XXX Simpozijum o operacionim istraživanjima
SYM-OP-IS 2003
Hotel PLAŽA, Herceg-Novi, 30.09.–03.10.2003
pp. 15–17.

APPLICATIONS OF THEOREMS OF ALTERNATIVE TO NUMERICAL METHODS

Yu.G. Evtushenko
Dorodnicyn Computing Centre of Russian Academy of Sciences, Vavilov str. 40,
Moscow, 119991, Russia, e-mail: evt@ccas.ru

Revised version 22 March 2004

Abstract: *New theorems of alternative are proved for systems of linear equalities and inequalities, which makes it possible to develop several efficient numerical methods. These methods are used to find normal solutions to systems of linear equations and inequalities, to construct separating hyperplanes, to simplify computations arising in the steepest descent methods, to propose new methods for solving LP problem.*

KEY WORD: THEOREMS OF ALTERNATIVE, LINEAR EQUALITIES AND INEQUALITIES, NORMAL SOLUTION, SEPARATING HYPERPLANE

1. INTRODUCTION

Theorems of the alternative (TA) lie at the heart of mathematical programming. TA were used to derive necessary optimality conditions for LP and NLP problems and for various other pure theoretical investigations. We show that TA give us an opportunity to construct new numerical methods for solving linear systems with equalities and inequalities, to simplify computations arising in the steepest descent method, to propose new methods for solving LP problem, to construct the separating plane and etc. 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. 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 variable 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.

2. BASIC THEOREMS

Let an $m \times n$ matrix A be given in the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11}, A_{12}, A_{21}, and A_{22} are rectangular matrices of dimensions $m_1 \times n_1$, $m_1 \times n_2$, $m_2 \times n_1$, and $m_2 \times n_2$, respectively. Let vectors $x \in \mathbb{R}^n$, $z, b, u \in \mathbb{R}^m$ be represented in partitioned form as $x^\top = [x_1^\top, x_2^\top]$, $z^\top = [z_1^\top, z_2^\top]$, $u^\top = [u_1^\top, u_2^\top]$, and $b^\top = [b_1^\top, b_2^\top]$, where $x_1 \in \mathbb{R}^{n_1}$, $x_2 \in \mathbb{R}^{n_2}$, $n = n_1 + n_2$, $z_1, u_1, b_1 \in \mathbb{R}^{m_1}$, $z_2, u_2, b_2 \in \mathbb{R}^{m_2}$, and $m = m_1 + m_2$. Let us introduce the auxiliary sets

$$\begin{aligned} \Pi_x &= \{[x_1, x_2] : x_1 \in \mathbb{R}^{n_1}_+, \ \ x_2 \in \mathbb{R}^{n_2}\}, \\ \Pi_u &= \{[u_1,\, u_2] : u_1 \in \mathbb{R}^{m_1}_+, \ \ u_2 \in \mathbb{R}^{m_2}\}; \end{aligned}$$

a vector $w \in \mathbb{R}^{n+1}$ represented as $w^\top = [w_1^\top, w_2^\top, w_3]$, 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

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

Consider the system of linear equalities and inequalities

$$A_{11}x_1 + A_{12}x_2 \geq b_1, \quad A_{21}x_1 + A_{22}x_2 = b_2, \qquad x_1 \geq 0_{n_1}. \tag{1}$$

We define the system adjoint to (1) as

$$A_{11}^\top z_1 + A_{21}^\top z_2 \leq 0_{n_1}, \quad A_{12}^\top z_1 + A_{22}^\top z_2 = 0_{n_2}, \qquad z_1 \geq 0_{m_1}, \tag{2}$$

and the system alternative to (1) as

$$A_{11}^\top u_1 + A_{21}^\top u_2 \leq 0_{n_1}, \quad A_{12}^\top u_1 + A_{22}^\top u_2 = 0_{n_2}, \qquad b_1^\top u_1 + b_2^\top u_2 = \rho, \quad u_1 \geq 0_{m_1}. \tag{3}$$

Here, $\rho > 0$ is an arbitrary fixed positive number.

The system adjoint to (3) has the form

$$A_{11}w_1 + A_{12}w_2 - b_1 w_3 \geq 0_{m_1}, \quad A_{21}w_1 + A_{22}w_2 - b_2 w_3 = 0_{m_2}, \qquad w_1 \geq 0_{n_1}. \tag{4}$$

We denote the solution sets of (1), (2), (3), and (4) by X, Z, U, and W, respectively. Unlike (1) and (3), systems (2) and (4) always have solutions, because $0_m \in Z$ and $0_{n+1} \in W$.

Лемма 1. *System* (1) *and* (3) *are not solvable simultaneously.*

Theorem 3 below imply that there always is a solution to precisely one system, (1) or (3). Therefore, these systems are alternative. The system alternative to (3) reduces to original system (1).

Let $\mathrm{pen}\,(x, X)$ denote the penalty for the violation of the condition $x \in X$ calculated at a point $x \in \Pi_x$. The quantity $\mathrm{pen}\,(u, U)$ is introduced by analogy. The penalties are calculated as the Euclidean norms of residual vectors for systems (1) and (3):

$$\begin{aligned} \mathrm{pen}(x, X) &= \left[\|(b_1 - A_{11}x_1 - A_{12}x_2)_+\|^2 + \|b_2 - A_{21}x_1 - A_{22}x_2\|^2\right]^{1/2}, \\ \mathrm{pen}(u, U) &= \left[\|(A_{11}^\top u_1 + A_{21}^\top u_2)_+\|^2 + \|A_{12}^\top u_1 + A_{22}^\top u_2\|^2 + (\rho - b_1^\top u_1 - b_2^\top u_2)^2\right]^{1/2}. \end{aligned}$$

Here, a_+ is nonnegative part of the vector a: i.e., the i-th component of the vector a_+ is equal to that of the vector a if the latter is nonnegative and is zero otherwise.

To find out whether a system is solvable and, if it is, to solve it, we apply methods of unconstrained minimization to either of the following problems:

$$I_1 = \min_{x \in \Pi_x} \left[\mathrm{pen}\,(x, X)\right]^2/2, \tag{5}$$

$$I_2 = \min_{u \in \Pi_u} \left[\mathrm{pen}\,(u, U)\right]^2/2. \tag{6}$$

In the strict sense, (5) and (6) are not unconstrained minimization problems, since they contain constraints on the signs of the components of the vectors x_1 and u_1. However, since most unconstrained minimization methods can easily be modified to allow for constraints on the signs of variables, we will keep this term for problems (5) and (6). Problems (5) and (6) are always solvable, since quadratic objective functions defined on nonempty feasible sets Π_x and Π_u are bounded from below by zero.

$$I_1^d = \max_{z \in Z} \left\{ b^\top z - \frac{\|z\|^2}{2} \right\}, \tag{7}$$

$$I_2^d = \max_{w \in W} \left\{ \rho w_3 - \frac{\|w\|^2}{2} \right\}. \tag{8}$$

Unlike systems (1) and (3), which may be consistent or inconsistent, problems (5) – (8) always have solutions. Moreover, problems (7) and (8) have unique solutions, since feasible sets Z and W in these problems are nonempty, and strictly concave quadratic objective functions are bounded from above. Problems (5) and (6) are dual to problems (7) and (8), respectively.

The projection of a point a onto a nonempty closed set X is a point $x^* \in X$ nearest to a, i.e., the point that minimizes the function

$$J = \min_{x \in X} \|a - x\| = \|a - x^*\|.$$

We write $x^* = \mathrm{pr}\,(a, X)$ and denote the distance from a point a to a set X as $\mathrm{dist}\,(a, X) = \|a - x^*\|$.

Теорема 1. *Any solution x^* of problem (5) determines a unique solution $z^{*\top} = [z_1^{*\top}, z_2^{*\top}]$ of problem (7) as*

$$z_1^* = (b_1 - A_{11}x_1^* - A_{12}x_2^*)_+, \qquad z_2^* = b_2 - A_{21}x_1^* - A_{22}x_2^* \tag{9}$$

and it holds that

$$\|z^*\|^2 = b^\top z^*, \tag{10}$$

$$z^* \perp Ax^*, \qquad z^* \perp (b - z^*), \tag{11}$$

$$z^* = \mathrm{pr}(b, Z), \qquad \|z^*\| = \mathrm{pen}\,(x^*, X), \qquad \|b - z^*\| = \mathrm{dist}\,(b, Z), \tag{12}$$

$$[\mathrm{pen}\,(x^*, X)]^2 + [\mathrm{dist}\,(b, Z)]^2 = \|b\|^2. \tag{13}$$

Relation (10) follows from equality of optimal values of the objective functions for the primal (7) and dual (5) problems. By virtue of (9), this equality is expressed in the terms of only z^*, which is a solution of problem (7).

Let $\hat{A} = [-A, b]$ be an $m \times (n+1)$ matrix and $r \in \mathbb{R}^{n+1}$ be a vector of the form $r^\top = [0_[^\top n], \rho]$.

Теорема 2. *Let* $u^{*\top} = \left[u_1^{*\top}, u_2^{*\top}\right]$ *be an arbitrary solution of problem* (6). *Then, a solution* $w^{*\top} = \left[w_1^{*\top}, w_2^{*\top}, w_3^*\right]$ *of problem* (6) *can be expressed in terms of* u^* *as*

$$w_1^* = (A_{11}^\top u_1^* + A_{21}^\top u_2^*)_+, \quad w_2^* = A_{12}^\top u_1^* + A_{22}^\top u_2^*, \quad w_3^* = \rho - b_1^\top u_1^* - b_2^\top u_2^* \tag{14}$$

and satisfies the following conditions:

$$\begin{aligned}
&\|w^*\|^2 = \rho w_3^*, \\
&w^* \perp \hat{A}^\top u^*, \qquad w^* \perp (r - w^*), \\
&w^* = \mathrm{pr}\,(r, W), \qquad \|w^*\| = \mathrm{pen}\,(u^*, U), \qquad \|r - w^*\| = \mathrm{dist}\,(r, W), \\
&[\mathrm{pen}\,(u^*, U)]^2 + [\mathrm{dist}\,(r, W)]^2 = \|r\|^2, \\
&\|w^*\| \le \rho, \quad 0 \le w_3^* \le \rho, \quad \|w_1^*\|^2 + \|w_2^*\|^2 \le \frac{\rho^2}{4}.
\end{aligned} \tag{15}$$

Relation (15) follows from the equality of the optimal value of the objective functions for the primal (8) and dual (6) problems. By virtue of (14), this equality is expressed in terms of only w^*, which is a solution to problem (8).

Теорема 3. *Let* x^* *and* u^* *be arbitrary solutions of problems* (5) *and* (6), *respectively, and let minimum residual vectors* z^* *and* w^* *be calculated by* (9) *and* (14). *Then, the following assertions are valid:*

i. *systems* (1) *and* (3) *are alternative; i.e., only one of them is solvable;*

ii. *if system* (1) *is inconsistent, then the normal solution* $\tilde{u}^*$ *of system* (3) *and the minimum residual vector* z^* *of system* (1) *are collinear, and*

$$\tilde{u}^* = \frac{\rho z^*}{\|z^*\|^2}, \qquad z^* = \frac{\rho \tilde{u}^*}{\|\tilde{u}^*\|^2};$$

iii. *if system* (3) *is inconsistent, then the components of the normal solution* $\tilde{x}^{*\top} = [\tilde{x}_1^{*\top}, \tilde{x}_2^{*\top}]$ *of system* (1) *are*

$$\tilde{x}_1^* = \frac{w_1^*}{w_3^*}, \qquad \tilde{x}_2^* = \frac{w_2^*}{w_3^*}.$$

Thus, Theorem 3 reduces the problem of solvability of system (1) or (3) to minimizing the residual of either system. If the norm of the minimum residual is nonzero, then this system is inconsistent, and, based on that residual, the normal solution of the consistent system can be found by simple formulas.

3. THE PROBLEM OF SEPARATING HYPERPLANES

Let us represent A, b, u, and z in the form

$$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}, \qquad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \qquad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \qquad z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix},$$

where A_1 and A_2 are $k \times n$ and $\ell \times n$ matrices, respectively; $b_1, u_1, z_1 \in \mathbb{R}^k$; $b_2, u_2, z_2 \in \mathbb{R}^\ell$; and $k + \ell = m$. Assuming that the set X consists of the nonempty sets

$$X_1 = \{x \in \mathbb{R}^n : A_1 x \ge b_1\}, \qquad X_2 = \{x \in \mathbb{R}^n : A_2 x \ge b_2\}$$

such that $X_1 \cap X_2 = \emptyset$, we consider the problem of finding a hyperplane that strictly separates X_1 and X_2. Let $\alpha \in [0, 1]$ be a scalar parameter.

Теорема 4. *Let $X_1 \neq \emptyset$, $X_2 \neq \emptyset$, and $X = X_1 \cap X_2 = \emptyset$, x^* be a solution of the problem*

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \left[\|b_1 - A_1 x)_+\|^2 + \|b_2 - A_2 x)_+\|^2 \right], \tag{16}$$

the components of the minimum residual vector $z^* = (b - Ax^*)_+$ are $z_1^* = (b_1 - A_1 x^*)_+$ and $z_2^* = (b_2 - A_2 x^*)_+$. Then the family of parallel hyperplanes separating X_1 and X_2 can be described by two equivalent equations

$$(z_1^*)^\top (A_1 x - b_1) + \alpha \|z^*\|^2 = 0, \qquad (z_2^*)^\top (b_2 - A_2 x) + (\alpha - 1)\|z^*\|^2 = 0,$$

when $0 < \alpha < 1$, these hyperplanes strictly separate X_1 and X_2.

The proof of Theorem 4 is similar to that of Eremin's theorem [1, Theorem 10.1], which is based on the Farkas lemma. The separating hyperplane in Eremin's theorem is described by the following equivalent equations:

$$(u_1^*)^\top (A_1 x - b_1) + \frac{\rho}{2} = 0, \qquad (u_2^*)^\top (b_2 - A_2 x) - \frac{\rho}{2} = 0,$$

where u_1^*, u_2^* is an arbitrary solution of system

$$A_1^\top u_1 + A_2^\top u_2 = 0_n, \qquad b_1^\top u_1 + b_2^\top u_2 = \rho > 0, \qquad u_1 \geq 0_k, \quad u_2 \geq 0_\ell. \tag{17}$$

By Theorem 4, to find a separating hyperplane, one must solve the problem (16) of unconstrained minimization of the residual of the inconsistent system in $\mathbb{R}^n$, whereas Eremin's theorem implies that one must solve the consistent system (17) in m unknowns. Since n, m our approach is more preferable.

These theorems and various close results are given in [2, 3].

4. ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 03-01-00465; under the State Program for Support of Leading Scientific Schools, project no. SCH-1737.2003.1.

5. REFERENCES

[1] Eremin, I.I., *Teoriya lineinoi optimizatsii* (Linear Optimization Theory), Yekaterinburg: Ural. Otd. Ross. Akad. Nauk, 1998.

[2] Golikov, A.I., and Evtushenko, Yu.G., “New Method for Solving Systems of Linear Equalities and Inequalities”, *Doklady Mathematics*, 64 (2001) 370–373. ; ю ’

[3] Golikov, A.I., and Evtushenko, Yu.G., “Theorems of Alternative and Their Applications in Numerical Methods”, *Computational Mathematics and Mathematical Physics*, 43 (2003) 338–358.