sciendo
Control and Cybernetics
vol. **50** (2021) No. 2
pages: 301-314
DOI: 10.2478/candc-2021-0015

A locally polynomial method for solving a system of linear inequalities*

by

Yuri Evtushenko [c], **Krzysztof Szkatuła** [a,b,*] **and Alexey Tretyakov** [a,b,c]

[a] Siedlce University of Natural Sciences and Humanities,
Faculty of Science, ul. Konarskiego 2, 08-110 Siedlce, Poland
[b] Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01-447 Warszawa, Poland
[c] Dorodnicyn Computing Centre of FRC CSC, Russian Academy of Sciences
Vavilov st. 40, 119333 Moscow, Russia
* Corresponding author: Krzysztof.Szkatula@ibspan.waw.pl

Abstract: The paper proposes a method for solving systems of linear inequalities. This method determines in a finite number of iterations whether the given system of linear ineqalities has a solution. If it does, the solution for the given system of linear inequalities is provided. The computational complexity of the proposed method is locally polynomial.

Keywords: linear programming, system of linear inequalities, computational complexity, locally-polynomial algorithm, convergence rate

1. Introduction

Let us consider a system of linear inequalities:

$$A \cdot x - b \leq 0_m, \tag{1}$$

with A is an $m \times n$ matrix, $A = \{a_{ij}\}$, $x \in R^n$, $x = \{x_j\}$, $b \in R^m$, $b = \{b_i\}$, where $i = 1, \ldots, m$; $j = 1, \ldots n$; 0_m being an m dimensional vector of zeroes.

The goal of this paper is to establish whether there exists a solution for (1) attainable in a finite number of iterations, with reasonable computational effort. It will be determined whether the set of solutions:

$$X = \{x \in R^n \mid A \cdot x - b \leq 0_m\} \tag{2}$$

*Submitted: April 2021; Accepted: June 2021.

is empty or not. In the case it is not empty, at least one solution of (1) will be established. The significant achievement of the paper is that the set of violated constraints is locally ordered, which results in the locally polynomial time complexity. In the general case, for all possible combinations of constraints, the computational complexity is exponential in the worst case.

Solving systems of linear inequalities is important from both theoretical and practical points of view. Systems of linear inequalities are often used for modelling and solving complex practical problems from very different domains, e.g., having economical or technological origins, and many others. The well-known linear programming problem is reduced to a system of linear inequalities, see Tretyakov (2010). Therefore, it is very important to first establish whether a given system of linear inequalities has a non-empty set of solutions X and, if so, to find at least one of them, $x \in X$. There is a well-known list of eighteen unsolved problems in mathematics that was presented by Smale in 1998. The problem of finding strongly-polynomial time algorithm which decides whether there exists a solution of (1) or, equivalently, whether the set X in (2) is non empty, constitutes the 9th of the Smale's problems and is still not solved. The ellipsoid methods and interior-point techniques are providing algorithms of weakly-polynomial time complexity.

In this paper, the method for establishing whether there exists a solution for (1) is proposed, and moreover, the number of iterations (equivalent to computational complexity), with respect to m and n, is locally polynomial and in the worst case it has a geometric convergence rate.

Let us define the set of pseudo-solutions of (1) as follows:

$$X^* = \left\{ x^* \middle| x^* = \arg\min_{x \in R^n} \left\| (A \cdot x - b)_+ \right\|^2 \right\}, \text{ where } c_+ = \max\{c, 0\}. \quad (3)$$

If some point, sufficiently close to the set X^* of solutions of (3) is known, then it is possible to a find a pseudo-solution of (1) in the polynomial number of computational iterations of the order of $O(m^3 \cdot n^3)$. It should be emphasized that the solution for (3) always exists and when $X \neq \varnothing$, it will be a solution for (1).

Many methods for solving (1) have been proposed (see Karmanov, 1989; Golikov and Evtushenko, 2003; Evtushenko and Golikov, 2003; Tretyakov, 2010; Tretyakov and Tyrtyshnikov, 2013, or Han, 1980) . All of those methods have reasonable computational complexity but, as mentioned above, up to date, no strongly-polynomial time algorithm for solving (1) was proposed. In Tretyakov and Tyrtyshnikov (2013) and in Mangasarian (2001) linear programming problems are solved by reducing them to the unconditional minimization of strongly convex piecewise quadratic function. A solution will be obtained in the finite polynomial number of iterations if the starting point of the algorithm belongs to the sufficiently close neighborhood of the unique solution of the problem. Unfortunately, there are severe limitations imposed on the function to be minimized. Namely, it should be strongly convex and the eigenvalues of the Hessian matrix should satisfy specific conditions, etc.

This results in substantial limitations on the classes of problems, which could be solved, e.g. it is required that (1) has only unique solution etc. The solution method, described in Tretyakov and Tyrtyshnikov (2013) is based on exploiting information on the problem being solved by analyzing sufficiently small neighborhood of an arbitrary solution of (1). Analogous methods were proposed in Facchinei, Fischer and Kanzow (1998) for the forecast (identification) of the active constraints in the sufficiently close neighborhood of the solution of the problem. In the papers by Tretyakov and Tyrtyshnikov (2015) and Wright (2005), locally polynomial methods for solving quadratic programming problems, based on the similar ideas, are presented. It was proved in Goffin (1982) that the well-known ellipsoid method is not polynomial in the worst case. Tretyakov (2010) proposed the gradient projection method for solving (1); this method is finding solution of (1) in the finite number of iterations and is a combination of iterational and straightforward (e.g. Gauss) methods.

The present paper proposes a computational method, which establishes the existence of a solution to (1) and finds it, if the solution exists. When the starting point for the proposed method is sufficiently close to the set X^*, i.e. the set of pseudo-solutions for (1), as defined in (3), then its computational complexity is locally polynomial, namely of the order $O(m^3 \cdot n^3)$.

Finding such a starting point from the sufficiently small neighbourhood of the set X^* is guaranteed by the gradient descent method with the special step choice when it minimizes the convex function $\varphi(x)$ with the Lipschitz gradient i.e.

$$\|\varphi'(x) - \varphi'(y)\| \leq L \cdot \|x - y\|,$$

where L is the Lipschitz constant. For the iterative sequence $\{x_k\}$, fulfilling the montonicity condition i.e.

$$\forall x^* \in X^* : \ \|x_{k+1} - x^*\| \leq \|x_k - x^*\|, \ k = 0, 1 \ldots,$$

the above is guaranteeing the convergence to x^* from any arbitrary point $x_0 \in R^n$ and getting into the neighbourhood X^* at a certain step $\overline{k}$.

The proposed method is working for the general formulation of (1); there are no restrictive assumptions made.

The rest of the paper is organized as follows: In Section 2 necessary mathematical apparatus, important definitions and theoretical results are presented, in Section 3 the idea of the monotonic gradient method with the special step choice is described, Section 4 is presenting main results of the paper and Section 5 is discussing some additional issues.

2. Definitions and theoretical results

Let

$$\varphi(x) = \frac{1}{2} \cdot \|(A \cdot x - b)_+\|^2 \text{, where } c_+ = \max\{c, 0\}. \tag{4}$$

Theorem 1 *Function $\varphi(x)$ is convex and has a non-empty set of minimal values*

$$X^* = \left\{ x^* \in R^n \;\middle|\; \varphi(x^*) = \min_{x \in R^n} \varphi(x) \right\}. \tag{5}$$

Proof. Theorem 1 follows immediately from the well-known features of the quadratic type convex functions. ∎

It is obvious that elements $x^* \in X^*$, cf. (5), will satisfy:

$$\varphi'(x^*) = \sum_{i=1}^{m} (\langle a_i, x^* \rangle - b_i)_+ \cdot a_i = 0_n = A^T \cdot (A \cdot x^* - b)_+, \tag{6}$$

where a_i^T is the $i-$th row of the matrix A, $i = 1, ..., m$. Therefore, in the general case, our goal is to solve the following equation:

$$\varphi'(x) = \sum_{i=1}^{m} (\langle a_i, x \rangle - b_i)_+ \cdot a_i = A^T \cdot (A \cdot x - b)_+ = 0_n, \text{ where } x \in R^n. \tag{7}$$

Let $x^* \in X$. It is obvious that if $A \cdot x^* - b \leq 0_m$ holds, then $X \neq \varnothing$. Otherwise, if $A \cdot x^* - b \nleq 0_m$, then $X = \varnothing$. Let us denote:

$$f_i(x) = \langle a_i, x \rangle - b_i, \; i \in D = \{1, \dots m\},$$

and

$$\begin{aligned} &J_0(x) = \{\, i \in D \mid f_i(x) = 0 \,\}, J_-(x) = \{\, i \in D \mid f_i(x) < 0 \,\}, \\ &J_+(x) = \{\, i \in D \mid f_i(x) > 0 \,\}, \end{aligned} \tag{8}$$

where $f_i(x)$ is introduced to simplify the definitions of the sets $J_0(x)$ and $J_+(x)$.

According to (6) and the above notations, $x^* \in X$ should satisfy the following equation.

$$\sum_{i \in J_0(x^*) \cup J_+(x^*)} (\langle a_i, x^* \rangle - b_i)_+ \cdot a_i = 0_n. \tag{9}$$

This, in turn, means that in the general case we should solve the following equations:

$$\sum_{i \in J_0(x) \cup J_+(x)} (\langle a_i, x \rangle - b_i)_+ \cdot a_i = 0_n, \tag{10}$$

or

$$\sum_{i \in J_+(x)} \left(\langle a_i, x\rangle - b_i\right)_+ \cdot a_i = 0_n, \tag{11}$$

since

$$\langle a_i, x\rangle - b_i = 0,\ i \in J_0(x).$$

Without loss of generality we may denote:

$$J_-(x^*) = \{1, \ldots, l\},\ J_0(x^*) = \{l+1, \ldots, p\},\ J_+(x^*) = \{p+1, \ldots, m\},$$

where $l \leq p \leq m$.

If the rank of a matrix B of size $r \times n$ is equal to r, then the pseudo inverse matrix (operator) B^+ may be defined as $B^+ = B^T \cdot (B \cdot B^T)^{-1}$. We will denote the quadratic matrix $n \times n$, orthogonally projected on the space of rows of matrix B as $\left(B^T\right)^{\shortparallel} = B^T \left(B \cdot B^T\right)^{-1} \cdot B = B^+ \cdot B$, and projection on the orthogonal complement as $\left(B^T\right)^{\perp} = I - \left(B^T\right)^{\shortparallel}$, where I is the identity matrix of size n. The main idea exploited in this paper is based on the following Lemma.

Lemma 1 *If for every $\varepsilon > 0$ there exists $x \in U_\varepsilon(x^*)$ such that $f_i(x) \geq 0$, then $f_i(x^*) \geq 0$.*

Proof. The statement of the Lemma immediately follows from the obvious fact that if $f_i(x^*) < 0$, then there is an $\varepsilon > 0$ such that $f_i(x) < 0$ for all $x \in U_\varepsilon(x^*)$, according to the continuity property of the function $f_i(x)$. ■

Due to the above, in the sufficiently small neighborhood of some fixed point $x^* \in X^*$, for every $\bar{x} \in U_\varepsilon(x^*)$, the following will hold

$$J_0(\bar{x}) \subseteq J_0(x^*) \text{ and } J_+(\bar{x}) \subseteq J_0(x^*) \cup J_+(x^*), \quad J_-(\bar{x}) \subseteq J_0(x^*) \cup J_-(x^*).$$

Conditions, which should be satisfied at the point x^* are as follows:

$$\begin{gathered} \sum_{i \in J_0(x^*) \cup J_+(x^*)} \left(\langle a_i, x^*\rangle - b_i\right)_+ \cdot a_i = \sum_{i \in J_0(x^*) \cup J_+(x^*)} \left(\langle a_i, x^*\rangle - b_i\right) \cdot a_i = 0_n, \\ \langle a_i, x^*\rangle - b_i < 0, \quad i \in J_-(x^*). \end{gathered} \tag{12}$$

In (12), it is taken into account that

$$\left(\langle a_i, x^*\rangle - b_i\right)_+ = \langle a_i, x^*\rangle - b_i, \quad i \in J_0(x^*) \cup J_+(x^*).$$

Now, our goal is to correctly define the sets $J_0(x^*)$, $J_+(x^*)$, based on the information acquired at the point $\bar{x} \in U_\varepsilon(x^*)$. Let us denote

$$\bar{J}_0(\bar{x}) := J_0(\bar{x}),\ \bar{J}_+(\bar{x}) := J_+(\bar{x}),\ \bar{J}_-(\bar{x}) := J_-(\bar{x}).$$

Let $\bar{x} \in R^n$:

$$M(\bar{x}) = \left\{ \sum_{i \in \overline{J}_0(\bar{x}) \cup \overline{J}_+(\bar{x})} (\langle a_i, x\rangle - b_i) \cdot a_i = 0_n, \;\; \langle a_j, x\rangle - b_j = 0, j \in \bar{J}_0(\bar{x}) \right\}. \tag{13}$$

Let point $z(\bar{x})$ be the projection of point $\bar{x}$ on the set $M(\bar{x})$. Let us observe that $x^* \in M(\bar{x})$ if $\bar{x} \in U_\varepsilon(x^*)$ and ε is sufficiently small.

Moreover, if at the point $z(\bar{x})$ the constraints $f_i(z(\bar{x})) \leq 0$ for a certain $i \in \overline{J}_+(\bar{x})$, then we will define the set I_- in the following way:

$$I_- = \left\{ i \in \bar{J}_+(\bar{x}) \mid f_i(z(\bar{x})) \leq 0 \right\}.$$

Otherwise, if at the point $z(\bar{x})$ the constraints $f_i(z(\bar{x})) \geq 0$ for a certain $i \in \overline{J}_-(\bar{x})$, we will define the set I_+ in the analogous way:

$$I_+ = \left\{ i \in \overline{J}_-(\bar{x}) \mid f_i(z(\bar{x})) \geq 0 \right\}.$$

It will always hold that $I_- \subseteq J_0(x^*)$ and $I_+ \subseteq J_0(x^*)$. It may also happen that the set I_- or I_+ will be empty.

Now, we will redefine $\overline{J}_0(\bar{x})$, $\overline{J}_+(\bar{x})$ and $\overline{J}_-(\bar{x})$ as follows:

$$\overline{J}_0(\bar{x}) := \overline{J}_0(\bar{x}) \cup I_- \cup I_+, \quad \overline{J}_+(\bar{x}) := \overline{J}_+(\bar{x}) \setminus I_-, \quad \overline{J}_-(\bar{x}) := \overline{J}_-(\bar{x}) \setminus I_+. \tag{14}$$

Next, we will again project point $\bar{x}$ on the new set $M(\bar{x})$, see (13), and the new point $z(\bar{x})$ will be obtained.

We will use this stage as the basic one in the Main Recursive Step of the Algorithm 1 when altering the sets $\overline{J}_0$, $\overline{J}_-$ and $\overline{J}_+$.

Let $A(\bar{x})$ and $b(\bar{x})$ denote the matrix and the vector obtained in this way from A and b, respectively. The rows of $A(\bar{x})$ and the coefficients of $b(\bar{x})$ correspond to the index set, defined by $\overline{J}_0(\bar{x}) \cup \overline{J}_+(\bar{x})$. In this case, equations (10)-(11) may be rewritten as:

$$\begin{aligned} &A^T(\bar{x}) \cdot (A(\bar{x}) \cdot x - b(\bar{x})) = 0_n \qquad (15) \\ &\langle a_i, x\rangle - b_i = 0, \; i \in \overline{J}_0(\bar{x}). \end{aligned}$$

Let $\bar{A}(\bar{x})$ denote the matrix in the equations in (15), corresponding to the maximum set of linearly independent rows and let $\bar{b}(\bar{x})$ denote the corresponding vector of constant terms in (15).

Equations in (15) may be formulated in the following way:

$$\bar{A}(\bar{x}) \cdot x - \bar{b}(\bar{x}) = 0_n. \tag{16}$$

Let:

$$z(x) = P_{M(\bar{x})}(x) = \left(\bar{A}^T(\bar{x})\right)^{\perp} \cdot x + \bar{A}^{+}(\bar{x}) \cdot \bar{b}(\bar{x}) \tag{17}$$

define the operator of the projection of point x on the set $M(\bar{x})$.

Let us observe that at the point x^* the following holds

$$A^T(x^*) \cdot (A(x^*) \cdot x^* - b(x^*))_+ = 0_n, \tag{18}$$

which, in turn, means that:

$$\bar{A}(x^*) \cdot x^* - \bar{b}(x^*) = 0_n. \tag{19}$$

3. Gradient method with special step choice

Let us consider the following problem

$$\min_{x \in R^n} \varphi(x). \tag{20}$$

The gradient scheme, solving (20) and ensuring monotonicity of the minimizing sequence, is as follows

$$x_{k+1} = x_k - \alpha \cdot \varphi'(x_k),\ \alpha = 1/L,\ x_0 \in R^n,\ k = 0, 1 \ldots, \tag{21}$$

where x_0 is an arbitrary point in R^n. Let

$$X^* = \arg\min_{x \in R^n} \varphi(x),$$

where $X^* \neq \varnothing$ and X^* could be, in the general case, unbounded.

Theorem 2 *Let $\varphi(x)$ be a convex function and $\varphi(x) \in C^{1,1}(R^n)$, i.e.*

$$\forall u, v \in R^n : \|\varphi'(u) - \varphi'(v)\| \leq L \cdot \|u - v\|,$$

where L is the Lipschitz constant and $\alpha = 1/L$. Then, for the gradient scheme (21), the following holds

$$\|x_{k+1} - y\| \leq \|x_k - y\|,\ \forall y \in X^*,\ k = 0, 1 \ldots, \tag{22}$$

and

$$x_k \to x^*,\ x^* \in X^*, k \to \infty.$$

Proof. For convex function $\varphi(x)$ and $\forall y \in X^*$ we have:

$$\langle \varphi'(x_k), x_k - y\rangle \geq \varphi(x_k) - \varphi(y) \geq \varphi(x_k) - \varphi(x_{k+1}) \geq \frac{\alpha}{2} \cdot \|\varphi'(x_k)\|^2, \quad (23)$$

where for $\varphi(x) \in C^{1,1}(R^n)$ the following holds:

$$\varphi(z) - \varphi(\xi) \geq \langle \varphi'(z), z - \xi\rangle - \frac{L}{2} \cdot \|z - \xi\|^2 \text{ and}$$
$$\varphi(x_k) - \varphi(x_{k+1}) \geq \frac{\alpha}{2} \cdot \|\varphi'(x_k)\|^2 .$$

Hence, from (23)

$$-2 \cdot \alpha \cdot \langle \varphi'(x_k), x_k - y\rangle + \alpha^2 \cdot \|\varphi'(x_k)\|^2 \leq 0$$

and therefore for $k = 0, 1 \ldots$

$$\|x_{k+1} - y\|^2 = \|x_k - y\|^2 - 2 \cdot \alpha \cdot \langle \varphi'(x_k), x_k - y\rangle + \alpha^2 \cdot \|\varphi'(x_k)\|^2 \leq \|x_k - y\|^2 . \quad (24)$$

The above means that both $\{\|x_k - y\|\}$ and $\{\|x_k\|\}$ are bounded, and therefore

$$\exists\ x^* = \lim_{j\to\infty} x_{k_j} \in X^*.$$

Hence

$$\varphi'(x^*) = \lim_{j\to\infty} \varphi'(x_{k_j}) = 0 \text{ and } \varphi(x_k) - \varphi(x_{k+1}) \geq \frac{\alpha}{2} \cdot \|\varphi'(x_k)\|^2, k = 0, 1 \ldots .$$

But

$$\lim_{k\to\infty} x_k = \lim_{j\to\infty} x_{k_j} \text{ and hence } \lim_{k\to\infty} x_k = x^* \in X^*.$$

■

Remark 3 *Property*

$$\|x_{k+1} - y\| \leq \|x_k - y\|, \forall\ y \in X^*$$

is called monotonicity of sequence $\{x_k\}$. This means that for the gradient method (21) there will always be a point $x_{\bar{k}}$, in the arbitrary small neighbourhood of a certain solution $x^* \in X^*$. This property will be significantly exploited in Algorithm 2 providing a solution of the (1).

4. Algorithm for finding the pseudo-solution of (1)

In this section, the algorithm designed to find the pseudo-solution for (1) is presented. The main idea of this algorithm is based on information, related to a current point $\bar{x}$, belonging to the sufficiently small neighborhood of the point $x^* \in X^*$. We will also show how to find such a point $z(\bar{x}) \in X^*$.

Algorithm 1

Initialization Step: For the current point $\bar{x}$, the sets of indices $J_0(\bar{x})$, $J_-(\bar{x})$ and $J_+(\bar{x})$ will be defined according to (8). If the set $J_+(\bar{x}) = \varnothing$, then $\bar{x}$ is the solution of (1) and Algorithm 1 is terminated. Otherwise, the **Main Recursive Step** will be performed.

Main Recursive Step: Let $z(\bar{x})$, the projection of point $\bar{x}$ on the set $M(\bar{x})$, be defined according to (17). We will check if the following condition is satisfied:

$$I_+ = \varnothing \text{ and } I_- = \varnothing. \tag{25}$$

Checking Step: If (25) holds, then $z(\bar{x}) \in X^*$, equation (9) is satisfied; $z(\bar{x})$ is the pseudo-solution of (1), as defined in (3), and Algorithm 1 is terminated. Otherwise, if for certain $i \in D$ the condition (25) is violated and $i \in I_+ \cup I_-$, we will define $\overline{J}_0(\bar{x})$, $\overline{J}_+(\bar{x})$ and $\overline{J}_-(\bar{x})$ according to (14), and $M(\bar{x})$ will be redefined according to (13), and the **Main Recursive Step** will be repeated.

Set D is finite, $|D| = m$, and therefore the number of changes in index sets $\overline{J}_0(\bar{x})$, $\overline{J}_+(\bar{x})$ and $\overline{J}_-(\bar{x})$ does not exceed m, and finally the point $z(\bar{x})$, fulfilling (10), will be established. This means that $z(\bar{x})$ is the pseudo-solution of (1), as defined in (3), $z(\bar{x}) \in X^*$. We will repeat the projection procedure no more than m times as this will be sufficient for finding the point $z(\bar{x}) \in X^*$.

It is of utmost importance that $\bar{x}$ should belong to the sufficiently small neighborhood of the point x^*, because otherwise $z(\bar{x})$ may not satisfy (10). If this is not the case, it is necessary to find another point $\bar{x}$ that is closer to x^*. The way we accomplish this is described below.

Theorem 4 *For sufficiently small $\varepsilon > 0$ and for every $\bar{x} \in U_\varepsilon(x^*)$ Algorithm 1 provides $z^* = z(\bar{x})$ as the solution for*

$$\varphi'(x) = A^T(x) \cdot (A(x) \cdot x - b)_+ = 0_n, \tag{26}$$

which is equivalent to finding the solution for (10) in the number iterations of the order $0(m^3 \cdot n^3)$.

Proof. Proof is based on the observation that for $\bar{x}$ belonging to a sufficiently small neighborhood of the point x^*, the constraints $f_i(\bar{x}) \geq 0$, according to Lemma 1, will correspond to constraints $f_i(x^*) \geq 0$. Therefore

$$\overline{J}_0(\bar{x}) \cup \overline{J}_+(\bar{x}) \subseteq J_0(x^*) \cup J_+(x^*).$$

Let us determine $z(\bar{x})$ as the projection of the point $\bar{x}$ on the set $M(\bar{x})$, defined according to (13). It may happen that the set $\overline{J}_0(\bar{x})$ will be enlarged. However, the number of iterations, in which $\overline{J}_0(\bar{x})$ may be enlarged does not exceed m, the number of elements of the set D. Therefore, at some iteration, (25) will be satisfied. This means that $z(\bar{x})$ satisfies (10) or, equivalently, $\varphi'(z(\bar{x})) = 0_n$. This demonstrates that $z(\bar{x})$ is the pseudo-solution for (1), as defined in (3). The computational complexity of establishing each projection $z(\bar{x})$ is of order $0(m^2 \cdot n^3)$, taking into account the computational effort, related to multiplications of matrices. The number of iterations does not exceed m, and therefore the overall computational complexity is of order $0(m^3 \cdot n^3)$. ■

On the basis of the results presented in Section 3, the gradient method for establishing $\bar{x}$, belonging to the sufficiently small neighborhood $U_\varepsilon(x^*)$ of some fixed solution $x^* \in X^*$ of (1) will be described. This gradient method has the following scheme:

$$x_{k+1} = x_k - \alpha \cdot \varphi'(x_k) \tag{27}$$

where gradient $\varphi'(x_k)$ fulfills the Lipschitz condition

$$\|\varphi'(x_{k+1}) - \varphi'(x_k)\| \leq L \cdot \|x_{k+1} - x_k\| \text{ where } L = 2 \cdot \left\|A^T \cdot A\right\|.$$

Convergence of the gradient method (27) is presented in Theorem 5 below and is based on Theorem 2.

Theorem 5 *Let $x_0 \in R^n$ and sequence $\{x_k\}$, $k = 0, 1, 2 \ldots$, be constructed according to (27), where $\alpha = 1/L$. Then*

$$x_k \to x^*,\ x^* \in X^*, \text{ where } k \to \infty \text{ and } \|x_{k+1} - y\| \leq \|x_k - y\|\ \forall\ y \in X^*.$$

Proof. Scheme (27) produces a sequence, which will converge to a certain $x^* \in X^*$. Moreover, for every sufficiently small $\varepsilon > 0$ there exists $\bar{k} = k(\varepsilon)$ such that $\{x_k\} \in U_\varepsilon(x^*)$, for all $k \geq \bar{k}$. This, in turn, means that on iteration $\bar{k}$ the hypothesis of Theorem 4 will be satisfied and we will obtain a pseudo-solution of (1). ■

Now we have all the necessary prerequisites to present the solving algorithm for (7).

Algorithm 2

Initialization Step: Let $k = 0$ and x_0 be an arbitrary point in R^n.

Main Recursive Step: Let

$$x_{k+1} = x_k - \alpha \cdot \varphi'(x_k).$$

Checking Step: If $z(x_k)$ is the solution for (7), then Algorithm 2 is terminated. Otherwise, we put $k := k + 1$ and the **Main Recursive Step** is repeated.

Theorem 6 *There exists a finite $\bar{k}$ such that $z(x_{\bar{k}}) \in X^*$ and $z(x_{\bar{k}})$ is the solution for (7).*

Proof. The sequence $\{x_k\}$ is converging to fixed $x^* \in X^*$ and therefore in a certain iteration $\bar{k}$ the hypothesis of Theorem 4 will be satisfied and we will obtain the solution $z^* = P_{M(x_{\bar{k}})} \in X^*$. ■

Theorem 6 allows us to establish whether (1) has a solution or not.

Corollary 1 *If*

$$z^* \in X,$$

then z^* *is the solution of (1). Otherwise, (1) has no solutions.*

5. Concluding remarks and appendix

As it was already mentioned, the locally-polynomial complexity estimate is valid only if the starting point belongs to a sufficiently small neighborhood of the set of pseudo-solutions X^*. For reaching such a desired point, the gradient method (27) is used. There are accelerated gradient methods, see, e.g., Nesterov (1984) and Poliak (1987), but these methods do not guarantee monotonic convergence to a set of pseudo-solutions X^*. The method, presented in this paper is monotonically converging to a certain point x^*, $x^* \in X^*$. It is obvious that the point x^* depends on the initial point x_0, and therefore the number of iterations required by the gradient method for entering into the proper neighborhood of point x^* depends on the position of the initial point x_0. Moreover, the ε radius of the neighborhood of point x^*, where the gradient method should get to, is, in the general case, unknown and depends on the specific problem being considered. However, it appears that we can guarantee a geometric convergence rate of the gradient method (27) while minimizing piecewise quadratic functions of the form (4).

Namely, for every strongly convex function $\psi(x)$, the gradient method (27) has a geometric convergence rate, i.e.

$$\psi(x_k) - \psi^* \leq c \cdot \delta^k, \text{ where } 0 < \delta < 1,\ c > 0,$$

where c is a constant, which is independent of the size of the problem, but it depends on the initial point x_0. In the general case, for the functions not convex in the strong sense, there is no proof of the geometric convergence of the gradient method (27). However, in the case of the function $\varphi(x)$, given by (4), it is possible to prove the geometric convergence of the gradient method (27). Let

$$l(x_k) = \{x^* + \beta \cdot (x_k - x^*)\,,\ \beta \geq 0\} \text{ and } M(s_k) = \{x^* + \beta \cdot s_k,\ \beta \geq 0\}\,,$$
$$s_k = \frac{x_k - x^*}{\|x_k - x^*\|}.$$

The theorem presented below proves the strong convexity of the function $\varphi(x)$ in the cone of convergence.

Theorem 7 *Elements of the sequence* $\{x_k\}$ *, defined by (27), belong to the cone of strong convexity of the function* $\varphi(x)$*, namely* $\forall\ x,\ y \in l(x_k)$ *the function* $\varphi(x)$ *will be uniformly strongly convex for the sequence* $\{x_k\}$*, i.e.*

$$\varphi(\lambda \cdot x + (1-\lambda)\cdot y) \leq \lambda \cdot \varphi(x) + (1-\lambda)\cdot \varphi(y) - \gamma \cdot \lambda \cdot (1-\lambda) \cdot \|x-y\|^2 \quad (28)$$

where $\lambda \in [0,1]$*,* $x, y \in l(x_k)$*,* $k = 0, 1 \ldots$*,* $\gamma > 0$*.*

Proof. First, it should be noted that because the second derivative of the function $\varphi(x)$ has a finite number of points of discontinuity in every direction $\bar{S} \in R^n$, i.e., on the ray $x^* + \lambda \cdot \bar{S}$, then there exists $\sigma > 0$ such that on the closed interval $[x^*, x^* + \sigma \cdot \bar{S}]$ the function $\varphi(x)$ has a continuous second derivative, obviously depending on $\bar{S}$. Let us assume that the Theorem does not hold, i.e. there does not exist $\gamma > 0$ such that (28) holds. This means that for

$$l(x_k) = \{x^* + \beta \cdot s_k,\ \beta \geq 0\}$$

the following will hold

$$\frac{\partial^2 \varphi(x^*)}{\partial s_k^2} = \gamma_k \to 0 \text{ when } k \to \infty, \quad (29)$$

or

$$\frac{\partial^2 \varphi(x^*)}{\partial s_k^2} = \left\langle A^T \cdot A \cdot s_k, s_k \right\rangle = \gamma_k \to 0 \text{ when } k \to \infty.$$

For vector $s = \lim_{k\to\infty} s_k$ the following condition $\left\langle A^T \cdot A \cdot s, s \right\rangle = 0$ will hold, or, due to the construction of $\varphi(x)$,

$$\varphi(x^* + \beta \cdot s) = 0 = \varphi(x^*) = \min \left\|(A \cdot x - b)_+\right\|^2,$$

where $\beta \in [0, \bar{\beta}]$, $\bar{\beta} > 0$ is a certain fixed constant. Let x_k^* be, obviously locally, the projection of x_k on the set $M(s) \in X^*$. Then, due to $s_k \to s$, $k \to \infty$, we have

$$\|x_k - x_k^*\| = \delta_k \cdot \|x_k - x^*\|, \text{ where } \delta_k \to 0,\ k \to \infty. \quad (30)$$

Let us set δ_k sufficiently small and consider points x_{k+r}, $r = 1, 2, \ldots$. Then, according to Theorem 5, we have:

$$\|x_{k+r} - x_k^*\| \leq \|x_k - x_k^*\|. \quad (31)$$

On the other hand, according to (30), when $r \to \infty$:

$$\begin{aligned} &\|x_{k+r} - x_k^*\| \geq \|x_k^* - x^*\| - \|x_{k+r} - x^*\| \geq \\ &\quad \|x_k - x^*\| - \|x_k - x_k^*\| - \|x_{k+r} - x^*\| \geq \\ &\frac{1}{\delta_k}\|x_k - x_k^*\| - \|x_k - x_k^*\| - \|x_{k+r} - x^*\| > \|x_k - x_k^*\|. \end{aligned}$$

This is contradictory to (31) and therefore Theorem 7 holds. ■

Theorem 7 allows for the estimation of the convergence rate of the gradient method (27).

THEOREM 8 *Under the assumptions of Theorem 7 for the sequence* $\{x_k\}$, *constructed according to (27), the following convergence rates will hold*

$$\varphi(x_k) - \varphi^* \leq c_1 \cdot \tau^k \text{ and } \|x_k - x^*\| \leq c_2 \cdot \tau^{\frac{k}{2}} \tag{32}$$

where $\tau \in (0,1)$, c_1, $c_2 > 0$, *the constants* c_1, c_2 *being independent of the value of* k, *but depending on the initial point* x_0.

PROOF. Let us denote

$$\mu_k = \varphi(x_k) - \varphi^*.$$

For the sequence $\{x_k\}$ and $q \in \left(\frac{1}{2}, 1\right)$, the following holds

$$\begin{aligned} \varphi(x_k) - \varphi(x_{k+1}) &\geq \alpha \cdot q \cdot \|\varphi'(x_k)\|^2 \geq \alpha \cdot q \cdot \langle \varphi'(x_k), s_k \rangle^2 = \frac{\partial^2 \varphi(x_k)}{\partial s_k^2} \geq \\ &\geq \alpha \cdot q \cdot \gamma^2 \cdot (\varphi(x_k) - \varphi^*), \end{aligned} \tag{33}$$

or, equivalently,

$$\mu_k - \mu_{k+1} \geq \alpha \cdot q \cdot \gamma^2 \mu_k.$$

Therefore, for $\tau \in (0,1)$ the following holds:

$$\mu_k \leq c_1 \cdot \tau^k \text{ or, equivalently, } \varphi(x_k) - \varphi^* \leq c_1 \cdot \tau^k$$

which proves the first part of (32), while the latter part of (32) follows from the strong convexity of the function $\varphi(x)$ in the cone of convergence. ■

Acknowledgements

The research of the third author was supported by the Russian Sciences Foundation (project No 21-71-30005).

The authors want to thank the anonymous referee for valuable comments and suggestions, which have allowed to improve the present version of the paper.

References

EVTUSHENKO, Y. G. AND GOLIKOV, A. (2003) New perspective on the theorems of alternative. In: *High Performance Algorithms and Software for Nonlinear Optimization*, Springer, 227–241.

Facchinei, F., Fischer, A. and Kanzow, C. (1998) On the accurate identification of active constraints. *SIAM Journal on Optimization*, **9**(1):14–32.

Goffin, J.-L. (1982) On the non-polynomiality of the relaxation method for systems of linear inequalities. *Mathematical Programming*, **22**(1):93–103.

Golikov, A. and Evtushenko, Y. G. (2003) Theorems of the alternative and their applications in numerical methods. *Computational Mathematics and Mathematical Physics*, **43**(3):338–358.

Han, S.-P. (1980) Least-squares solution of linear inequalities. Technical report, University of Wisconsin – Madison, Mathematical Research Center. November 1980.

Karmanov, V. G. (1989) *Mathematical Programming.* Mir Publishers, Moscow.

Mangasarian, O. (2001) A finite Newton method for classification problems. Technical Report 01-11, Data Mining Institute, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin.

Nesterov, Y. (1984) One class of methods of unconditional minimization of a convex function, having a high rate of convergence. *USSR Computational Mathematics and Mathematical Physics*, **24**(4):80–82.

Poliak, B. (1987) *Introduction to Optimization.* Optimization Software, Inc., New York.

Smale, S. (1998) Mathematical problems for the next century. *The Mathematical Intelligencer*, **20**(2):7–15.

Tretyakov, A. (2010) A finite-termination gradient projection method for solving systems of linear inequalities. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **25**(3):279–288.

Tretyakov, A. and Tyrtyshnikov, E. (2013) A finite gradient-projective solver for a quadratic programming problem. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **28**(3):289–300.

Tretyakov, A. and Tyrtyshnikov, E. (2015) Exact differentiable penalty for a problem of quadratic programming with the use of a gradient-projective method. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **30**(2):121–128.

Wright, S. J. (2005) An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM Journal on Optimization*, **15**(3):673–696.