f(x + \dot{x}) \leq 2^{-1}|z - x|^2 + \alpha f(z) - 2^{-1}|x + \dot{x} - z|^2 \tag{3.6}$$

for all $z \in Q$. Now we show that process (3.1) allows us to find the solution to the original problem in a finite time $t = t_f$, moreover, the solution has the form $x(t_f) + \dot{x}(t_f) = x^* \in X^*$.

Theorem 1. *If the set of solutions to problem* (2.1) *is nonempty and satisfies sharpness condition* (2.5) *or* (2.10)*, the goal function $f(x)$ is convex, and the set Q is convex and closed, then the trajectory $x(t)$ of proximal process* (3.1) *with a parameter $\alpha > 0$ converges to the solution in a finite time, i.e., there exists t_f such that $x(t_f) + \dot{x}(t_f) = x^* \in X^*$.*

Proof. We put $z = x^*$ in (3.6) and obtain

$$2^{-1}|x + \dot{x} - x|^2 + \alpha f(x + \dot{x}) \leq 2^{-1}|x^* - x|^2 + \alpha f(x^*) - 2^{-1}|x + \dot{x} - x^*|^2, \tag{3.7}$$

where x^* is an arbitrary point of the solution set X^*. Hence,

$$2^{-1}|x - x^*|^2 + 2^{-1}|\dot{x}|^2 + \langle \dot{x}, x - x^* \rangle + 2^{-1}|\dot{x}|^2 + \alpha f(x + \dot{x}) \leq 2^{-1}|x - \dot{x}|^2 + \alpha f(x^*),$$

or

$$2^{-1}\frac{d}{dx}|x - x^*|^2 + |\dot{x}|^2 + \alpha\{f(x + \dot{x}) - f(x^*)\} \leq 0. \tag{3.8}$$

Since the third term in the left-hand side of (3.8) is nonnegative, we can integrate it from t_0 to t to get

$$|x - x^*|^2 + 2\int_{t_0}^{t} |\dot{x}|^2 d\tau \leq (x^0 - x^*)^2. \tag{3.9}$$

It follows from (3.9) that $\lim\limits_{t\to\infty} |\dot{x}(t)| = 0$. Indeed, if this were not the case, i.e., $|\dot{x}(t)|^2 \geq \varepsilon$ for all $t \geq t_0$, then we arrived at a contradiction with convergence of the integral. Hence there exists a subsequence of time moments $t_i \to \infty$ such that $|\dot{x}(t_i)| \to 0$. This implies

$$\lim_{t\to\infty} |\dot{x}(t)| = 0, \qquad t_i \to \infty. \tag{3.10}$$

We return to the inequality (3.8) and rewrite it as

$$\langle \dot{x}, x + \dot{x} - x^* \rangle + \alpha\{f(x + \dot{x}) - f(x^*)\} \leq 0 \tag{3.11}$$

for all $x^* \in X^*$. In particular, this inequality holds for $x^* = \pi_{X^*}(x + \dot{x})$, whence, because of sharpness condition (2.5) or (2.10), we derive

$$\langle \dot{x}, x + \dot{x} - \pi_{X^*}(x + \dot{x}) \rangle + \alpha\gamma|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{1+\nu} \leq 0,$$

where probably $\nu = 0$ (a sharp minimum). Therefore,

$$\alpha\gamma|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{1+\nu} \leq |\dot{x}|\,|x + \dot{x} - \pi_{X^*}(x + \dot{x})|. \tag{3.12}$$

Assume that $|x + \dot{x} - \pi_{X^*}(x + \dot{x})| \neq 0$ for all $t \geq t_0$, i.e., $x + x \neq \pi_{X^*}(x + x)$. Then (3.12) yields

$$\alpha\gamma|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{\nu} \leq |\dot{x}|. \tag{3.13}$$

In particular, if $\nu = 0$ (a sharp minimum), we have

$$\alpha\gamma \leq |\dot{x}|. \tag{3.14}$$

We compare inequalities (3.10) and (3.14) and arrive at a contradiction. If the parameter ν belongs to the interval $(0, 1)$, we transform inequality (3.13) as follows:

$$\alpha\gamma \leq |\dot{x}|^{1-\nu}\frac{|\dot{x}|^{\nu}}{|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{\nu}} = |\dot{x}|^{1-\nu}\left|\frac{x - x^*}{|\dot{x}|} + \frac{\dot{x}}{|\dot{x}|}\right|^{-\nu},$$

where $x^* = \pi_{X^*}(x+\dot{x})$. In the latter inequality the expression in parentheses is bounded and does not depend on the behavior of $(x - x^*)/|\dot{x}|$ as $t \to \infty$ and $|\dot{x}|^{1-\nu} \to 0$ as $t_i \to \infty$. Consequently, we also arrive at a contradiction.

To get out of this, we should not require the condition $|x+\dot{x}-\pi_{X^*}(x)(x+\dot{x})| \neq 0$ for all $t \geq t_0$. Consequently, the condition $x(t_f) + \dot{x}(t_f) = \pi_{X^*}(x(t_f) + \dot{x}(t_f)) = x^* \in X^*$ is satisfied with some t_f. In other words, process (3.1) allows us to find the solution to the problem in a finite time (in the sense of the quantity of $x(t_f) + \dot{x}(t_f)$). The theorem is proved. $\square$

4. CONVERGENCE OF THE CONTINUOUS METHOD OF GRADIENT PROJECTION IN A FINITE TIME

We consider the residual $\pi_Q(x - \alpha\nabla f(x)) - x$ as a transformation of the space $\mathbb{R}^n$ to $\mathbb{R}^n$. This transformation defines a vector field. We pose a problem to find a trajectory such that its tangent vector coincides with the field vector at the same point. Formally, this problem is described by the system of differential equations

$$\frac{dx}{dt} + x = \pi_Q(x - \alpha\nabla f(x)), \qquad x(t_0) = x^0. \tag{4.1}$$

The continuous right-hand side of system (4.1) provides existence and uniqueness of a trajectory on an infinite interval, i.e., for all $t \geq t_0$, as it follows from general theorems.

Recall that the operator $\pi_Q(b)$ projecting a vector b onto the set Q satisfies the inequality

$$\langle \pi_Q(b) - b, z - \pi_Q(b)\rangle \geq 0 \tag{4.2}$$

for all $z \in Q$.

We represent Eq. (4.1) as an inequality (4.2), namely,

$$\langle \dot{x} + x - (x - \alpha\nabla f(x)), z - \dot{x} - x\rangle \geq 0 \tag{4.3}$$

for all $z \in Q$.

We also represent the original problem (2.1) as a variational inequality

$$\langle \nabla f(x^*), z - x^*\rangle \geq 0 \tag{4.4}$$

for all $z \in Q$ and all $x^* \in X^*$.

We now cite a theorem on finite convergence of the gradient projection method, assuming that $f(x)$ is a convex function differentiable everywhere in its domain. Moreover, the gradient of this function satisfies the Hölder condition [7]

$$|\nabla f(y) - \nabla f(x)| \le L|y-x|^{\kappa} \tag{4.5}$$

for all x and y of some set, where L is a constant and $0 < \kappa \ge 1$. When $\kappa = 1$, we have the Lipschitz condition. The class of function with a sharp minimum whose gradients satisfy the Lipschitz condition is narrow. The Hölder condition allows us to consider a wider class.

Theorem 2. *Assume that the set of solutions to problem* (2.1) *is nonempty and satisfies the sharpness condition* (2.5) *or* (2.10), *$f(x)$ is a convex and differentiable function whose gradient satisfies Hölder condition* (4.5), *where $\kappa > \nu$, and Q is a convex closed set. Then the trajectory $x(t)$ of the gradient projection method* (4.1) *with a parameter $\alpha > 0$ converges to the solution to the problem in a finite time, i.e., there exists t_f such that $x(t_f) + \dot{x}(t_f) = x^* \in X^*$.*

Proof. We put $z = x^*$ in (4.3) and $z = x + \dot{x}$ in (4.4) and combine the resultants. Then we get

$$\langle \dot{x} + \alpha(\nabla f(x) - \nabla f(x^*), x^* - x - \dot{x})\rangle \ge 0 \tag{4.6}$$

for all $x^* \in X^*$. We represent (4.6) as

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

Since the gradient is monotonic, the latter inequality yields

$$\frac{d}{dt}\,|x - x^*|^2 + |\dot{x}|^2 + \alpha\,\frac{d}{dt}(f(x) - f(x^*) - \langle \nabla f(x^*), x - x^*\rangle) \le 0. \tag{4.7}$$

Let us integrate inequality (4.7) from t_0 to t. Since $f(x)$ is a convex function, i.e., $(f(x) - f(x^*) - -\langle \nabla f(x^*), x - x^*\rangle) \ge 0$, the latter inequality can be rewritten as

$$|x - x^*|^2 + \int_{t_0}^{t} |\dot{x}|^2 d\tau \le C_0. \tag{4.8}$$

As shown in Theorem 1, inequality (4.8) implies that

$$\lim_{t\to\infty} |\dot{x}(t)| = 0. \tag{4.9}$$

We return to the inequality (4.3), put in it $z = x^*$, and then rewrite it in the form

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

We add and subtract $\nabla f(x+\dot{x})$ to the left-hand side of this inequality and then set $x^* = \pi_{X^*}(x+\dot{x})$. This is possible because the inequality holds for all $x^* \in X^*$. As a result, we obtain

$$\langle \dot{x}, x + \dot{x} - \pi_{X^*}(x + \dot{x})\rangle + \alpha\langle \nabla f(x + \dot{x}) + \nabla f(x) - \nabla f(x + \dot{x}), x + \dot{x} - \pi_{X^*}(x + \dot{x})\rangle \le 0. \tag{4.10}$$

On the other hand, using the sharpness condition for the minimum (2.5) or (2.10) and convexity of the function $f(x)$ we write

$$\gamma|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{1+\nu} \le f(x + \dot{x}) - f(\pi_{X^*}(x + \dot{x})) \le \langle \nabla f(x + \dot{x}), x + \dot{x} - \pi_{X^*}(x + \dot{x})\rangle, \tag{4.11}$$

where probably $\nu = 0$ (a sharp minimum). We compare (4.10) and (4.11) and use the condition (4.5) to derive

$$\langle \dot{x}, x + \dot{x} - \pi_{X^*}(x + \dot{x})\rangle + \alpha\gamma|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{1+\nu} \leq \alpha L|\dot{x}|^{\kappa}|x + \dot{x} - \pi_{X^*}(x + \dot{x})|.$$

Therefore,

$$\alpha\gamma|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{1+\nu} \leq |\dot{x}|\,|x + \dot{x} - \pi_{X^*}(x + \dot{x})| + \alpha L|\dot{x}|^{\kappa}|x + \dot{x} - \pi_{X^*}(x + \dot{x})|. \qquad (4.12)$$

Assume that $|x + \dot{x} - \pi_{X^*}(x + \dot{x})| \neq 0$ for all $t \geq t_0$, then from (4.12) we have

$$\alpha\gamma|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{\nu} \leq |\dot{x}| + \alpha L|\dot{x}|^{\kappa}. \qquad (4.13)$$

In particular, if $\nu = 0$ (a sharp minimum), we obtain

$$\alpha\gamma \leq |\dot{x}| + \alpha L|\dot{x}|^{\kappa}. \qquad (4.14)$$

Comparing inequalities (4.9) and (4.14), we arrive at a contradiction. If the parameter ν belongs to the interval $(0, 1)$, then we transform (4.13) in the following way

$$\begin{aligned} \alpha\gamma &\leq |\dot{x}|^{1-\nu}\frac{|\dot{x}|^{\nu}}{|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{\nu}} + \alpha L|\dot{x}|^{\kappa-\nu}\frac{|\dot{x}|^{\nu}}{|x + \dot{x} - \pi_{X^*}(x + \dot{x})|^{\nu}} = \\ &= \left(|\dot{x}|^{1-\nu} + \alpha L|\dot{x}|^{\kappa-\nu}\right)\left|\frac{x - x^*}{|\dot{x}|} + \frac{\dot{x}}{|\dot{x}|}\right|^{-\nu}, \end{aligned}$$

where $x^* = \pi_{X^*}(x + \dot{x})$. The second factor in the obtained inequality is always bounded and does not depend on the behavior of $(x - x^*)/|\dot{x}|$ as $t \to \infty$. Since $|\dot{x}| \to 0$ as $t_i \to \infty$, we also arrive at a contradiction in this case.

The latter implies that our assumption is false, i.e., there exists a moment t_f such that $x(t_f) + \dot{x}(t_f) = \pi_{X^*}(x(t_f) + \dot{x}(t_f)) = x^* \in X^*$. In other words, process (4.1) allows us to find the solution to the problem in a finite time. Theorem is proved. □

The gradient method can be applied to optimize a convex function if this function is differentiable. In the general case, it is necessary to use the generalized differential equation

$$\frac{dx}{dt} \in \pi_Q(x - \alpha\nabla f(x)) - x, \qquad x(t_0) = x^0, \qquad (4.15)$$

to find the minimum, where $\nabla f(x)$ is any subgradient which belongs to the subdifferential. It is difficult to examine convergence of this process. We note works [8, 9] for more detail. The method of subgradient projection [3] is an iterative analog of (4.15). Advantages and disadvantages of the proximal method are clearly seen in the above example. The proximal method is more universal for complicated situations and that is why it is perspective for the solution of large-scale problems. The gradient method is efficient in simple (smooth) situations.

5. ITERATIVE PROCESSES

In this section we consider iterative analogs of the above continuous processes. It is necessary to emphasize that we consider iterative analogs rather than discrete approximations of continuous processes. There is an essential difference between them. The latter have a small step of discretization with respect to time and, therefore, they approximate continuous trajectories well. However, a small time step causes a low speed of the process outside a neighborhood of the solution to the problem. Processes with large time steps ($\Delta t = 1$) converge fast to a neighborhood of the solution, but their convergence should be established separately.

We first consider an iterative proximal method described by the recursion relations

$$x^{n+1} = \text{argmin}\, \{2^{-1}|z - x^n|^2 + \alpha f(z) : z \in Q\}. \tag{5.1}$$

Many authors investigated this approach applied in various situations. Of most interest are works by Rockafellar and Ferris [10, 4] who proved finite convergence of the proximal method for the minimization problem for a convex function with a sharp minimum. Let us prove this using a technique for deriving bounds we developed in a series of works.

We first represent process (5.1) in the form of inequality (3.2), namely,

$$2^{-1}|x^{n+1} - x^n|^2 + \alpha f(x^{n+1}) \le 2^{-1}|z - x^n|^2 + \alpha f(z) - 2^{-1}|z - x^{n+1}|^2 \tag{5.2}$$

for all $z \in Q$.

Theorem 3. *If the set of solutions to problem* (2.1) *is nonempty and satisfies the sharpness condition* (2.5) *or* (2.10)*, the goal function $f(x)$ is convex, and set Q is convex and closed, then the sequence x^n of the proximal process* (5.1) *with a parameter $\alpha > 0$ converges to the solution to the problem in a finite number of iterations, i.e., there exists a number n_f such that $x^{n_f+1} = x^* \in X^*$.*

Proof. We set $z = x^*$ in (5.2) to get

$$2^{-1}|x^{n+1} - x^n|^2 + \alpha f(x^{n+1}) \le 2^{-1}|x^* - x^n|^2 + \alpha f(x^*) - 2^{-1}|x^* - x^{n+1}|^2 \tag{5.3}$$

for all $z \in Q$, where x^* is an arbitrary point of the solution set X^*. We take into account the estimate $f(x^{n+1}) - f(x^*) \ge 0$ and sum up (5.3) from $n = 0$ to $n = N$ to obtain

$$|x^{N+1} - x^*|^2 + \sum_{k=0}^{k=N} |x^{k+1} - x^k|^2 \le |x^0 - x^*|^2. \tag{5.4}$$

The partial sums in the left-hand side of inequality (5.4) are bounded for all N. Consequently, the series $\sum_{k=0}^{\infty} |x^{k+1} - x^k|^2$ converges, which implies that

$$|x^{n+1} - x^n|^2 \to 0, \qquad n \to \infty. \tag{5.5}$$

Let us transform the sum of squares in (5.3) to an inner product using identity (3.5). We thus derive

$$\langle x^{n+1} - x^n, x^{n+1} - x^* \rangle + \alpha(f(x^{n+1}) - f(x^*)) \le 0.$$

Since the latter inequality holds for all $x^* \in X^*$, we put $x^* = \pi_{X^*}(x^{n+1})$ and use the sharpness condition for minimum (2.5) or (2.10), i.e.,

$$\langle x^{n+1} - x^n, x^{n+1} - \pi_{X^*}(x^{n+1}) \rangle + \alpha\gamma |x^{n+1} - \pi_{X^*}(x^{n+1})|^{1+\nu} \le 0,$$

where the parameter ν is probably zero (a sharp minimum). As a result we get

$$\alpha\gamma |x^{n+1} - \pi_{X^*}(x^{n+1})|^{1+\nu} \le |x^{n+1} - x^n|\, |x^{n+1} - \pi_{X^*}(x^{n+1})|. \tag{5.6}$$

Assume that $|x^{n+1} - \pi_{X^*}(x^{n+1})| \ne 0$ for all n, then inequality (5.6) yields

$$\alpha\gamma |x^{n+1} - \pi_{X^*}(x^{n+1})|^{\nu} \le |x^{n+1} - x^n|. \tag{5.7}$$

In particular, when $\nu = 0$ (a sharp minimum), we obtain

$$\alpha\gamma \le |x^{n+1} - x^n|. \tag{5.8}$$

Clearly, inequalities (5.5) and (5.8) contradict each other.

If the parameter ν belongs to the interval $(0, 1)$, we transform (5.7) in the following way:

$$\alpha\gamma \le |x^{n+1} - x^n|^{1-\nu} \frac{|x^{n+1} - x^n|^\nu}{|x^{n+1} - \pi_{X^*}(x^{n+1})|^\nu} = |x^{n+1} - x^n|^{1-\nu} \left| \frac{x^n - \pi_{X^*}(x^{n+1})}{|x^{n+1} - x^n|} + \frac{x^{n+1} - x^n}{|x^{n+1} - x^n|} \right|^{-\nu}.$$

The expression in brackets in this inequality is bounded and does not depend on the behavior of $(x^n - \pi_{X^*}(x^{n+1}))/|x^{n+1} - x^n|$ as $n \to \infty$, moreover $|x^{n+1} - x^n|^{1-\nu} \to 0$ as $n \to \infty$. We also arrive at a contradiction in this case.

Consequently, the assumption $|x^{n+1} - \pi_{X^*}(x^{n+1})| \ne 0$ for all n is false, and there exists a number n_f such that $x^{n_f+1} = \pi_{X^*}(x^{n_f+1}) = x^* \in X^*$. The theorem is proved. $\square$

Now we investigate the behavior of the iterative method of gradient projection

$$x^{n+1} = \pi_Q(x^n - \alpha\nabla f(x^n)), \qquad x^0 \in \mathbb{R}^n, \tag{5.9}$$

applied to problems with a sharp minimum. We write this method in the variational form as

$$\langle x^{n+1} - x^n + \alpha\nabla f(x^n), z - x^{n+1}\rangle \ge 0 \tag{5.10}$$

for all $z \in Q$.

Let us formulate a theorem on finite convergence of the gradient projection method, assuming that $f(x)$ is a convex function differentiable everywhere in its domain whose gradient satisfies the Lipschitz condition. A special case of this theorem was proved in [3].

Theorem 4. *Assume that the set of solutions to problem* (2.1) *is nonempty and satisfies the sharpness condition* (2.5)*, $f(x)$ is a convex differentiable function whose gradient satisfies the Lipschitz condition with a constant L, and Q is a convex closed set. Then the sequence x^n of the gradient projection method* (5.9) *with a parameter $\alpha < 2/L$ converges to the solution to the problem in a finite number of iterations, i.e., there exists a number n_f such that $x^{n_f+1} = x^* \in X^*$.*

Proof. We set $z = x^*$ in (5.10) and $z = x^{n+1}$ in (4.4) and sum up both inequalities to find

$$\langle x^{n+1} - x^n + \alpha(\nabla f(x^n) - \nabla f(x^*)), x^* - x^{n+1}\rangle \ge 0. \tag{5.11}$$

Hence,

$$\langle x^{n+1} - x^n, x^{n+1} - x^*\rangle + \alpha\langle\nabla f(x^n) - \nabla f(x^*), x^{n+1} - x^*\rangle \le 0. \tag{5.12}$$

We transform the first term in (5.12) using identity (3.5) and transform the second one using the inequality [11]

$$\langle\nabla f(x_1) - \nabla f(x_3), x_3 - x_2\rangle \le (L/4)|x_1 - x_2|^2, \tag{5.13}$$

which holds for all x_1, x_2, and x_3 in Q, where L is the Lipschitz constant. Then

$$|x^{n+1} - x^*|^2 + d|x^{n+1} - x^n|^2 \le |x^n - x^*|^2, \tag{5.14}$$

where $d = (1 - \alpha L/2) > 0$ because $\alpha < 2/L$. Let us sum up inequality (5.14) from $n = 0$ to $n = N$. We thus obtain

$$|x^{N+1} - x^*|^2 + d\sum_{k=0}^{N} |x^{k+1} - x^k|^2 \le |x^0 - x^*|^2. \tag{5.15}$$

The partial sums in the left-hand side of (5.15) are bounded for all N. Consequently, the series $\sum\limits_{k=0}^{\infty} |x^{k+1} - x^k|^2$ converges, whence we conclude that $|x^{k+1} - x^k|^2 \to 0$ as $n \to \infty$.

We again consider inequality (5.10) and set $z = x^*$. The inequality becomes $\langle x^{n+1} - x^n, x^{n+1} - x^* \rangle + \alpha\langle \nabla f(x^n), x^{n+1} - x^* \rangle \leq 0$. Therefore,

$$\langle x^{n+1} - x^n, x^{n+1} - x^* \rangle + \alpha\langle \nabla f(x^{n+1}) + \nabla f(x^n) - \nabla f(x^{n+1}), x^{n+1} - x^* \rangle \leq 0. \tag{5.16}$$

On the other hand, condition (2.5) and convexity of $f(x)$ yield

$$\gamma|x^{n+1} - \pi_{X^*}(x^{n+1})| \leq f(x^{n+1}) - f(\pi_{X^*}(x^{n+1})) \leq \langle \nabla f(x^{n+1}), x^{n+1} - \pi_{X^*}(x^{n+1}) \rangle. \tag{5.17}$$

Since inequality (5.16) holds for all $x^* \in X^*$, we put $x^* = \pi_{X^*}(x)(x^{n+1})$ and compare (5.16) and (5.17). According to the Lipschitz condition we obtain

$$\alpha\gamma|x^{n+1} - \pi_{X^*}(x^{n+1})| \leq (1 + \alpha L)|x^{n+1} - x^n|\,|x^{n+1} - \pi_{X^*}(x^{n+1})|. \tag{5.18}$$

Assuming that $|x^{n+1} - \pi_{X^*}(x^{n+1})| \neq 0$ for all n in (5.18), we deduce

$$\alpha\gamma \leq (1 + \alpha L)|x^{n+1} - x^n|. \tag{5.19}$$

Since $|x^{k+1} - x^k|^2 \to 0$ as $n \to \infty$, inequality (5.19) leads to a contradiction.

Consequently, the assumption $|x^{n+1} - \pi_{X^*}(x^{n+1})| \neq 0$ for all n is false, and there exists a number n_f such that $x^{n_f+1} = \pi_{X^*}(x^{n_f+1}) = x^* \in X^*$. The theorem is proved. $\square$

The author is grateful to F.P. Vasiliev for a fruitful discussion of this work.

LITERATURE CITED

1. F.P. Vasiliev, *Methodes of Solving Extremal Problems* [in Russian], Moscow (1981).

2. F.P. Vasiliev, *Zh. Vych. Mat. Mat. Fiz.*, **27**, No.10, 1443–1450 (1987).

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

4. M.C. Ferris, *Mathematical Programming*, **50**, 359–366 (1991).

5. E.S. Levitin, *Perturbation Theory in Mathematical Programming and Its Applications* [in Russian], Moscow (1992).

6. A.S. Antipin, *Zh. Vych. Mat. Mat. Fiz.*, **27**, No.3, 368–376 (1987).

7. Yu.E. Nesterov, *Effective Methods in Nonlinear Programming* [in Russian], Moscow (1988).

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

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

10. R.T. Rockafellar, *SIAM Journal on Control and Optimization*, **14**, 877–898 (1976).

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