entropy

Article

On the Locally Polynomial Complexity of the Projection-Gradient Method for Solving Piecewise Quadratic Optimisation Problems

Agnieszka Prusińska [1,*,†], Krzysztof Szkatuła [1,2,†] and Alexey Tret'yakov [1,2,3,†]

[1] Faculty of Exact and Natural Sciences, Siedlce University, 08-110 Siedlce, Poland; Krzysztof.Szkatula@ibspan.waw.pl (K.S.); tret@uph.edu.pl (A.T.)
[2] Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland
[3] Dorodnicyn Computing Centre of FRC CSC, Russian Academy of Sciences, 119333 Moscow, Russia
* Correspondence: aprus@uph.edu.pl
† These authors contributed equally to this work.

Abstract: This paper proposes a method for solving optimisation problems involving piecewise quadratic functions. The method provides a solution in a finite number of iterations, and the computational complexity of the proposed method is locally polynomial of the problem dimension, i.e., if the initial point belongs to the sufficiently small neighbourhood of the solution set. Proposed method could be applied for solving large systems of linear inequalities.

Keywords: optimisation problem; piecewise quadratic function; gradient method; system of linear inequalities; locally polynomial complexity

check for updates

Citation: Prusińska, A.; Szkatuła, K.; Tret'yakov, A. On the Locally Polynomial Complexity of the Projection-Gradient Method for Solving Piecewise Quadratic Optimisation Problems. *Entropy* **2021**, *23*, 465. https://doi.org/10.3390/e23040465

Academic Editors: José A. Tenreiro Machado and Ivanka Stamova

Received: 13 March 2021
Accepted: 14 April 2021
Published: 15 April 2021

1. Introduction

Let us consider the following optimisation problem:

$$\min_{x \in R^n} \left\| (A \cdot x - b)_+ \right\|^2, \tag{1}$$

where $c_+ := \max\{c, 0\}$, A is an $m \times n$ matrix , $A = \{a_{ij}\}$, $x \in R^n$, $x = \{x_i\}$, $b \in R^m$, $b = \{b_j\}$, $i = 1, \ldots, n$, $j = 1, \ldots m$ and $\| \cdot \|$ is the Euclidean norm of R^n.

In this paper, a method for solving the problem in (1) is proposed; moreover, the number of iterations (equivalent to the computational complexity) required by the proposed method with respect to m and n is locally polynomial, and in the worst-case scenario, it has a geometric convergence rate.

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

$$X^* := \left\{ x^* \,\middle|\, x^* = \arg\min_{x \in R^n} \left\| (A \cdot x - b)_+ \right\|^2 \right\}. \tag{2}$$

If some point sufficiently close to the set X^* of solutions to (1) is known, then it is possible to find a solution of (1) within a polynomial number of computational iterations; thus, the computational complexity is of the order of $O(m^3 \cdot n^3)$.

Many methods for solving (1) have been proposed (cf. Karmanov [1], Golikov and Evtushenko [2], Evtushenko and Golikov [3], Tretyakov [4], Tretyakov and Tyrtyshnikov [5] and Han [6]). All of these methods have reasonable computational complexity but, as mentioned above, to date, no strongly polynomial-time algorithm for solving (1) has been proposed. In studies by Tretyakov and Tyrtyshnikov [7] and Mangasarian [8], linear programming problems were solved by reducing them to the unconditional minimisation of strongly convex piecewise quadratic functions. A solution is obtained within a finite polynomial number of iterations if the starting point of the algorithm belongs to a sufficiently close neighbourhood of the unique solution to the problem. Unfortunately, the authors imposed severe limitations on the functions to be minimised: they should be strongly convex, the eigenvalues of the Hessian matrices should satisfy specific conditions, etc.

These results create significant limitations on the class of problems that can be solved: it is required that (1) has only one unique solution, etc. The solution method described by Tretyakov and Tyrtyshnikov [7] is based on exploiting information about the problem being solved by analysing a sufficiently small neighbourhood of an arbitrary solution of (1). Analogous methods were proposed by Facchinei et al. [9] for the forecasting (identification) of the active constraints in a sufficiently close neighbourhood of the solution to the problem. In papers by Tretyakov and Tyrtyshnikov [5] and Wright [10], locally polynomial methods for solving quadratic programming problems based on similar ideas were presented. Tretyakov [4] proposed the gradient projection method for solving (1); this method involves finding a solution of (1) in a finite number of iterations and is a combination of iterative and straightforward (e.g., Gaussian) methods.

This paper proposes a computational method for solving (1). When the starting point of the proposed method is sufficiently close to the set X^* of solutions to (1), then its computational complexity is locally polynomial, i.e., it is of the order of $O(m^3 \cdot n^3)$.

We point out that solving a system of linear inequalities involves

$$A \cdot x - b \leq 0_m, \tag{3}$$

where the 0_m – m-dimensional vectors of zeroes can be reduced to solve the problem (1). This means that the number of computations required for establishing a solution (if a given system of linear inequalities has one) is locally polynomial.

Let us denote

$$X := \{x \in R^n | \; A \cdot x - b \leq 0_m\}. \tag{4}$$

It is obvious that the set X might be empty in general, but our method, presented in this paper, either determines this situation in a locally polynomial number of computations or provides a solution to the system (3). The proposed method could be applied when solving large systems of linear inequalities, which appear in many practical, industrial applications, e.g., the simplex method (Pan [11]), Karmarkar's method (Wright, [12]), Chubanov's method (Roos [13]), and Fourier–Motzkin elimination method (Khachiyan [14], I. Šimeček, R. Fritsch, D. Langr, R. Lórencz [15]).

2. Definitions and Theoretical Results

Let

$$\varphi(x) = \frac{1}{2} \cdot \|(A \cdot x - b)_+\|^2. \tag{5}$$

Theorem 1. *The function $\varphi(x)$ is convex and has a nonempty set of minimal values*

$$X^* := \left\{ x^* \in R^n \;\middle|\; \varphi(x^*) = \min_{x \in R^n} \varphi(x) \right\}. \tag{6}$$

Proof. Theorem 1 follows immediately from the well-known features of quadratic-type convex functions (see, e.g., [16]). □

It is obvious that the elements $x^* \in X^*$, cf. (6) satisfy

$$\varphi'(x^*)^\top = \sum_{i=1}^{m} (\langle a_i, x^* \rangle - b_i)_+ \cdot a_i = 0_n = A^T \cdot (A \cdot x^* - b)_+, \tag{7}$$

where a_i^T is the ith row of matrix A.

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

$$\varphi'(x)^\top = \sum_{i=1}^{m} (\langle a_i, x\rangle - b_i)_+ \cdot a_i = A^T \cdot (A \cdot x - b)_+ = 0_n, \text{ where } x \in R^n. \tag{8}$$

In the sequel, x^* stands for an arbitrary element of X^* (the minimum point of φ). If the minimum value of φ is equal to zero, then $X = X^*$, and if the minimum value of φ is positive, then $X = \emptyset$. Let us denote

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

and

$$\begin{aligned} J_0(x) &:= \{i \in D \mid f_i(x) = 0\}, \quad J_-(x) := \{i \in D \mid f_i(x) < 0\}, \\ J_+(x) &:= \{i \in D \mid f_i(x) > 0\}, \end{aligned} \tag{9}$$

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

According to (7) and the above notations, x^* should satisfy the formula

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

The formula (10) is equivalent to a condition that should be satisfied at point x^*

$$\sum_{i \in J_0(x^*) \cup J_+(x^*)} (\langle a_i, x^*\rangle - b_i) \cdot a_i = 0_n, \tag{11}$$

In (11), it is considered that

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

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

$$\sum_{i \in J_0(x) \cup J_+(x)} (\langle a_i, x\rangle - b_i)_+ \cdot a_i = 0_n, \tag{12}$$

or

$$\begin{gathered} \sum_{i \in J_+(x)} (\langle a_i, x\rangle - b_i)_+ \cdot a_i = 0_n, \\ \langle a_i, x\rangle - b_i = 0, \ i \in J_0(x). \end{gathered} \tag{13}$$

Without loss of generality, we may denote

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

where $l \le p \le m$.

The main idea exploited in this paper is based on the following Lemma. For $\varepsilon > 0$ we set $U_\varepsilon(x^*) := \{x \in R^n \ : \ \|x - x^*\| \le \varepsilon\}$.

Lemma 1. *Let x^* be a solution to the problem (1). Then, there exists $\varepsilon > 0$, such that for any $x \in U_\varepsilon(x^*)$, the inequality $f_i(x) \ge 0$ implies the inequality $f_i(x^*) \ge 0$.*

Proof. If $i \in J_-(x^*)$, that is $f_i(x^*) < 0$, then, by continuity of the function f_i, there exists $\varepsilon_i > 0$, such that $f_i(x) < 0$ for all $x \in U_{\varepsilon_i}(x^*)$. Set

$$\varepsilon = \min_{i \in J_-(x^*)} \varepsilon_i.$$

Then, for all $i \in J_-(x^*)$ and for all $x \in U_\varepsilon(x^*)$, we have $f_i(x) < 0$. Consequently, if there exists $x \in U_\varepsilon(x^*)$ such that $f_i(x) \geq 0$ with some $i \in \{1, \dots, m\}$, then $i \notin J_-(x^*)$, that is $f_i(x^*) \geq 0$. □

By virtue of the above lemma, in a sufficiently small neighbourhood of some fixed point $x^* \in X^*$, for every $\bar{x} \in U_\varepsilon(x^*)$, the following hold

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

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

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

Let $A(\bar{x})$ and $b(\bar{x})$ represent the matrix and vector obtained from A and b, respectively. The rows of $A(\bar{x})$ and the coefficients of $b(\bar{x})$ correspond to the index set, which is defined by $\overline{J}_0(\bar{x}) \cup \overline{J}_+(\bar{x})$. In this case, Equations (12)–(13) may be rewritten as

$$\begin{aligned} &A^T(\bar{x}) \cdot (A(\bar{x}) \cdot x - b(\bar{x})) = 0_n, \\ &\langle a_i, x \rangle - b_i = 0, \ i \in \overline{J}_0(\bar{x}). \end{aligned} \tag{14}$$

Let $\bar{A}(\bar{x})$ denote the matrix in the equations in (14) corresponding to the maximum set of linearly independent rows, and let $\bar{b}(\bar{x})$ denote the corresponding vector of constant terms in (14).

The equations in (14) may be reformulated in the following way

$$\bar{A}(\bar{x}) \cdot x - \bar{b}(\bar{x}) = 0_n. \tag{15}$$

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

$$A^T(x^*) \cdot (A(x^*) \cdot x^* - b(x^*))_+ = 0_n. \tag{16}$$

This, in turn, means that

$$\bar{A}(x^*) \cdot x^* - \bar{b}(x^*) = 0_n. \tag{17}$$

$$\begin{aligned} M(\bar{x}) := \{x \in \mathbb{R}^n \mid &\sum_{i \in \overline{J}_0(\bar{x}) \cup \overline{J}_+(\bar{x})} (\langle a_i, x \rangle - b_i) \cdot a_i = 0_n \\ &\text{and } \langle a_j, x \rangle - b_j = 0, \text{ where } j \in \overline{J}_0(\bar{x})\}. \end{aligned} \tag{18}$$

If the rank of a matrix B of size $r \times n$ is equal to r, then the pseudoinverse matrix (operator) B^+ may be defined as $B^+ := B^T \cdot (B \cdot B^T)^{-1}$. We denote the quadratic matrix $n \times n$ orthogonally projected on the space containing the rows of matrix B as $\left(B^T\right)^{\text{II}} := B^T\left(B \cdot B^T\right)^{-1} \cdot B = B^+ \cdot B$, and its projection on the orthogonal complement of the matrix is denoted as $\left(B^T\right)^{\perp} := I - \left(B^T\right)^{\text{II}}$, where I is an all-ones matrix of size $n \times n$.

Let a 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 the constraints at point $z(\bar{x})$ are $f_i(z(\bar{x})) \leq 0$ for a certain $i \in \overline{J}_+(\bar{x})$, then we 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 the constraints at point $z(\bar{x})$ are $f_i(z(\bar{x})) \geq 0$ for a certain $i \in \overline{J}_-(\bar{x})$, we define the set I_+ in an 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 redefine $\overline{J}_0(\bar{x})$, $\overline{J}_+(\bar{x})$ and $\overline{J}_-(\bar{x})$ as follows

$$\overline{J}_0(\bar{x}) := \overline{J}_0(\bar{x}) \cup I_- \cup I_+, \;\; \overline{J}_+(\bar{x}) := \overline{J}_+(\bar{x}) \setminus I_-, \;\; \overline{J}_-(\bar{x}) := \overline{J}_-(\bar{x}) \setminus I_+. \tag{19}$$

Next, we project point $\bar{x}$ on the new set $M(\bar{x})$, cf. (18), and a new point $z(\bar{x})$ is obtained.

Let

$$z(x) = P_{M(\bar{x})}(x) = \left(A^T(\bar{x})\right)^{\perp} \cdot x + \bar{A}^{+}(\bar{x}) \cdot \bar{b}(\bar{x}) \tag{20}$$

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

3. Algorithm for Finding the Solution of (1)

In this section, the algorithm designed to find the solution to (1) is presented. The main idea of this algorithm is based on information related to a current point $\bar{x}$ belonging to a sufficiently small neighbourhood of the point $x^* \in X^*$. We also demonstrate how to find such a point. The proposed method comprises two algorithms. The starting point of the method can be arbitrary, because Algorithm 2 (gradient method with a special step selection) starts at an arbitrary point and, on a certain iteration, it will provide a point arbitrarily close to the solution set. Therefore, Algorithm 1 could start at the point specified by Algorithm 2.

Algorithm 1. ***Initialisation Step****: For the current point $\bar{x}$, the sets of indices $J_0(\bar{x})$, $J_-(\bar{x})$ and $J_+(\bar{x})$ are defined according to (9). If set $J_+(\bar{x}) = \emptyset$, then $\bar{x}$ is the solution of (1) and Algorithm 1 is terminated. Otherwise, the Main Recursive Step is performed.*

Main Recursive Step*: Let $z(\bar{x})$, the projection of point $\bar{x}$ on the set $M(\bar{x})$, be defined according to (20). We check if the following condition is satisfied*

$$I_+ = \emptyset \quad \text{and} \quad I_- = \emptyset. \tag{21}$$

Checking Step*: If (21) holds, then $z(\bar{x}) \in X^*$, and equation (10) is satisfied; $z(\bar{x})$ is the solution of (1), as defined in (2), and Algorithm 1 is terminated. Otherwise, if for certain values of $i \in D$ the condition (21) is violated and $i \in I_+ \cup I_-$, we define $\overline{J}_0(\bar{x})$, $\overline{J}_+(\bar{x})$ and $\overline{J}_-(\bar{x})$ according to (19), $M(\bar{x})$ is redefined according to (18), and the Main Recursive Step is repeated.*

The set D is finite, and $|D| = m$; therefore, the number of changes to the 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 (12) is established. This means that $z(\bar{x})$ is the solution of (1), as defined in (2).

It is of utmost importance that $\bar{x}$ belongs to a sufficiently small neighbourhood of the point x^*, because otherwise, $z(\bar{x})$ may not satisfy (12). If this is not the case, it is necessary to find another point $\bar{x}$ that is closer to x^*. The process for accomplishing this is described below.

Theorem 2. *For a sufficiently small $\varepsilon > 0$ and for every $\bar{x} \in U_\varepsilon(x^*)$, Algorithm 1 provides $z^* = z(\bar{x})$ as the solution for*

$$\varphi'(x) = A^T(x) \cdot (A(x) \cdot x - b)_+ = 0_n, \tag{22}$$

and this is equivalent to finding the solution for (12) within a number of iterations of the order of $0(m^3 \cdot n^3)$.

Proof. The proof is based on the observation that for $\bar{x}$ belonging to a sufficiently small neighbourhood of the point x^*, according to Lemma 1, the constraints $f_i(\bar{x}) \geq 0$ 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 $M(\bar{x})$, which is defined according to (18). It may happen that the set $\overline{J}_0(\bar{x})$ becomes enlarged. However, the number of iterations required when $\overline{J}_0(\bar{x})$ becomes enlarged does not exceed m, the number of elements in the set D. Therefore, at some iteration, (21) is satisfied. This means that $z(\bar{x})$ satisfies (12) or, equivalently, $\varphi'(z(\bar{x})) = 0_n$. This demonstrates that $z(\bar{x})$ is the solution for (1), as defined in (2). The computational complexity of establishing each projection $z(\bar{x})$ is of order $0(m^2 \cdot n^3)$; this process takes the computational effort related to the multiplications of matrices into account. 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 neighbourhood $U_\varepsilon(x^*)$ of some fixed solution $x^* \in X^*$ to (1) is described. This gradient method has the following scheme

$$x_{k+1} = x_k - \alpha \cdot \varphi'(x_k) \tag{23}$$

where $\alpha = \frac{1}{L}$ and gradient $\varphi'(x_k)$ fulfils the Lipschitz condition

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

The convergence of the gradient method (23) is considered in the following theorem, cf. Karmanov [1].

Theorem 3. *Let $x_0 \in R^n$ and the sequence $\{x_k\}$, $k = 0, 1, 2\ldots$, be constructed according to (23). Then,*

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

Proof. The scheme in (23) produces a sequence that converges 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 at iteration $\bar{k}$, the hypothesis of Theorem 2 is satisfied, and we obtain a solution to (1). □

Now, we have all the necessary prerequisites to present the solution algorithm for (3).

Algorithm 2. ***Initialisation Step***: *Let $k = 0$, and let 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 (3), then Algorithm 2 is terminated. Otherwise, we set $k := k + 1$, and the Main Recursive Step is repeated.*

Theorem 4. *There exists a finite $\bar{k}$ such that $z(x_{\bar{k}}) \in X^*$and $z(x_{\bar{k}})$ is the solution for (3).*

Proof. The sequence $\{x_k\}$ converges to a fixed $x^* \in X^*$ and therefore, at a certain iteration $\bar{k}$, the hypothesis of Theorem 2 is satisfied, and we obtain the solution $z^* = P_{M(x_{\bar{k}})} \in X^*$. □

Theorem 4 allows us to establish whether (3) has a solution or not.

Corollary 1. *If*

$$z^* \in X,$$

then z^* is the solution of (3). Otherwise (3) has no solutions.

4. Conclusions and Appendix

As previously mentioned, the locally polynomial complexity estimate is valid only if the starting point of the proposed method belongs to a sufficiently small neighbourhood of the set of solutions X^*. To reach such a desired point, the gradient method (23) is used. There are accelerated gradient methods (see those of Nesterov [17] and Poliak [18]), but these methods do not guarantee monotonic convergence to a set of solutions X^*. The method presented in this paper monotonically converges 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 the proper neighbourhood of point x^* depends on the position of the initial point x_0. Moreover, the ε radius of the neighbourhood of point x^*, which the gradient method should reach, is unknown in the general case and depends on the specific problem being considered. However, it appears that we can guarantee a geometric convergence rate for the gradient method (23) while minimising piecewise quadratic functions of the form (5).

Namely, for every strongly convex function $\psi(x)$, the gradient method (23) 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 that is independent of the size of the problem but depends on the initial point x_0. In the general case, for functions that are not convex in the strongest sense, there is no proof of the geometric convergence of the gradient method (23). However, in the case where the function $\varphi(x)$ is given by (5), it is possible to prove the geometric convergence of the gradient method (23). 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. *The elements of the sequence $\{x_k\}$, defined by (23), belong to the cone of strong convexity of the function $\varphi(x)$, namely $\forall\ x, y \in l(x_k)$, the function $\varphi(x)$ is uniformly strongly convex for the sequence $\{x_k\}$, i.e.,*

$$\varphi(\lambda \cdot x + (1-\lambda) \cdot y) \leq \lambda \cdot \varphi(x) + (1-\lambda) \cdot \varphi(y) - \gamma \cdot \lambda \cdot (1-\lambda) \cdot \|x - y\|^2 \tag{24}$$

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

Proof. First, it should be pointed out that because the second derivative of the function $\varphi(x)$ has a finite number of points of discontinuity $\bar{S} \in R^n$ in every direction, i.e., on the ray $x^* + \lambda \cdot \bar{S}$, 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 that obviously depends on $\bar{S}$. Let us assume that the theorem does not hold, i.e., there is not $\gamma > 0$, such that (24) holds. This means that for

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

the following must hold

$$\frac{\partial^2 \varphi(x^*)}{\partial s_k^2} = \gamma_k \to 0 \text{ when } k \to \infty, \tag{25}$$

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 $s = \lim\limits_{k\to\infty} s_k$, the following condition $\langle A^T \cdot A \cdot s, s\rangle = 0$ holds, or, due to the construction of $\varphi(x)$,

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

where $\beta \in [0, \bar{\beta}]$, $\bar{\beta} > 0$ is a certain fixed constant. Let x_k^* be (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

$$\|x_k - x_k^*\| = \delta_k \cdot \|x_k - x^*\|, \text{ where } \delta_k \to 0, k \to \infty. \tag{26}$$

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

$$\|x_{k+r} - x_k^*\| < \|x_k - x_k^*\|. \tag{27}$$

On the other hand, according to (26), when $r \to \infty$

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

This is contradictory to (27), and therefore Theorem 5 holds. □

Theorem 5 allows for the estimation of the convergence rate of the gradient method (23).

Theorem 6. *Under the assumptions of Theorem 5 for the sequence $\{x_k\}$, constructed according to (23), the following convergence rates hold*

$$\varphi(x_k) - \varphi^* \leq c_1 \cdot \tau^k \textit{ and } \|x_k - x^*\| \leq c_2 \cdot \tau^{\frac{k}{2}}, \tag{28}$$

where $\tau \in (0, 1)$, and c_1 and $c_2 > 0$; the constants c_1, c_2 are independent of the value of k but depend 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

$$\begin{gathered}\varphi(x_k) - \varphi(x_{k+1}) \geq \alpha \cdot q \cdot \|\varphi'(x_k)\|^2 \geq \alpha \cdot q \cdot \langle \varphi'(x_k), s_k\rangle^2 = \frac{\partial^2 \varphi(x_k)}{\partial s_k^2} \geq \\ \geq \alpha \cdot q \cdot \gamma^2 \cdot (\varphi(x_k) - \varphi^*)\end{gathered} \tag{29}$$

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

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

This proves the first part of (28), while the latter part of (28) follows from the strong convexity of the function $\varphi(x)$ in the cone of convergence. □

The conduction computational experiments and comparison of the presented method with other methods in the literature remains a topic for future research.

Author Contributions: Conceptualisation, A.T. and K.S.; methodology, A.P., K.S. and A.T.; validation, A.P., K.S. and A.T.; formal analysis, A.P.; investigation, A.P., K.S. and A.T.; resources, A.T.; writing—original draft preparation, A.P. and K.S.; supervision, A.T.; project administration, A.P.; funding acquisition, A.P. and A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Higher Education, grant number 61/20/B.

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

Conflicts of Interest: The authors declare no conflict of interest. Formal analysis, funding acquisition, investigation, methodology, project administration, validation, writing—original draft, A.P.; conceptualization, investigation, methodology, validation, writing—original draft, K.S.; conceptualization, funding acquisition, investigation, methodology, resources, supervision, validation, A.T.

References

1. Karmanov, V.G. *Mathematical Programming*; Mir Publishers: Moscow, Russia, 1989.
2. Golikov, A.; Evtushenko, Y.G. Theorems of the alternative and their applications in numerical methods. *Comput. Math. Math. Phys.* **2003**, *43*, 338–358.
3. Evtushenko, Y.G.; Golikov, A. New perspective on the theorems of alternative. In *High Performance Algorithms and Software for Nonlinear Optimization*; Springer: Berlin/Heidelberger, Germany, 2003; pp. 227–241.
4. Tretyakov, A. A finite-termination gradient projection method for solving systems of linear inequalities. *Russ. J. Numer. Anal. Model.* **2010**, *25*, 279–288. [CrossRef]
5. Tretyakov, A.; Tyrtyshnikov, E. Exact differentiable penalty for a problem of quadratic programming with the use of a gradient-projective method. *Russ. J. Numer. Anal. Model.* **2015**, *30*, 121–128. [CrossRef]
6. Han, S.-P. *Least-Squares Solution of Linear Inequalities*; Technical Report; Wisconsin Univ-Madison Mathematics Research Center: Madison, WI, USA, 1980.
7. Tretyakov, A.; Tyrtyshnikov, E. A finite gradient-projective solver for a quadratic programming problem. *Russ. J. Numer. Anal. Model.* **2013**, *28*, 289–300. [CrossRef]
8. Mangasarian, O. *A Finite Newton Method for Classification Problems*; Technical Report, Technical Report 01-11; Data Mining Institute, Computer Sciences Department, University of Wisconsin: Madison, WI, USA, 2001.
9. Facchinei, F.; Fischer, A.; Kanzow, C. On the accurate identification of active constraints. *SIAM* **1998**, *9*, 14–32. [CrossRef]
10. Wright, S.J. An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM* **2005**, *15*, 673–696. [CrossRef]
11. Pan, P.Q. A Projective Simplex Algorithm Using LU Decomposition. *Comput. Math. Appl.* **2000**, *39*, 187–208. [CrossRef]
12. Wright, M.H. The interior-point revolution in optimization: History, recent developments, and lasting consequences. *Bull. Am. Math. Soc.* **2005**, *42*, 39–56. [CrossRef]
13. Roos, K. An improved version of Chubanov's method for solving a homogeneous feasibility problem. *Optim. Methods. Softw.* **2018**, *33*, 26–44. [CrossRef]
14. Khachiyan, L. Fourier-Motzkin elimination method. In *Encyclopedia of Optimization*; Floudas, C.A., Pardalos, P.M., Eds.; Springer: Berlin/Heidelberger, Germany, 2009; pp. 1074–1077
15. Šimeček, I.; Fritsch, R.; Langr, D.; Lórencz, R. Paralel solver of large systems of linear inequalities using Fourier-Motzkin elimination. *Comput. Inform.* **2016**, *35*, 1307–1337.
16. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.
17. Y. Nesterov. One class of methods of unconditional minimization of a convex function, having a high rate of convergence. *USSR* **1984**, *24*, 80–82.
18. Poliak, B. *Introduction to Optimization*; Optimization Software, Inc.: New York, NY, USA, 1987.