Computational Mathematics and Mathematical Physics, Vol. 45, No. 2, 2005, pp. 227–242.
Translated from Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki, Vol. 45, No. 2, 2005, pp. 238–253.

On Families of Hyperplanes That Separate Polyhedra

A.I. Golikov, Yu.G. Evtushenko, and S. Ketabchi

Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119991 Russia
e-mail: gol@ccas.ru, evt@ccas.ru

Received April 19, 2004; in final form, September 20, 2004

Abstract — The problem of constructing a family of hyperplanes that separate two disjoint nonempty polyhedra is examined. The polyhedra are given by systems of linear inequalities or by systems of linear equalities with nonnegative variables. Constructive algorithms for solving this problem are presented. The construction of separating hyperplanes relies heavily on theorems of the alternative.

Keywords: theorems of alternative hyperplanes that separate two given polyhedra

1. INTRODUCTION

The theorem on the existence of a separating hyperplane plays a key role in functional analysis, optimization theory, and operations research. In solving practical problems, one should not only know that there exists a separating hyperplane but also be able to constructively determine it. We examine the problem of numerical construction of a family of hyperplanes that separate two disjoint nonempty polyhedra given by systems of linear inequalities or by systems of linear equalities with nonnegative variables. Our considerations are based on the theorem on a separating hyperplane of two polyhedra given by systems of inequalities that was proved by I.I. Eremin (see [1, Theorem 10.1]). We use a specific form of a separating hyperplane, where the normal vector and the shift vector are expressed in terms of an arbitrary solution to a certain system that is alternative to an inconsistent system. The inconsistent system is formed of two consistent subsystems, each of which defines a nonempty polyhedron. The entire system is inconsistent because these polyhedra do not intersect. The construction of separating hyperplanes relies heavily on theorems of the alternative.

In Section 2, we consider an application of the normal solution of the alternative system to the construction of a family of separating hyperplanes. The results of [2] allow us to find the normal solution by solving the unconstrained minimization problem for the residual of the inconsistent inequality system that determines both polyhedra. The latter problem normally has a much lower number of variables than the alternative consistent system. Hence, the proposed method is less labor-consuming than solving the alternative system.

In Section 3, we examine the following problem: how can one distinguish a solution to the alternative system that generates a family of separating hyperplanes with a maximal thickness, which coincides with the minimal distance between the polyhedra?

In Section 4, we construct families of separating hyperplanes for two polyhedra given by systems of linear equalities with nonnegative variables. In contrast to the case of polyhedra given by systems of linear inequalities, every solution to the alternative system now determines two distinct families of separating hyperplanes.

In Section 5, we give a brief review of the generalized Newton method for calculating the normal solution to the alternative system. The normal solution is used for constructing a family of separating hyperplanes for polyhedra given by systems of linear inequalities. The generalized

Newton method was implemented in Matlab and showed good performance in solving large-scale test problems.

2. CONSTRUCTION OF SEPARATING HYPERPLANES WITH THE HELP OF THE GALE THEOREM OF THE ALTERNATIVE

Let $x \in \mathbb{R}^n$ and $b \in \mathbb{R}^m$, where $\|b\| \neq 0$, be given vectors and $A \in \mathbb{R}^{m\times n}$ be a given rectangular matrix. Define the two sets

$$X = \{x \in \mathbb{R}^n : Ax \geq b\} \ \text{ and } \ U = \{u \in \mathbb{R}^m : A^\top u = 0_n, \ \ b^\top u = \rho, \ \ u \geq 0_m\},$$

where $\rho > 0$ is an arbitrary fixed positive number and 0_i is the zero vector of dimension i.

The linear systems

$$Ax \geq b, \tag{1}$$

and

$$A^\top u = 0_n, \ \ b^\top u = \rho, \ \ u \geq 0_m, \tag{2}$$

which determine the sets X and U, respectively, are alternative for any strictly positive value of ρ, which means that exactly one of them is consistent (this is the Gale theorem; e.g., see [2]). We take the scalar products of both sides of the first equality in (2) with the vector x and then subtract the second equality from the resulting relation. This yields

$$u^\top(Ax - b) = -\rho < 0. \tag{3}$$

This equality is a key tool for constructing a family of hyperplanes that separate two polyhedra given as intersections of half-spaces.

We write A, b, and u in the form

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

where A_1 and A_2 are matrices of sizes m_1-by-n and m_2-by-n, respectively; $b_1, u_1 \in \mathbb{R}^{m_1}$; $b_2, u_2 \in \mathbb{R}^{m_2}$; and $m_1 + m_2 = m$. Define the two nonempty sets

$$X_1 = \{x \in \mathbb{R}^n : A_1 x \geq b_1\} \ \text{ and } \ X_2 = \{x \in \mathbb{R}^n : A_2 x \geq b_2\},$$

which determine two polyhedra (or polyhedral sets; see [3]) such that $X = X_1 \cap X_2 = \emptyset$.

Define the hyperplane $c^\top x - \gamma = 0$, where $c \in \mathbb{R}^n$, $\|c\| \neq 0$ is a normal vector and γ is a scalar. We say that this hyperplane $c^\top x - \gamma = 0$ separates X_1 and X_2 if $c^\top x - \gamma \geq 0$ for all $x \in X_1$ and $c^\top x - \gamma \leq 0$ for all $x \in X_2$. If we have the strict inequalities in both conditions, then we say that this hyperplane strictly separates X_1 and X_2.

Consider the problem of calculating the hyperplanes that separate X_1 and X_2. Taking into account the partition introduced above, we can rewrite systems (1) and (2) and relation (3) as follows:

$$A_1 x \geq b_1, \qquad A_2 x \geq b_2, \tag{4}$$

$$A_1^\top u_1 + A_2^\top u_2 = 0_n, \qquad b_1^\top u_1 + b_2^\top u_2 = \rho, \qquad u_1 \geq 0_{m_1}, \quad u_2 \geq 0_{m_2}, \tag{5}$$

$$u_1^\top(A_1 x - b_1) + u_2^\top(A_2 x - b_2) = -\rho < 0. \tag{6}$$

Define a linear function $\varphi(x, \alpha)$ of variable $x \in \mathbb{R}^n$ and a scalar parameter α ranging on the interval $[0, 1]$:

$$\varphi(x, \alpha) = u_1^\top(A_1 x - b_1) + \alpha\rho. \tag{7}$$

Relation (6) implies that $\varphi(x,\alpha)$ can be equivalently defined as

$$\varphi(x,\alpha) = u_2^\top(b_2 - A_2x) + (\alpha - 1)\rho. \tag{8}$$

The equality $\varphi(x,\alpha) = 0$, where u_1 and u_2 satisfy (5) and α belongs to $[0,1]$, determines the hyperplane that separates the sets X_1 and X_2. Indeed, if $x \in X_1$, then, according to (7), we have $\varphi(x,\alpha) \geq 0$, while if $x \in X_2$, then, according to (8), we have $\varphi(x,\alpha) \leq 0$. The separating hyperplane $\varphi(x,\alpha) = 0$ with $\alpha = 1/2$ was first introduced and studied by Eremin (e.g., see [1]).

In view of system (5), the hyperplane $\varphi(x,\alpha) = 0$ determined by a function of form (7) or (8) can be written as

$$\varphi(x,\alpha) = c^\top x - \gamma = 0,$$

where

$$c = A_1^\top u_1 = -A_2^\top u_2, \qquad \gamma = b_1^\top u_1 - \alpha\rho = -b_2^\top u_2 - (\alpha - 1)\rho.$$

Here, u_1 and u_2 are arbitrary solutions to system (5).

For fixed distinct vectors $u^\top = [u_1^\top, u_2^\top]$ that satisfy system (5), we examine the family of parallel hyperplanes given by the following equivalent definitions:

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : u_1^\top A_1x - b_1^\top u_1 + \alpha\rho = 0\} = \{x \in \mathbb{R}^n : \varphi(x,\alpha) = 0\}, \tag{9}$$

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : -u_2^\top A_2x + b_2^\top u_2 + (\alpha - 1)\rho = 0\} \tag{10}$$

All the hyperplanes belonging to this family are parallel, because they have the common normal vector $c = A_1^\top u_1 = -A_2^\top u_2$.

The hyperplane $\Gamma(1)$ can be obtained from $\Gamma(0)$ with the help of the shift vector y:

$$\Gamma(1) = \Gamma(0) + y.$$

The norm of y (i.e., the distance between the hyperplanes $\Gamma(1)$ and $\Gamma(0)$) will be called the *thickness of the family of hyperplanes.*

According to [2, 3], the projection $\bar{x}^*$ of a point $\bar{x}$ onto the hyperplane $\Gamma(\alpha)$ is determined by the formula

$$\bar{x}^* = \bar{x} + c(b_1^\top u_1 - c^\top\bar{x} - \alpha\rho)/\|c\|^2. \tag{11}$$

Denote by $\mathrm{pr}(0_n, \Gamma(\alpha))$ the projection of the origin onto the hyperplane $\Gamma(\alpha)$. Setting $\bar{x} = 0_n$ in (11), we obtain $\mathrm{pr}(0_n, \Gamma(\alpha)) = c(b_1^\top u_1 - \alpha\rho)/\|c\|^2$. From this, we find the shift vector y and the thickness $\|y\|$ of the family of hyperplanes $\Gamma(\alpha)$:

$$y = \mathrm{pr}(0_n, \Gamma(1)) - \mathrm{pr}(0_n, \Gamma(0)) = -c\rho/\|c\|^2, \tag{12}$$

$$\|y\| = \rho/\|c\|. \tag{13}$$

System (5) may have many solutions. In this section, we examine the properties of the family of separating hyperplanes, where u is the normal solution $\tilde{u}^*$ to system (5). The results of [2] allow us to relatively easily construct the normal solution, i.e., to solve the following quadratic programming problem:

$$\min_{u \in U} \frac{1}{2}\|u\|^2, \qquad U = \{u \in \mathbb{R}^m : A^\top u = 0_n, \;\; b^\top u = \rho, \;\; u \geq 0_m\}. \tag{14}$$

Henceforth, we use the Euclidean norm of vectors.

We introduce the following unconstrained minimization problem for the residual vector of system (1):

$$I_1 = \min_{x \in \mathbb{R}^n} \frac{1}{2}\|(b - Ax)_+\|^2, \tag{15}$$

Here, a_+ is the nonnegative part of a vector a; i.e., the ith component of a_+ is the same as the ith component of a if the latter is nonnegative; otherwise, the ith component of a_+ is zero.

The unconstrained minimization problem (15) is dual to the following quadratic programming problem:

$$
\begin{aligned}
I_2 &= \max_{z \in Z} \left\{ b^\top z - \frac{1}{2}\|z\|^2 \right\}, \\
Z &= \{ z^\top = [z_1^\top, z_2^\top] \in \mathbb{R}^m : A_1^\top z_1 + A_2^\top z_2 = 0_n, \;\; z_1 \geq 0_{m_1}, \;\; z_2 \geq 0_{m_2} \}.
\end{aligned} \tag{16}
$$

Problems (15) and (16) are always solvable. Moreover, problem (16) has a unique solution, because its feasible set is nonempty and its objective function, which is quadratic and strictly concave, is bounded above by $\|b\|^2/2$. Theorem 1, given below, asserts the equivalence between the quadratic programming problems (14) and (16) in the sense that the solution to one problem determines the solution to the other. The solution $z^* \in \mathbb{R}^m$ to the quadratic programming problem (16) can be expressed in terms of the solution $x^* \in \mathbb{R}^n$ to the simpler problem (15) of the unconstrained minimization of a piecewise quadratic function. Usually, we have $n \ll m$ in the problem of separating polyhedra (4).

Theorem 1. *Let X_1 and X_2 be nonempty polyhedra with an empty intersection. Every solution x^* to problem* (15) *determines a unique solution ${z^*}^\top = [{z_1^*}^\top, {z_2^*}^\top]$ to problem* (16) *given by the formulas*

$$
z_1^* = (b_1 - A_1 x^*)_+, \qquad z_2^* = (b_2 - A_2 x^*)_+. \tag{17}
$$

The normal solution $\tilde{u}^*$ to system (5) *can be obtained from the solution z^* to problem* (16) *by the formula*

$$
\tilde{u}^* = \rho z^* / \|z^*\|^2, \tag{18}
$$

while the solution z^* to problem (16) *can be obtained from the solution $\tilde{u}^*$ to problem* (14) *by the formula*

$$
z^* = \rho \tilde{u}^* / \|\tilde{u}^*\|^2. \tag{19}
$$

It holds that $\|\tilde{u}^*\|\|z^*\| = \rho$. The optimal values of the objective functions in problems (15) *and* (16) *are the same: $I_1 = I_2 = \|z^*\|^2/2$.*

The assertions of Theorem 1 follow from the results of [2]. The vector ${z^*}^\top = [{z_1^*}^\top, {z_2^*}^\top]$ will be called the *vector of minimal residuals* of system (4).

Consider the family of hyperplanes (9), (10) that uses the normal solution $\tilde{u}^*$ to system (5). This family is given by the two equivalent definitions

$$
\Gamma(\alpha) = \{ x \in \mathbb{R}^n : {\tilde{u}_1^*}^\top (A_1 x - b_1) + \alpha\rho = 0 \}, \tag{20}
$$

$$
\Gamma(\alpha) = \{ x \in \mathbb{R}^n : -{\tilde{u}_2^*}^\top (A_2 x - b_2) + (\alpha - 1)\rho = 0 \}. \tag{21}
$$

Note that if $\tilde{u}^*$ is replaced by z^* using (19), then families (20) and (21) can be written in yet another equivalent form

$$
\begin{aligned}
\Gamma(\alpha) &= \{ x \in \mathbb{R}^n : {z_1^*}^\top (A_1 x - b_1) + \alpha \|z^*\|^2 = 0 \}, \\
\Gamma(\alpha) &= \{ x \in \mathbb{R}^n : {z_2^*}^\top (b_2 - A_2 x) + (\alpha - 1) \|z^*\|^2 = 0 \}.
\end{aligned}
$$

Theorem 2 (on family (20), (21) of parallel separating hyperplanes). *Let X_1 and X_2 be nonempty polyhedra with an empty intersection. Assume that x^* is a solution to problem* (15), *while the vectors z^* and $\tilde{u}^*$ are determined by* (17) *and* (18). *Then, the following is true:*

(1) *There exists a solution to system* (5); *for every solution to this system, it holds that $\|u_1\| \neq 0$, $\|u_2\| \neq 0$, and $\|A_1^\top u_1\| = \|A_2^\top u_2\| \neq 0$.*

(2) *When* $0 \le \alpha \le 1$, *the set* $\Gamma(\alpha)$ *determines a family of parallel hyperplanes that separate* X_1 *and* X_2; *if* $0 < \alpha < 1$, *then the hyperplanes* $\Gamma(\alpha)$ *strictly separate* X_1 *and* X_2.

(3) *If* α *is equal to* $\alpha_* = \|z_1^*\|^2/\|z^*\|^2$, *then the point* x^* *belongs to the separating hyperplane corresponding to this value of* α.

(4) *The shift vector* $y = \Gamma(1) - \Gamma(0)$ *and the thickness of the family* $\Gamma(\alpha)$ *are determined by the formulas*

$$y = \frac{-\rho A_1^\top \tilde{u}_1^*}{\|A_1^\top \tilde{u}_1^*\|^2}, \qquad \|y\| = \frac{\rho}{\|A_1^\top \tilde{u}_1^*\|}.$$

(5) *If* $\alpha > 0$, *then* $X_1 \cap \Gamma(\alpha) = \emptyset$; *if* $\alpha < 1$, *then* $X_2 \cap \Gamma(\alpha) = \emptyset$.

(6) *If* $X_1 \cap \Gamma(0) \neq \emptyset$, *then* $\Gamma(0)$ *is a supporting hyperplane of the set* X_1; *if* $X_2 \cap \Gamma(1) \neq \emptyset$, *then* $\Gamma(1)$ *is a supporting hyperplane of the set* X_2.

(7) *Every solution* x^* *to problem* (15) *belongs to neither* X_1 *nor* X_2.

Proof. 1. System (5) is alternative to the inconsistent system (4); hence, there exists a solution to (5); moreover, $\|A_1^\top u_1\| = \|A_2^\top u_2\|$. We show that these norms cannot vanish. The relation $b_1^\top u_1 + b_2^\top u_2 = \rho > 0$ implies that at least one of the two summands on the left-hand side is strictly positive. Without loss of generality, we can assume that

$$b_1^\top u_1 = \rho_1 > 0. \tag{22}$$

By the condition of the theorem, $X_1 \neq \emptyset$. Hence, the system $A_1^\top x_1 = 0_n$, $b_1^\top u_1 = \rho_1$, $u_1 \ge 0_{m_1}$, which is alternative to the system $A_1 x \ge b_1$, is inconsistent. Thus, if (22) is fulfilled and $u_1 \ge 0_{m_1}$, then the vector $A_1^\top u_1$ cannot be zero. Therefore, $A_2^\top u_2$ is not a zero vector as well. It follows that the solutions u_1 and u_2 to system (5) are nonzero.

2. The necessary condition for a minimum in problem (15), combined with (17) and (18), leads to $A^\top z^* = 0_n$ and $A^\top \tilde{u}^* = 0_n$. Define the vector c as follows:

$$c = A_1^\top \tilde{u}_1^* = -A_2^\top \tilde{u}_2^*. \tag{23}$$

Since assertion 1 has already been proved, we have $\|c\| \neq 0$. Taking the normal solution $\tilde{u}^*$ as the vector u in formulas (7) and (8), we arrive at the relations

$$\begin{aligned} \varphi(x,\alpha) &= \tilde{u}_1^{*\top}(A_1^\top x - b_1) + \alpha\rho = c^\top x - b_1^\top \tilde{u}_1^* + \alpha\rho, & (24)\\ \varphi(x,\alpha) &= \tilde{u}_2^{*\top}(b_2 - A_2^\top x) + (\alpha - 1)\rho = c^\top x + b_2^\top \tilde{u}_2^* + (\alpha-1)\rho. & (25) \end{aligned}$$

If $x \in X_1$ and $\alpha \ge 0$, then $\varphi(x,\alpha) \ge 0$. If $x \in X_2$ and $\alpha \le 1$, then $\varphi(x,\alpha) \le 0$. Hence, $\Gamma(\alpha)$, where $0 \le \alpha \le 1$, is indeed a family of separating hyperplanes.

If $\alpha > 0$ and $x \in X_1$, then, by (24), we have $\varphi(x,\alpha) > 0$. Similarly, if $\alpha < 1$ and $x \in X_2$, then (25) implies that $\varphi(x,\alpha) < 0$. Thus, in this case, $\Gamma(\alpha)$ defines a family of strictly separating hyperplanes.

3. We set $\bar{x}$ in (11) equal to the vector x^* and set α equal to α_*. Then, we find from (11) that

$$\bar{x}^* - x^* = c\left(b_1^\top \tilde{u}_1^* - \rho\frac{\|z_1^*\|^2}{\|z^*\|^2} - c^\top x^*\right).$$

Using this relation and taking into account (17) and (18), we arrive at the equality $\bar{x}^* - x^* = 0_n$; i.e., in this case, the vector x^* belongs to the separating hyperplane $\Gamma(\alpha_*)$. Similarly, substituting α_* into (21) and taking into account the relation $1-\alpha_* = \|z_1^*\|/\|z^*\|^2$, we conclude that x^* belongs to the separating hyperplane $\Gamma(\alpha_*)$.

4. This assertion follows from formulas (12) and (13).

5. The conditions $x_1 \in X_1$ and $\tilde{u}_1^* \geq 0_{m_1}$ imply that $\tilde{u}_1^{*\top}(A_1x_1 - b_1) \geq 0$. On the other hand, from the condition $x_1 \in \Gamma(\alpha)$, we have $\tilde{u}_1^{*\top}(A_1x_1 - b) + \alpha\rho = 0$, which is impossible if $\alpha\rho > 0$. Hence, the intersection of X_1 and $\Gamma(\alpha)$ is empty for any $\alpha > 0$. The case $\alpha < 1$ is treated analogously.

6. The set X_1 has at least one common point x_1 with $\Gamma(0)$. Moreover, X_1 belongs to the half-space $c^\top x_1 - \tilde{u}_1^{*\top} b_1 \geq 0$, because this inequality can be rewritten as $\tilde{u}_1^{*\top}(A_1x_1 - b_1) \geq 0$. It follows that $\Gamma(0)$ is a separating hyperplane of the set X_1 at its point x_1. The second assertion is proved analogously.

7. Assume the contrary; i.e., there exists a solution x^* to problem (15) such that $x^* \in X_1$. This means that $z_1^* = 0_{m_1}$. Then, by (18), the solution to system (5) is such that $\|\tilde{u}_1^*\| = 0$, which contradicts assertion 1. The theorem is proved. □

Theorem 2 suggests that the simplest method for constructing a family of separating hyperplanes is as follows. First, one solves in $\mathbb{R}^n$ the unconstrained minimization problem (15) for the residual of the inconsistent system (1) and calculates the normal solution $\tilde{u}^*$ to system (5). Then, one constructs $\Gamma(\alpha)$ using (20) or (21). The approach of Eremin is to find an arbitrary solution to the consistent system (5), where the number of unknowns is m. Since we usually have $n \ll m$ in the problem of constructing a separating hyperplane, the approach suggested by Theorem 2 is preferable.

Note that the normal solution $\tilde{u}^*$ to system (5) can be found by a different method, namely, by solving the dual problem to the quadratic programming problem (14). The dual problem is the following unconstrained maximization problem for a piecewise quadratic function:

$$\max_{\beta \in \mathbb{R}^1} \max_{x \in \mathbb{R}^n} \left\{ \beta\rho - \frac{1}{2}\|(\beta b - Ax)_+\|^2 \right\}. \tag{26}$$

The number of variables in this problem is $n+1$.

If β', x' is a solution to problem (26), then the normal solution $\tilde{u}^*$ to system (5) is given by

$$\tilde{u}^* = (\beta' b - Ax')_+.$$

In the following theorem, we determine the distance between the supporting hyperplanes constructed with the help of the normal solution $\tilde{u}^*$ to system (5).

Theorem 3. *Let the conditions of Theorem 2 be fulfilled. Then, there exist $\hat{\alpha} \leq 0$ and $\tilde{\alpha} \geq 1$ such that the family of parallel hyperplanes* (20), (21), *where $\hat{\alpha} \leq \alpha \leq \tilde{\alpha}$, separates X_1 and X_2. The hyperplanes $\Gamma(\hat{\alpha})$ and $\Gamma(\tilde{\alpha})$ are supporting hyperplanes of the sets X_1 and X_2, respectively. The hyperplane $\Gamma(\tilde{\alpha})$ can be obtained from $\Gamma(\hat{\alpha})$ by the formula $\Gamma(\tilde{\alpha}) = \Gamma(\hat{\alpha}) + y$, where the shift vector y and its norm are given by $y = (\hat{\alpha} - \tilde{\alpha})c/\|c\|^2$ and $\|y\| = (\tilde{\alpha} - \hat{\alpha})/\|c\|$.*

Proof. The form of $\varphi(x, \alpha)$ implies that the inequalities

$$\begin{aligned} \varphi(x,\alpha) &= \tilde{u}_1^{*\top}(A_1^\top x - b_1) + \alpha\rho = c^\top x - b_1^\top \tilde{u}_1^* + \alpha\rho > 0, && (27)\\ c^\top x &\geq b_1^\top \tilde{u}_1^* - \alpha\rho && (28) \end{aligned}$$

are fulfilled for all $x \in X_1$ and $\alpha \geq 0$. Similarly, the inequalities

$$\begin{aligned} \varphi(x,\alpha) &= \tilde{u}_2^{*\top}(b_2 - A_2^\top x) + (\alpha-1)\rho = c^\top x + b_2^\top \tilde{u}_2^* + (\alpha-1)\rho \leq 0, && (29)\\ c^\top x &\leq -b_2^\top \tilde{u}_2^* - (\alpha-1)\rho && (30) \end{aligned}$$

hold for all $x \in X_2$ and $\alpha \le 1$. According to (28) and (30), there exist $\hat{x} \in X_1$ and $\tilde{x} \in X_2$ such that

$$c^\top \hat{x} = \min_{x \in X_1} c^\top x, \qquad c^\top \tilde{x} = \max_{x \in X_2} c^\top x. \tag{31}$$

Setting $x = \hat{x}$ in (27) yields

$$c^\top \hat{x} - b_1^\top \tilde{u}_1^* + \alpha\rho \ge 0. \tag{32}$$

If $\alpha = 0$, then we have

$$b_1^\top \tilde{u}_1^* - c^\top \hat{x} \le 0. \tag{33}$$

Therefore, (32) is valid for any $\alpha \ge \hat{\alpha}$, where

$$\hat{\alpha} = (b_1^\top \tilde{u}_1^* - c^\top \hat{x})/\rho \le 0. \tag{34}$$

Relation (29) implies that

$$\varphi(x, \alpha) = c^\top \tilde{x} + b_2^\top \tilde{u}_2^* + \rho(\alpha - 1) \le 0$$

for $x \in X_2$ and $\alpha \le 1$. If $\alpha = 1$, then we have

$$c^\top \tilde{x} + b_2^\top \tilde{u}_2^* \le 0. \tag{35}$$

Hence, the inequality $\varphi(x, \alpha) \le 0$ holds for all $x \in X_2$ and α such that $\alpha \le \tilde{\alpha}$, where

$$\tilde{\alpha} = 1 - (c^\top \tilde{x} + b_2^\top \tilde{u}_2^*)/\rho \ge 1. \tag{36}$$

The hyperplane $\Gamma(\hat{\alpha}) = \{x \in \mathbb{R}^n : c^\top x = c^\top \hat{x}\}$ has the common point $\hat{x}$ with the set X_1. In view of (31), every point of X_1 belongs to the half-space $c^\top(\hat{x} - x) \le 0$. It follows that $\Gamma(\hat{\alpha})$ is a supporting hyperplane of X_1. The vector c is a supporting vector of X_1 at the point $\hat{x}$. In particular, if $\hat{\alpha} = 0$, then $\Gamma(0)$ is a supporting hyperplane. In a similar way, we show that $\Gamma(\tilde{\alpha})$ is a supporting hyperplane of X_2 at the point $\tilde{x}$. The shift vector y is obtained by simple calculations similar to those in (12) and (13). The theorem is proved. $\square$

In certain cases, knowledge of the normal solution $\tilde{u}^*$ makes it possible to easily determine the optimal values of the objective functions in problems (31) and find out whether the hyperplanes $\Gamma(\alpha)$ corresponding to $\alpha = 0$ and $\alpha = 1$ are supporting hyperplanes for X_1 and X_2, respectively. Denote by $w_1 \in \mathbb{R}^{m_1}_+$ and $w_2 \in \mathbb{R}^{m_2}_+$ the Lagrange multipliers, and define the Lagrangian functions for problems (31):

$$L_1(x, w_1) = c^\top x + w_1^\top (b_1 - A_1 x), \qquad L_2(x, w_2) = -c^\top x + w_2^\top (b_2 - A_2 x).$$

The pair $[x_1, w_1]$ is a Kuhn–Tucker point for the first problem in (31) if it holds that

$$c = A_1^\top w_1, \quad D(w_1)(b_1 - A_1 x_1) = 0_{m_1}, \quad w_1 \ge 0_{m_1}, \quad A_1 x_1 \ge b_1. \tag{37}$$

Analogously, the pair $[x_2, w_2]$ is a Kuhn–Tucker point for the second problem in (31) if

$$c = -A_2^\top w_2, \quad D(w_2)(b_2 - A_2 x_2) = 0_{m_2}, \quad w_2 \ge 0_{m_2}, \quad A_2 x_2 \ge b_2. \tag{38}$$

We take $\tilde{u}_1^*$ as the vector w_1 in (37). If there exists a vector x_1 that satisfies (37), then $[x_1 \tilde{u}_1^*]$ is a Kuhn–Tucker point. Moreover, (34) implies that $\hat{\alpha} = 0$; i.e., $\Gamma(0)$ is a supporting hyperplane of X_1 at the point x_1. Similarly, let us set $w_2 = \tilde{u}_2^*$ in the second problem in (31). If there exists a vector x_2 that satisfies (38), then, in view of (36), we conclude that $\tilde{\alpha} = 1$. Thus, $\Gamma(1)$ is a supporting hyperplane of X_2 at the point x_2. If $\hat{\alpha} < 0$ or $\tilde{\alpha} > 1$, then we seek the optimal Lagrange multipliers w_1^* or w_2^* for the corresponding problems in (31). The first conditions in (37) and (38) imply that these vectors satisfy the relation

$$A_1^\top w_1^* + A_2^\top w_2^* = 0.$$

Setting $\hat{x} = x_1$ and $\tilde{x} = x_2$ in (33) and (35), respectively, we obtain

$$c^\top x_1 \geq b_1 \tilde{u}_1^*, \qquad -c^\top x_2 \geq b_2 \tilde{u}_2^*.$$

Adding these inequalities, we find that

$$b_1^\top w_1^* + b_2^\top w_2^* = c^\top (x_1 - x_2) \geq b_1^\top \tilde{u}_1^* + b_2^\top \tilde{u}_2^* = \rho \geq 0.$$

It follows that the vector $w^{*\top} = [w_1^{*\top}, w_2^{*\top}]$, together with $\tilde{u}^*$, satisfies system (5). The family of separating hyperplanes can be represented as

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : w_1^{*\top}(A_1 x - b_1) + \alpha\rho = 0\}, \tag{39}$$

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : -w_2^{*\top}(A_2 x - b_2) + (\alpha - 1)\rho = 0\}. \tag{40}$$

Thus, we have constructed two families of separating hyperplanes of form (20), (21) and (39), (40), respectively. For the first family, we use the normal solution to system (5); for the second, the optimal Lagrange multipliers for the linear programming problems (31). Both vectors $\tilde{u}^*$ and w^* satisfy (5). The following theorem asserts that the normal vector c and the scalar γ, which determine an arbitrary strictly separating hyperplane, can be expressed in terms of a solution to system (5).

Theorem 4. *Let the hyperplane $c^\top x = \gamma$ strictly separate two nonempty disjoint polyhedra X_1 and X_2. Then, there exists a solution u_1, u_2 to the system*

$$A_1^\top u_1 + A_2^\top u_2 = 0, \qquad b_1^\top u_1 + b_2^\top u_2 = \rho > 0, \qquad u_1 \geq 0, \quad u_2 \geq 0, \tag{41}$$

such that the vector c and the scalar γ are given by

$$c = A_1^\top u_1 = -A_2^\top u_2, \qquad \gamma = b_1^\top u_1 - \rho_1 = -b_2^\top u_2 + \rho_2,$$

where ρ_1 and ρ_2 are arbitrary positive constants such that $\rho_1 + \rho_2 = \rho$.

Proof. For definiteness, we assume that the given strictly separating hyperplane is such that all $x \in X_1$ satisfy the inequality $c^\top x > \gamma$, while all $x \in X_2$ satisfy the inequality $c^\top x < \gamma$. Then, the system

$$A_1 x \geq b_1, \qquad c^\top x \leq \gamma$$

is unsolvable, whereas the alternative system is consistent. Hence, there exist a vector $q \geq 0_{m_1}$ and a scalar $\eta \geq 0$, $\eta \in \mathbb{R}^1$ such that

$$A_1^\top q - c\eta = 0_n, \qquad b_1^\top q - \gamma\eta = \rho_1 > 0. \tag{42}$$

Here, ρ_1 is an arbitrary positive constant. The scalar η cannot vanish, since, otherwise, the consistent system (42) has the form

$$A_1^\top q = 0_n, \qquad b_1^\top q = \rho_1 > 0, \qquad q \geq 0_{m_1}.$$

Therefore, the alternative system $A_1 x \geq b_1$ is inconsistent, which contradicts the condition $X_1 \neq \emptyset$. Thus, (42) yields $c = A_1^\top q/\eta$ and $\gamma = b_1^\top q/\eta - \rho_1/\eta$. Using the notation $u_1 = q/\eta$ and $\rho_1/\eta = \rho_1$, we obtain

$$A_1^\top u_1 = c, \qquad b_1^\top u_1 - \gamma = \rho_1 > 0, \qquad u_1 \geq 0_{m_1}. \tag{43}$$

In a similar manner, we arrive at the relations

$$A_2^\top u_2 = -c, \qquad b_2^\top u_2 + \gamma = \rho_2 > 0. \tag{44}$$

Adding (43) and (44), we obtain the consistent system (41), which is alternative to (4). The theorem is proved. $\square$

Example 1. Let $n = 2$, $m = 6$, $\rho = 1$,

$$\begin{aligned} X_1 &= \{x \in \mathbb{R}^2 : -x^1 \geq 2,\ \ -x^1 - x^2 \geq 1,\ \ -x^1 + x^2 \geq 2,\ \ x^1 \geq -4\}, \\ X_2 &= \{x \in \mathbb{R}^2 : x^1 \geq 1,\ \ x^1 - x^2 \geq 0,\ \ 5x^1 + x^2 \geq 2\ \ -x^1 \geq -2\}. \end{aligned}$$

Solving problem (15) and using formulas (17) and (23), we obtain

$$x^* = \begin{bmatrix} 0.11 \\ 0.63 \end{bmatrix}, \quad \|z^*\| = 3.13, \quad c = \begin{bmatrix} -5.34 \\ -0.26 \end{bmatrix}.$$

Solving problems (31) and using formulas (34) and (36), we find that

$$\tilde{x} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}, \quad \hat{x} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \quad \tilde{\alpha} = 1.18, \quad \hat{\alpha} = -0.13.$$

From (18) and systems (37), (38), we have

$$\begin{aligned} \tilde{u}_1^{*\top} &= [0.18\ \ 0.15\ \ 0.13\ \ 0.00], \qquad & \tilde{u}_2^{*\top} &= [0.08\ \ 0.04\ \ 0.07\ \ 0.00], \\ w_1^{*\top} &= [0.44\ \ 0.02\ \ 0.00\ \ 0.00], \qquad & w_2^{*\top} &= [0.35\ \ 0.00\ \ 0.02\ \ 0.00]. \end{aligned}$$

Figure 1 shows the sets X_1 and X_2; the separating hyperplanes corresponding to $\alpha = \hat{\alpha}$, $\alpha = 0$, $\alpha = 1$ and $\alpha = \tilde{\alpha}$; the vector x^*; and the unit vector $Q = c/\|c\|$. The set X_1 belongs to the positive half-space with respect to c, whereas X_2 belongs to the negative half-space. The thickness of the family of hyperplanes corresponding to $0 \leq \alpha \leq 1$ is $\|y\| = 2.13$. The extension of this family with the use of $\hat{\alpha}$ and $\tilde{\alpha}$ results in $\|y\| = 2.80$.

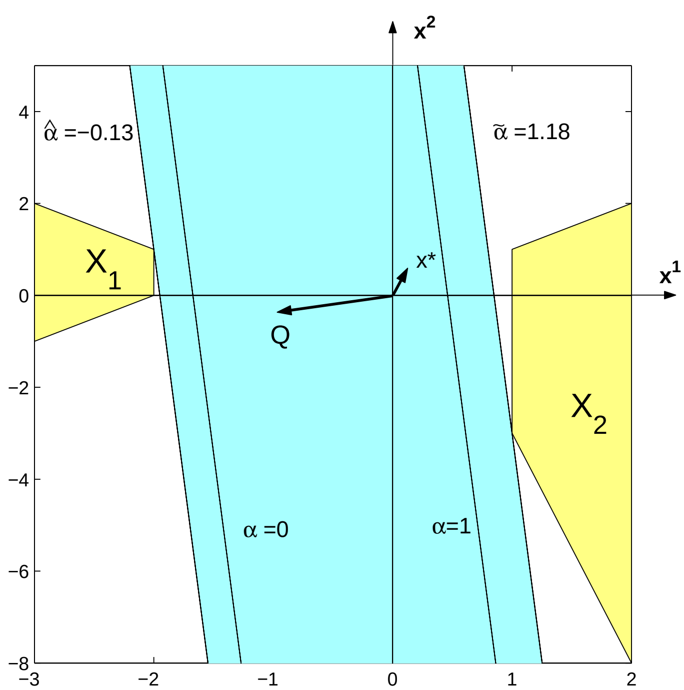

Fig. 1

3. FAMILY OF SEPARATING HYPERPLANES WITH A MAXIMAL THICKNESS

The problem of finding the minimal distance between two disjoint sets can be written in the form

$$\min_{x_1 \in X_1} \min_{x_2 \in X_2} \frac{1}{2}\|x_1 - x_2\|^2. \tag{45}$$

We change the variable to $p = x_1 - x_2$ and rewrite problem (45) as

$$\min_{p \in \mathbb{R}^n} \min_{x_2 \in X_2} \frac{1}{2}\|p\|^2 \tag{46}$$

subject to

$$A_1 x_2 + A_1 p \geq b_1, \qquad A_2 x_2 \geq b_2. \tag{47}$$

The norm $\|p\|$ is the same as the distance between the convex sets X_1 and X_2. The vector p obtained by solving problem (46), (47) will be called the vector determining the distance between these sets. The vector y introduced above is not always the same as the vector p produced by solving problem (46).

The Lagrangian function for problem (46) has the form

$$L(p, x_2, v) = \|p\|^2/2 + v_1^\top (b_1 - A_1 p - A_1 x_2) + v_2^\top (b_2 - A_2 x_2).$$

Using this function, we can write the dual problem:

$$\max_{v_1 \in \mathbb{R}^{m_1}_+} \max_{v_2 \in \mathbb{R}^{m_2}_+} \min_{x_2 \in \mathbb{R}^n} \min_{p \in \mathbb{R}^n} L(p, x_2, v). \tag{48}$$

The optimality conditions for the inner problem in (48) are as follows:

$$L_p(p, x_2, v) = p - A_1^\top v_1 = 0_n, \tag{49}$$

$$L_{x_2}(p, x_2, v) = -A_1^\top v_1 - A_2^\top v_2 = 0_n. \tag{50}$$

Relations (49) and (50) imply that $p = A_1^\top v_1 = -A_2^\top v_2$. The substitution of this expression into the Lagrangian function results in the dual Lagrangian function

$$\begin{aligned} \tilde{L}(x_2, v) &= \|A_1^\top v_1\|^2/2 + v_1^\top (b_1 - A_1 A_1^\top v_1 - A_1 x_2) + v_2^\top (b_2 - A_2 x_2) = \\ &= b_1^\top v_1 + b_2^\top v_2 - \|A_1^\top v_1\|^2/2 - x_2^\top (A_1^\top v_1 + A_2^\top v_2). \end{aligned}$$

Taking into account (50), we obtain the dual problem to problem (46), (47):

$$\max_{v_1 \in \mathbb{R}^{m_1}_+} \max_{v_2 \in \mathbb{R}^{m_2}_+} \{b_1^\top v_1 + b_2^\top v_2 - \|A_1^\top v_1\|^2/2\} \tag{51}$$

subject to

$$A_1^\top v_1 + A_2^\top v_2 = 0_n, \qquad v_1 \geq 0_{m_1}, \quad v_2 \geq 0_{m_2}. \tag{52}$$

Denote by $[p^*, x_2^*]$ a solution to problem (46), (47) and by $[v_1^*, v_2^*]$ a solution to the dual problem (51), (52). By the duality theorem, we have

$$b_1^\top v_1^* + b_2^\top v_2^* - \|A_1^\top v_1^*\|^2/2 = \|p^*\|^2/2. \tag{53}$$

Substituting $p^* = A_1^\top v_1^*$ into (53), we obtain

$$b_1^\top v_1^* + b_2^\top v_2^* = \|A_1^\top v_1^*\|^2. \tag{54}$$

Thus, we have the following assertion.

Theorem 5. *Every solution* $v^{*\top} = [v_1^{*\top}, v_2^{*\top}]$ *to the dual problem* (51), (52) *determines the unique first component* p^* *in a solution* $[p^*, x_2^*]$ *to problem* (48)*, which is given by the formulas*

$$p^* = A_1^\top v_1^* = -A_2^\top v_2^*. \tag{55}$$

Moreover, it holds that

$$b^\top v^* = \|A_1^\top v_1^*\|^2 = \|A_2^\top v_2^*\|^2 = \|p^*\|^2.$$

Theorem 5 implies that the vector v^* found from (51), (52) satisfies system (5) for $\rho = \|p^*\|^2$.

Note that a solution v^* to the dual problem (51), (52) determines only the first component p^* in a solution $[p^*, x_2^*]$ to the primal problem (46), (47). To determine x_2^*, one should substitute p^* into the constraints of the primal problem and solve the resulting system of inequalities

$$A_1 x_2 \geq b_1 - A_1 p^*, \qquad A_2 x_2 \geq b_2,$$

with respect to the vector x_2. Thus, the situation here is different from that of a pair of mutually dual problems, which was examined above.

Theorem 6 (on a family of parallel separating hyperplanes). *Let* X_1 *and* X_2 *be nonempty polyhedra with an empty intersection, and let* v^*, p^*, x_2^* *be a solution to problem* (48)*. Then, the family of parallel hyperplanes that separate* X_1 *and* X_2 *can be represented in the form*

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : p^{*\top} x - b_1^\top v_1^* + \alpha \|p^*\|^2 = 0\}, \tag{56}$$

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : p^{*\top} x + b_2^\top v_2^* - (1-\alpha) \|p^*\|^2 = 0\}, \tag{57}$$

where $\alpha \in [0, 1]$*. Moreover, if* $0 < \alpha < 1$*, then these hyperplanes strictly separate* X_1 *and* X_2*. The hyperplanes* $\Gamma(0)$ *and* $\Gamma(1)$ *are supporting hyperplanes for the sets* X_1 *and* X_2*, respectively. The thickness of the family* $\Gamma(\alpha)$ *is equal to* $\|p^*\|$*; it is the same as the distance between the polyhedra* X_1 *and* X_2*.*

Proof. Since v^* is a solution to problem (51), (52), we have

$$A_1^\top v_1^* + A_2^\top v_2^* = 0_n.$$

Multiplying this relation on the left by $x^\top$, where x is an arbitrary vector in $\mathbb{R}^n$, and subtracting the resulting equality from (54), we obtain

$$x^\top A_1^\top v_1^* + x^\top A_2^\top v_2^* - b_1^\top v_1^* - b_2^\top v_2^* = -\|A_1^\top v_1^*\|^2. \tag{58}$$

Using the coefficient $\alpha \in [0, 1]$ and the equality $\|A_1^\top v_1^*\| = \|A_2^\top v_2^*\|$, we can rewrite (58) as

$$v_1^{*\top}(A_1 x - b_1) + \alpha \|A_1^\top v_1^*\|^2 = v_2^{*\top}(b_2 - A_2 x) + (\alpha - 1)\|A_2^\top v_2^*\|^2.$$

Taking into account (55), we arrive at the equivalent representations (56) and (57) of the hyperplane $\Gamma(\alpha)$.

The triple $[v^*, p^*, x_2^*]$ is a Kuhn–Tucker point for problem (46). Therefore, the following complementary slackness conditions must hold:

$$v_1^{*\top} L_{v_1}(p^*, x_2^*, v^*) = 0, \qquad v_2^{*\top} L_{v_2}(p^*, x_2^*, v^*) = 0, \qquad v_1^* \geq 0_{m_1}, \quad v_2^* \geq 0_{m_2},$$
$$L_{v_1}(p^*, x_2^*, v^*) \leq 0_n, \qquad L_{v_2}(p^*, x_2^*, v^*) \leq 0_{m_2}.$$

From these conditions, we derive

$$v_1^{*\top}(b_1 - A_1 x_1^*) = 0, \qquad v_2^{*\top}(b_2 - A_2 x_2^*) = 0, \qquad A_1 x_1^* \geq b_1, \qquad A_2 x_2^* \geq b_2. \tag{59}$$

The relations obtained imply that $x_1^* \in X_1$, $x_1^* \in \Gamma(0)$, $x_2^* \in X_2$, and $x_2^* \in \Gamma(1)$.

For an arbitrary point x in X_1, we have $A_1 x \geq b_1$. Taking the scalar product of this inequality with the non-negative vector v_1^*, we obtain $v_1^{*\top}(A_1 x - b_1) \geq 0$. Taking into account (55), we arrive at the relation

$$p^{*\top} x - b_1^\top v_1^* \geq 0. \tag{60}$$

From (59), we find that

$$b_1^\top v_1^* = p^{*\top} x_1^*. \tag{61}$$

From (60), we conclude that $p^{*\top} x \geq p^{*\top} x_1^*$ for any $x \in X_1$ and that at least one point $x_1^* \in X_1$ belongs to the separating hyperplane $p^{*\top} x = p^{*\top} x_1^*$. Hence, X_1 belongs to one of the half-spaces determined by the hyperplane $\Gamma(0)$, and $\Gamma(0)$ is a supporting hyperplane for this set at its point x_1^*. In a similar way, we show that $\Gamma(1)$ is a supporting hyperplane for X_2 at the point x_2^*. It holds that

$$b_2^\top v_2^* = -p^{*\top} x_2^*. \tag{62}$$

All the points in X_2 satisfy the inequality $p^{*\top} x \leq p^{*\top} x_2^*$ and at least one point $x_2^* \in X_2$ belongs to the hyperplane $p^{*\top} x = p^{*\top} x_2^*$.

The distance between the supporting hyperplanes is the same as the distance between X_1 and X_2; both are equal to $\|p^*\|$, which follows from the formulation of problem (46), (47). The theorem is proved. $\square$

Note that, if relations (61) and (62) are taken into account, then hyperplanes $\Gamma(\alpha)$ of form (56), (57) can be represented as

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : p^{*\top} x - (1-\alpha) p^{*\top} x_1^* - \alpha p^{*\top} x_2^* = 0\}, \tag{63}$$

i.e., each member of the family of separating hyperplanes can be represented as a convex combination of supporting hyperplanes for X_1 and X_2. To construct a family of hyperplanes of form (63), one needs to solve problem (46), (47) in a space of variables of dimension $2n$. To represent the same family in form (56), (57), it is required to solve the dual problem (51), (52) in a space of variables of dimension m. The vector p^* appearing in this representation is expressed in terms of v^* by formula (55).

Now, we examine the following issue: is it possible to distinguish a solution to system (5) that determines a family of hyperplanes whose thickness is equal to the distance between the sets X_1 and X_2? According to formulas (12) and (13), the shift vector y and the thickness $\|y\|$ of the family of separating hyperplanes $\Gamma(\alpha)$ are given by $y = -\rho A_1^\top u_1 / \|A_1^\top u_1\|^2$ and $\|y\| = \rho / \|A_1^\top u_1\|$, respectively; here, $u^\top = [u_1^\top, u_2^\top]$ is a solution to system (5). It is then natural to pose the problem of finding a solution $u^{*\top} = [u_1^{*\top}, u_2^{*\top}] \in U$ to system (5) for which the thickness of the family of separating hyperplanes is maximal:

$$\frac{1}{2}\|A_1^\top u_1^*\|^2 = \min_{u \in U} \frac{1}{2}\|A_1^\top u_1\|^2, \tag{64}$$

$$U = \{u \in \mathbb{R}^m : A_1^\top u_1 + A_2^\top u_2 = 0_n,\ \ b_1^\top u_1 + b_2^\top u_2 = \rho,\ \ u_1 \geq 0_{m_1},\ \ u_2 \geq 0_{m_2}\}. \qquad (65)$$

In this case, the shift vector $y = \Gamma(1) - \Gamma(0)$ yields the thickness of the family of hyperplanes, which is identical to the minimal distance between the polyhedra X_1 and X_2. This is explained by the fact that the solution u^* to problem (64), (65) allows us to find the minimal distance between X_1 and X_2. It turns out that problems (64), (65) and (51), (52) are equivalent in the sense that the solution to one of them can be found from the solution to the other.

Theorem 7. *Let X_1 and X_2 be nonempty polyhedra with an empty intersection. Then, the solution v^* to problem* (51), (52) *and the solution u^* to problem* (64), (65) *satisfy the relations*

$$v^* = \frac{\rho u^*}{\|A_1^\top u_1^*\|^2}, \qquad u^* = \frac{\rho v^*}{b^\top v^*}. \qquad (66)$$

The family of separating hyperplanes can be represented in each of the following forms:

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : {u_1^*}^\top A_1 x - b_1^\top u_1^* + \alpha\rho = 0\}, \qquad (67)$$

$$\Gamma(\alpha) = \{x \in \mathbb{R}^n : -{u_2^*}^\top A_2 x + b_2^\top u_2^* + (\alpha - 1)\rho = 0\}, \qquad (68)$$

where $0 \leq \alpha \leq 1$. The thickness of this family is identical to the minimal distance between X_1 and X_2.

Proof. The vector u^* satisfies system (5); hence, $\|A_1^\top u_1^*\| \neq 0$ (see the first assertion in Theorem 2). By Theorem 5, it holds that $b^\top v^* \neq 0$. Formulas (66) are obtained by comparing the Kuhn–Tucker conditions for problems (51), (52) and (64), (65). Using (55), (66), and (12), we have

$$p^* = A_1^\top v_1^* = \frac{\rho A_1^\top u_1^*}{\|A_1^\top u_1^*\|^2} = -y.$$

It follows that, for the family of separating hyperplanes (67), (68), the thickness is identical to the minimal distance between X_1 and X_2:

$$\|p^*\| = \frac{\rho}{\|A_1^\top u_1^*\|} = \|y\|.$$

The theorem is proved. □

The dual problem to problem (64), (65) is as follows:

$$\max_{q \in \mathbb{R}^n} \ \max_{x \in \mathbb{R}^n} \ \max_{\xi \in \mathbb{R}^1} \left\{\rho\xi - \frac{1}{2}\|q\|^2\right\}$$

subject to

$$A_1(q + x) - b_1\xi \geq 0_{m_1}, \qquad A_2 x - b_2 \xi \geq 0_{m_2}.$$

The unknown vector q can be obtained from the solution u^* to the primal problem (64), (65) by using the formula $q^* = A_1^\top u_1^* = -A_2^\top u_2^*$. By the duality theorem, the solutions u^* and $[q^*, x^*, \xi^*]$ to the primal and dual problems satisfy the relation $\rho\xi^* = \|A_1^\top u_1^*\|^2$. Therefore, $\xi^* > 0$, and $p^* = q^*/\xi^*$, $x_2^* = x^*/\xi^*$.

Thus, we have constructed the three equivalent representations (56), (57), (63), and (67), (68) for the same family of separating hyperplanes whose thickness is equal to the minimal distance between the polyhedra. Each representation requires solving its own optimization problem.

4. CONSTRUCTION OF SEPARATING HYPERPLANES WITH THE HELP OF FARKAS' LEMMA

Consider the case where two polyhedra are represented by equality systems given on the nonnegative orthant; i.e., we have two nonempty sets

$$X_1 = \{x \in \mathbb{R}^n : A_1 x = b_1,\ \ x \geq 0_n\}, \qquad X_2 = \{x \in \mathbb{R}^n : A_2 x = b_2,\ \ x \geq 0_n\}$$

such that $X = X_1 \cap X_2 = \emptyset$.

According to Farkas' lemma, the inconsistency of the system

$$A_1 x = b_1, \qquad A_2 x = b_2, \qquad x \geq 0_n,$$

where the variables are nonnegative, implies that the system

$$A_1^\top u_1 + A_2^\top u_2 \leq 0_n, \qquad b_1^\top u_1 + b_2^\top u_2 = \rho \tag{69}$$

is consistent. Here, ρ is a positive constant.

System (69) is solvable, and every of its solutions satisfies the inequalities $\|u_1\| \neq 0$, $\|u_2\| \neq \neq 0$, $\|A_1^\top u_1\| \neq 0$, and $\|A_2^\top u_2\| \neq 0$. Indeed, assume the contrary; namely, let $A_1^\top u_1 = 0_n$. Since X_1 is nonempty, the alternative system $A_1^\top u_1 \leq 0$, $b_1^\top u_1 = \rho_1 \neq 0$ is inconsistent. By assumption, $A_1^\top u_1 = 0_n$; hence, $b_1^\top u_1 = 0$. In this case, (69) converts into the consistent system $A_2^\top u_2 \leq 0_n$, $b_2^\top u_2 = \rho$. However, this is the alternative system to $A_2 x = b_2$, $x \geq 0_n$. By assumption, the latter system is consistent, because X_2 is nonempty. The contradiction obtained proves that the equality $A_1^\top u_1 = 0_n$ is impossible. If $u_1 = 0_{m_1}$, then $A_1^\top u_1 = 0_n$, which is impossible.

Taking the scalar product of the inequality part in (69) with a nonnegative vector x and then subtracting the equality part, we obtain

$$u_1^\top (A_1 x - b_1) + u_2^\top (A_2 x - b_2) \leq -\rho < 0. \tag{70}$$

Define two linear functions of variable $x \in \mathbb{R}^n$ and a parameter $\alpha \in [0, 1]$ as follows:

$$\varphi_1(x, \alpha) = u_1^\top (A_1 x - b_1) + \alpha\rho, \qquad \varphi_2(x, \alpha) = -u_2^\top (A_2 x - b_2) - (1 - \alpha)\rho.$$

Then, inequality (70) can be rewritten as

$$\varphi_1(x, \alpha) = u_1^\top (A_1 x - b_1) + \alpha\rho \leq -u_2^\top (A_2 x - b_2) - (1 - \alpha)\rho = \varphi_2(x, \alpha). \tag{71}$$

Fix vectors u_1 and u_2 constituting an arbitrary solution to system (69). Using $\varphi_1(x, \alpha)$ and $\varphi_2(x, \alpha)$ we obtain two families of hyperplanes that correspond to $\alpha \in [0, 1]$ and separate X_1 and X_2.

If $x \in X_1$, then $\varphi_1(x, \alpha) \geq 0$ for $\alpha \in [0, 1]$. If $x \in X_2$, then $\varphi_2(x, \alpha) \leq 0$ for $\alpha \in [0, 1]$. Then, (71) implies that $\varphi_1(x, \alpha) \leq 0$. It follows that the hyperplanes in the family $\varphi_1(x, \alpha) = 0$, $\alpha \in [0, 1]$ separate X_1 and X_2. If $0 < \alpha < 1$, then inequality (71) shows that the hyperplane $\varphi_1(x, \alpha) = 0$ strictly separates these sets.

Now, we show that the condition $\varphi_2(x, \alpha) = 0$ determines the family of hyperplanes that separate X_1 and X_2 for $\alpha \in [0, 1]$ and strictly separate these sets if $0 < \alpha < 1$. Indeed, if $x \in X_2$, then $\varphi_2(x, \alpha) \leq 0$ for $\alpha \in [0, 1]$ and $\varphi_2(x, \alpha) < 0$ if $0 \leq \alpha < 1$. If $x \in X_1$, then (71) implies that $\varphi_2(x, \alpha) \geq 0$ for $\alpha \in [0, 1]$ and $\varphi_2(x, \alpha) > 0$ if $0 < \alpha \leq 1$.

Thus, by solving the alternative consistent system (69), we obtain two families of separating hyperplanes determined by $\varphi_1(x,\alpha)$ and $\varphi_2(x,\alpha)$. It follows that the case under analysis differs from the case of polyhedra given by inequality systems, which was examined above.

Define a nonnegative linear combination of $\varphi_1(x,\alpha)$ and $\varphi_2(x,\alpha)$ by setting

$$\varphi_3(x,\alpha) = \lambda_1\varphi_1(x,\alpha) + \lambda_2\varphi_2(x,\alpha).$$

Here, $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$.

Consider the following three families of separating hyperplanes and their unions:

$$\Gamma_i(\alpha) = \{x \in \mathbb{R}^n : \varphi_i(x,\alpha) = 0\}, \qquad \Gamma_i = \bigcup_{\alpha=0}^{1} \Gamma_i(\alpha), \qquad i = 1,2,3.$$

It is easy to show that the hyperplanes in $\Gamma_3(\alpha)$ separate X_1 and X_2 for any nonnegative scalars λ_1 and λ_2 of which at least one is nonzero.

As in the preceding section, we denote by p^* the vector joining the two nearest points in X_1 and X_2; then, the distance between these sets is $\|p^*\|$. It is often possible to choose λ_1 and λ_2 such that the vector

$$p^* = \lambda_1 A_1^\top u_1 - \lambda_2 A_2^\top u_2,$$

where $u^\top = [u_1^\top, u_2^\top]$, satisfies condition (69). In this case, p^* is the normal vector of $\Gamma_3(x,\alpha)$, and the thickness of this family is equal to the minimal distance between X_1 and X_2. The following theorem is an analogue of Theorem 4.

Theorem 8. *Let the hyperplane $c^\top x - \gamma = 0$ strictly separate two nonempty disjoint polyhedra X_1 and X_2. Then, there exists a solution u_1, u_2 to the system*

$$A_1^\top u_1 + A_2^\top u_2 \leq 0, \qquad b_1^\top u_1 + b_2^\top u_2 = \rho > 0$$

such that

$$A_1^\top u_1 \leq c, \quad A_2^\top u_2 \leq -c, \quad \gamma = b_1^\top u_1 - \rho_1 = -b_2^\top u_2 + \rho_2,$$

where ρ_1 and ρ_2 are arbitrary positive constants such that $\rho_1 + \rho_2 = \rho$.

The proof is an almost word-for-word repetition of the proof of Theorem 4.

Theorem 8 asserts that the polyhedra given by equality systems on the nonnegative orthant are different from the polyhedra given by inequality systems in the sense that it is not always possible to find u_1 and u_2 that satisfy the consistent alternative system (69) and, at the same time, satisfy either the condition $c = A_1^\top u_1$ or the condition $c = -A_2^\top u_2$. In other words, there may not exist vectors u_1 and u_2 that satisfy (69) and have the property that the separating hyperplane $c^\top x - \gamma = 0$ belongs to either $\Gamma_1(\alpha)$ or $\Gamma_2(\alpha)$.

Example 2. Let the polyhedra be given by the conditions

$$\begin{aligned} X_1 &= \{x \in \mathbb{R}^2 : x^1 + x^2 = 1,\ x \leq 0_2\}, \\ X_2 &= \{x \in \mathbb{R}^2 : 2x^1 - x^2 = 6,\ x \geq 0_2\}. \end{aligned}$$

We set $\rho = 1$ in system (69). Figure 2 shows three unions of the families of separating hyperplanes: Γ_1, Γ_2, and Γ_3; it also shows the vectors

$$c_1^\top = [-1/2 \;\; -1/2], \quad c_2^\top = [-1/2 \;\; 1/4], \quad c_3^\top = [-1 \;\; 0], \quad x^{*\top} = [2.6 \;\; 0].$$

From the formulas given above, we obtain

$$\|p^*\| = 2, \quad \lambda_1 = 4/3, \quad \lambda_2 = 8/3, \quad u_1^* = -1/2, \quad u_2^* = 1/4,$$

$$
\begin{aligned}
\varphi_1(x,\alpha) &= -x^1/2 - x^2/2 + 1/2 + \alpha, \\
\varphi_2(x,\alpha) &= -x^1/2 + x^2/4 + 3/2 - (1-\alpha), \\
\varphi_3(x,\alpha) &= -2x^1 + 2 + 4\alpha.
\end{aligned}
$$

Using the last formula, we find the family with a maximal thickness:

$$\Gamma_3(\alpha) = \{x \in \mathbb{R}^n : x^1 = 1 + 2\alpha\}.$$

On the other hand, we arrive at the same family $\Gamma_3(\alpha)$ by setting $x_1^{*\top} = [1\ 0]$, $x_2^{*\top} = [3\ 0]$, and $p^{*\top} = [-2\ 0]$ in (63).

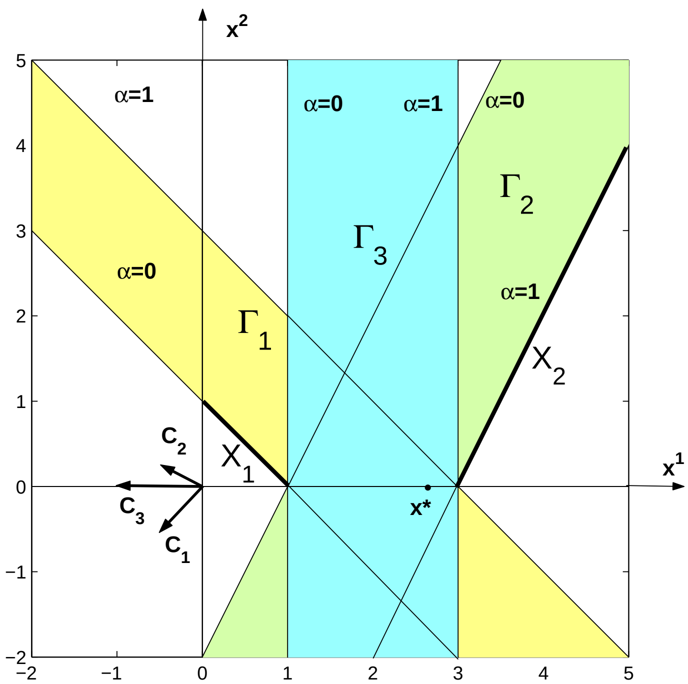

Fig. 2

5. THE GENERALIZED NEWTON METHOD

Since we usually have $n \ll m$ in the problem of separating polyhedra given by inequality systems (4), it is preferable to solve problem (25): minimize the function $F(x) = \|(b - Ax)_+\|^2/2$, which depends on n variables. The unconstrained minimization of $F(x)$ can be performed by any method, such as the conjugate gradient method. However, Mangasarian showed that the generalized Newton method is especially efficient for the unconstrained optimization of a piecewise quadratic function (see [4, 5]). We give a brief description of this method.

The objective function $F(x)$ of problem (15) is convex, piecewise quadratic, and differentiable. Such a function does not have the conventional Hessian matrix. Indeed, the gradient

$$F_x(x) = -A^\top (b - Ax)_+$$

of $F(x)$ is not differentiable. However, for this function, one can define the generalized Hessian matrix, which is an n-by-n symmetric positive semidefinite matrix of the form

$$\partial^2 F(x) = A^\top D^\sharp(z) A.$$

Here, $D^\sharp(z)$ denotes the m-by-m diagonal matrix whose ith diagonal entry z^i is equal to one if $(b - Ax)^i > 0$; z^i is equal to zero if $(b - Ax)^i \leq 0$ $(i = 1, 2, \ldots, m)$. Since the generalized Hessian matrix can be singular, the following modified Newton direction is used:

$$-\left[\partial^2 F(x) + \delta I_n\right]^{-1} F_x(x),$$

where δ is a small positive number (in our calculations, we typically set $\delta = 10^{-4}$) and I_n is the identity matrix of order n. In this case, the modified Newton method has the form

$$x_{s+1} = x_s - \left[\partial^2 F(x_s) + \delta I_n\right]^{-1} F_x(x_s). \tag{72}$$

We used the following stopping criterion for this method:

$$\|x_{s+1} - x_s\| \leq \text{tol}.$$

Mangasarian has studied the convergence of the generalized Newton method as applied to the unconstrained minimization of a convex piecewise quadratic function of this type with the step size chosen by the Armijo rule. The proof of the finite global convergence of the generalized Newton method can be found in [4] – [6].

The generalized Newton method as applied to the unconstrained minimization problem (15) was implemented in Matlab and showed good performance in solving large-scale test problems. For instance, problem (15) with $n = 500$ and $m = 10^4$, whose matrix A was fully filled with nonzero entries, was solved in less than one minute on a 2.24 GHz Pentium-IV computer.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 03-01-00465) and by the Program of the State Support of Leading Scientific Schools (project no. NSh-1737.2003.1).

REFERENCES

1. I.I. Eremin, *Theory of Linear Optimization* (Ural. Otd. Ross. Akad. Nauk, Yekaterinburg, 1998) [in Russian].

2. A.I. Golikov and Yu.G. Evtushenko, “Theorems of the Alternative and Their Application in Numerical Methods”, Zh. Vychisl. Mat. Mat. Fiz. **43**, 354–375 (2003) [Comput. Math. Math. Phys. **43**, 338–358 (2003)].

3. F.P. Vasil’ev and A.Yu. Ivanitskii, *Linear Programming* (Faktorial, Moscow, 2003) [in Russian].

4. O.L. Mangasarian, “A Finite Newton Method for Classification”, Optimizat. Methods Software **17**, 913–930 (2002).

5. O.L. Mangasarian, “A Newton Method for Linear Programming”, J. Optimizat. Theory Appl. **121**, 1–18 (2004).

6. C. Kanzow, H. Qi, and L. Qi, “On the Minimum Norm Solution of Linear Programs”, J. Optimizat. Theory Appl. **116**, 333–345 (2003).