algorithms

MDPI

Article

Properties of the Quadratic Transformation of Dual Variables

Vladimir Krutikov [1,2], Elena Tovbis [3], Anatoly Bykov [4], Predrag Stanimirovic [1,5], Ekaterina Chernova [2] and Lev Kazakovtsev [1,3,*]

1 Laboratory "Hybrid Methods of Modeling and Optimization in Complex Systems", Siberian Federal University, 79 Prosp. Svobodny, Krasnoyarsk 660041, Russia
2 Department of Applied Mathematics, Kemerovo State University, 6 Krasnaya Street, Kemerovo 650043, Russia
3 Institite of Informatics and Telecommunications, Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarskii Rabochii Prospekt, Krasnoyarsk 660037, Russia
4 Institute of Computational Technologies SB RAS, 21 Rukavishnikov Street, Kemerovo 650025, Russia
5 Faculty of Sciences and Mathematics, University of Nis, 18000 Nis, Serbia
* Correspondence: levk@bk.ru

Abstract: We investigate a solution of a convex programming problem with a strongly convex objective function based on the dual approach. A dual optimization problem has constraints on the positivity of variables. We study the methods and properties of transformations of dual variables that enable us to obtain an unconstrained optimization problem. We investigate the previously known method of transforming the components of dual variables in the form of their modulus (modulus method). We show that in the case of using the modulus method, the degree of the degeneracy of the function increases as it approaches the optimal point. Taking into account the ambiguity of the gradient in the boundary regions of the sign change of the new dual function variables and the increase in the degree of the function degeneracy, we need to use relaxation subgradient methods (RSM) that are difficult to implement and that can solve non-smooth non-convex optimization problems with a high degree of elongation of level surfaces. We propose to use the transformation of the components of dual variables in the form of their square (quadratic method). We prove that the transformed dual function has a Lipschitz gradient with a quadratic method of transformation. This enables us to use efficient gradient methods to find the extremum. The above properties are confirmed by a computational experiment. With a quadratic transformation compared to a modulus transformation, it is possible to obtain a solution of the problem by relaxation subgradient methods and smooth function minimization methods (conjugate gradient method and quasi-Newtonian method) with higher accuracy and lower computational costs. The noted transformations of dual variables were used in the program module for calculating the maximum permissible emissions of enterprises (MPE) of the software package for environmental monitoring of atmospheric air (ERA-AIR).

Keywords: convex programming; strongly convex functions; nonlinear programming; dual problem; subgradient; subgradient methods; optimization algorithms

Citation: Krutikov, V.; Tovbis, E.; Bykov, A.; Stanimirovic, P.; Chernova, E.; Kazakovtsev, L. Properties of the Quadratic Transformation of Dual Variables. *Algorithms* **2023**, *16*, 148. https://doi.org/10.3390/a16030148

Academic Editor: Frank Werner

Received: 30 January 2023
Revised: 4 March 2023
Accepted: 6 March 2023
Published: 7 March 2023

1. Introduction

Convex optimization [1–4] has become an efficient and important method to solve multi-dimensional problems in many fields, such as energy management [5–7], transportation management [8], trajectory optimization [9–13], image restoration [14], text summarization [15], cryptography [16], robotics [17], machine learning [18], and many others.

Methods for solving optimization problems using the dual approach are being intensively developed [19–31] as they can greatly simplify the solution of certain types of nonlinear programming problems. The dual approach is efficient at low computational costs for obtaining the characteristics of the dual function, which is typical, for example, for minimizing separable functions with linear constraints. When the number of constraints is significantly less than the number of variables, the transition from the original problem to the dual one enables us to obtain a problem of much smaller dimensionality.

In the article, we consider methods for solving a convex programming problem with a strongly convex objective function based on the dual approach. The complexity of optimizing the dual function in the problem of convex programming, in contrast to the dual functions considered in [19], lies in the presence of constraints on the positivity of the variables. It seems reasonable to develop and investigate ways to reduce such a problem to an unconstrained optimization problem, where modern efficient numerical methods for optimizing smooth and non-smooth functions are applicable.

The standard way to reduce such a problem to an unconstrained optimization problem is to replace the components of the dual variables by their moduli [32] (modulus method). Given a strongly convex objective function in a convex programming problem, the dual function is differentiable with a Lipschitz gradient in the range of positivity of the variables [33–37]. Replacing the components of the dual function variables by their moduli [32], we obtain a multi-extremal function that is differentiable with a Lipschitz gradient in the range of the sign constancy of the components of the generated variables (SCV) and having a local extremum determining the extremal dual variables. The ambiguity of the gradient arises in the regions of the sign change of the components of the transformed dual variables corresponding to inactive constraints. The components of the gradients corresponding to such variables at the boundary of adjacent SCV regions are equal in magnitude and have opposite signs.

We show that the cone of possible directions (CPD) for improving the dual function performance index, determined by a set of gradients in the neighborhood of a point from the boundary region for adjacent SCV regions, narrows as it approaches the extremum. RSM methods are suitable for such a situation [38–43]. The organization of RSM methods is to find the descent direction that enables us to go beyond a particular neighborhood of the current extremum [40–43], i.e., they find a descent direction that is gradient-consistent with all gradients in the neighborhood of the current extremum approximation. The narrower the CPD, the lower the convergence rate and the higher the computational cost of finding the RSM optimum [40–43]. Thus, with the modulus method of transforming dual variables, the dual function is non-smooth, and its degree of degeneracy increases as it approaches the extremum. This negatively affects the convergence rate of the minimization method and leads to a decrease in the accuracy of the resulting solution. In this case, it becomes necessary to involve complex subgradient methods of accelerated convergence for solving the problem.

Considering the noted shortcomings of the modulus transformation, it is important to search for a transformation of dual variables without gradient discontinuities. Our contribution is as follows. We propose a quadratic transformation of dual variables for the optimization problems with the number of constraints much less than the number of variables. The quadratic transformation method has proved successful in the problem of imposing constraints on the pollutant emissions of enterprises [33], which aroused interest in studying the properties of the function transformed in this way. In this work, we prove that in the case of a quadratic transformation, the gradient of the newly formed function is Lipschitz. In addition to subgradient methods, the noted property enables us to exploit methods of unconstrained minimization of smooth functions in solving the problem. As a result of a quadratic method compared to a modulus one, this is an increase in the convergence speed of the minimization method and the accuracy of the resulting solution when using well-studied and proven fast-converging methods for minimizing smooth functions.

We conducted a computer experiment for practical analysis of the methods for transforming dual variables. Models of convex programming problems similar in properties to those used in the software package for environmental monitoring of atmospheric air (ERA-AIR) [33] were taken as tests. Based on the extremum conditions, convex programming problems with two-sided constraints were generated with a number of variables up to 10,000 and a number of linear constraints up to 1000. The number of active two-sided and linear constraints varied.

To solve test problems, we used the multi-step subgradient method [43], subgradient methods with a change in the space metric [40–42], the BFGS (Broyden–Fletcher–Goldfarb-Shanno) quasi-Newtonian method [44–47], and the conjugate gradient method [34]. In the case of a quadratic transformation, the number of iterations required to solve the problem was reduced and the accuracy of its solution was increased. Problems with a quadratic transformation of the dual function variables were also successfully solved by rapidly converging methods for minimizing smooth functions: the conjugate gradient method [34] and the quasi-Newtonian BFGS method [44–47].

The noted transformations of dual variables were used in the program module for calculating the MPE, which is the part of the software package for environmental monitoring of atmospheric air, ERA-AIR. In related literature, the authors present different types of models for this applied problem: regression models [48,49], the particle swarm optimization model [50], the fuzzy c-means model [50], and a nonlinear programming model [51–53].

The rest of the paper is organized as follows. In Section 2, we formulate the dual problem for a convex programming problem with a strongly convex objective function and a summary of the results of the duality theory necessary for further research. In Section 3, we study the previously known modulus method for reducing the dual problem to an unconstrained optimization, and we propose the quadratic transformation of dual variables in Section 4. In Section 5, we provide a description of test problems, and we present the results of their solution in Section 6. In Section 7, we present the results of solving the applied problem of calculating the MPE of enterprises using the software package for environmental monitoring of atmospheric air. In Section 8, we summarize the work.

2. The Problem Formulation

Consider the convex programming problem:

$$\min_{x\in Q} f(x),\ Q\subset R^n, \tag{1}$$

$$g_i(x)\le 0,\ i=1,2,\dots,m. \tag{2}$$

In contrast to [15], where linear equality constraints are used, constraints (2) have the form of inequalities, which imposes restrictions on the positivity of dual variables.

Assumption 1. *In problem (1)–(2), assume that f is a strongly convex function, g_i are convex functions, set $Q\subset R^n$ is a convex closed bounded set, and the Slater condition is satisfied, i.e., there is an interior point x^0 of the set Q, such that the following conditions are satisfied:*

$$g_j(x^0)<0,\ j=1,2,\dots,m. \tag{3}$$

We form the Lagrangian function

$$L(x,y)=f(x)+\langle y,g(x)\rangle,\ y\in R^m,\ y\ge 0,\ x\in Q, \tag{4}$$

where $y^{\mathrm{T}}=(y_1,y_2,\dots,y_m)$ are the Lagrange multipliers and $g^{\mathrm{T}}(x)=(g_1(x),g_2(x),\dots,g_m(x))$. Here and in the following, we denote by $\langle y,g\rangle$ the scalar product of vectors y,g.

According to [34], we introduce the dual function

$$\psi(y)=\inf_{x\in Q} L(x,y)=\min_{x\in Q}[f(x)+\langle y,g(x)\rangle],\ y\ge 0 \tag{5}$$

and formulate the dual problem

$$\max\psi(y),\ y\ge 0. \tag{6}$$

Let us introduce the function $\phi(y) = -\psi(y)$, $y \geq 0$. Instead of (6), we get the problem of minimizing a convex function

$$\min \phi(y),\ y \geq 0. \tag{7}$$

If Assumption 1 holds, then the function $\psi(y)$ is differentiable [34]. The gradients of the functions $\psi(y)$ and $\phi(y)$ are [19]:

$$\nabla\psi(y) = g(\ x(y)\ ),\ \nabla\phi(y) = -g(\ x(y)\ ),\ y \geq 0, \tag{8}$$

where $x(y) = \underset{x \in Q}{\mathrm{argmin}} L(x, y)$. The gradient $\nabla\phi(y)$ in the area $y \geq 0$ satisfies the Lipschitz condition [34,54]

$$||\nabla\phi(y_1) - \nabla\phi(y_2)|| \leq L_\phi ||y_1 - y_2||. \tag{9}$$

Here and below, $||a|| = (a, a)^{0.5}$ is the norm of the vector a.

For $m << n$ and given a simple method for finding the minimum point of the Lagrangian with respect to x, the transition to the dual problem may simplify the solution. The complexity of solving the problem (7) lies in the presence of the constraint $y \geq 0$, which prevents the use of efficient numerical methods of unconstrained optimization. In this paper, we consider two ways to reduce the constrained optimization problem (7) to the unconstrained minimization problem and study their properties theoretically and experimentally.

3. Qualitative Analysis of the Modulus Transformation of Dual Variables

Assume that Assumption 1 holds for the primal convex programming problem (1), (2). To reduce problem (7) to an unconstrained optimization problem, we use the following transformation of variables [32]:

$$y^T(u) = (|u_1|, |u_2|, \ldots, |u_m|),\ i = 1, 2, \ldots, m. \tag{10}$$

The transformation method (10) was previously called the modulus method. Let us denote $\theta(u) = \phi(y(u))$. The minimization problem (7) under such a variables substitution takes the form

$$\min \theta(u),\ u \in R^m. \tag{11}$$

The problem (11) is multi-extremal, but in each of the quadrants determined by the sign constancy of the components of the variable u, the function is convex, and the components of its gradient are given by the expression

$$\nabla\theta_i(u) = -sign(u_i) \cdot g_i(\ x(y(u))\ ),\ i = 1, 2, \ldots, m. \tag{12}$$

In each of the quadrants, the gradient satisfies the Lipschitz condition. Note that, because of the convexity of the function, the gradients in the corresponding quadrants are also subgradients. At points on the boundaries of adjacent quadrants, we have gradients of adjacent areas. For this reason, to solve the problem of minimizing the function (11), we can use relaxation subgradient methods, in which the descent direction is chosen so that leaving the neighborhood of the current minimum is possible by one-dimensional minimization along this direction.

Using a two-dimensional problem as an example, let us consider the problem of increasing the degree of conditionality of the problem at the extremum area. Let us assume that the number of constraints in the problem (1), (2) is $m = 2$, and the minimum in the corresponding problem (11) is at the point

$$(u*)^T = (\ u_1^* > 0,\quad u_2^* = 0). \tag{13}$$

Here, we assume that only the first condition is active. The gradients (12) at the point (13) at u_2 = 0, depending on whether they belong to the area $u_2 \geq 0$ or $u_2 \leq 0$, respectively, have the form

$$(g_+^*)^T = (0,\ \ -g_2^*),\ (g_-^*)^T = (0,\ \ g_2^*). \tag{14}$$

Since constraint $g_2(x)$ is inactive, $g_2^* < 0$.

For a small $\delta_1 > 0$, let us consider the gradients of problem (11) on the boundary of the domains at u_2 = 0 in the point

$$(u^0)^T = (u_1^* - \delta_1 > 0\,,\ \ \ u_2^0 = 0), \tag{15}$$

determined by belonging to half-planes $u_2 \geq 0$ and $u_2 \leq 0$, respectively:

$$(g_+^0)^T = (\varepsilon_1,\ \ -g_2^* + \varepsilon_2),\ (g_-^0)^T = (\varepsilon_1,\ \ g_2^* - \varepsilon_2). \tag{16}$$

Since the gradients in the half-planes $u_2 \geq 0$ and $u_2 \leq 0$ satisfy the Lipschitz condition and $||u^0 - u^*|| = \delta_1$, the value δ_1 determines the smallness of the values ε_1 and ε_2. Here ε_1 < 0, since the gradients (16) are subgradients of the function $\theta(u)$ that is convex in the corresponding half-planes. Therefore, each subgradient represents a normal on the tangential hyperplane to the level surface, and the reverse direction of the subgradients shows the half-plane of the minimum point for this hyperplane.

In plane u_1, u_2, Figure 1 shows points u^*, u^0, gradients g_+^0, g_-^0 and the line segments L_+, L_-, whose normals are the corresponding gradients g_+^0, g_-^0. The set G is the convex hull of the gradients g_+^0, g_-^0. To reduce the function from the point u^0, we need to find the direction falling into the interior of the cone with the opening angle α, bounded by the lines L_+, L_-. The value of the angle α is proportional to the component ε_1. Since the gradients of each of the half-planes satisfy the Lipschitz condition, the values of ε_1 and ε_2 decrease as the minimum is approached, i.e., as the value δ_1 in (15) decreases. Thus, the cone of possible directions for decreasing the function narrows as it approaches the minimum.

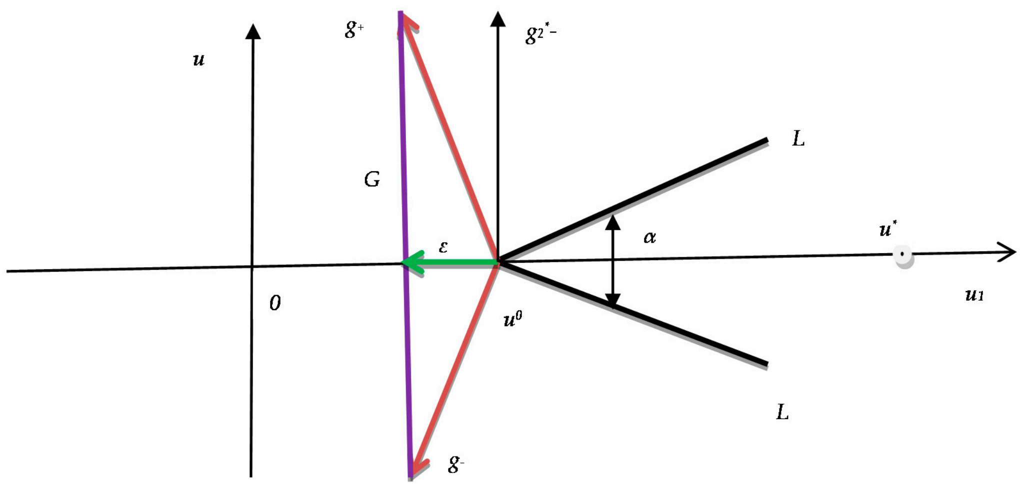

Figure 1. Gradients for adjacent half-planes.

Let us consider the effect of the above feature on the convergence rate of relaxation subgradient methods. According to research [40–43], an RSM contains a built-in algorithm for finding the descent direction that forms an acute angle with all subgradients in the neighborhood of the current minimum. Minimizing the function along the opposite direction (with a minus sign) allows us to go beyond the neighborhood of the current minimum. RSMs are formed as follows. Let us consider the problem of minimizing a function $f(x)$ in R^n solved.

In the RSM, successive approximations are built according to the equations [40–43]:

$$x_{k+1} = x_k - \gamma_k s_{k+1}, \quad \gamma_k = \underset{\gamma \in \mathbf{R}}{\operatorname{argmin}} f(x_k - \gamma s_{k+1}), \tag{17}$$

where the descent direction s_k is chosen as a solution to the system of inequalities [40–43]:

$$\langle s, g \rangle > 0, \quad \forall g \in G. \tag{18}$$

Here, G is the subgradient set in the neighborhood of the current minimum point xi. The system of inequalities (18) in the RSM is solved by iterative methods, using subgradients calculated on the descent trajectory of the minimization algorithm. As experience shows, RSMs are suitable for solving smooth and non-smooth non-convex minimization problems. Theoretical studies show that the speed of finding the descent direction is determined by the ratio of the lengths of the minimum to the length of the maximum vectors of the convex hull of subgradients in the current minimum domain. It decreases as the value of this ratio decreases. In the example, this ratio for set G at the point u^0 is equal to

$$\tau = \frac{|\varepsilon_1|}{||g_-^0||} = \frac{|\varepsilon_1|}{||g_+^0||} \to 0 \text{ , by } \delta_1 \to 0. \tag{19}$$

This means that the cost of finding the direction of descent in the RSM method is increased at $\delta_1 \to 0$.

Since the components of the gradients along u_2 near the extremum are practically fixed and non-zero in gradient-type methods, the region near $u_2 = 0$ is the area of attraction for the sequence of points searched for a minimum when approaching the extreme value by subgradient methods. Therefore, it is appropriate to use the described problem of narrowing the cone of permissible directions to decrease the objective function as one approaches the extremum.

The narrowing of the cone of admissible directions leads to an increase in the computational cost of subgradient methods, which is confirmed by a computational experiment. To improve the accuracy in solving the problem, it is necessary to apply more complex subgradient methods with a change in the space metric [40–42] and use parameter settings that significantly increase the computational cost.

4. Qualitative Analysis of the Quadratic Transformation of Dual Variables

Let assumption 1 be satisfied with respect to the primal problem of convex programming (1), (2). To reduce the problem (7) to an unconstrained optimization problem, we use the transformation of variables in the following form:

$$y^T(u) = (u_1^2, u_2^2, \ldots, u_m^2), \; i = 1, 2, \ldots, m. \tag{20}$$

The transformation method (20) was previously called the quadratic method. Let us denote $\Phi(u) = \phi(y(u))$. The minimization problem (7) under such a change of variables becomes

$$\min \; \Phi(u), \; u \in R^m. \tag{21}$$

Problem (21) is multiextremal, with each minimum u^* giving the optimal value of the dual variables $y(u^*)$. Given the expressions (8) for the function $\phi(y)$, the gradient components of the function are given by

$$\nabla \Phi_i(u) = -2 u_i g_i(\; x(y(u)) \;), \; i = 1, 2, \ldots, m. \tag{22}$$

We want to show that the gradient of the function $\Phi(u)$ satisfies the Lipschitz condition.

Theorem 1. *Let Assumption 1 hold for the primal convex programming problem (1), (2) and, then, in the bounded domain*

$$||u|| \leq R \tag{23}$$

the gradient of the function satisfies the Lipschitz condition.

Proof of Theorem 1. Denote $a,\ \ b \in R^m$, $y^a = y(a)$, $y^b = y(b)$, $g^a = g(y(a))$, $g^b = g(y(b))$ and write the components of the gradient difference

$$(\nabla\Phi(a) - \nabla\Phi(b))_i/2 = b_i g_i^b - a_i g_i^a = a_i(g_i^b - g_i^a) + g_i^b(b_i - a_i),\ i = 1, 2, \ldots, m. \tag{24}$$

Form vectors $v,\ \ z \in R^m$ with components $v_i = a_i(g_i^b - g_i^a),\ \ z_i = g_i^b(b_i - a_i), i = 1, 2, \ldots, m$. Taking into account the notation and properties of the norm, we obtain from (24)

$$||\nabla\Phi(a) - \nabla\Phi(b)||/2 = ||v + z|| \leq ||v|| + ||z||. \tag{25}$$

Let us estimate $||v||^2$ taking into account (9):

$$\begin{gathered} ||v||^2 = \sum_{i=1}^{m} a_i^2 (g_i^b - g_i^a)^2 \leq \sum_{i=1}^{m} a_i^2 \sum_{i=1}^{m} (g_i^b - g_i^a)^2 = ||a||^2 \cdot ||g^b - g^a||^2 \leq \\ L_\phi ||a||^2 \cdot ||y^b - y^a||^2 \leq L_\phi ||a||^2 \cdot ||a - b||^2 \cdot ||a + b||^2. \end{gathered}$$

Here, the last inequality is derived on the basis of the estimate

$$\begin{gathered} ||y^b - y^a||^2 = \sum_{i=1}^{m} (a_i^2 - b_i^2)^2 = \sum_{i=1}^{m} (a_i - b_i)^2 (a_i + b_i)^2 \leq \\ \sum_{i=1}^{m} (a_i - b_i)^2 \sum_{i=1}^{m} (a_i + b_i)^2 = ||a - b||^2 \cdot ||a + b||^2. \end{gathered}$$

Get an estimate of $||z||^2$:

$$||z||^2 = \sum_{i=1}^{m} (g_i^b)^2 (b_i - a_i)^2 \leq \sum_{i=1}^{m} (g_i^b)^2 \sum_{i=1}^{m} (b_i - a_i)^2 \leq ||g^b||^2 \cdot ||b - a||^2.$$

Finally, taking into account (23), we obtain the proof

$$\begin{gathered} ||\nabla\Phi(a) - \nabla\Phi(b)|| \leq 2(||v|| + ||z||) \leq \\ 2L_\phi ||a|| \cdot ||a - b|| \cdot ||a + b|| + 2||g^b|| \cdot ||b - a|| \leq \\ 4L_\phi R^2 \cdot ||a - b|| + 2C_g \cdot ||b - a|| \leq L_\Phi ||a - b||, \end{gathered}$$

where $C_g = \max_{x \in Q} ||g(x)||$. □

Gradient minimization methods are suitable for solving minimization problems with a Lipschitz gradient [34].

5. Test Problems

We formed the following problems for testing:

$$\min_{x \in Q}\ \ f(x, c),\ c,\ \ x \in R^n, \tag{26}$$

$$Ax \leq b,\ b \in R^m,\ A \in R^{m \times n}, \tag{27}$$

$$Q = \{x | a \leq x \leq d\},\ a, d \in R^n \tag{28}$$

Here, the parameters c, A, b, d are defined when tests are created. The constraints (27) together with (28) can be represented as follows

$$Ax - b \leq 0,\ x - d \leq 0, -x + a \leq 0. \tag{29}$$

Denote $\hat{A}^T = (A^T,\ I,\ -I)$, $\hat{b}^T = (b^T,\ d,\ -a)$. Constraints (29) can be rewritten as

$$\hat{A}x - \hat{b} \leq 0. \tag{30}$$

For problem (26) and (30), we write the Lagrange function

$$L(x, \hat{y}) = f(x, c) + \left\langle \hat{y}, \hat{A}x - \hat{b} \right\rangle, \tag{31}$$

where $\hat{y}$ is the vector of Lagrange multipliers of the combined group of linear constraints (30)

$$\hat{y} \in R^{m+n+n},\ \hat{y}^T = (y_b^T,\ \ y_d^T,\ y_a^T\ ),\ y_b \in R^m,\ \ y_d,\ y_a \in R^n\ \hat{y} \geq 0. \tag{32}$$

The optimality conditions for the problem (26), (30) includes the stationarity conditions

$$\nabla L_x(x, \hat{y}) = \nabla f(x, c) + \hat{y}^T \hat{A} \ \ = 0 \tag{33}$$

and conditions of complementary slackness

$$\hat{y}_i(\hat{A}x - \hat{b})_i = 0,\ i = 1, 2, \ldots, m + 2n. \tag{34}$$

In a linear programming problem, conditions (32)–(34) are sufficient for a feasible point x to be a minimum point [30]. In a non-linear programming problem, conditions (32)–(34) and $\nabla^2 L_x x\langle x, \hat{y}\rangle > 0$ are sufficient for a feasible point x to be a local minimum [19]. Based on the conditions (32)–(34), we will generate our test problems.

Algorithm for generating the test problem is as follows.

(1) Determine the number m of constraints, dimension n of the problem, the number mb of active constraints in (27), the number md of active constraints from the set of constraints $x - d \leq 0$, the number ma of active constraints from the set of constraints $-x + a \leq 0$. We assume that the active constraints in each group are the first constraints.

(2) Define the elements of the matrix A according to the equation $A_{ij} = r(p,q)$, where $r(p,q)$ is a uniformly distributed random number on the interval $[p,q]$, $0 < p < q$.

```
for i: =1 to m do
 for j: =1 to n do
  A[i,j]: =r(5,10).
```

(3) Set the boundaries $0 \leq a \leq d$ of the set Q, the optimal point x and the Lagrange multipliers y_d,y_a:$y_d = 0$, $y_a = 0$. Set the optimal point x by a generator of uniformly distributed numbers $a < x < d$. Set the boundaries of set Q for i: =1 to n do

```
begin
 a[i]: =5;
  ya[i]: =0;
  d[i]: =15;
  yd[i]: =0;
end;
```

(4) To determine the first mb active constraints from the set of constraints $Ax - b \leq 0$, we set the mb components of the vectors y_b and b ($y_{b,i} > 0$, $b_i = (Ax)_i$). The remaining components of the vector b are set according to the rule $b_i = (Ax)_i(1 + 1/i + 0.1)$, and the Lagrange multipliers for them are assumed to be zero: $y_{b,i} = 0$. Calculate $b = Ax$.

```
for i: =1 to n do
  b[i]: =0;
for i: =1 to m do
  for j: =1 to n do
    b[i]: =b[i] + A[i,j] × x[j];
```

Set Lagrange multipliers for the active constraints

```
for i: =1 to m do
```

  $yb[i]$: =0;
for i: =1 to mb do
  $yb[i]$: =1;
Define the right side b for inactive constraints
for i: =mb + 1 to m do
  $b[i]$: =$b[i]$ × (1 + $1/I$ + 0.1);

(5) Find the parameters c in the representation of the function $f(x,c)$ from (26), solving the system of Equation (32) with respect to c.

Since the elements of the matrix A are positive, the problem generated by the presented algorithm can be interpreted as a cost minimization problem under resource constraints. The resulting test problems are of the form (26)–(28) with the parameters c, A, b, d determined by given: the number m of constraints, the dimension n, the number mb of active constraints (27), the number of variables md lying on the upper boundary of the set Q, and the number ma of variables lying on the lower boundary of the set Q.

Table 1 shows the form of the test functions $f(x)$ and the corresponding names of the dual functions.

Table 1. Test functions and their names.

F(x,c)	θ(u) (11)	Φ(u) (31)
$f_1 = \langle c, x \rangle$	LP_R_Y1	LP_R_Y2
$f_2 = \langle c, x \rangle + \varepsilon(x, x)/2$	L2_Y1	L2_Y2
$f_3 = \sum_{i=1}^{n} c_i / x_i$	1/x_Y1	1/x_Y2

A linear programming problem for the function f_1 was generated. Since some of the constraints (27) may not be satisfied for known dual variables, a regularized function [34] $f_1 = \langle c, x \rangle + \alpha(x, x)/2$ was used to solve the generated problem, with α = 0.01. The function f_2 has the parameter ε = 0.01.

Our functions are separable. Therefore, the solution of the problem $x(y) = \underset{x \in Q}{\operatorname{argmin}} L(x, y)$ necessary to obtain the dual function decomposes into a series of one-dimensional minimization problems.

Consider the solution of the problem (21) with the function $f(x,c) = f_2$. Variables u and y are related according to (20), which means

$$\Phi(u) = \phi(y(u)) = -\psi(y(u)) = -\min_{x \in Q} L(x, y(u)) = -\min_{x \in Q}[f(x, c) + \langle y(u), g(x) \rangle] = -\min_{x \in Q}[\langle c, x \rangle + \frac{\varepsilon}{2}\sum_{i=1}^{n} x_i^2 + \langle y(u), Ax - b \rangle]$$

Under the condition that there are no constraints on the variable x, having solved the system

$$\frac{\partial L(x, y(u))}{\partial x_i} = c_i + \varepsilon x_i + [A^T y(u)]_i = 0$$

we obtain

$$x_i^* = -\frac{[A^T y(u) + c]_i}{\varepsilon},\ i = 1, 2, \ldots, n. \tag{35}$$

Given the constraints $a \le x \le d$, the solution $x(y(u))$ has the form

$$x_i(y(u)) = \begin{cases} a_i, \ \text{if} x_i^* \le a_i, \\ d_i, \ \text{if} x_i^* \ge d_i, \\ x_i^*, \ \text{if} a_i < x_i^* < d_i, \end{cases} \quad i = 1, 2, \ldots, n. \tag{36}$$

The parameters u, x, y resulting from the solution of the minimization problem are denoted by $\hat{u}$, $\hat{x}$, $\hat{y}$, and the optimal parameters are $u*$, $x*$, $y*$. Having solved the unconstrained minimization problem (21), we find the parameters $\hat{u}$ and Lagrange multipliers $\hat{y} = y(\hat{u})$. The solution of the primal problem follows from the Equations (35) and (36).

For the dual function (5), the following relations

$$f(x) \ge \psi(y),\ f(x*) = \psi(y*) \tag{37}$$

are fulfilled for any admissible x and y. Therefore, to assess the quality of the obtained solution, we will use the indicator

$$\Delta\psi = (f(x*) - \psi(\hat{y}))/|f(x*)|. \tag{38}$$

The second indicator will be obtained based on the assessment of the quality of the point $\hat{x}$. The obtained solution of the primal problem may not satisfy the constraints (27). Therefore, the value of the objective function $f(\hat{x})$ may turn out to be less than $f(x*)$ not suitable. In order to evaluate the quality of the approximation of the minimum value, we need to obtain a solution in the admissible area. Taking into account the specifics of the test problems, we will make its rough correction by moving to the admissible region along the ray $a - \hat{x}$, where the vector a is the lower bound of the set Q. Let us denote $v(\beta) = a + \beta\,(\hat{x} - a)$. We set $x_D = v(\beta_D)$, where β_D is the maximum value β at which the point satisfies the constraints (27). Taking into account that the inequality

$$f_D \geq f(x*), \tag{39}$$

is fulfilled, the quality of the approximation of the objective function is estimated using the indicator

$$\Delta f = (f_D - f(x*))/|f(x*)| \tag{40}$$

6. Solving Test Problems

Tables 2–5 show the results for dimensions n = 1000, m = 100. Each table contains the results of calculations for three functions with two ways of dual variable transformation. The number mb of active constraints is varied.

Table 2. Results of calculations for dimensions n = 1000, m = 100 with mb = 5.

Method	$\Delta\psi$	Δf	Number of Iterations	Number of Evaluations of Function and Gradient
Function LP_R_Y2 mb = 5, ma = 10, md = 10				
sub	$4.860 \times 10^{--1}$	1.098×10^{-5}	104	208
Shor	4.860×10^{-1}	1.561×10^{-5}	323	644
yv	4.860×10^{-1}	1.081×10^{-7}	64	130
MsgrPPR	4.860×10^{-1}	2.023×10^{-8}	259	559
BFGS	4.860×10^{-1}	1.770×10^{-5}	102	263
Function LP_R_Y1 mb = 5, ma = 10, md = 10				
sub	4.860×10^{-1}	7.999×10^{-4}	438	878
Shor	4.860×10^{-1}	3.335×10^{-5}	1471	3127
yv	4.860×10^{-1}	1.747×10^{-4}	443	843
Function L2_Y2 mb = 5, ma = 10, md = 10				
sub	3.011×10^{-9}	1.402×10^{-3}	100	202
Shor	1.223×10^{-11}	1.327×10^{-5}	279	563
yv	1.380×10^{-12}	1.019×10^{-6}	58	120
MsgrPPR	2.328×10^{-8}	2.394×10^{-4}	107	242
BFGS	9.582×10^{-12}	5.955×10^{-6}	73	183
Function L2_Y1 mb = 5, ma = 10, md = 10				
sub	1.635×10^{-6}	3.738×10^{-3}	427	856
Shor	7.250×10^{-9}	5.664×10^{-4}	1418	3001
yv	4.695×10^{-9}	2.193×10^{-4}	506	975
Function 1/x_Y2 mb = 5, ma = 10, md = 10				
sub	9.478×10^{-13}	3.952×10^{-7}	95	192
Shor	1.307×10^{-11}	1.086×10^{-6}	69	130
yv	2.215×10^{-13}	7.348×10^{-8}	49	100
MsgrPPR	2.489×10^{-15}	8.798×10^{-11}	62	128
BFGS	0.129×10^{-12}	1.825×10^{-7}	40	86
Function 1/x_Y1 mb = 5, ma = 10, md = 10				
sub	8.000×10^{-5}	1.142×10^{-4}	212	427
Shor	4.560×10^{-7}	1.228×10^{-5}	737	1560
yv	6.589×10^{-6}	3.316×10^{-5}	173	323

Table 3. Results of calculations for dimensions $n = 1000$, $m = 100$ with $mb = 25$.

Method	$\Delta\psi$	Δf	Number of Iterations	Number of Evaluations of Function and Gradient
		Function LP_R_Y2 $mb = 25$, $ma = 10$, $md = 10$		
sub	4.853×10^{-1}	2.032×10^{-4}	242	486
Shor	4.853×10^{-1}	8.940×10^{-6}	414	784
yv	4.853×10^{-1}	1.611×10^{-5}	97	194
MsgrPPR	4.853×10^{-1}	8.500×10^{-5}	113	271
BFGS	4.853×10^{-1}	6.074×10^{-7}	164	362
		Function LP_R_Y1 $mb = 25$, $ma = 10$, $md = 10$		
sub	4.853×10^{-1}	8.649×10^{-3}	439	881
Shor	4.853×10^{-1}	3.526×10^{-5}	1070	2261
yv	4.853×10^{-1}	5.491×10^{-5}	568	1111
		Function L2_Y2 $mb = 25$, $ma = 10$, $md = 10$		
sub	2.819×10^{-11}	2.345×10^{-4}	249	500
Shor	2.116×10^{-12}	1.049×10^{-5}	404	778
yv	1.691×10^{-12}	6.200×10^{-6}	91	177
MsgrPPR	7.367×10^{-11}	6.708×10^{-6}	103	245
BFGS	5.803×10^{-13}	1.865×10^{-6}	140	313
		Function L2_Y1 $mb = 25$, $ma = 10$, $md = 10$		
sub	6.277×10^{-7}	3.469×10^{-3}	445	893
Shor	3.265×10^{-13}	1.607×10^{-4}	1082	2269
yv	1.143×10^{-9}	1.793×10^{-4}	435	859
		Function 1/x_Y2 $mb = 25$, $ma = 10$, $md = 10$		
sub	1.958×10^{-12}	2.068×10^{-8}	207	417
Shor	4.996×10^{-13}	4.330×10^{-8}	81	150
yv	2.571×10^{-13}	7.876×10^{-9}	56	114
MsgrPPR	3.721×10^{-16}	1.670×10^{-11}	94	192
BFGS	7.797×10^{-13}	3.390×10^{-8}	50	108
		Function 1/x_Y1 $mb = 25$, $ma = 10$, $md = 10$		
sub	3.233×10^{-7}	1.502×10^{-5}	206	415
Shor	2.434×10^{-3}	6.075×10^{-4}	53	109
yv	4.642×10^{-7}	3.565×10^{-5}	155	292

Table 4. Results of calculations for dimensions $n = 1000$, $m = 100$ with $mb = 50$.

Method	$\Delta\psi$	Δf	Number of Iterations	Number of Evaluations of Function and Gradient
		Function LP_R_Y2 $mb = 50$, $ma = 10$, $md = 10$		
sub	4.849×10^{-1}	2.214×10^{-5}	274	553
Shor	4.849×10^{-1}	1.421×10^{-5}	383	740
yv	4.849×10^{-1}	1.481×10^{-5}	113	223
MsgrPPR	4.849×10^{-1}	1.562×10^{-4}	129	329
BFGS	4.849×10^{-1}	9.510×10^{-6}	167	372
		Function LP_R_Y1 $mb = 50$, $ma = 10$, $md = 10$		
sub	4.849×10^{-1}	5.526×10^{-3}	1388	2724
Shor	4.849×10^{-1}	2.283×10^{-5}	1002	2066
yv	4.849×10^{-1}	4.728×10^{-5}	588	1163
		Function L2_Y2 $mb = 50$, $ma = 10$, $md = 10$		
sub	1.683×10^{-10}	4.578×10^{-5}	302	610
Shor	1.867×10^{-12}	9.584×10^{-6}	471	873
yv	3.489×10^{-13}	2.129×10^{-6}	113	223
MsgrPPR	5.736×10^{-8}	3.649×10^{-4}	103	262
BFGS	1.282×10^{-12}	2.686×10^{-6}	161	345

Table 4. *Cont.*

Method	$\Delta\psi$	Δf	Number of Iterations	Number of Evaluations of Function and Gradient
		Function L2_Y1 mb = 50, ma = 10, md = 10		
sub	3.784×10^{-8}	8.079×10^{-4}	535	1077
Shor	1.238×10^{-12}	3.893×10^{-6}	921	1901
yv	2.465×10^{-12}	6.771×10^{-6}	444	869
		Function 1/x_Y2 mb = 50, ma = 10, md = 10		
sub	2.081×10^{-13}	8.761×10^{-9}	220	442
Shor	4.956×10^{-13}	3.592×10^{-8}	57	115
yv	4.816×10^{-14}	1.217×10^{-8}	60	119
MsgrPPR	7.175×10^{-6}	6.305×10^{-6}	7	17
BFGS	1.600×10^{-13}	1.933×10^{-7}	50	106
		Function 1/x_Y1 mb = 50, ma = 10, md = 10		
sub	1.413×10^{-7}	6.486×10^{-6}	204	410
Shor	1.207×10^{-7}	3.374×10^{-6}	381	803
yv	1.584×10^{-7}	3.113×10^{-6}	149	296

Table 5. Results of calculations for dimensions n = 1000, m = 100 with mb = 75.

Method	$\Delta\psi$	Δf	Number of Iterations	Number of Evaluations of Function and Gradient
		Function LP_R_Y2 mb = 75, ma = 10, md = 10		
sub	4.850×10^{-1}	3.150×10^{-4}	429	860
Shor	4.850×10^{-1}	3.128×10^{-5}	341	655
yv	4.850×10^{-1}	2.136×10^{-5}	122	248
MsgrPPR	4.850×10^{-1}	1.912×10^{-7}	167	420
BFGS	4.850×10^{-1}	1.163×10^{-5}	170	372
		Function LP_R_Y1 mb = 75, ma = 10, md = 10		
sub	4.850×10^{-1}	3.967×10^{-3}	1016	2059
Shor	4.850×10^{-1}	1.649×10^{-5}	566	1141
yv	4.850×10^{-1}	1.444×10^{-5}	393	773
		Function L2_Y2 mb = 75, ma = 10, md = 10		
sub	1.957×10^{-11}	4.027×10^{-4}	348	703
Shor	2.678×10^{-11}	1.657×10^{-5}	468	881
yv	3.674×10^{-11}	1.159×10^{-5}	95	184
MsgrPPR	3.536×10^{-11}	1.361×10^{-5}	204	445
BFGS	1.440×10^{-12}	3.770×10^{-6}	159	370
		Function L2_Y1 mb = 75, ma = 10, md = 10		
sub	1.104×10^{-6}	5.373×10^{-3}	544	1097
Shor	6.602×10^{-11}	1.385×10^{-5}	580	1199
yv	8.253×10^{-11}	3.199×10^{-5}	369	710
		Function 1/x_Y2 mb = 75, ma = 10, md = 10		
sub	2.684×10^{-13}	5.858×10^{-9}	207	417
Shor	2.118×10^{-13}	1.804×10^{-8}	55	114
yv	1.320×10^{-15}	2.850×10^{-9}	64	125
MsgrPPR	1.651×10^{-16}	8.098×10^{-11}	101	207
BFGS	8.277×10^{-13}	3.196×10^{-8}	34	74
		Function 1/x_Y1 mb = 75, ma = 10, md = 10		
sub	1.611×10^{-7}	1.764×10^{-5}	214	430
Shor	1.076×10^{-8}	2.246×10^{-6}	210	450
yv	6.973×10^{-8}	2.496×10^{-6}	106	214

The tables use labels n for the number of variables of the primal problem (1), m for the number of constraints in (2), mb for the number of active constraints in (2), md for the number of active constraints from the set of constraints $x - d \leq 0$ for Q, and ma for the number of active constraints from the set of constraints $-x + a \leq 0$ for Q. Each of the problems is specified by the name of the function from Table 1 and the parameters (N_F, nx, ny, mb, ma, md).

To solve problems (11), (31), we used the subgradient methods: sub [43], Shor [32], Orty [42], and yv [41] implemented with the one-dimensional search from [41–43]. To solve problem (31), we used additional methods for minimizing smooth functions: the Polak–Ribière–Polyak conjugate gradient method (Msgr_PPR) [34] and the quasi-Newtonian method BFGS [44–47].

The following conclusions have been made:

1. When varying the number *mb* of active constraints, no important differences in accuracy and number of function and gradient evaluations were found.
2. In the case of a modulus transformation of dual variables, due to manifestations of the degeneracy of the function, the calculation accuracy by the multi-step method *sub* is significantly inferior compared to the method *yv* with a change in the space metric of the RSM.
3. Due to the low dimensionality of the dual problem, the quality of the solution is somewhat lower for a modulus transformation of the dual variables than for the quadratic transformation for RSM with a change in the space metric.
4. The cost of solving the problem with a modulus transformation of dual variables increases significantly compared to a quadratic transformation.
5. The methods used for minimizing smooth functions (the conjugate gradient method and the quasi-Newtonian method) have excellent results both in terms of accuracy and number of function and gradient evaluations.

Tables 6–8 show the results for large-scale problems. The number *mb* of active constraints is varied.

Table 6. Results of calculations for dimensions n = 10,000, m = 1000 with mb = 250.

Method	$\Delta\psi$	Δf	**Number of Iterations**	**Number of Evaluations of Function and Gradient**
		Function LP_R_Y2 mb = 250, ma = 25, md = 25		
sub	4.949×10^{-1}	8.922×10^{-6}	149	305
yv	4.949×10^{-1}	7.680×10^{-5}	231	490
MsgrPPR	4.949×10^{-1}	9.844×10^{-6}	468	1109
BFGS	4.949×10^{-1}	2.817×10^{-6}	468	1072
		Function LP_R_Y1 mb = 250, ma = 25, md = 25		
sub	1.203×10^{-1}	1.248×10^{-2}	1130	2257
yv	4.949×10^{-1}	4.213×10^{-4}	2503	4606
		Function L2_Y2 mb = 250, ma = 25, md = 25		
sub	9.330×10^{-12}	1.708×10^{-5}	119	243
yv	4.623×10^{-11}	8.016×10^{-6}	205	419
MsgrPPR	2.295×10^{-11}	7.233×10^{-6}	1297	2739
BFGS	8.929×10^{-11}	5.816×10^{-5}	297	636
		Function L2_Y1 mb = 250, ma = 25, md = 25		
sub	9.525×10^{-5}	1.267×10^{-2}	237	479
yv	2.157×10^{-9}	3.622×10^{-4}	3035	5534
		Function 1/x_Y2 mb = 250, ma = 25, md = 25		
sub	1.656×10^{-12}	4.226×10^{-8}	139	284
yv	7.935×10^{-15}	3.233×10^{-9}	113	233
MsgrPPR	8.434×10^{-12}	8.817×10^{-8}	19	45
BFGS	9.680×10^{-15}	6.361×10^{-8}	137	286
		Function 1/x_Y1 mb = 250, ma = 25, md = 25		
sub	1.334×10^{-5}	7.513×10^{-6}	112	229
yv	3.241×10^{-5}	1.181×10^{-6}	394	717

Table 7. Results of calculations for dimensions n = 10,000, m = 250 with mb = 50.

Method	$\Delta\psi$	Δf	Number of Iterations	Number of Evaluations of Function and Gradient
		Function LP_R_Y2 mb = 50, ma = 25, md = 25		
sub	3.134×10^{-8}	3.134×10^{-8}	971	1907
Shor	4.980×10^{-1}	1.706×10^{-5}	2495	4554
Orty	5.280×10^{-10}	3.259×10^{-10}	1864	3535
yv	4.952×10^{-9}	3.044×10^{-9}	507	1037
MsgrPPR	5.002×10^{-1}	3.355×10^{-6}	126	319
BFGS	2.701×10^{-1}	7.262×10^{-11}	656	1524
		Function LP_R_Y1 mb = 50, ma = 25, md = 25		
sub	4.980×10^{-1}	6.439×10^{-3}	1280	2540
Shor	4.980×10^{-1}	2.210×10^{-6}	2972	6250
Orty	4.980×10^{-1}	2.057×10^{-6}	1179	2189
yv	4.980×10^{-1}	7.798×10^{-6}	1121	2087
		Function L2_Y2 mb = 50, ma = 25, md = 25		
sub	3.967×10^{-10}	4.126×10^{-5}	178	371
Shor	1.065×10^{-12}	1.988×10^{-5}	2481	4541
Orty	5.418×10^{-12}	3.259×10^{-5}	909	1772
yv	8.057×10^{-11}	1.533×10^{-5}	284	609
MsgrPPR	2.745×10^{-5}	4.235×10^{-3}	58	151
BFGS	2.865×10^{-13}	8.488×10^{-6}	439	966
		Function L2_Y1 mb = 50, ma = 25, md = 25		
sub	2.252×10^{-7}	9.320×10^{-3}	2221	4269
Shor	5.502×10^{-11}	4.765×10^{-6}	2780	5871
Orty	3.657×10^{-15}	4.047×10^{-7}	2589	4639
yv	5.850×10^{-12}	1.390×10^{-6}	1668	3208
		Function 1/x_Y2 mb = 50, ma = 25, md = 25		
sub	2.771×10^{-13}	1.434×10^{-8}	130	267
Shor	5.433×10^{-13}	2.314×10^{-8}	71	149
Orty	1.521×10^{-12}	1.813×10^{-7}	193	399
yv	2.544×10^{-15}	1.207×10^{-9}	84	167
MsgrPPR	$3.782\text{E} \times 10^{-15}$	8.183×10^{-9}	41	90
BFGS	9.539×10^{-15}	1.098×10^{-8}	86	181
		Function 1/x_Y1 mb = 50, ma = 25, md = 25		
sub	5.754×10^{-8}	3.910×10^{-7}	218	441
Shor	3.289×10^{-4}	8.393×10^{-5}	622	1340
Orty	7.004×10^{-5}	2.455×10^{-5}	289	512
yv	1.760×10^{-4}	5.199×10^{-5}	221	416

Table 8. Results of calculations for dimensions n = 5000, m = 500 with mb = 50.

Method	$\Delta\psi$	Δf	Number of Terations	Number of Evaluations of Function and Gradient
		Function LP_R_Y2 mb = 50, ma = 25, md = 25		
sub	5.050×10^{-1}	6.241×10^{-3}	139	284
Shor	5.050×10^{-1}	1.594×10^{-5}	1926	3767
Orty	5.050×10^{-1}	1.493×10^{-5}	335	693
yv	5.050×10^{-1}	2.622×10^{-6}	181	348
MsgrPPR	5.079×10^{-1}	1.973×10^{-3}	200	448
BFGS	5.050×10^{-1}	1.464×10^{-5}	286	665

Table 8. *Cont.*

Method	$\Delta\psi$	Δf	Number of Terations	Number of Evaluations of Function and Gradient
		Function LP_R_Y1 mb = 50, ma = 25, md = 25		
sub	5.050×10^{-1}	6.822×10^{-5}	2497	4998
Shor	5.050×10^{-1}	1.832×10^{-4}	4300	9000
Orty	5.050×10^{-1}	8.383×10^{-4}	1360	2374
yv	5.050×10^{-1}	3.862×10^{-4}	1751	3234
		Function L2_Y2 mb = 50, ma = 25, md = 25		
sub	1.915×10^{-8}	4.225×10^{-4}	154	313
Shor	6.302×10^{-12}	1.926×10^{-5}	2165	4259
Orty	4.738×10^{-12}	8.416×10^{-6}	361	741
yv	3.793×10^{-13}	1.865×10^{-6}	165	330
MsgrPPR	1.000×10^{-15}	3.116×10^{-8}	414	897
BFGS	1.359×10^{-12}	3.216×10^{-5}	124	272
		Function L2_Y1 mb = 50, ma = 25, md = 25		
sub	2.569×10^{-10}	8.985×10^{-5}	2577	5157
Shor	2.541×10^{-9}	2.008×10^{-4}	4317	9058
Orty	4.461×10^{-10}	1.431×10^{-4}	2516	4331
yv	4.310×10^{-10}	1.019×10^{-4}	2024	3708
		Function 1/x_Y2 mb = 50, ma = 25, md = 25		
sub	1.200×10^{-15}	2.118×10^{-11}	117	240
Shor	1.279×10^{-13}	2.000×10^{-8}	46	101
Orty	2.271×10^{-13}	1.907×10^{-7}	99	199
yv	3.357×10^{-13}	1.299×10^{-7}	80	144
MsgrPPR	1.375×10^{-13}	1.243×10^{-8}	53	110
BFGS	4.152×10^{-13}	3.220×10^{-8}	59	118
		Function 1/x_Y1 mb = 50, ma = 25, md = 25		
sub	2.023×10^{-6}	6.824×10^{-5}	1122	2249
Shor	6.456×10^{-4}	1.764×10^{-4}	1407	2951
Orty	1.790×10^{-4}	6.067×10^{-5}	284	523
yv	4.916×10^{-5}	2.513×10^{-6}	392	712

The performed computer experiment confirms the efficiency of the quadratic transformation.

7. Solving Applied Problem

For calculating MPE in the software package ERA-AIR [55], two optimization models of the form (26)–(28) with objective functions $f_1 = (c, x)$ and $f_3 = \sum_{i=1}^{n} c_i/x_i$ are implemented. For f_1 components $c_i = -1$, and for f_3 components $c_i = d_i$, $i = 1,2, \ldots, n$, where d_i are vector d components from (28). The variables x_i of the problem are the volumes of atmospheric pollution from pollution sources, the number of which can reach about 10,000. Constraints of the problem (27) are:

$$\sum_{i=1}^{n} A_{ji} x_i \le b_j, \ j = 1, 2, \ldots, m, \tag{41}$$

where the left part in (41) is the total concentration for all sources of atmospheric pollution, b_j is permissible concentration at the j-th point, and A_{ji} is concentration per volume of emission from source i in quota location j. It is assumed that the vector x components satisfy two-sided constraints (28). In (28), vector d is the current state of MPE limitations for sources, and a is the technical capability to reduce emissions.

As an example, we present the data collected as a consolidated volume of MPE for Kemerovo city in 2003 [56] and give a solution to the problem of emission quotas for 1085 sources. Sources emit nitrogen dioxide into the atmosphere (MPC = 0.2 mg/m^3). For air pollution and emission quotas calculations (vector x), we used the ERA-AIR software complex [55]. The calculations were carried out on a regular grid of 22 km by 24 km with a step of 250 m.

Figure 2 shows the area exceeding the maximum one-time surface concentrations, obtained by calculating, according to the Russian regulatory methodology in search mode, the maximum in terms of the wind speed and direction. It can be seen that the pollution zone with an excess of the MPC (inside the black line) covers the living sector closest to the industrial site, where compliance with MPC is required. To select points for the constraints (27) on the concentration, an additional calculation was performed to cover the living sector with 300 calculated points, which will be the set of quota points.

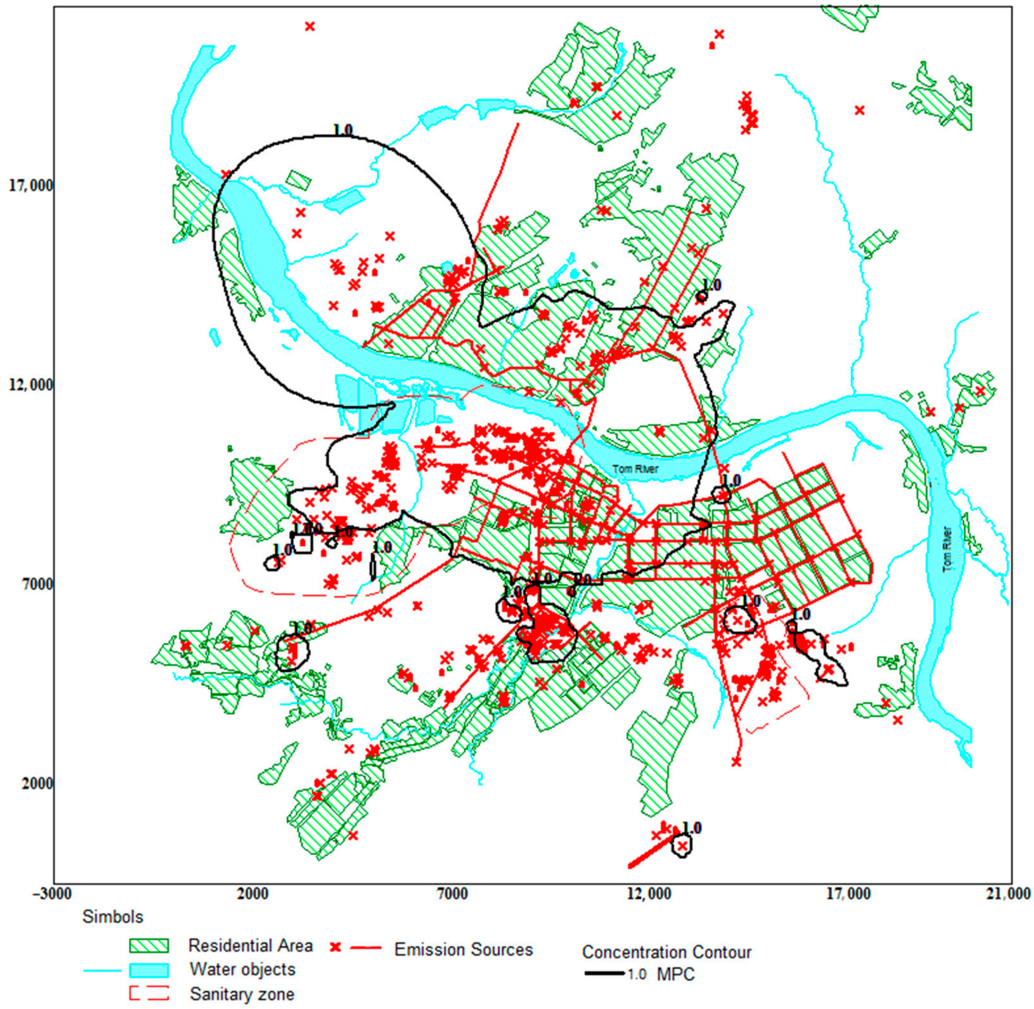

Figure 2. The area of excess (inside the black contours) of the calculated maximum one-time concentrations of nitrogen dioxide from the set of emission sources of Kemerovo city.

To solve the problem in the selected quota points, the influence coefficients A_{ji} are calculated, which form the matrix A. All elements of the b_j vector in this example are equal to MPC, but their individual values can be set. The vectors a and d of constraints (28) on the solution search area x are given as follows: the vector d is a set of emissions from all 1085 sources for which the initial calculation was carried out, and the vector $a = 0.1d$. Thus, the MPE for each source cannot exceed its existing emission, and the allowable emission reduction at the source cannot be more than 90%. The problem was solved by various methods. The coefficients of socio-economic significance [57] were chosen to be the same and were given above, and, in this example, these are the components of the vector C.

The ERA-AIR software package includes modules for calculating source emission reductions (necessary to achieve MPC) using methods [58,59]. The first of them [58] enables one to find a unique solution to the system of inequalities (41), but it is not an optimization one. The second [59] makes it possible to achieve MPC with a minimum reduction in emissions, with a total for the entire set of emission sources in the city. Recently, a method for solving problem (3–5) [57] considered in [33] has been added to them.

In this work, we used the modulus and quadratic methods of transforming the variables of the dual function. We used the linear programming model with regularization $f_1 = (c, x)$ [59] and the non-linear emission quota model $f_3 = \sum_{i=1}^{n} c_i / x_i$ [57]. Due to property (37) for the function f_3, we can use a criterion for evaluating the quality of the solution, similar to (38)

$$\Delta\psi = (f_D - \psi(\hat{y}))/|f_D|, \tag{42}$$

where f_d is calculated by the previously described method (39), (40), and the point $\hat{y}$ is the solution found for the dual problem.

Using the quadratic method, we obtained solutions for the model f_3 with an accuracy $\Delta\psi \le 10^{-9}$ by each of the above methods. For the modulus method $\Delta\psi \le 10^{-6}$, at the same time, the cost of the function and the gradient calculations is approximately two times higher than with the quadratic method.

Table 9 shows the results of changes in total emissions when MPC is reached in a living sector.

Table 9. Existing cumulative emission and cumulative emission value (MPE) to achieve MPC.

Method	Total Value of Existing Emissions before Reduction, *g/s*	Total Value of Emissions That Ensure the Achievement of MPC, *g/s*
Equal quota [58]	1042.89	607.83
Linear programming model with regularization $f_1 = (c, x)$ [59]	1042.89	833.56
Nonlinear model $f_3 = \sum_{i=1}^{n} c_i / x_i$ [57]	1042.89	689.73

The solution of the optimization problem with the function f_1 makes enables us to provide limits on pollution while leaving a noticeably larger total value of emissions compared to the other two methods. The main purpose of the numerical experiment is to demonstrate the efficiency of the proposed methods in solving the problems of calculating quotas for a fairly large industrial city. The necessary preliminary testing of the methods was carried out in order to ensure the trouble-free operation of the minimization methods built into the system when calculating numerous variants of the problem that arise due to the dependence of the coefficients of the matrix A on air temperature, wind strength, and direction.

8. Conclusions

In the problem of convex programming with a strongly convex objective function, we studied the possibility of reducing the dual problem to an unconstrained minimization problem under the constraint of positivity of the variables. Two methods of transforming the components of the dual variables were studied. In the case of a modulus transformation of dual variables, it is shown that the dual function is non-smooth, and its degree of degeneracy increases as it approaches the extremum. This fact negatively affects the convergence rate of the minimization method and leads to a decrease in the accuracy of the resulting solution. This leads to the necessity of involving complex subgradient methods of accelerated convergence with a change in the space metric for solving the problem.

The proposed quadratic method for transforming dual variables has been investigated theoretically. We have proved that in the case of a quadratic transformation of dual variables, the dual function has a Lipschitz gradient. This property of the gradient makes it possible to apply efficient methods of unconstrained minimization of smooth functions. In this case, in comparison with the modulus method, the convergence rate of the minimization method and the accuracy of solving the problem increase. Our computer experiment confirms the efficiency of the quadratic transformation.

Based on the research conducted, the following general conclusions can be drawn.

Firstly, in the case of modulus transformation of dual variables at high dimensions, due to the high degree of the function degeneracy, in order to obtain sufficient computational accuracy, it is necessary to use subgradient minimization methods with a change in the space metric, which is confirmed by comparing the results of the multi-step sub method and the yv method.

Secondly, the quality of the solution of the dual minimization problem with the modulus method is significantly reduced in comparison with the quadratic transformation.

Thirdly, the cost of solving the dual minimization problem with the modulus method increases significantly compared to the quadratic transformation.

Fourthly, the methods used for minimizing smooth functions (the conjugate gradient method and the quasi-Newtonian method) have results both in accuracy and in the cost of the function and gradient calculations commensurate with the best results for RSMs.

The results of solving an applied problem are commensurate in accuracy and quality with the results for similar test problems.

Author Contributions: Conceptualization, V.K. and E.T.; methodology, V.K., E.T., A.B., E.C. and P.S.; software, V.K. and A.B.; validation, L.K., E.T. and P.S.; formal analysis, E.T., P.S. and A.B.; investigation, V.K., E.T. and E.C.; resources, L.K.; data curation, P.S. and A.B.; writing—original draft preparation, V.K. and E.T.; writing—review and editing, E.T., P.S. and L.K.; visualization, V.K. and E.T.; supervision, V.K. and P.S.; project administration, P.S.; funding acquisition, P.S. and L.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2022-1121).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Zhao, T.H.; Shi, L.; Chu, Y.M. Convexity and concavity of the modified Bessel functions of the first kind with respect to Hölder means. *Rev. De La Real Acad. De Cienc. Exactas Físicas Y Naturales. Ser. A. Matemáticas* **2020**, *114*, 96. [CrossRef]
2. Zhao, T.H.; Wang, M.K.; Chu, Y.M. Monotonicity and convexity involving generalized elliptic integral of the first kind. *Rev. De La Real Acad. De Cienc. Exactas Físicas Y Naturales. Ser. A. Matemáticas* **2021**, *115*, 46. [CrossRef]
3. Khan, M.A.; Adnan; Saeed, T.; Nwaeze, E.R. A New Advanced Class of Convex Functions with Related Results. *Axioms* **2023**, *12*, 195. [CrossRef]
4. Khan, M.A.; Ullah, H.; Saeed, T. Some estimations of the Jensen difference and applications. *Math. Methods Appl. Sci.* **2022**, 1–30. [CrossRef]
5. Li, Y.; Tang, X.; Lin, X.; Grzesiak, L.; Hu, X. The role and application of convex modeling and optimization in electrified vehicles. *Renew. Sustain. Energy Rev.* **2022**, *153*, 111796. [CrossRef]
6. Gbadega, P.A.; Sun, Y. Primal–dual interior-point algorithm for electricity cost minimization in a prosumer-based smart grid environment: A convex optimization approach. *Energy Rep.* **2022**, *8*, 681–695. [CrossRef]
7. Liu, B.; Sun, C.; Wang, B.; Liang, W.; Ren, Q.; Li, J.; Sun, F. Bi-level convex optimization of eco-driving for connected Fuel Cell Hybrid Electric Vehicles through signalized intersections. *Energy* **2022**, *252*, 123956. [CrossRef]
8. Huotari, J.; Manderbacka, T.; Ritari, A.; Tammi, K. Convex optimisation model for ship speed profile: Optimisation under fixed schedule. *J. Mar. Sci. Eng.* **2021**, *9*, 730. [CrossRef]
9. Dong, C.; Yang, H.; Li, S.; Li, B. Convex optimization of asteroid landing trajectories driven by solar radiation pressure. *Chin. J. Aeronaut.* **2022**, *35*, 200–211. [CrossRef]
10. Li, Y.; Wei, C.; He, Y.; Hu, R. A convex approach to trajectory optimization for boost back of vertical take-off/vertical landing reusable launch vehicles. *J. Frankl. Inst.* **2021**, *358*, 3403–3423. [CrossRef]
11. Sagliano, M.; Mooij, E. Optimal drag-energy entry guidance via pseudospectral convex optimization. *Aerosp. Sci. Technol.* **2021**, *117*, 106946. [CrossRef]
12. Zhang, Z.; Zhao, D.; Li, X.; Kong, C.; Su, M. Convex Optimization for Rendezvous and Proximity Operation via Birkhoff Pseudospectral Method. *Aerospace* **2022**, *9*, 505. [CrossRef]
13. Li, W.; Li, W.; Cheng, L.; Gong, S. Trajectory Optimization with Complex Obstacle Avoidance Constraints via Homotopy Network Sequential Convex Programming. *Aerospace* **2022**, *9*, 720. [CrossRef]
14. Shen, Z.; Chen, Q.; Yang, F. A convex relaxation framework consisting of a primal–dual alternative algorithm for solving $\ell 0$ sparsity-induced optimization problems with application to signal recovery based image restoration. *J. Comput. Appl. Math.* **2023**, *421*, 114878. [CrossRef]
15. Popescu, C.; Grama, L.; Rusu, C. A Highly scalable method for extractive text summarization using convex optimization. *Symmetry* **2021**, *13*, 1824. [CrossRef]
16. Yu, W. Convex optimization of random dynamic voltage and frequency scaling against power attacks. *Integration* **2022**, *82*, 7–13. [CrossRef]
17. Bolívar-Nieto, E.A.; Summers, T.; Gregg, R.D.; Rezazadeh, S. A convex optimization framework for robust-feasible series elastic actuators. *Mechatronics* **2021**, *79*, 102635. [CrossRef]
18. Bottou, L.; Curtis, F.E.; Nocedal, J. Optimization methods for large-scale machine learning. *SIAM Rev.* **2018**, *60*, 223–311. [CrossRef]
19. Anikin, A.S.; Gasnikov, A.V.; Dvurechensky, P.E.; Tyurin, A.I.; Chernov, A.V. Dual approaches to the minimization of strongly convex functionals with a simple structure under affine constraints. *Comput. Math. Math. Phys.* **2017**, *57*, 1262–1276. [CrossRef]
20. Nesterov, Y. Primal-dual subgradient methods for convex problems. *Math. Program. Ser. B* **2009**, *120*, 261–283. [CrossRef]

21. Nemirovski, A.; Onn, S.; Rothblum, U.G. Accuracy certificates for computational problems with convex structure. *Math. Oper. Res.* **2010**, *35*, 52–78. [CrossRef]
22. Devolder, O. Exactness, Inexactness and Stochasticity in First-Order Methods for Large-Scale Convex Optimization. Ph.D. Thesis, CORE UCLouvain, Louvain-la-Neuve, Belgium, 2013.
23. Nesterov, Y. New primal-dual subgradient methods for convex optimization problems with functional constraints. In Proceedings of the International Workshop "Optimization and Statistical Learning", Les Houches, France, 11–16 January 2015.
24. Nesterov, Y. Complexity Bounds for Primal-Dual Methods Minimizing the Model of Objective Function. CORE Discussion Papers. 2015. Available online: https://uclouvain.be/en/research-institutes/lidam/core/core-discussion-papers.html (accessed on 27 January 2023).
25. Gasnikov, A.V.; Gasnikova, E.V.; Dvurechensky, P.E.; Ershov, E.I.; Lagunovskaya, A.A. Search for the stochastic equilibria in the transport models of equilibrium flow distribution. *Proc. MIPT* **2015**, *7*, 114–128.
26. Komodakis, N.; Pesquet, J.-C. Playing with Duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Process. Mag.* **2015**, *32*, 31–54. [CrossRef]
27. Dvurechensky, P.; Shtern, S.; Staudigl, M. First-order methods for convex optimization. *EURO J. Comput. Optim.* **2021**, *9*, 100015. [CrossRef]
28. Cohen, M.B.; Sidford, A.; Tian, K. Relative lipschitzness in extragradient methods and a direct recipe for acceleration. In Proceedings of the Innovations in Theoretical Computer Science Conference (ITCS 2021), Cambridge, MA, USA, 6–8 January 2021; pp. 62:1–62:18. [CrossRef]
29. Dvurechensky, P.E.; Gasnikov, A.V.; Nurminski, E.A.; Stonyakin, F.S. Advances in Low-Memory Subgradient Optimization. In *Numerical Nonsmooth Optimization*; Bagirov, A., Gaudioso, M., Karmitsa, N., Mäkelä, M., Taheri, S., Eds.; Springer: Cham, Switzerland, 2020; pp. 19–59.
30. Ghadimi, S.; Lan, G.; Zhang, H. Generalized uniformly optimal methods for non-linear programming. *J. Sci. Comput.* **2019**, *79*, 1854–1881. [CrossRef]
31. Tang, Y.; Qu, G.; Li, N. Semi-global exponential stability of augmented primal–dual gradient dynamics for constrained convex optimization. *Syst. Control Lett.* **2020**, *144*, 104754. [CrossRef]
32. Shor, N. *Minimization Methods for Nondifferentiable Functions*; Springer: Berlin/Heidelberg, Germany, 1985.
33. Krutikov, V.; Bykov, A.; Tovbis, E.; Kazakovtsev, L. Method for calculating the air pollution emission quotas. In *Mathematical Optimization Theory and Operations Research. MOTOR 2021*; Communications in Computer and Information Science; Strekalovsky, A., Kochetov, Y., Gruzdeva, T., Orlov, A., Eds.; Springer: Cham, Switzerland, 2021; Volume 1476, pp. 342–357. [CrossRef]
34. Polyak, B.T. *Introduction to Optimization*; Optimization Software: New York, NY, USA, 1987.
35. Dvurechensky, P.; Gasnikov, A.; Ostroukhov, P.; Uribe, C.; Ivanova, A. Near-Optimal Tensor Methods for Minimizing the Gradient Norm of Convex Function. 2019. Available online: https://arxiv.org/abs/1912.03381v3 (accessed on 27 January 2023).
36. Nesterov, Y. Inexact accelerated high-order proximal-point methods. *Math. Program.* **2023**, *197*, 1–26. [CrossRef]
37. Nesterov, Y. Inexact high-order proximal-point methods with auxiliary search procedure. *SIAM J. Comput.* **2021**, *31*, 2807–2828. [CrossRef]
38. Wolfe, P. Note on a method of conjugate subgradients for minimizing nondifferentiable functions. *Math. Program.* **1974**, *7*, 380–383. [CrossRef]
39. Lemarechal, C. An extension of Davidon methods to non-differentiable problems. *Math. Program. Study* **1975**, *3*, 95–109.
40. Krutikov, V.; Meshechkin, V.; Kagan, E.; Kazakovtsev, L. Machine Learning Algorithms of Relaxation Subgradient Method with Space Extension. In *Mathematical Optimization Theory and Operations Research. MOTOR 2021*; Lecture Notes in Computer Science; Pardalos, P., Khachay, M., Kazakov, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2021; Volume 12755, pp. 477–492.
41. Krutikov, V.N.; Stanimirovic, P.S.; Indenko, O.N.; Tovbis, E.M.; Kazakovtsev, L.A. Optimization of Subgradient Method Parameters Based on Rank-Two Correction of Metric Matrices. *J. Appl. Ind. Math.* **2022**, *16*, 427–439. [CrossRef]
42. Krutikov, V.; Gutova, S.; Tovbis, E.; Kazakovtsev, L.; Semenkin, E. Relaxation subgradient algorithms with machine learning procedures. *Mathematics* **2022**, *10*, 3959. [CrossRef]
43. Krutikov, V.N.; Vershinin, Y.N. The subgradient multistep minimization method for nonsmooth high-dimensional problems. *Vestn. Tomsk. Gos. Univ. Mat. I Mekhanika* **2014**, *3*, 5–19. (In Russian)
44. Broyden, C.G. The convergence of a class of double-rank minimization algorithms. *J. Inst. Math. Appl.* **1970**, *6*, 76–90. [CrossRef]
45. Fletcher, R. A new approach to variable metric algorithms. *Comp. J.* **1970**, *13*, 317–322. [CrossRef]
46. Goldfarb, D.F. A family of variable-metric methods derived by variational means. *Math. Comp.* **1970**, *24*, 23–26. [CrossRef]
47. Shanno, D. Conditioning of quasi-Newton methods for function minimization. *Math. Comp.* **1970**, *24*, 647–656. [CrossRef]
48. Yu, B.; Lee, W.S.; Rafiq, S. Air Pollution Quotas and the Dynamics of Internal Skilled Migration in Chinese Cities. IZA Discussion Paper Series. Available online: https://www.iza.org/publications/dp/13479/air-pollution-quotas-and-the-dynamics-of-internal-skilled-migration-in-chinese-cities (accessed on 2 February 2023).
49. Zhao, C.; Kahn, M.E.; Liu, Y.; Wang, Z. The consequences of spatially differentiated water pollution regulation in China. *J. Environ. Econ. Manag.* **2018**, *88*, 468–485. [CrossRef]
50. Yu, S.; Wei, Y.M.; Wang, K. Provincial allocation of carbon emission reduction targets in China: An approach based on improved fuzzy cluster and Shapley value decomposition. *Energy Policy* **2014**, *66*, 630–644. [CrossRef]

51. Pozo, C.; Galan-Martin, A.; Cortes-Borda, D.; Sales-Pardo, M.; Azapagic, A.; Guimera, R.; Guillen-Gosalbez, G. Reducing global environmental inequality: Determining regional quotas for environmental burdens through systems optimization. *J. Clean. Prod.* **2020**, *270*, 121828. [CrossRef]
52. Liu, H.; Lin, B. Cost-based modeling of optimal emission quota allocation. *J. Clean. Prod.* **2017**, *149*, 472e484. [CrossRef]
53. Wang, X.; Cai, Y.; Xu, Y.; Zhao, H.; Chen, J. Optimal strategies for carbon reduction at dual levels in China based on a hybrid nonlinear grey-prediction and quota-allocation model. *J. Clean. Prod.* **2014**, *83*, 185–193. [CrossRef]
54. Nugroho, S.A.; Taha, A.F.; Hoang, V. Nonlinear Dynamic Systems Parameterization Using Interval-Based Global Optimization: Computing Lipschitz Constants and Beyond. *IEEE Trans. Automat. Contr.* **2022**, *67*, 3836–3850. [CrossRef]
55. The ERA-AIR Software Complex. Available online: https://lpp.ru/ (accessed on 18 February 2023).
56. Azhiganich, T.E.; Alekseichenko, T.G.; Bykov, A.A. Conducting summary calculations of atmospheric pollution in Kemerovo for emission regulation and diagnostic assessments. In Proceedings of the V City Scientific and Practical Conference, Kemerovo, Russia, 17–18 November 2003; pp. 41–45.
57. Ministry of Natural Resources of the Russian Federation. *The Order of the Ministry of Natural Resources of the Russian Federation Dated 29.11.2019 N 814 "On Approval of the Rules for Emissions Quotas of Pollutants (except Radioactive Substances) into the atmosphere" (Registered with the Ministry of Justice of the Russian Federation on 24.12.2019 n 56956)*; Ministry of Natural Resources of the Russian Federation: Moscow, Russia, 2019.
58. Russian State Ecological Committee. *The Order 16.02.1999 N 66. Regulations on Using the Sistem of Aggregate Calculations of Atmospheric Pollution for Finding Admissible Quotas of Industrial and Motor Transport Emissions*; Russian State Ecological Committee: Moscow, Russia, 1999.
59. Meshechkin, V.V.; Bykov, A.A.; Krutikov, V.N.; Kagan, E.S. Distributive Model of Maximum Permissible Emissions of Enterprises into the Atmosphere and Its Application. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *224*, 012019. [CrossRef]