Raport Badawczy

Research Report

RB/18/2017

A locally-polynomial method for establishing the existence of a solution of a system of linear inequalities

K. Szkatuła, A. Tretyakov

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences

POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. dr hab. inż. Zbigniew Nahorski

Warszawa 2017

A locally-polynomial method for establishing the existence of a solution of a system of linear inequalities

Krzysztof SZKATUŁA [a,b,*], 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
E-mail: Krzysztof.Szkatula@ibspan.waw.pl

November 26, 2017

Abstract

The paper proposes a method for solving systems of linear inequalities. This method establishes in finite number of iterations if a given system of linear inequalities 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 and in the worst case it has a geometric convergence rate.

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

1 Introduction

Let us consider a system of linear inequalities:

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

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

The goal of this paper is to establish if there is a solution for (1) attainable in a finite number of iterations, with reasonable computational effort. In the case when the set of solutions:

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

is not empty, at least one solution of (1) will be established.

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 to optimize, i.e. find a maximum or minimum value of a so-called objective function subject to a number of constraints on the variables, which are usually in the form of linear inequalities (cf. Angel and Porter [1]). 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 [13]. The problem of finding strongly-polynomial time algorithm which decides whether there exists a solution of (1) or, equivalent, is set X in (2) non empty, is the 9th of the Smale's problems and is still not solved.

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 (cf. Karmanov [8], Golikov and Evtushenko [7], Evtushenko and Golikov [4], Tretyakov [15], Tretyakov and Tyrtyshnikov [17]). 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 [16] and Mangasarian [9] linear programming problems are solved by reducing 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 fulfill 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. Solution methods described in Tretyakov and Tyrtyshnikov [16] and Mangasarian [9] are based on exploiting information on the problem being solved by analyzing sufficiently small neighborhood of the unique solution of (1). Analogous methods were proposed in Facchinei et al. [5] for the forecast (identification) of the active constraints in the sufficiently close neighborhood of the solution of the problem. In Mangasarian [10] a 2-factor method for solving degenerated systems of nonlinear equations was proposed. Similar approaches were exploited for constructing computational methods aimed at solving degenerated problems in nonlinear programming, cf. Belash and Tretyakov [2], Brezneva and Tretyakov [3] and Szczepanik and Tretyakov [14]. In the papers by Tretyakov and Tyrtyshnikov [17] and Wright [18], locally polynomial methods for solving quadratic programming problems, based on the similar ideas, are presented. It was proven in Goffin [6] that the well-known ellipsoid method is not polynomial in the worst case. Tretyakov [15] 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.

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

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^* of the set X^*, $x^* \in X^*$, cf. (5) will fulfill:

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

Therefore, in the general case, our goal is to solve the following equation:

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

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

$$J_0 \{i \in D \mid f_i(x) = 0\}, \; J_- \{i \in D \mid f_i(x) < 0\}, \; J_+ \{i \in D \mid f_i(x) > 0\}. \tag{8}$$

According to (6) and the above notations, x^* should fulfill the following equations:

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

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

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

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

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 $\left(B^T\right)'' = B^T \left(B \cdot B^T\right)^{-1} \cdot B = B^+ \cdot B$, and projection on the orthogonal complement as $\left(B^T\right)^{\perp} = I - \left(B^T\right)''$, where I is an all-ones matrix of the size $n \times n$. The main idea exploited in this paper is based on the following Lemma.

Lemma 1 *Let $x^* \in X^*$ be the pseudo-solution of (1). For every sufficiently small $\varepsilon > 0$ there exist $x \in U_\varepsilon(x^*)$, $U_\varepsilon(x^*) = \{x \in R^n \, | \|x - x^*\| \leq \varepsilon\}$, such that if $f_i(x) \geq 0$ then $f_i(x^*) \geq 0$.*

Proof. *From the construction of the set X^*, cf. (5), it follows that either $\langle a_i, x^* \rangle - b_i \geq 0$ or $\langle a_i, x^* \rangle - b_i < 0$, $x^* \in X^*$. In the first case, the Lemma proposition holds. In the second case, from the well-known features of the linear functions we obtain:*

$$f_i(x) \leq f_i(x^*) + \|x - x^*\| \cdot \|a_i\| .$$

From $0 < \varepsilon < -\frac{f_i(x^*)}{\|a_i\|}$ it follows that $f_i(x) < 0$, which is contradictory to the Lemma proposition that in $U_\varepsilon(x^*)$ there are points for which $f_i(x) \geq 0$. Therefore, $i \in J_0(x^*) \cup J_+(x^*)$. ■

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 fulfilled in the point x^* are as follows:

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

$$\langle a_i, x^* \rangle - b_i < 0, \quad i \in J_-(x^*). \tag{12}$$

In (12), it is taken into account that

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

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

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

$$M(\bar{x}) = \left\{ x \in R^n \mid \sum_{i \in \overline{J}_0(\bar{x}) \cup \overline{J}_+(\bar{x})} (\langle a_i, x\rangle - b_i) \cdot a_i = 0_n \text{ and } \langle a_j, x\rangle - b_j = 0 \right\}, \tag{13}$$

where $j \in \overline{J}_0(\bar{x})$.

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 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:

$$I_- = \left\{ i \in \overline{J}_+(\bar{x}) \mid f_i(z(\bar{x})) \leq 0 \right\};\ I_- \subseteq J_0(x^*).$$

Otherwise, if at 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:

$$I_+ = \left\{ i \in \overline{J}_-(\bar{x}) \mid f_i(z(\bar{x})) \geq 0 \right\};\ I_+ \subseteq J_0(x^*).$$

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

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

Next, we will again project point $\bar{x}$ on the new set $M(\bar{x})$, cf. (13), and the new point $z(\bar{x})$ will be obtained. Let denote as $A(\bar{x})$ and $b(\bar{x})$ the matrix and vector obtained thereby from A and b, respectively. The rows of $A(\bar{x})$ and 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:

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

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:

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

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

Let us observe that at 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 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.

Algorithm 1

Initialization Step: For the current point $\bar{x}$, the sets of indexes $J_0(\bar{x})$, $J_-(\bar{x})$ and $J_+(\bar{x})$ will be defined according to (8). If 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 fulfilled:

$$I_+ = \varnothing \text{ and } I_- = \varnothing. \tag{20}$$

Checking Step: If (20) holds, then $z(\bar{x}) \in X^*$, equation (9) is fulfilled; $z(\bar{x})$ is the pseudo-solution of (1), as defined in (3), and Algorithm 1 is terminated. Otherwise, if for certain $i \in D$ the condition (20) is violated, 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).

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 fulfill (10). If this is not the case, it is necessary to find another point $\bar{x}$ that is closer to x^*. How we accomplish this is described below.

Theorem 2 *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*

$$\varphi'(x) = A^T(x) \cdot (A(x) \cdot x - b)_+ = 0_n, \tag{21}$$

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 when $\overline{J}_0(\bar{x})$ may be enlarged does not exceed m, the number of elements of the set D. Therefore, at some iteration, (20) will be fulfilled. 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 computational effort related to multiplications of matrices. The number of iterations does not exceed m and therefore the overall computational complexity is of order $0(m^3 \cdot n^3)$. ■

To complement the presentation of this chapter, 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:

$$x_{k+1} = x_k - \alpha \cdot \varphi'(x_k) \qquad (22)$$

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 (22) is considered in the following theorem, cf. Karmanov [8].

Theorem 3 *Let $x_0 \in R^n$ and sequence $\{x_k\}$, $k = 0, 1, 2 \ldots$, be constructed according to (22), where $\alpha = \left\|A^T \cdot A\right\|$. Then*

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

Proof. Scheme (22) 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 2 will be fulfilled and we will obtain a pseudo-solution of (1). ■

Now we have all 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 4 *There exists 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 2 will be fulfilled and we will obtain the solution $z^* = P_{M(x_{\bar{k}})} \in X^*$. ■

Theorem 4 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.

4 Concluding remarks

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^*. To reach such a desired point, the gradient method (22) is used. There are accelerated gradient methods, of, see Nesterov [11] and Poliak [12], but these methods do not guarantee monotonic convergence to a set of pseudo-solutions X^*. Presented in this paper method 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 (22) while minimizing piecewise quadratic functions of the form (4).

Namely, for every strongly convex function $\psi(x)$, the gradient method (22) 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 (22). However, in the case of the function $\varphi(x)$ given by (4) it is possible to prove the geometric convergence of the gradient method (22). 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 5 *Elements of the sequence $\{x_k\}$, defined by (22), 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.*

$$\varphi(\lambda \cdot x + (1-\lambda) \cdot y) \leq \lambda \cdot \varphi(x) + (1-\lambda) \cdot \varphi(y) - \gamma \cdot \lambda \cdot (1-\lambda) \cdot \|x - y\|^2 \quad (23)$$

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

Proof. Let us assume that the Theorem does not hold, i.e. there does not exist $\gamma > 0$ such that (23) holds. This means that for

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

the following will hold

$$\frac{\partial^2 \varphi(x^*)}{\partial s_k^2} = \gamma_k \to 0 \text{ when } k \to \infty, \quad (24)$$

or

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

For vector $s = \lim_{k\to\infty} s_k$ will $\langle A^T \cdot A \cdot s, s \rangle = 0$ 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

$$\|x_k - x_k^*\| = \delta_k \cdot \|x_k - x^*\| , \text{ where } \delta_k \to 0,\ k \to \infty. \tag{25}$$

Let us set δ_k sufficiently small and consider points x_{k+r}, $r = 1, 2, \ldots$. Then, according to Theorem 3 we have:

$$\|x_{k+r} - x_k^*\| < \|x_k - x_k^*\| . \tag{26}$$

On the other hand, according to (25), when $r \to \infty$:

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

This is contradictory to (26) and therefore Theorem 5 holds. ■

Theorem 5 allows for the estimation of the convergence rate of the gradient method (22).

Theorem 6 *Under the assumptions of Theorem 5 for the sequence $\{x_k\}$, constructed according to (22), the following convergence rates will hold*

$$\varphi(x_k) - \varphi^* \leq c_1 \cdot \tau^k \text{ and } \|x_k - x^*\| \leq c_2 \cdot \tau^{\frac{k}{2}} \tag{27}$$

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

$$\begin{aligned} \varphi(x_k) - \varphi(x_{k+1}) \geq \alpha \cdot q \cdot \|\varphi'(x_k)\|^2 \geq \alpha \cdot q \cdot \langle \varphi'(x_k), s_k \rangle^2 = \frac{\partial^2 \varphi(x_k)}{\partial s_k^2} \geq \\ \geq \alpha \cdot q \cdot \gamma^2 \cdot (\varphi(x_k) - \varphi^*) \end{aligned} \tag{28}$$

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 (27), while the latter part of (27) follows from the strong convexity of the function $\varphi(x)$ in the cone of convergence. ■

Acknowledgements

This research was partially supported by the Russian Foundation for Basic Research Grant No17-07-00510, by the Leading Research Schools Grant No. NSH-8860.2016.1 and by the Russian Academy of Sciences Presidium Program 1-33 P RAS.

References

[1] A. R. Angel and S. R. Porter. *A survey of mathematics with applications.* Pearson Addison Wesley, 2009.

[2] K. Belash and A. A. Tretyakov. Methods for solving degenerate problems. *USSR Computational Mathematics and Mathematical Physics*, 28(4):90–94, 1988.

[3] O. A. Brezhneva and A. A. Tretyakov. The p-factor-lagrange methods for degenerate nonlinear programming. *Numerical Functional Analysis and Optimization*, 28(9-10):1051–1086, 2007.

[4] Y. G. Evtushenko and A. Golikov. New perspective on the theorems of alternative. In *High Performance Algorithms and Software for Nonlinear Optimization*, pages 227–241. Springer, 2003.

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

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

[7] A. Golikov and Y. G. Evtushenko. Theorems of the alternative and their applications in numerical methods. *Computational mathematics and mathematical physics*, 43(3):338–358, 2003.

[8] V. G. Karmanov. *Mathematical programming.* Mir Publishers, Moscow, 1989.

[9] O. Mangasarian. A finite newton method for classification problems. Technical report, Technical Report 01-11, Data Mining Institute, Computer Sciences Department, University of Wisconsin, Madison, Wisconsin, 2001.

[10] O. Mangasarian. A newton method for linear programming. *Journal of Optimization Theory and Applications*, 121(1):1–18, 2004.

[11] Y. Nesterov. 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, 1984.

[12] B. Poliak. *Introduction to optimization.* Optimization Software, Inc., New York, 1987.

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

[14] E. Szczepanik and A. Tretyakov. p-factor methods for nonregular inequality-constrained optimization problems. *Nonlinear Analysis: Theory, Methods & Applications*, 69(12):4241–4251, 2008.

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

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

[17] A. Tretyakov and E. Tyrtyshnikov. 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, 2015.

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