High Performance Algorithms and Software for Nonlinear Optimization, pp. 223–237
G. Di Pillo and A. Murli, Editors

New perspective on the theorems of alternative

Yu.G. Evtushenko (evt@ccas.ru)

A.I. Golikov (gol@ccas.ru)
Computing Centre of Russian Academy of Sciences,
Moscow, Russia

Abstract

New general theorems of the alternative are presented. The constructive proofs based on the duality theory are given. From these results many well-known theorems of the alternative are obtained by simple substitutions. Computational applications of theorems of the alternative to solving linear systems, LP and NLP problems are given. A linear systems of possibly unsolvable equalities and inequalities are considered. With original linear system an alternative system is associated such that one and only one of these systems is consistent. If the original system is solvable then numerical method for solving this system consists of minimization of the residual of the alternative inconsistent system. From the results of this minimization the normal solution of the original system is determined.

Keywords: theorems of the alternative, duality theory, alternative system, normal solution, inconsistent system, steepest descent, linear programming.

1. Introduction

Among vast literature on theorems of the alternative we mention only few publications [1]–[7]. First papers in this field was published in 1873 by P. Gordan. Theorems of the alternative (TA) were extended and widely used for proving existence and uniqueness of some theorems of linear algebra, differential and integral equations. In [7] C. Broyden writes: "Theorems of the alternative lie at the heart of mathematical programming". TA were employed to derive necessary optimality conditions for LP and NLP problems and for various other pure theoretical investigations.

In this paper we present new quite general TA. The constructive proofs based on the duality theory is given. From our results many well-known TA can be obtained by simple substitutions (Fredholm's TA, Gale's TA, Jordan's TA, Farkas's TA and many others). We consider these results as a by-product and suppose that the main result of this paper is computational applications of TA. Below we will show that TA have computational implications. TA give us an opportunity to construct new numerical methods for solving linear systems, to simplify computations arising in NLP when the steepest descent method is used, to reduce the finding of normal (least-square) solution of LP problem to unconstrained minimization of convex piecewise-quadratic differentiable function.

We consider a linear system of possibly inconsistent equalities and inequalities. Systems solving are covered in an extensive literature. We mention only [1, 2]. Usually, these systems are solved by the reduction to unconstrained minimization of the residual (a measure of satisfaction of feasibility of the original system). It is not known a priori whether an original system has a solution. Thus, the problem is, firstly, to determine whether the given system is solvable and, secondly, in case if it is solvable to find its normal solution (a solution with minimal Euclidean norm).

In Section 2 we prove 3 key theorems of the alternative. With original linear system we associate an alternative system such that one and only one of these systems is consistent. Moreover an alternative system is such that the dimension of its variable equals to the total amount of equalities and inequalities (except constraints on the signs of variables) in the original system.

If the original system is solvable then numerical method for solving this system consists of minimization of the residual of the alternative inconsistent system. From the results of this minimization we determine a normal solution of the original system. If the original system is not solvable then numerical method for solving alternative system consists of minimization of the residual of the original inconsistent system. From the results of this minimization we determine a normal solution of the alternative system.

Since the dimensions of the variables in original and alternative systems are different, the passage from the original consistent system to the minimization problem for the residual of the alternative inconsistent system may be very reasonable. This reduction may lead to the minimization problem with respect to variables of lower dimension and makes it possible to determine easy a normal solution of the original system. Proposed technique does not need an a priori assumption regarding the consistency of the original system. The essence of this approach is based on the duality theory.

In Section 3 we show that the proposed method can be utilized to determine a steepest descent direction in the feasible direction method. In Section 4 the theorems of alternative are used for solving LP problems.

2. Main theorems

Let A be an $m \times n$ matrix
$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$
where A_{11}, A_{12}, A_{21}, and A_{22} are arbitrary rectangular matrices of sizes $m_1 \times n_1$, $m_1 \times n_2$, $m_2 \times n_1$, and $m_2 \times n_2$, respectively. Suppose that vectors $x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, and $b \in \mathbb{R}^m$ admit decompositions $x^\top = [x_1^\top, x_2^\top]$, $u^\top = [u_1^\top, u_2^\top]$, and $b^\top = [b_1^\top, b_2^\top]$, where $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$, $n = n_1 + n_2$, $u_1 \in \mathbb{R}^{m_1}$, $u_2 \in \mathbb{R}^{m_2}$, $b_1 \in \mathbb{R}^{m_1}$, $b_2 \in \mathbb{R}^{m_2}$, and $m = m_1 + m_2$. Let us introduce the auxiliary sets
$$\begin{aligned} \Pi_x &= \{[x_1, x_2] : x_1 \in \mathbb{R}^{n_1}_+, \;\; x_2 \in \mathbb{R}^{n_2}\}, \\ \Pi_u &= \{[u_1, u_2] : u_1 \in \mathbb{R}^{m_1}_+, \;\; u_2 \in \mathbb{R}^{m_2}\}, \\ \Pi_z &= \{[z_1, z_2] : z_1 \in \mathbb{R}^{m_1}_+, \;\; z_2 \in \mathbb{R}^{m_2}\}. \end{aligned}$$

Consider the system of linear equalities and inequalities
$$A_{11}x_1 + A_{12}x_2 \geq b_1, \quad A_{21}x_1 + A_{22}x_2 = b_2, \qquad x_1 \geq 0_{n_1}. \tag{I}$$

We define the system conjugate to (I) as
$$A_{11}^\top z_1 + A_{21}^\top z_2 \leq 0_{n_1}, \quad A_{12}^\top z_1 + A_{22}^\top z_2 = 0_{n_2}, \qquad z_1 \geq 0_{m_1}, \tag{I$'$}$$

and the system alternative to (I) as
$$A_{11}^\top u_1 + A_{21}^\top u_2 \leq 0_{n_1}, \quad A_{12}^\top u_1 + A_{22}^\top u_2 = 0_{n_2}, \qquad b_1^\top u_1 + b_2^\top u_2 = \rho, \quad u_1 \geq 0_{m_1}. \tag{II}$$

Here, $\rho > 0$ is an arbitrary fixed positive number. Note that the requirement that ρ be positive automatically implies that $\|b\| \neq 0$.

Let us introduce a vector $w \in \mathbb{R}^{n+1}$ representable in the form

$$w^\top = [w_1^\top, w_2^\top, w_3],$$

where $w_1 \in \mathbb{R}^{n_1}$, $w_2 \in \mathbb{R}^{n_2}$, and $w_3 \in \mathbb{R}^1$, and the auxiliary set

$$\Pi_w = \{[w_1, w_2, w_3] : w_1 \in \mathbb{R}^{n_1}_+,\ w_2 \in \mathbb{R}^{n_2},\ w_3 \in \mathbb{R}^1\}.$$

The system conjugate to (II) has the form

$$A_{11}w_1 + A_{12}w_2 - b_1w_3 \geq 0_{m_1}, \quad A_{21}w_1 + A_{22}w_2 - b_2w_3 = 0_{m_2}, \qquad w_1 \geq 0_{n_1}. \qquad \text{(II}'\text{)}$$

We denote the solution sets of (I), (I′), (II), and (II′) by X, Z, U, and W, respectively. Unlike (I) and (II), systems (I′) and (II′) always have solutions, because $0_m \in Z$ and $0_{n+1} \in W$.

Let $\text{pen}\,(x, X)$ denote the penalty at a point $x \in \Pi_x$ for violation of the condition $x \in X$. As the penalty, we use the Euclidean norm of the residual vector:

$$\text{pen}\,(x, X) = \left[\|(b_1 - A_{11}x_1 - A_{12}x_2)_+\|^2 + \|b_2 - A_{21}x_1 - A_{22}x_2\|^2\right]^{1/2}.$$

Similarly, we define

$$\text{pen}\,(u, U) = \left[\|(A_{11}^\top u_1 + A_{21}^\top u_2)_+\|^2 + \|A_{12}^\top u_1 + A_{22}^\top u_2\|^2 + (\rho - b_1^\top u_1 - b_2^\top u_2)^2\right]^{1/2}.$$

Hereinafter follows, a_+ is the nonnegative part of the vector a; i.e., the ith component of the vector a_+ coincides with the ith component of the vector a if it is nonnegative; otherwise, this component is zero.

Consider the following four quadratic problems:

$$I_1 = \min_{x \in \Pi_x} [\text{pen}\,(x, X)]^2/2, \qquad (2.1)$$

$$I_2 = \min_{u \in \Pi_u} [\text{pen}\,(u, U)]^2/2, \qquad (2.2)$$

$$I_1^d = \max_{z \in Z} \{b^\top z - \|z\|^2/2\}, \qquad (2.3)$$

$$I_2^d = \max_{w \in W} \{\rho w_3 - \|w\|^2/2\}. \qquad (2.4)$$

The sets Z and W are always nonempty, because they contain zero vectors. Unlike systems (I) and (II), which may be solvable or not, problems (2.1) – (2.4) always have solutions. Moreover, problems (2.3) and (2.4) always have unique solutions, because their feasible sets Z and W are nonempty and strictly concave quadratic objective functions are bounded from above.

Lemma 1. *The problems dual to problems* (2.3) *and* (2.4) *are the problems* (2.1) *and* (2.2), *respectively. Each problem* (2.1) *and* (2.2) *can be converted to an equivalent quadratic constrained minimization problems, which are dual to* (2.3) *and* (2.4), *respectively.*

Proof. The first statement follows immediately from traditional representations of the dual problem.

The second statement follows from two-step representation of problems (2.1) and (2.2). With the help of additional variables it is possible to construct artificial quadratic constrained minimization problems which are equivalent to problems (2.1) and (2.2) and dual to (2.3) and (2.4), respectively. Such an approach was used in papers [8, 9].

Let us introduce additional variables $y \in \mathbb{R}^m$, $y^\top = [y_1^\top, y_2^\top]$, where $y_1 \in \mathbb{R}^{m_1}$, $y_2 \in \mathbb{R}^{m_2}$ be such that

$$y_1 = b_1 - A_{11}x_1 - A_{12}x_2, \qquad y_2 = b_2 - A_{21}x_1 - A_{22}x_2.$$

Then the problem (2.1) can be replaced by equivalent problem

$$I_1 = \min_{[x,y]\in G} f(y), \tag{2.5}$$

where the goal function and feasible set are

$$f(y) = \|(y_1)_+\|^2/2 + \|y_2\|^2/2,$$

$$G = \{[x, y] : A_{11}x_1 + A_{12}x_2 + y_1 = b_1,\ A_{21}x_1 + A_{22}x_2 + y_2 = b_2,\ x \in \Pi_x\}.$$

In contrast to set X, the set G is always nonempty.

For quadratic programming problem (2.5) we introduce the Lagrange multiplier vector $z \in$ $\in \Pi_z$ and define the Lagrange function as

$$L(x, y, z) = f(y) + z_1^\top(b_1 - A_{11}x_1 - A_{12}x_2 - y_1) + z_2^\top(b_2 - A_{21}x_1 - A_{22}x_2 - y_2).$$

By the simple rearrangement, we can rewrite the Lagrange function as the following:

$$\begin{aligned} L(x, y, z) &= f(y) - x_1^\top(A_{11}^\top z_1 + A_{21}^\top z_2) - x_2^\top(A_{12}^\top z_1 + A_{22}^\top z_2) + \\ &+ z_1^\top(b_1 - y_1) + z_2^\top(b_2 - y_2). \end{aligned} \tag{2.6}$$

We define the Lagrange dual function as the minimum value of the Lagrange function (2.6) over $x \in \Pi_x$ and $y \in \mathbb{R}^m$:

$$F(z) = \min_{x\, \in\, \Pi_x} \min_{y\in\mathbb{R}^n} L(x, y, z). \tag{2.7}$$

Consider the Lagrange dual problem associated with the problem (2.5)

$$\max_{z\in\Pi_z} F(z).$$

Necessary and sufficient saddle point optimality conditions for problem (2.7) can be written as

$$\begin{aligned} L_{x_1}(x, y, z) &= -A_{11}^\top z_1 - A_{21}^\top z_2 \geq 0_{n_1}, \quad D(x_1)(A_{11}^\top z_1 + A_{21}^\top z_2) = 0_{n_1}, \quad x_1 \geq 0_{n_1}, \quad (2.8)\\ L_{x_2}(x, y, z) &= -A_{12}^\top z_1 - A_{22}^\top z_2 = 0_{n_2}, \quad (2.9)\\ L_{y_1}(x, y, z) &= (y_1)_+ - z_1 = 0_{m_1}, \qquad L_{y_2}(x, y, z) = y_2 - z_2 = 0_{m_2}. \quad (2.10) \end{aligned}$$

Hereinafter $D(z)$ denotes the diagonal matrix whose entries are the components of the vector z.

If $z \in \Pi_z$ then from (2.10) we obtain that $z = y$. Let's substitute this relation in (2.6) and impose the condition $z \in Z$.

Then due to definition (2.7) and optimality conditions (2.8), (2.9) the Lagrange dual function takes the form $F(z) = b^\top z - \|z\|^2/2$. Thus, we obtain problem (2.3), which is dual to (2.5). Taking into account that (2.5) is equivalent to (2.1), we can say that (2.3) is dual to (2.1). More generally we say that problems (2.1) and (2.3), problems (2.2) and (2.4) are mutually dual respectively. □

According to strong duality theorem the optimum objective values are the same in problems (2.1) and (2.3), (2.2) and (2.4), therefore

$$I_1 = I_1^d, \qquad I_2 = I_2^d. \tag{2.11}$$

We refer to systems (I) and (II) as *alternative*, if either system (I) has a solution, or system (II) has a solution, but never both.

Lemma 2. *Systems* (I) *and* (II) *are not solvable simultaneously.*

Theorem 3 below implies that there exists a solution of precisely (I) or (II). Therefore, these systems are alternative. The original system (I) belongs to the class of systems which are alternative to (II).

We define a projection of a point $\bar{x}$ on a nonempty closed set X as a point $x_* \in X$ nearest to $\bar{x}$, i.e., such that x_* solves the problem

$$\min_{x \in X} \|\bar{x} - x\|^2/2 = \|\bar{x} - x^*\|^2/2.$$

Hence, we can write $x_* = \mathrm{pr}\,(\bar{x}, X)$; the distance from $\bar{x}$ to X is denoted by $\mathrm{dist}\,(\bar{x}, X) = \|x^* - \bar{x}\|$.

Theorem 1. *Each solution* x^* *to problem* (2.1) *determines a unique solution* $z^{*\top} = [z_1^{*\top}, z_2^{*\top}]$ *to problem* (2.3) *as*

$$z_1^* = (b_1 - A_{11}x_1^* - A_{12}x_2^*)_+, \qquad z_2^* = b_2 - A_{21}x_1^* - A_{22}x_2^*, \tag{2.12}$$

and the following assertions are valid

$$\|z^*\|^2 = b^\top z^*, \tag{2.13}$$

$$z^* \perp Ax^*, \qquad z^* \perp (b - z^*), \tag{2.14}$$

$$z^* = \mathrm{pr}\,(b, Z), \qquad \|z^*\| = \mathrm{pen}\,(x^*, X), \qquad \|b - z^*\| = \mathrm{dist}\,(b, Z), \tag{2.15}$$

$$[\mathrm{pen}\,(x^*, X)]^2 + [\mathrm{dist}\,(b, Z)]^2 = \|b\|^2. \tag{2.16}$$

Proof. Necessary and sufficient conditions for optimality in problem (2.1) are following

$$\begin{array}{l} -A_{11}^\top(b_1 - A_{11}x_1^* - A_{12}x_2^*)_+ - A_{21}^\top(b_2 - A_{21}x_1^* - A_{22}x_2^*) \ge 0_{n_1}, \\ D(x_1^*)[(A_{11}^\top(b_1 - A_{11}x_1^* - A_{12}x_2^*)_+ + A_{21}^\top(b_2 - A_{21}x_1^* - A_{22}x_2^*)] = 0_{n_1}, \qquad x_1^* \ge 0_{n_1}, \\ A_{12}^\top(b_1 - A_{11}x_1^* - A_{12}x_2^*)_+ + A_{22}^\top(b_2 - A_{21}x_1^* - A_{22}x_2^*) = 0_{n_2}. \end{array} \tag{2.17}$$

In formulas (2.17) we introduce the notations

$$z_1^* = (b_1 - A_{11}x_1^* - A_{12}x_2^*)_+, \qquad z_2^* = b_2 - A_{21}x_1^* - A_{22}x_2^*. \tag{2.18}$$

Now we show, that $z^{*\top} = [z_1^{*\top}, z_2^{*\top}]$ is a unique solution of a constrained quadratic problem (2.3). Let us rewrite the optimality conditions (2.17) as

$$A_{11}^\top z_1^* + A_{21}^\top z_2^* \le 0_{n_1}, \qquad A_{12}^\top z_1^* + A_{22}^\top z_2^* = 0_{n_2}, \tag{2.19}$$

$$D(x_1^*)(A_{11}^\top z_1^* + A_{21}^\top z_2^*) = 0_{n_1}, \qquad x_1^* \ge 0_{n_1}. \tag{2.20}$$

From (2.18) and (2.19) we conclude that $z^* \in Z$. Premultiplying the first formula in (2.18) by z_1^*, second formula by z_2^*, we obtain

$$\begin{array}{l} \|z_1^*\|^2 = z_1^{*\top}(b_1 - A_{11}x_1^* - A_{12}x_2^*)_+ = z_1^{*\top}(b_1 - A_{11}x_1^* - A_{12}x_2^*) = b_1^\top z_1^* - x_1^{*\top}A_{11}^\top z_1^* - x_2^{*\top}A_{12}^\top z_1^*, \\ \|z_2^*\|^2 = z_2^{*\top}(b_2 - A_{21}x_1^* - A_{22}x_2^*) \;\; = b_2^\top z_2^* - x_1^{*\top}A_{21}^\top z_2^* - x_2^{*\top}A_{22}^\top z_2^*. \end{array}$$

Combining these results, we obtain

$$\begin{aligned}\|z^*\|^2 = \|z_1^*\|^2 + \|z_2^*\|^2 &= b_1^\top z_1^* + b_2^\top z_2^* - x_1^{*\top}(A_{11}^\top z_1^* + A_{21}^\top z_2^*) - x_2^{*\top}(A_{12}^\top z_1^* + A_{22}^\top z_2^*) = \\ &= (b - Ax^*)^\top z^* = b^\top z^*.\end{aligned} \tag{2.21}$$

In (2.21) we took into account (2.19) and (2.20), therefore $x_1^{*\top}(A_{11}^\top z_1^* + A_{21}^\top z_2^*) + x_2^{*\top}(A_{12}^\top z_1^* + A_{22}^\top z_2^*) = x^{*\top} A^\top z^* = 0$. Hence the equality (2.13) is proved and it is shown, that

$$z^{*\top} A x^* = 0,$$

i.e. the vectors z^* and Ax^* are orthogonal. The condition (2.13) is equivalent to $z^{*\top}(z^* - b) = 0$. We have proved both statements in (2.14).

The Lagrange function associated with problem (2.3) is

$$L(z, x) = b^\top z - \|z\|^2/2 - x_1^\top (A_{11}^\top z_1 + A_{21}^\top z_2) - x_2^\top (A_{12}^\top z_1 + A_{22}^\top z_2) \tag{2.22}$$

Kuhn–Tucker optimality conditions yield

$$L_{z_1}(z, x) = b_1 - z_1 - A_{11}x_1 - A_{12}x_2 \le 0_{m_1}, \tag{2.23}$$
$$D(z_1)(b_1 - z_1 - A_{11}x_1 - A_{12}x_2) = 0_{m_1}, \qquad z_1 \ge 0_{m_1}, \tag{2.24}$$
$$L_{z_2}(z, x) = b_2 - z_2 - A_{21}x_1 - A_{22}x_2 = 0_{m_2}, \tag{2.25}$$
$$L_{x_1}(z, x) = -(A_{11}^\top z_1 + A_{21}^\top z_2) \ge 0_{n_1}, \qquad x_1 \ge 0_{m_1}, \quad D(x_1)(A_{11}^\top z_1 + A_{21}^\top z_2) = 0_{n_1}, \tag{2.26}$$
$$L_{x_2}(z, x) = -(A_{12}^\top z_1 + A_{22}^\top z_2) = 0_{n_2}. \tag{2.27}$$

Let us compare necessary and sufficient conditions (2.18) – (2.20) for a problem (2.1) with necessary and sufficient optimality conditions (2.23) – (2.27) for quadratic programming problem (2.3). If in conditions (2.23) – (2.27) we substitute vector x^* for x and substitute the vector z^* defined in (2.18) for z, then (2.26), (2.27) transform into (2.19), (2.20). Hence, formulas (2.18) ensure conditions (2.23) – (2.25). Thus, the saddle point $[z^*, x^*]$ of the Lagrange function (2.22) consists of vector z^* which is the solution of problem (2.3) and vector x^* which is the solution of problem (2.1), and between these vectors exists relation given by formulas (2.12).

From (2.3), (2.11), (2.13) we obtain

$$I_1 = I_1^d = \|z^*\|^2/2 = [\text{pen}\,(x^*, X)]^2/2.$$

From this, we get the second statement in (2.15).

We shall transform problem (2.3) to the following equivalent problems:

$$I_1^d = \max_{z \in Z} \frac{[-\|b - z\|^2 + \|b\|^2]}{2} = \frac{\|b\|^2}{2} - \min_{z \in Z} \frac{\|b - z\|^2}{2}.$$

Since z^* is a unique solution of a problem (2.3), it follows that $z^* = \text{pr}\,(b, Z)$ and $\|b - z^*\| = \text{dist}\,(b, Z)$. Hence all statements (2.15) are proved.

From (2.13) we conclude that the vector z^* is orthogonal to the vector $b - z^*$. Therefore, three points: an origin of $\mathbb{R}^m$, points z^* and b form a rectangular triangle, at which vector b defines a hypotenuse, vector z^* is a cathetus which length is equal to pen (x^*, X), vector $b - z^*$ is the second cathetus with length equal to dist (b, Z). Pyphagoras' law yields (2.16). □

From Theorem 1 follows

Criterion 1. *System* (I) *is solvable if and only if problem* (2.3) *has a zero solution* $z^* = 0_m$.

This criterion is possible to reformulated it as: system (I) is unsolvable if and only if problem (2.3) has a nonzero solution $\|z^*\| = 0$.

The analysis of problems (2.2) and (2.4) is similar, therefore we give only the formulation of corresponding theorem. Consider the vector $r \in \mathbb{R}^{n+1}$ such that $r^\top = [0_n^\top, \rho]$ and the $m \times (n+1)$ matrix $\bar{A} = [-A, b]$.

Theorem 2. *Suppose that $u^{*\top} = [u_1^{*\top}, u_2^{*\top}]$ is an arbitrary solution of problem (2.2). Then the solution $w^{*\top} = [w_1^{*\top}, w_2^{*\top}, w_3^*]$ of problem (2.4) depends on u^* as*

$$w_1^* = (A_{11}^\top u_1^* + A_{21}^\top u_2^*)_+, \quad w_2^* = A_{12}^\top u_1^* + A_{22}^\top u_2^*, \quad w_3^* = \rho - b_1^\top u_1^* - b_2^\top u_2^* \tag{2.28}$$

and the following assertions are valid:

$$\|w^*\|^2 = \rho w_3^*, \tag{2.29}$$

$$w^* \perp \hat{A}^\top u^*, \qquad w^* \perp (r - w^*), \tag{2.30}$$

$$w^* = \mathrm{pr}\,(r, W), \qquad \|w^*\| = \mathrm{pen}\,(u^*, U), \qquad \|r - w^*\| = \mathrm{dist}\,(r, W), \tag{2.31}$$

$$[\mathrm{pen}\,(u^*, U)]^2 + [\mathrm{dist}\,(r, W)]^2 = \|r\|^2, \tag{2.32}$$

$$\|w^*\| \leq \rho, \quad 0 \leq w_3^* \leq \rho, \quad \|w_1^*\|^2 + \|w_2^*\|^2 \leq \rho^2/4. \tag{2.33}$$

Criterion 2. *System (II) is solvable (is unsolvable) if and only if problem (2.4) has a zero (nonzero) solution w^*.*

Theorem 3. *Systems (I) and (II) are alternative, i.e. one and only one of them is consistent, moreover*

1) *if system (I) has no solutions then the normal solution $\tilde{u}^* = \mathrm{pr}\,(0_m, U)$ of system (II) is defined by formula*

$$\tilde{u}^* = \rho z^* / \|z^*\|^2, \tag{2.34}$$

where the unique vector z^* solves problem (2.3) and is related with an arbitrary solution x^* of problem (2.1) by (2.12);

2) *if system (II) has no solutions then the normal solution $\tilde{x}^*$ of system (I) is defined by $\tilde{x}^* = \mathrm{pr}\,(0_n, X)$ and has components*

$$\tilde{x}_1^* = w_1^*/w_3^*, \qquad \tilde{x}_2^* = w_2^*/w_3^*, \tag{2.35}$$

where the unique vector w^* solves problem (2.4) and is related with arbitrary solution u^* of problem (2.2) by (2.28).

Proof. Due to Lemma 2 the systems (I) and (II) are not solvable simultaneously. Let us show, that one of them is necessarily solvable. We consider two cases.

1. Let $X = \emptyset$, then $\mathrm{pen}\,(x^*, X) \neq 0$. Vector $z^* \in Z$ defined from (2.12) is such that $\|z^*\| \neq 0$. Multiplying both sides in (2.34) by b, and taking into account (2.13), we obtain $b^\top \tilde{u}^* = \rho$, therefore, $\tilde{u}^* \in U$, $U \neq \emptyset$. Let's show, that $\tilde{u}^*$ is the normal unique solution of system (II), i.e. $\tilde{u}^*$ solves the following quadratic problem:

$$\min_{u \in U} \|u\|^2/2. \tag{2.36}$$

The Lagrange function associated with problem (2.36) is expressed as

$$L(u, \hat{x}) = \|u\|^2/2 + \hat{x}_1^\top (A_{11}^\top u_1 + A_{21}^\top u_2) + \hat{x}_2^\top (A_{12}^\top u_1 + A_{22}^\top u_2) + \hat{x}_3(\rho - b_1^\top u_1 - b_2^\top u_2).$$

Maximizing the Lagrange dual function with respect to multiplier vector we obtain the following problem, which is dual to (2.36).

$$\max_{\hat{x}_1 \in \mathbb{R}^{n_1}_+} \max_{\hat{x}_2 \in \mathbb{R}^{n_2}} \max_{\hat{x}_3 \in \mathbb{R}^1} \left[\rho \hat{x}_3 - \frac{\|(b_1 \hat{x}_3 - A_{11}\hat{x}_1 - A_{12}\hat{x}_2)_+\|^2}{2} - \frac{\|b_2 \hat{x}_3 - A_{21}\hat{x}_1 - A_{22}\hat{x}_2\|^2}{2} \right]. \quad (2.37)$$

Let $u^{*\top} = [u_1^{*\top}, u_2^{*\top}]$ be solution of problem (2.36) and $\hat{x}^{*\top} = [\hat{x}_1^{*\top}, \hat{x}_2^{*\top}, \hat{x}_3^*]$ be solution of problem (2.37). The pair $[u^*, \hat{x}^*]$ forms a saddle point of the Lagrange function. This pair satisfies the Kuhn–Tucker conditions that is

$$u_1^* + A_{11}\hat{x}_1^* + A_{12}\hat{x}_2^* - b_1\hat{x}_3^* \geq 0_{m_1}, \quad D(u_1^*)(u_1^* + A_{11}\hat{x}_1^* + A_{12}\hat{x}_2^* - b_1\hat{x}_3^*) = 0, \quad u_1^* \geq 0_{m_1}, \quad (2.38)$$

$$u_2^* + A_{21}\hat{x}_1^* + A_{22}\hat{x}_2^* - b_2\hat{x}_3^* = 0_{m_2}, \quad (2.39)$$

$$A_{11}^\top u_1^* + A_{21}^\top u_2^* \leq 0_{n_1}, \quad D(\hat{x}_1^*)(A_{11}^\top u_1^* + A_{21}^\top u_2^*) = 0_{n_1}, \quad \hat{x}_1^* \geq 0_{n_1}, \quad (2.40)$$

$$A_{12}^\top u_1^* + A_{22}^\top u_2^* = 0_{n_2}, \quad (2.41)$$

$$\rho - b_1^\top u_1^* - b_2^\top u_2^* = 0. \quad (2.42)$$

From (2.38) and (2.39) we obtain, that the solution u^* of problem (2.36) is related to solution $\hat{x}^*$ of problem (2.37) by

$$u_1^* = (b_1\hat{x}_3^* - A_{11}\hat{x}_1^* - A_{12}\hat{x}_2^*)_+, \quad u_2^* = b_2\hat{x}_3^* - A_{21}\hat{x}_1^* - A_{22}\hat{x}_2^*.$$

Using these relations and taking into account the equality of optimal values of the goal functions of primal problem (2.36) and dual problem (2.37) we have $\|u^*\|^2 = \rho\hat{x}_3^*$. Since $U \neq \emptyset$ and $u^* \in U$, from condition $b^\top u^* = \rho > 0$ we get $\|u^*\| \neq 0$, which yields $\hat{x}_3^* > 0$.

Let us transform variables in (2.38) – (2.41)

$$u^* = \hat{x}_3^* z^*, \qquad \hat{x}_1^* = \hat{x}_3^* x_1^*, \qquad \hat{x}_2^* = \hat{x}_3^* x_2^*$$

Dividing expressions (2.38) – (2.41) by positive value $\hat{x}_3^*$ we obtain Kuhn–Tucker conditions (2.23) – (2.27) for problem (2.3) which are calculated at point $[z^*, x^*]$. Substituting the value $u^* = \hat{x}_3^* z^*$ into (2.42) and using (2.13) we obtain

$$\frac{\rho}{\hat{x}_3^*} - b^\top z^* = \frac{\rho}{\hat{x}_3^*} - \|z^*\|^2.$$

Hence, if $\hat{x}_3^* = \rho/\|z^*\|^2$, then $u^* = \hat{x}_3^* z^* = \rho z^*/\|z^*\|^2 = \tilde{u}^*$, i.e. the normal solution of (II) is expressed by formula (2.34).

2. Let $U = \emptyset$, then $\mathrm{pen}\,(u^*, U) \neq 0$. Vector $w^* \in W$ defined by (2.28) is such that $\|w^*\| \neq 0$. Owing to conditions (2.29) we have $w_3^* > 0$. Vector $\tilde{x}^*$ defined by (2.35) satisfies (I).

Now we prove that $\tilde{x}^*$ is the normal solution of system (I), i.e. it solves

$$\min_{x \in X} \|x\|^2.$$

Let us define the Lagrange function for this problem by

$$L(x, \mu) = \frac{\|x\|^2}{2} + \mu_1^\top (b_1 - A_{11}x_1 - A_{12}x_2) + \mu_2^\top (b_2 - A_{21}x_1 - A_{22}x_2)$$

We denote $\tilde{x}^{*\top} = [\tilde{x}_1^{*\top}, \tilde{x}_2^{*\top}]$, $\tilde{\mu}^{*\top} = [\tilde{\mu}_1^{*\top}, \tilde{\mu}_2^{*\top}]$. Let $[\tilde{x}^*, \tilde{\mu}^*]$ be the saddle point of the Lagrange function $L(x, \mu)$. Then the Kuhn–Tucker conditions can be written as

$$\begin{array}{lll} \tilde{x}_1^* - A_{11}^\top \mu_1^* - A_{21}^\top \mu_2^* \geq 0_{n_1}, & D(\tilde{x}_1^*)(\tilde{x}_1^* - A_{11}^\top \mu_1^* - A_{21}^\top \mu_2^*) = 0_{n_1}, & \tilde{x}_1^* \geq 0_{n_1}, \\ \tilde{x}_2^* - A_{12}^\top \mu_1^* - A_{22}^\top \mu_2^* = 0_{n_2}, & & \\ b_1 - A_{11}\tilde{x}_1^* - A_{12}\tilde{x}_2^* \leq 0_{m_1}, & D(\mu_1^*)(b_1 - A_{11}\tilde{x}_1^* - A_{12}\tilde{x}_2^*) = 0_{m_1}, & \mu_1^* \geq 0_{m_1}, \\ b_2 - A_{21}\tilde{x}_1^* - A_{22}\tilde{x}_2^* = 0_{m_2}. & & \end{array}$$

These conditions can be found if we write the Kuhn–Tucker conditions for problem (2.4) and denote $x_1^* = w_1^*/w_3^*$, $x_2^* = w_2^*/w_3^*$, $\mu_1^* = u_1^*/w_3^*$, $\mu_2^* = u_2^*/w_3^*$. Thus, we obtain that vector $\tilde{x}^*$, with components (2.35) is a normal solution of systems (I). □

Alternative system (II) admit various representations. As follows from the theorems stated above, the system alternative to (I) is obtained from conjugate system (I′) by adding a condition excluding the trivial solution of problem (2.3). For example, we can require that the solutions to conjugate system (I′) satisfy the condition $b^\top u > 0$ (as in the Farkas alternative) or the condition $b^\top u = 1$ (as in the Gale alternative), and so on. If system (I) has no inequalities then instead of condition $\rho > 0$ we can confine ourselves only by condition $\rho \neq 0$.

3. Finding of the steepest descent in feasible directions method

Let z^* be a nonzero solution of problem (2.3). Introduce normalized vectors $z_n = z/\|z^*\|$, $z_n^* = z^*/\|z^*\|$ and define a feasible set of the normalized vectors

$$Z_n = \{z_n \in \mathbb{R}^m : z_n \in Z,\ \|z_n\| = 1\},$$

where set Z is the conjugate set (I′).

Consider the problem

$$I_3 = \max_{z_n \in Z_n} b^\top z_n. \tag{3.1}$$

Theorem 4. *Let x^* be an arbitrary solution of problem* (2.1)*, the vector z^*, defined by formula* (2.12)*, is the corresponding unique solution of problem* (2.3)*, moreover $\|z^*\| \neq 0$. Then the vector $z_n^* = z^*/\|z^*\|$ is a solution of problem* (3.1) *and $I_3 = b^\top z_n^* = \|z^*\|$.*

Proof. We apply strong duality theorem to mutually dual problems (2.1) and (2.3). It follows that:

$$\frac{1}{2} = \max_{z_n \in Z} \left[\frac{b^\top z_n}{\|z^*\|} - \frac{\|z_n\|^2}{2} \right].$$

Since $Z_n \subset Z$ this relation implies

$$\|z^*\| \geq \max_{z_n \in Z_n} b^\top z_n. \tag{3.2}$$

If as z_n we take vector $z^*/\|z^*\|$, then owing to (2.13) we obtain equality in (3.2). So, we conclude that z_n^* solves (3.1). □

The problem (3.1) arises when feasible directions methods is applied to the following nonlinear programming problem:

$$\min_{p \in P} f(p), \qquad P = \{p \in \mathbb{R}^m : h(p) \leq 0_{n_1},\ g(p) = 0_{n_2}\}, \tag{3.3}$$

where $f : \mathbb{R}^m \to \mathbb{R}^1$, $h : \mathbb{R}^m \to \mathbb{R}^{n_1}$, $g : \mathbb{R}^m \to \mathbb{R}^{n_2}$, functions $f(p)$, $h(p)$, $g(p)$ are continuously differentiable, the set P is not empty and the problem (3.3) has a solution.

We assume, that the arbitrary feasible point $p \in P$ is fixed. We define the Lagrange multiplier $x \in \mathbb{R}^n$, $x^\top = [x_1^\top, x_2^\top]$, where $x_1 \in \mathbb{R}^{n_1}_+$, $x_2 \in \mathbb{R}^{n_2}$, $n = n_1 + n_2$. Let us introduce the Lagrange function

$$L(p, x) = f(p) + h^\top(p)x_1 + g^\top(p)x_2.$$

and the complementary slackness conditions

$$x_1^i h^i(p) = 0, \qquad 1 \leq i \leq n_1. \tag{3.4}$$

The component $h^i(p)$ of a vector $h(p)$ is *active* at a point $p \in P$, if $h^i(p) = 0$. Due to (3.4) all components of a vector x_1, corresponding to nonactive components of vector $h(p)$ are equal to zero. For simplicity we suppose that all components of vector $h(p)$ are active. The Kuhn–Tucker conditions for problem (3.3) evaluated at the point $[p, x]$, where $p \in P$, are written as

$$L_p(p, x) = f_p(p) + h_p(p)x_1 + g_p(p)x_2 = 0_m, \qquad x_1 \geq 0_{n_1}. \tag*{(I)$_3$}$$

Let us introduce a vector $p' = p + \tau z$, where τ is step-size, $z \in \mathbb{R}^m$, $\|z\| = 1$ is descent direction of problem (3.3). We linearize the goal function and the functions which define the constraints at the point p. This results in problem of finding

$$I_4 = \min_{z \in \hat{Z}_n} z^\top f_p(p), \qquad \hat{Z}_n = \{z \in \mathbb{R}^m : h_p^\top(p)z \leq 0_{n_1},\ \ g_p^\top(p)z = 0_{n_2},\ \ \|z\| = 1\}. \tag{3.5}$$

If this problem has the solution z_n^* such that $I_4 < 0$, then we say that z_n^* is the steepest descent direction. It means that at least locally the point p can be improved by taken a new vector p'. If τ is small enough then the vector p' remains in feasible set P and the value of goal function $f(p') < f(p)$. If the problem (3.5) has no such solution, then it is impossible to improve the point p locally.

Let $h_p^\top(p)A_{21}^\top$, $g_p^\top(p) = A_{22}^\top$, $f_p(p) = -b_2$. According to previous section the system alternative to (I) is following:

$$u^\top h_p(p) \leq 0_{n_1}^\top, \qquad u^\top g_p(p) = 0_{n_2}^\top, \qquad -u^\top f_p(p) = \rho > 0. \tag*{(II)$_3$}$$

If the system (II)$_3$ is solvable, then due to the theorem 3 its normal solutions is $\tilde{u}^* = = \rho z^*/\|z^*\|^2$, where z^* is defined by (2.12), x^* can be found from the following unconstrained minimization problem:

$$\min_{x_1 \in \mathbb{R}^{n_1}_+} \min_{x_2 \in \mathbb{R}^{n_2}} \frac{\|L_p(p, x)\|^2}{2}. \tag{3.6}$$

We normalize vector $\tilde{u}^*$ and receive $\tilde{u}_n^* = z^*/\|z^*\| = z_n^*$. Vector $-z_n^*$ belongs to Z_n and according to the Theorem 3 we have $I_4 = -I_3 = -\|z^*\|$, i.e. $-z_n^*$ is a direction of the steepest descent in the linearized problem (3.5). This direction exists if and only if $[p, x^*]$ is not Kuhn–Tucker point, whereas $p \in P$, $x^* \in \Pi_x$. Owing to Theorem 4 we do not need to solve auxiliary problem (3.5) for finding the steepest descent direction. It is enough to solve unconstrained probqlem (3.6).

4. Application to LP problems

Let us apply the results stated above to linear programming problems. Consider the primal linear programming problem in the form

$$\min_{x \in X} c^\top x, \qquad X = \{x \in \mathbb{R}^n : Ax = b,\ x \geq 0_n\}. \tag{P}$$

Here, A is an $m \times n$ matrix of rank m; $m < n$; $\nu = n - m$ is the defect of the matrix A; c, $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$ are vectors.

Instead of the traditional necessary and sufficient optimality conditions for linear programming problems, we apply the conditions given in [10]. For this purpose, we introduce a $\nu \times n$ matrix K of rank ν such that $\operatorname{im} K^\top = \ker A$, $AK^\top = 0$, and $\mathbb{R}^n = \operatorname{im} A^\top \oplus \operatorname{im} K^\top$. Let us introduce $d = Kc \in \mathbb{R}^\nu$, the dual slack vector $v = c - A^\top u$, and two affine sets

$$\bar{X} = \{x \in \mathbb{R}^n : Ax = b\}, \qquad \bar{V} = \{v \in \mathbb{R}^n : Kv = d\}.$$

Let $\bar{x}$ and $\bar{v}$ be arbitrary fixed n-vectors satisfying the conditions $\bar{x} \in \bar{X}$ and $\bar{v} \in \bar{V}$, respectively. According to [10, 11], the necessary and sufficient minimum conditions for problem (P) are

$$\begin{bmatrix} A & 0_{mn} \\ 0_{\nu n} & K \\ \bar{v}^\top & \bar{x}^\top \end{bmatrix} \begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} b \\ d \\ \bar{x}^\top \bar{v} \end{bmatrix}, \qquad x \geq 0_n, \quad v \geq 0_n. \tag{4.1}$$

If problem (P) has a solution, then system (4.1) is consistent, and solving it gives the solutions of problem (P) and the conjugate problem

$$\min_{v \in V} \bar{x}^\top v, \qquad V = \{v \in \mathbb{R}^n : Kv = d,\ v \geq 0_n\}. \tag{C}$$

System (4.1) comprises $n+1$ equalities and $2n$ inequalities in $2n$ unknowns. The alternative system has only $n+1$ unknowns and comprises $2n$ linear inequalities and one equality, namely,

$$\begin{bmatrix} A^\top & 0_{n\nu} & \bar{v} \\ 0_{nm} & K^\top & \bar{x} \end{bmatrix} \begin{bmatrix} p \\ q \\ \alpha \end{bmatrix} \leq 0_{2n}, \qquad b^\top p + d^\top q + \bar{x}^\top \bar{v} \alpha = \rho, \tag{4.2}$$

where $\rho > 0$ is an arbitrary positive constant.

Since system (4.1) is consistent, the alternative system (4.2) is inconsistent. Problem (2.2) can be written as the following unconstrained minimization problem:

$$\min_{p \in \mathbb{R}^m} \min_{q \in \mathbb{R}^\nu} \min_{\alpha \in \mathbb{R}^1} \frac{\left[\|(A^\top p + \bar{v}\alpha)_+\|^2 + \|(K^\top q + \bar{x}\alpha)_+\|^2 + (\rho - b^\top p - d^\top q - \bar{x}^\top \bar{v}\alpha)^2\right]}{2}.$$

Solving this problem yields the optimal vectors p^*, q^*, and α^*, which determine the residuals of inconsistent system (4.2) according to

$$w_x^* = (A^\top p^* + \bar{v}\alpha^*)_+, \qquad w_v^* = (K^\top q^* + \bar{x}\alpha^*)_+, \qquad w_3^* = \rho - b^\top p^* - d^\top q^* - \bar{x}^\top \bar{v}\alpha^*.$$

By Theorem 4, normal solutions of system (4.1) are given by the formulas

$$\tilde{x}^* = \frac{w_x^*}{w_3^*}, \qquad \tilde{v}^* = \frac{w_v^*}{w_3^*}$$

and they are simultaneously normal solutions to problems (P) and (C).

Thus, we have reduced solving a linear programming problem to unconstrained minimization of a convex parameter-free globally differentiable piecewise-quadratic function of $n+1$ variables with a Lipschitz continuous gradient. As the result of such minimization we obtain a unique least-norm solution of system (4.1), which defined the sufficient optimality conditions in (P) and (C).

Acknowledgements.

This Research has been supported Russian Foundation for Basic Research (Grants No. 01-01-00804 and No. 00-15-96080). Authors are very thankful to Zhang Jian Zhong for attention to this paper, useful advice and opportunity to visit Hong Kong City University for scientific research.

References

1. Chernikov, S.N. (1968) *Linear Inequalities*, Moscow, Nauka.

2. Razumikhin, B.S. (1975) *Physical Models and the Methods of Equilibrium Theory in Programming and Economics*, Moscow, Nauka, 1975. *Physical Models and Equilibrium Methods in Programming and Economics*, Dordrecht, Boston, D. Reidel Publishing, 1984.

3. Gale, D. (1960) *The Theory of Linear Economic Models*, New York, Toronto, London, McGraw-Hill Book Company.

4. Mangasarian, O.L. (1994) *Nonlinear Programming*, Philadelphia, SIAM.

5. Dax, A. (1993) "The relationship between theorems of the alternative, least norm problems, steepest descent directions, and degeneracy: A review", *Annals of Operation Research*, vol. 46, pp. 11–60.

6. Giannessi, F. (2001) "Theorems of the Alternative and Optimization", *Encyclopedia of Optimization*, Dordrecht, Kluwer, vol. 5, pp. 437–444.

7. Broyden, C.G. (2001) "On Theorems of the Alternative", *Optimization Methods and Software*, vol. 16, pp. 101–111.

8. Evtushenko, Yu.G. (1998) "Computation of Exact Gradients in Distributed Dynamic Systems", *Optimization Methods and Software*, vol. 9, pp. 45–75.

9. Eremin, I.I. (2001) *Duality in Linear Optimization*, Ekaterinburg, Ural Branch of RAS.

10. Golikov, A.I., and Evtushenko, Yu.G. (2000) "Search for Normal Solutions in Linear Programming Problems", *Computational Mathematics and Mathematical Physics*, vol. 40, No.12, pp. 1694–1714.

11. Golikov, A.I., and Evtushenko, Yu.G. (2002) "Two Parametric Families of LP Problems and Their Applications", *Proceedings of the Steklov Institute of Mathematics*, Suppl. 1, pp. S52-S66.