ELSEVIER
Available online at www.sciencedirect.com
SCIENCE DIRECT®
Linear Algebra and its Applications 416 (2006) 184–213
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

The distance between two convex sets

Achiya Dax *

Hydrological Service, P.O. Box 36118, Jerusalem 91360, Israel

Received 2 February 2005; accepted 19 March 2006

Submitted by A. Berman

Abstract

In this paper we explore the duality relations that characterize least norm problems. The paper starts by presenting a new Minimum Norm Duality (MND) theorem, one that considers the distance between two convex sets. Roughly speaking the new theorem says that the shortest distance between the two sets is equal to the maximal "separation" between the sets, where the term "separation" refers to the distance between a pair of parallel hyperplanes that separates the two sets.

The second part of the paper brings several examples of applications. The examples teach valuable lessons about the role of duality in least norm problems, and reveal new features of these problems. One lesson exposes the polar decomposition which characterizes the "solution" of an inconsistent system of linear inequalities. Another lesson reveals the close links between the MND theorem, theorems of the alternatives, steepest descent directions, and constructive optimality conditions.

Keywords: A new minimum norm duality theorem; The distance between two convex polytopes; The distance between two ellipsoids; Linear least norm problems; Inconsistent systems of linear inequalities; The polar decomposition of the least deviation problem; Theorems of the alternatives; Steepest descent directions; Constructive optimality conditions; The double role of duality in least norm problems

1. Introduction

Let $\mathscr{Y}$ be a nonempty closed convex set and let $\mathbf{z}$ be some point outside $\mathscr{Y}$. The best approximation problem is to find a point $\mathbf{y}^* \in \mathscr{Y}$ which is closest to $\mathbf{z}$ among all the points of $\mathscr{Y}$. The interest in this problem is common to several branches of mathematics, such as Approximation Theory, Functional Analysis, Convex Analysis, Optimization, Numerical Linear Algebra, Statistics, and

* Tel.: +972 26442506; fax: +972 26442529.
E-mail address: dax20@water.gov.il

0024-3795/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2006.03.022

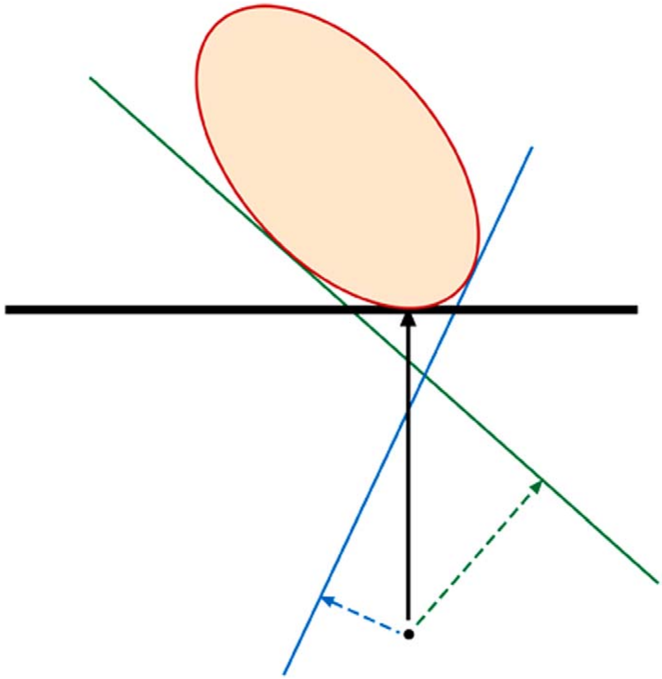

Fig. 1. The Nirenberg–Luenberger MND theorem.

other fields. However, the duality properties of this problem are not quite well known behind the Hahn–Banach theory on bounded linear functionals in normed linear vector spaces. One aim of this paper is, therefore, to point attention to the elegancy of the duality relations which characterize the best approximation problem. A second aim is to show that in $\mathscr{R}^m$ these results can be derived from simple geometric arguments, without invoking the Hahn–Banach theorem or the duality theorems of Lagrange and Fenchel. A third aim of this paper is to establish a new duality theorem, one that considers the distance between two convex sets.

The Minimum Norm Duality (MND) theorem considers the distance between a point $\mathbf{z}$ and a convex set $\mathscr{Y}$. It says that the shortest distance from $\mathbf{z}$ to $\mathscr{Y}$ is equal to the maximum of the distances from $\mathbf{z}$ to any hyperplane separating $\mathbf{z}$ and $\mathscr{Y}$ (see Fig. 1). This fundamental observation gives rise to several useful duality relations in best approximation problems, linear least norm problems, and theorems of the alternative. As far as we know, the first statement of the MND theorem is due to Nirenberg [40], who established this assertion in any normed linear space by applying the Hahn–Banach theorem. The name "MND theorem" was coined by Luenberger [29], who also derived the "alignment" relation between primal and dual solutions.

The new MND theorem considers the distance between two convex sets. Roughly speaking it says that the shortest distance between the two sets is equal to the maximal "separation" between the sets, where the term "separation" refers to the distance between a pair of parallel hyperplanes that separates the two sets (see Fig. 2).

In order to distinguish between the two MND theorems we refer to the first one as the Nirenberg–Luenberger MND theorem. (Although it is quite possible that the theorem was known earlier as result of the geometric Hahn–Banach theorem.) It is noted in the last section that the new MND theorem can be derived from the first one. However, while former proofs of the

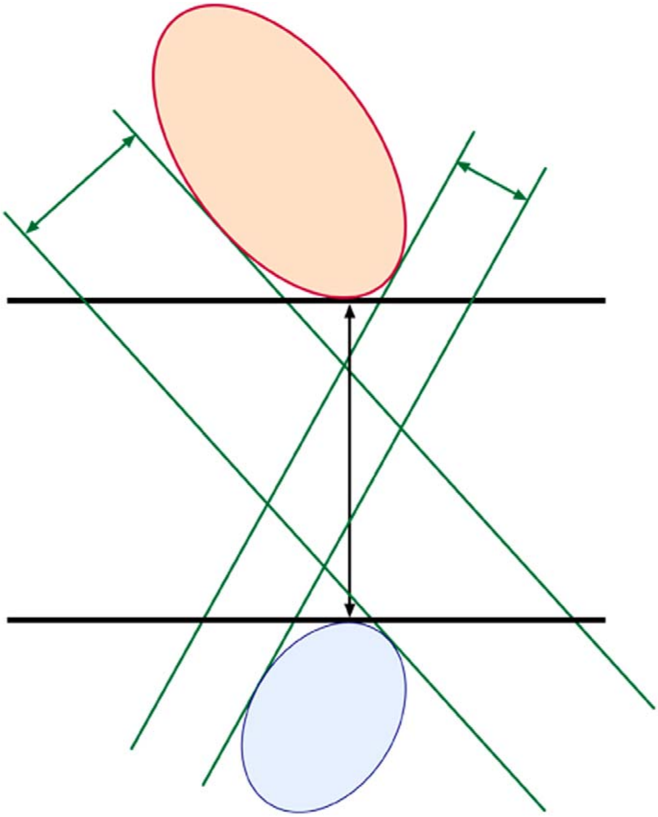

Fig. 2. The new MND theorem.

Nirenberg–Luenberger theorem rely on the Hahn–Banach theorem in a general normed linear vector space, e.g., [19,20,27,29,40], there are several practical applications which consider convex sets in $\mathscr{R}^m$. See the second part of the paper. This raises the question of whether we can find a direct proof that does not rely on the Hahn–Banach theory. A further aim of this paper is, therefore, to provide an elementary proof which is based on simple geometric ideas. For this purpose the coming discussion is restricted to convex sets in $\mathscr{R}^m$.

The plan of our paper is as follows. The first part contains necessary background. Section 2 explains the basic facts on dual norms and alignment. Section 3 derives an explicit formula for the distance between a point and hyperplane. This formula is used to calculate the distance between two parallel hyperplanes. The necessary facts on support functions and separating hyperplanes are given in Section 4. The reader who is familiar with these issues may skip to Section 5, in which the new MND theorem is proved.

The second part of the paper brings several examples of applications. The examples teach valuable lessons about duality in linear least norm problems, and reveal several new features of these problems. One lesson exposes the role of the polar decomposition when "solving" an inconsistent system of linear inequalities. Another lesson reveals the close links between the MND theorem, theorems of the alternative, steepest descent directions, and constructive optimality conditions.

Part 1: A new minimum norm duality theorem

2. Norms and duality

A norm on $\mathscr{R}^m$ is a real-valued function $\|\cdot\|$ which maps each vector $\mathbf{x}$ in $\mathscr{R}^m$ into a real number $\|\mathbf{x}\|$ and satisfies the following three requirements:

(a) $\|\mathbf{x}\| \geqslant 0 \; \forall \mathbf{x} \in \mathscr{R}^m$ and $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$;
(b) $\|\mathbf{x} + \mathbf{z}\| \leqslant \|\mathbf{x}\| + \|\mathbf{z}\| \; \forall \mathbf{x}, \mathbf{z} \in \mathscr{R}^m$;
(c) $\|a\mathbf{x}\| = |a| \cdot \|\mathbf{x}\| \; \forall a \in \mathscr{R}, \mathbf{x} \in \mathscr{R}^m$.

Recall that $\|\mathbf{x}\|$ is a convex function while the unit norm ball

$$\mathscr{B} = \{\mathbf{x} \mid \|\mathbf{x}\| \leqslant 1\}$$

is a closed bounded convex set which contains the origin in its interior; see Refs. [25,49]. The **dual norm** to $\|\cdot\|$ is denoted by $\|\cdot\|'$. This norm is obtained from $\|\cdot\|$ in the following way. Given a vector $\mathbf{z} \in \mathscr{R}^m$, then $\|\mathbf{z}\|'$ is defined by the following rule:

$$\|\mathbf{z}\|' = \max_{\|\mathbf{x}\| \leqslant 1} \mathbf{x}^{\mathrm{T}}\mathbf{z}.$$

For the sake of clarity we mention that

$$\mathbf{x}^{\mathrm{T}}\mathbf{z} = \sum_{i=1}^{m} x_i z_i \quad \text{for all } \mathbf{x} = (x_1, \ldots, x_m)^{\mathrm{T}} \in \mathscr{R}^m \text{ and } \mathbf{z} = (z_1, \ldots, z_m)^{\mathrm{T}} \in \mathscr{R}^m.$$

Since $\mathscr{B}$ is a compact set, the optimal value is attained for some vector $\mathbf{x} \in \mathscr{B}$. The above definition implies the general Hölder inequality

$$|\mathbf{x}^{\mathrm{T}}\mathbf{z}| \leqslant \|\mathbf{x}\| \|\mathbf{z}\|', \quad \forall \mathbf{x}, \mathbf{z} \in \mathscr{R}^m,$$

and that $\|\cdot\|$ is the dual norm to $\|\cdot\|'$. A pair of vectors that satisfy the equality

$$\mathbf{x}^{\mathrm{T}}\mathbf{z} = \|\mathbf{x}\| \|\mathbf{z}\|'$$

is said to be **aligned** with respect to $\|\cdot\|$ or $\|\cdot\|'$.

A well known example of dual norms in $\mathscr{R}^m$ is related to the ℓ_p norm

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p},$$

where $1 < p < \infty$ is a real number. The dual norm is the ℓ_q norm

$$\|\mathbf{z}\|_q = \left(\sum_{i=1}^{m} |z_i|^q \right)^{1/q},$$

where $q = p/(p-1)$. That is, $1/p + 1/q = 1$. This pair of dual norms satisfies the Hölder inequality

$$\sum_{i=1}^{m} |x_i z_i| \leqslant \|\mathbf{x}\|_p \|\mathbf{z}\|_q,$$

and the vectors $\mathbf{x}$ and $\mathbf{z}$ are aligned if and only if

$$\operatorname{sign}(x_i) = \operatorname{sign}(z_i) \quad \text{and} \quad (|x_i|/\|\mathbf{x}\|_p)^{1/q} = (|z_i|/\|\mathbf{z}\|_q)^{1/p} \quad \text{for } i = 1, \ldots, m,$$

e.g., [29] or [54]. Similar duality relations hold between the ℓ_1 norm

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{m} |x_i|$$

and the ℓ_∞ norm

$$\|\mathbf{z}\|_\infty = \max_{i=1,\ldots,m} |z_i|,$$

but the alignment relations need some corrections. When $p = q = 2$ we obtain the Euclidean norm,

$$\|\mathbf{x}\|_2 = (\mathbf{x}^{\mathrm{T}}\mathbf{x})^{1/2} = \left(\sum_{i=1}^{m} x_i^2\right)^{1/2},$$

which is privileged to be its own dual.

Let $\mathbf{z}$ be a given point in $\mathscr{R}^m$. Then, $\mathbf{x} \in \mathscr{R}^m$ is a **dual vector** of $\mathbf{z}$ with respect to $\|\cdot\|$ if it satisfies $\|\mathbf{x}\| = 1$ and $\mathbf{x}^{\mathrm{T}}\mathbf{z} = \|\mathbf{z}\|'$. The uniqueness of the dual vector is related to the question of whether $\|\cdot\|$ is a strictly convex norm. Recall that a norm $\|\cdot\|$ is said to be **strictly convex** if the unit sphere $\{\mathbf{x} | \|\mathbf{x}\| = 1\}$ contains no line segment. Now, one can verify that a norm $\|\cdot\|$ is strictly convex if and only if each $\mathbf{z} \in \mathscr{R}^m$ has a unique dual vector.

A norm $\|\cdot\|$ is called **smooth** if, at each boundary point of $\mathscr{B}$, there is a unique hyperplane that supports $\mathscr{B}$. Of course a smooth norm is not necessarily strictly convex and vice versa. However, there are interesting links between these properties. Let $\{\mathbf{u} | \mathbf{n}_0^{\mathrm{T}}\mathbf{u} = \mathbf{n}_0^{\mathrm{T}}\mathbf{x}_0\}$ be a supporting hyperplane of $\mathscr{B}$ at a point $\mathbf{x}_0$, $\|\mathbf{x}_0\| = 1$, with a normal vector $\mathbf{n}_0 \in \mathscr{R}^m$. That is, $\mathbf{n}_0^{\mathrm{T}}\mathbf{x} \leqslant \mathbf{n}_0^{\mathrm{T}}\mathbf{x}_0 \ \forall \mathbf{x} \in \mathscr{B}$. (The existence of a supporting hyperplane is established in Section 4.) The last inequality means that $\mathbf{x}_0$ is a dual vector of $\mathbf{n}_0$ with respect to $\|\cdot\|$. This shows that $\mathbf{x}_0$ and $\mathbf{n}_0$ are aligned and that $\mathbf{n}_0/\|\mathbf{n}_0\|'$ is a dual vector of $\mathbf{x}_0$ with respect to $\|\cdot\|'$ (see Fig. 3). Therefore, if there is more than one supporting hyperplane at $\mathbf{x}_0$, then the dual vector of $\mathbf{x}_0$ with respect to $\|\cdot\|'$ is not unique. In other words, $\|\cdot\|'$ is strictly convex if and only if $\|\cdot\|$ is smooth. It is also worthwhile mentioning that a norm is smooth if and only if it is continuously differentiable at each point $\mathbf{x} \neq 0$; e.g., [44, pp. 113–118]. For further discussion of dual norms and their properties see [18,25,26,46,54].

3. The distance between parallel hyperplanes

Let $\|\cdot\|$ be some arbitrary norm in $\mathscr{R}^m$ and let $\|\cdot\|'$ denote the corresponding dual norm. A hyperplane in $\mathscr{R}^m$ is a set of the form

$$\mathscr{H} = \{\mathbf{x} | \mathbf{a}^{\mathrm{T}}\mathbf{x} = \alpha\},$$

where $\mathbf{a} \in \mathscr{R}^m$ and $\alpha \in \mathscr{R}$. The distance between $\mathscr{H}$ and a point $\mathbf{z} \in \mathscr{R}^m$ is defined as

$$\operatorname{dist}(\mathbf{z}, \mathscr{H}) = \inf_{\mathbf{x} \in \mathscr{H}} \|\mathbf{z} - \mathbf{x}\|.$$

Assume first that $\mathbf{z}$ lies in the positive side of $\mathscr{H}$. That is, $\mathbf{a}^{\mathrm{T}}\mathbf{z} > \alpha$. In this case one can verify that

$$\operatorname{dist}(\mathbf{z}, \mathscr{H}) = (\mathbf{a}^{\mathrm{T}}\mathbf{z} - \alpha)/\|\mathbf{a}\|'. \tag{3.1}$$

Otherwise, when $\mathbf{z}$ lies in the negative side of $\mathscr{H}$,

$$\operatorname{dist}(\mathbf{z}, \mathscr{H}) = (\alpha - \mathbf{a}^{\mathrm{T}}\mathbf{z})/\|\mathbf{a}\|'.$$

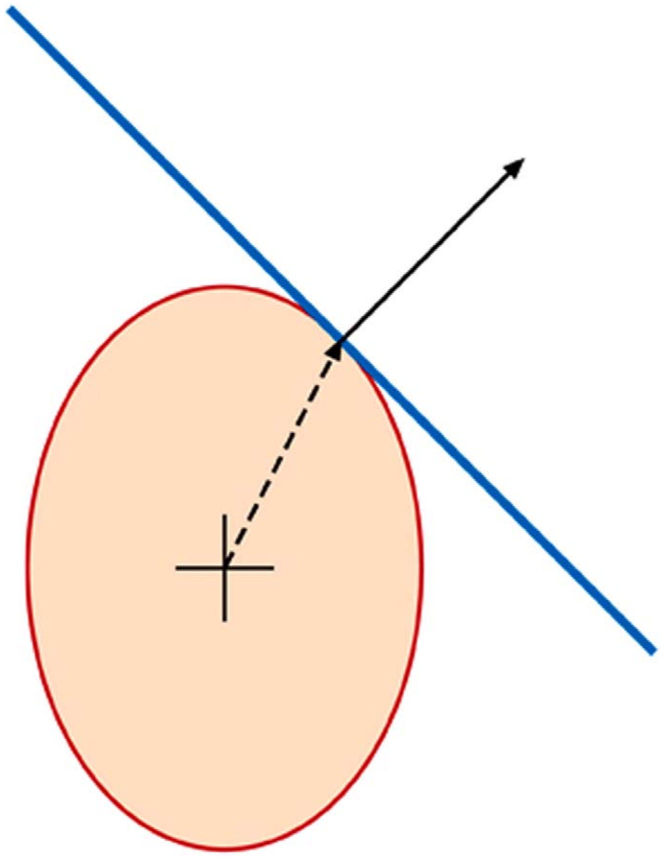

Fig. 3. Alignment relations on a unit norm ball.

See [17] or [34] for detailed proof of these assertions.

Let $\mathscr{H}_1 = \{\mathbf{x}|\mathbf{a}^T\mathbf{x} = \alpha_1\}$ and $\mathscr{H}_2 = \{\mathbf{x}|\mathbf{a}^T\mathbf{x} = \alpha_2\}$ be a pair of parallel hyperplanes such that $\alpha_1 < \alpha_2$. The distance between these sets is defined as

$$\text{dist}(\mathscr{H}_1, \mathscr{H}_2) = \inf\{\|\mathbf{x}_1 - \mathbf{x}_2\| | \mathbf{x}_1 \in \mathscr{H}_1, \mathbf{x}_2 \in \mathscr{H}_2\}.$$

Using (3.1) we see that

$$\text{dist}(\mathbf{x}_2, \mathscr{H}_1) = (\mathbf{a}^T\mathbf{x}_2 - \alpha_1)/\|\mathbf{a}\|' \quad \forall \mathbf{x}_2 \in \mathscr{H}_2.$$

Therefore, since $\mathbf{a}^T\mathbf{x}_2 = \alpha_2 \quad \forall \mathbf{x}_2 \in \mathscr{H}_2$,

$$\text{dist}(\mathscr{H}_1, \mathscr{H}_2) = (\alpha_2 - \alpha_1)/\|\mathbf{a}\|'. \tag{3.2}$$

4. Separating hyperplanes

In this section we mention some useful relations between hyperplanes and convex sets. The reader is referred to [17,24,41,43–45] or [49], for detailed proofs and discussions of these properties. Let $\mathscr{K}$ be a nonempty convex set in $\mathscr{R}^m$. Here and henceforth $\overline{\mathscr{K}}$ denotes the closure of $\mathscr{K}$, $\mathscr{K}^\circ$ denotes the interior of $\mathscr{K}$, and $ri\mathscr{K}$ denotes the relative interior of $\mathscr{K}$. A hyperplane $\mathscr{H} = \{\mathbf{x}|\mathbf{a}^T\mathbf{x} = \alpha\}$ is said to be a **supporting hyperplane** of $\mathscr{K}$ if

$$\sup_{\mathbf{x}\in\mathscr{K}} \mathbf{a}^{\mathrm{T}}\mathbf{x} = \alpha.$$

If, in addition, there exists a point $\mathbf{x}^* \in \overline{\mathscr{K}}$ such that $\mathbf{a}^{\mathrm{T}}\mathbf{x}^* = \alpha$, then $\mathscr{H}$ is said to be a supporting hyperplane of $\mathscr{K}$ at $\mathbf{x}^*$. The function

$$\alpha(\mathbf{a}) = \sup_{\mathbf{x}\in\mathscr{K}} \mathbf{a}^{\mathrm{T}}\mathbf{x}$$

is called the **support function** of $\mathscr{K}$. (If $\mathscr{K}$ is not bounded it is possible to have $\alpha(\mathbf{a}) = \infty$.)

Theorem 1 (The Projection Theorem). *Let* $\mathbf{z}$ *be some point of* $\mathscr{R}^m$ *which does not belong to* $\overline{\mathscr{K}}$. *Then there exists a unique point* $\mathbf{x}^* \in \overline{\mathscr{K}}$ *that satisfies the following properties*:

$$\|\mathbf{z} - \mathbf{x}^*\|_2 = \inf_{\mathbf{x}\in\mathscr{K}} \|\mathbf{z} - \mathbf{x}\|_2 \tag{4.1}$$

and

$$(\mathbf{z} - \mathbf{x}^*)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*) \leqslant 0 \quad \forall \mathbf{x} \in \mathscr{K}, \tag{4.2}$$

where $\|\mathbf{y}\|_2 = (\mathbf{y}^{\mathrm{T}}\mathbf{y})^{1/2}$ *denotes the Euclidean norm.*

Corollary 2 (The Separating Hyperplane Theorem). *Define*

$$\mathbf{a} = \mathbf{z} - \mathbf{x}^* \quad \textit{and} \quad \alpha = \mathbf{a}^{\mathrm{T}}\mathbf{x}^* = \sup_{\mathbf{x}\in\mathscr{K}} \mathbf{a}^{\mathrm{T}}\mathbf{x}.$$

Then the hyperplane $\mathscr{H} = \{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x} = \alpha\}$ *separates* $\overline{\mathscr{K}}$ *and* $\mathbf{z}$. *More precisely*, $\overline{\mathscr{K}}$ *is contained in the negative halfspace* $\mathscr{H}_- = \{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x} \leqslant \alpha\}$, *while* $\mathbf{z}$ *satisfies* $\mathbf{a}^{\mathrm{T}}\mathbf{z} > \alpha$. *Note also that* $\mathscr{H}$ *is a supporting hyperplane of* $\mathscr{K}$ *at* $\mathbf{x}^*$.

The point $\mathbf{x}^*$ is called the **Euclidean projection** of $\mathbf{z}$ on $\mathscr{K}$. The next theorem considers the case when $\mathbf{z}$ is a boundary point of $\mathscr{K}$. Recall that the boundary of $\mathscr{K}$ consists of all the points $\mathbf{z} \in \mathscr{R}^m$ that have the following property: Every neighborhood of $\mathbf{z}$ contains at least one point of $\mathscr{K}$ and one point not in $\mathscr{K}$. Recall also that $\mathscr{K}$ and $\overline{\mathscr{K}}$ share the same boundary.

Theorem 3 (The Existence of Supporting Hyperplane). *Let* $\mathbf{z}$ *be a boundary point of* $\overline{\mathscr{K}}$. *Then there exists a vector* $\mathbf{a} \in \mathscr{R}^m$ *such that*

$$\sup_{\mathbf{x}\in\mathscr{K}} \mathbf{a}^{\mathrm{T}}\mathbf{x} = \mathbf{a}^{\mathrm{T}}\mathbf{z}$$

and the hyperplane $\mathscr{H} = \{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x} = \mathbf{a}^{\mathrm{T}}\mathbf{z}\}$ *supports* $\mathscr{K}$ *at the point* $\mathbf{z}$.

Corollary 4. *If* $\mathbf{0}$ *is a boundary point of* $\mathscr{K}$ *then there exists a nonzero vector* $\mathbf{a} \in \mathscr{R}^m$ *that satisfies*

$$\mathbf{a}^{\mathrm{T}}\mathbf{x} \leqslant 0 \quad \forall \mathbf{x} \in \mathscr{K}.$$

Corollary 5. *Let* $\mathscr{Y}$ *and* $\mathscr{Z}$ *be two convex sets in* $\mathscr{R}^m$ *such that*

$$\mathrm{ri}\,\mathscr{Y} \cap \mathrm{ri}\,\mathscr{Z} = \emptyset \quad \textit{but} \quad \overline{\mathscr{Y}} \cap \overline{\mathscr{Z}} \neq \emptyset.$$

Then for any point $\mathbf{x}^* \in \overline{\mathscr{Y}} \cap \overline{\mathscr{Z}}$ *there exists a nonzero vector* $\mathbf{a} \in \mathscr{R}^m$ *such that*

$$\sup_{\mathbf{y}\in\mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y} = \mathbf{a}^{\mathrm{T}}\mathbf{x}^* = \inf_{\mathbf{z}\in\mathscr{Z}} \mathbf{a}^{\mathrm{T}}\mathbf{z}. \tag{4.3}$$

5. The distance between two convex sets

Let $\|\cdot\|$ be some arbitrary norm on $\mathscr{R}^m$ and let $\|\cdot\|'$ denote the corresponding dual norm. Let $\mathscr{Y}$ and $\mathscr{Z}$ be two nonempty convex sets in $\mathscr{R}^m$ such that $\mathscr{Y} \cap \mathscr{Z} = \emptyset$. The distance between $\mathscr{Y}$ and $\mathscr{Z}$ is defined as

$$\text{dist}(\mathscr{Y}, \mathscr{Z}) = \inf\{\|\mathbf{y} - \mathbf{z}\| \,|\, \mathbf{y} \in \mathscr{Y}, \mathbf{z} \in \mathscr{Z}\}.$$

Recall that the difference set,

$$\mathscr{X} = \mathscr{Y} - \mathscr{Z} = \{\mathbf{y} - \mathbf{z} | \mathbf{y} \in \mathscr{Y}, \mathbf{z} \in \mathscr{Z}\}$$

is convex. Note also that $\mathbf{0} \notin \mathscr{X}$ and

$$\text{dist}(\mathscr{Y}, \mathscr{Z}) = \text{dist}(\mathbf{0}, \mathscr{X}) = \inf_{\mathbf{x} \in \mathscr{X}} \|\mathbf{x}\|.$$

The motivation behind this formulation lies in the following observations.

Lemma 6. *The least norm problem*

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{x}\| \\ &\text{subject to} \quad \mathbf{x} \in \overline{\mathscr{X}} \end{aligned} \tag{5.1}$$

is always solvable. In other words, there exists a point $\mathbf{x}^* \in \overline{\mathscr{X}}$ *such that* $\|\mathbf{x}^*\| \leqslant \|\mathbf{x}\| \quad \forall \mathbf{x} \in \overline{\mathscr{X}}$. *Of course, if* $\|\cdot\|$ *is a strictly convex norm then the problem has a unique solution.*

Proof. Define $\mathscr{B}_1 = \{\mathbf{x} | \|\mathbf{x}\| \leqslant \|\mathbf{x}_1\|\}$ where $\mathbf{x}_1 \neq \mathbf{0}$ is some point of $\mathscr{X}$. Then the intersection set $\mathscr{X}_1 = \overline{\mathscr{X}} \cap \mathscr{B}_1$ is a closed bounded convex set. Therefore, since the objective function $f(\mathbf{x}) = \|\mathbf{x}\|$ is continuous in $\mathscr{R}^m$, it attains a minimizer $\mathbf{x}^*$ on $\mathscr{X}_1$, and this point solves (5.1). □

The last assertion ensures the existence of a point $\mathbf{x}^* \in \overline{\mathscr{X}}$ such that

$$\text{dist}(\mathscr{Y}, \mathscr{Z}) = \inf_{\mathbf{x} \in \mathscr{X}} \|\mathbf{x}\| = \|\mathbf{x}^*\|.$$

However, the existence of points $\mathbf{y}^* \in \overline{\mathscr{Y}}$ and $\mathbf{z}^* \in \overline{\mathscr{Z}}$ such that $\text{dist}(\mathscr{Y}, \mathscr{Z}) = \|\mathbf{y}^* - \mathbf{z}^*\|$ is not always guaranteed. Consider for example the case when $\mathscr{Y} = \{(x, y) \in \mathscr{R}^2 | y \geqslant \mathrm{e}^x + 1\}$ and $\mathscr{Z} = \{(x, y) \in \mathscr{R}^2 | y \leqslant -\mathrm{e}^x - 1\}$. Nevertheless, since $\mathbf{x}^* \in \overline{\mathscr{X}}$, we have the following corollary.

Corollary 7. *There exist two sequences*, $\mathbf{y}_k \in \mathscr{Y}$, $k = 1, 2, \ldots,$ *and* $\mathbf{z}_k \in \mathscr{Z}$, $k = 1, 2, \ldots,$ *such that*

$$\lim_{k \to \infty} (\mathbf{y}_k - \mathbf{z}_k) = \mathbf{x}^*$$

and

$$\lim_{k \to \infty} \|\mathbf{y}_k - \mathbf{z}_k\| = \|\mathbf{x}^*\| = \text{dist}(\mathscr{Y}, \mathscr{Z}).$$

Let us turn now to construct the dual problem of (5.1). For this purpose we introduce the support functions

$$\alpha(\mathbf{a}) = \sup_{\mathbf{y} \in \mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y} \quad \text{and} \quad \beta(\mathbf{a}) = \inf_{\mathbf{z} \in \mathscr{Z}} \mathbf{a}^{\mathrm{T}}\mathbf{z}, \tag{5.2}$$

which are well defined for any vector $\mathbf{a} \in \mathscr{R}^m$. If $\mathbf{a} \neq \mathbf{0}$ and $\alpha(\mathbf{a}) \leqslant \beta(\mathbf{a})$ then the parallel hyperplanes

$$\mathscr{H}_\alpha = \{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x} = \alpha(\mathbf{a})\} \quad \text{and} \quad \mathscr{H}_\beta = \{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x} = \beta(\mathbf{a})\}$$

separate $\mathscr{Y}$ and $\mathscr{Z}$. Also, as we have seen,

$$\mathrm{dist}(\mathscr{H}_\beta, \mathscr{H}_\alpha) = (\beta(\mathbf{a}) - \alpha(\mathbf{a}))/\|\mathbf{a}\|'. \tag{5.3}$$

The existence of separating hyperplanes is ensured by the next assertion.

Lemma 8. *Let $\mathscr{A}$ denote the set of all the points $\mathbf{a} \in \mathscr{R}^m$, $\mathbf{a} \neq \mathbf{0}$, for which the hyperplanes $\mathscr{H}_\alpha$ and $\mathscr{H}_\beta$ separate $\mathscr{Y}$ and $\mathscr{Z}$. That is,*

$$\mathscr{A} = \{\mathbf{a} \in \mathscr{R}^m|\mathbf{a} \neq \mathbf{0} \quad \textit{and} \quad \alpha(\mathbf{a}) \leqslant \beta(\mathbf{a})\}.$$

Then $\mathscr{A}$ is not empty.

Proof. The assumption $\mathscr{Y} \cap \mathscr{Z} = \phi$ implies that $\mathbf{0} \notin \mathscr{X}$. Hence there exists a vector $\mathbf{a} \in \mathscr{R}^m$, $\mathbf{a} \neq \mathbf{0}$, such that

$$\mathbf{a}^{\mathrm{T}}\mathbf{x} \leqslant 0 \quad \forall \mathbf{x} \in \mathscr{X}.$$

Consequently

$$\mathbf{a}^{\mathrm{T}}\mathbf{y} \leqslant \mathbf{a}^{\mathrm{T}}\mathbf{z} \quad \forall \mathbf{y} \in \mathscr{Y}, \ \ \mathbf{y} \in \mathscr{Z}$$

and

$$\sup_{\mathbf{y}\in\mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y} \leqslant \inf_{\mathbf{z}\in\mathscr{Z}} \mathbf{a}^{\mathrm{T}}\mathbf{z}. \qquad \square$$

The **maximal separation problem** is to find a vector $\mathbf{a} \in \mathscr{A}$ for which $\mathrm{dist}(\mathscr{H}_\alpha, \mathscr{H}_\beta)$ attains maximal value. The value of the distance function (5.3) is not effected by the size of $\mathbf{a}$. Hence there is no loss of generality in assuming that $\|\mathbf{a}\|' = 1$. In this case $\mathrm{dist}(\mathscr{H}_\alpha, \mathscr{H}_\beta)$ is given by the **separation function**

$$\sigma(\mathbf{a}) = \beta(\mathbf{a}) - \alpha(\mathbf{a}). \tag{5.4}$$

Observe that $\sigma(\mathbf{a})$ is a continuous function on $\mathscr{A}$. Consequently it attains a maximum on the compact set

$$\mathscr{A}' = \{\mathbf{a} \in \mathscr{A}|\|\mathbf{a}\|' \leqslant 1\}.$$

Furthermore, since $\sigma(\rho\mathbf{a}) = \rho\sigma(\mathbf{a})\ \forall \rho > 0$, any maximizer $\mathbf{a}^*$ must satisfy $\|\mathbf{a}^*\|' = 1$. The last equality means that $\mathbf{a}^*$ solves the maximal separation problem, which proves the following conclusions.

Lemma 9. *The* **maximal separation problem** *can be formulated as*

$$\begin{array}{ll} \text{maximize} & \sigma(\mathbf{a}) = \beta(\mathbf{a}) - \alpha(\mathbf{a}) \\ \text{subject to} & \mathbf{a} \in \mathscr{A}', \end{array} \tag{5.5}$$

and this problem is always solvable. Moreover, let $\mathbf{a}^*$ solve (5.5) then

$$\|\mathbf{a}^*\|' = 1 \quad \textit{and} \quad \sigma(\mathbf{a}^*) = \sup_{\mathbf{a}\in\mathscr{A}} (\beta(\mathbf{a}) - \alpha(\mathbf{a}))/\|\mathbf{a}\|'.$$

The next assertions are aimed to show that (5.1) and (5.5) are dual problems.

Lemma 10. *Let $\mathbf{x}^* \in \overline{\mathscr{X}}$ solve (5.1), then*

$$\sigma(\mathbf{a}) \leqslant \|\mathbf{x}^*\| \quad \forall \mathbf{a} \in \mathscr{A}'. \tag{5.6}$$

Proof. Let $\mathbf{a}$ be some vector in $\mathscr{A}'$ and let the hyperplanes $\mathscr{H}_\alpha$ and $\mathscr{H}_\beta$ be as above. Let the sequences $\{\mathbf{y}_k\}$ and $\{\mathbf{z}_k\}$ be as in Corollary 7. Then for $k = 1, 2, \ldots, \mathbf{y}_k$ belongs to the halfspace $\mathscr{H}_\alpha^- = \{\mathbf{x}|\mathbf{a}^\mathrm{T}\mathbf{x} \leqslant \alpha(a)\}$ while $\mathbf{z}_k$ belongs to the halfspace $\mathscr{H}_\beta^+ = \{\mathbf{x}|\mathbf{a}^\mathrm{T}\mathbf{x} \geqslant \beta(\mathbf{a})\}$. Consequently the line segment that connects $\mathbf{y}_k$ with $\mathbf{z}_k$ crosses the two hyperplanes. Let $\hat{\mathbf{y}}_k$ and $\hat{\mathbf{z}}_k$ denote the corresponding crossing points on $\mathscr{H}_\alpha$ and $\mathscr{H}_\beta$, respectively. Then, clearly,

$$\sigma(\mathbf{a}) = \mathrm{dist}(\mathscr{H}_\alpha, \mathscr{H}_\beta) \leqslant \|\hat{\mathbf{y}}_k - \hat{\mathbf{z}}_k\| \leqslant \|\mathbf{y}_k - \mathbf{z}_k\|.$$

Thus when passing to the limit we obtain that

$$\sigma(\mathbf{a}) \leqslant \|\mathbf{x}^*\|. \qquad \square$$

Lemma 11. *Let $\mathbf{x}^* \in \overline{\mathscr{X}}$ solve (5.1). Then there exists a vector $\mathbf{a}^* \in \mathscr{A}'$ such that*

$$\sigma(\mathbf{a}^*) = \|\mathbf{x}^*\|.$$

In other words, the maximal separation between $\mathscr{Y}$ and $\mathscr{Z}$ equals the shortest distance between these sets.

Proof. Recall that $0 \leqslant \sigma(\mathbf{a}) \leqslant \|\mathbf{x}^*\|\ \forall \mathbf{a} \in \mathscr{A}'$. Therefore $\|\mathbf{x}^*\| = 0$ implies $\sigma(\mathbf{a}) = 0\ \forall \mathbf{a} \in \mathscr{A}'$. Hence it is left to consider the case when $\mathbf{x}^* \neq \mathbf{0}$. Define

$$\mathscr{B} = \{\mathbf{x}|\|\mathbf{x}\| \leqslant \|\mathbf{x}^*\|\},$$

then $ri\mathscr{B} = \mathscr{B}^\circ$, $\mathscr{B}^\circ \cap \mathscr{X} = \emptyset$ but $\overline{\mathscr{B}} \cap \overline{\mathscr{X}} \neq \emptyset$ since $\mathbf{x}^* \in \overline{\mathscr{B}} \cap \overline{\mathscr{X}}$. Hence by Corollary 5 there exists a point $\mathbf{a} \in \mathscr{R}^m$ such that

$$\sup_{\mathbf{x}\in\mathscr{X}} \mathbf{a}^\mathrm{T}\mathbf{x} = \mathbf{a}^\mathrm{T}\mathbf{x}^* = \inf_{\mathbf{x}\in\mathscr{B}} \mathbf{a}^\mathrm{T}\mathbf{x}.$$

Define $\mathbf{a}^* = \mathbf{a}/\|\mathbf{a}\|'$. Then $\mathscr{H} = \{\mathbf{x}|(-\mathbf{a}^*)^\mathrm{T}\mathbf{x} = (-\mathbf{a}^*)^\mathrm{T}\mathbf{x}^*\}$ is a supporting hyperplane of $\mathscr{B}$ at the point $\mathbf{x}^*$. Therefore, since $\mathscr{B}$ is a "norm ball", $\mathbf{x}^*$ and $-\mathbf{a}^*$ are aligned. That is,

$$(-\mathbf{a}^*)^\mathrm{T}\mathbf{x}^* = \|-\mathbf{a}^*\|'\|\mathbf{x}^*\| = \|\mathbf{x}^*\|.$$

Consequently

$$\sup_{\mathbf{x}\in\mathscr{X}}(\mathbf{a}^*)^\mathrm{T}\mathbf{x} = -\|\mathbf{x}^*\|,$$

$$(\mathbf{a}^*)^\mathrm{T}\mathbf{y} - (\mathbf{a}^*)^\mathrm{T}\mathbf{z} \leqslant -\|\mathbf{x}^*\| \quad \forall \mathbf{y} \in \mathscr{Y},\ \ \mathbf{z} \in \mathscr{Z},$$

$$(\mathbf{a}^*)^\mathrm{T}\mathbf{z} \geqslant \|\mathbf{x}^*\| + (\mathbf{a}^*)^\mathrm{T}\mathbf{y} \quad \forall \mathbf{y} \in \mathscr{Y},\ \ \mathbf{z} \in \mathscr{Z},$$

$$\inf_{\mathbf{z}\in\mathscr{Z}}(\mathbf{a}^*)^\mathrm{T}\mathbf{z} \geqslant \|\mathbf{x}^*\| + \sup_{\mathbf{y}\in\mathscr{Y}}(\mathbf{a}^*)^\mathrm{T}\mathbf{y},$$

and

$$\sigma(\mathbf{a}^*) = \inf_{\mathbf{z}\in\mathscr{Z}}(\mathbf{a}^*)^\mathrm{T}\mathbf{z} - \sup_{\mathbf{y}\in\mathscr{Y}}(\mathbf{a}^*)^\mathrm{T}\mathbf{y} \geqslant \|\mathbf{x}^*\|.$$

On the other hand (5.6) implies $\sigma(\mathbf{a}^*) \leqslant \|\mathbf{x}^*\|$, so $\sigma(\mathbf{a}^*) = \|\mathbf{x}^*\|$. $\square$

Lemma 12. *Let $\mathbf{a}^* \in \mathscr{A}'$ be any point that solves (5.5) and let $\mathbf{x}^*$ be any point that solves (5.1). If $\mathbf{x}^* \neq \mathbf{0}$ then $\mathbf{x}^*$ and $-\mathbf{a}^*$ are aligned. That is,*

$$(-\mathbf{a}^*)^\mathrm{T}\mathbf{x}^* = \|-\mathbf{a}^*\|'\|\mathbf{x}^*\| = \|\mathbf{x}^*\|. \tag{5.7}$$

Proof. Let $\mathbf{a}$ be some vector in $\mathscr{A}'$. Then $\|\mathbf{a}\|' = 1$ and

$$\mathscr{H} = \{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x} = -\sigma(\mathbf{a})\}$$

is a supporting hyperplane of $\mathscr{X}$ which separates between $\mathscr{X}$ and the origin point $\mathbf{0}$. The distance between $\mathbf{0}$ and $\mathscr{H}$ is $\sigma(\mathbf{a})$. Hence the same hyperplane

$$\mathscr{H} = \{\mathbf{x}|(-\mathbf{a})^{\mathrm{T}}\mathbf{x} = \sigma(\mathbf{a})\}$$

supports the norm ball

$$\mathscr{B} = \{\mathbf{x}|\|\mathbf{x}\| \leqslant \sigma(\mathbf{a})\}.$$

Let us turn now to consider the special case when $\mathbf{a}^* \in \mathscr{A}'$ and $\sigma(\mathbf{a}^*) = \|\mathbf{x}^*\|$. In this case the hyperplane

$$\mathscr{H}^* = \{\mathbf{x}|(\mathbf{a}^*)^{\mathrm{T}}\mathbf{x} = -\|\mathbf{x}^*\|\}$$

supports $\mathscr{X}$, while the same hyperplane

$$\mathscr{H}^* = \{\mathbf{x}|(-\mathbf{a}^*)^{\mathrm{T}}\mathbf{x} = \|\mathbf{x}^*\|\}$$

supports the norm ball

$$\mathscr{B}^* = \{\mathbf{x}|\|\mathbf{x}\| \leqslant \|\mathbf{x}^*\|\}.$$

Therefore, since $\mathbf{x}^* \in \mathscr{B}^* \cap \overline{\mathscr{X}}$, $\mathbf{x}^* \in \mathscr{H}^*$ and (5.7) holds. □

In practice it is convenient to write the **maximal separation problem** in the form

$$\begin{aligned} &\text{maximize} \quad \sigma(\mathbf{a}) = \inf_{\mathbf{z}\in\mathscr{Z}} \mathbf{a}^{\mathrm{T}}\mathbf{z} - \sup_{\mathbf{y}\in\mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y} \\ &\text{subject to} \quad \|\mathbf{a}\|' \leqslant 1. \end{aligned} \tag{5.8}$$

The justification for replacing (5.5) with (5.8) lies in the definition of $\mathscr{A}$. Recall that $\mathbf{a} \in \mathscr{A}$ whenever $\sigma(\mathbf{a}) \geqslant 0$, and $\mathbf{a} \notin \mathscr{A}$ whenever $\sigma(\mathbf{a}) < 0$. Consequently any solution of (5.8) solves (5.5) and vice versa. The next theorem summarizes our results.

Theorem 13 (The new MND theorem). *The dual of the least norm problem* (5.1) *is the maximum separation problem* (5.8) *and both problems are solvable. Let* $\mathbf{x}^* \in \overline{\mathscr{X}}$ *solve* (5.1) *and let* $\mathbf{a}^*$ *solve* (5.8). *Then*

$$\|\mathbf{a}^*\|' = 1 \quad \textit{and} \quad \sigma(\mathbf{a}^*) = \|\mathbf{x}^*\| = \mathrm{dist}(\mathscr{Y}, \mathscr{Z}).$$

Furthermore, if $\mathbf{x}^* \neq \mathbf{0}$ *then* $\mathbf{x}^*$ *and* $-\mathbf{a}^*$ *are aligned. That is,*

$$(-\mathbf{a}^*)^{\mathrm{T}}\mathbf{x}^* = \|-\mathbf{a}^*\|'\|\mathbf{x}^*\| = \|\mathbf{x}^*\|.$$

The geometric interpretation of the new MND theorem is quite straightforward. On one hand

$$\|\mathbf{x}^*\| = \mathrm{dist}(\mathscr{Y}, \mathscr{Z}) = \inf\{\|\mathbf{y} - \mathbf{z}\||\mathbf{y} \in \mathscr{Y}, \mathbf{z} \in \mathscr{Z}\},$$

so $\|\mathbf{x}^*\|$ equals the shortest distance between $\mathscr{Y}$ and $\mathscr{X}$. On the other hand, for $\|\mathbf{a}\|' = 1$ the separation function $\sigma(\mathbf{a})$ measures the distance between two parallel hyperplanes that separate $\mathscr{Y}$ and $\mathscr{Z}$. The distance between the two hyperplanes is always smaller than $\mathrm{dist}(\mathscr{Y}, \mathscr{Z})$. Yet the maximum separation is attained when $-\mathbf{a}$ is aligned to $\mathbf{x}^*$, and in this case the maximal separation equals the shortest distance between $\mathscr{Y}$ and $\mathscr{Z}$ (see Figs. 2 and 7).

The power of the new MND theorem comes from the fact that $\mathscr{Y}$ and $\mathscr{Z}$ are arbitrary disjoint convex sets in $\mathscr{R}^m$, while $\|\cdot\|$ can be any norm in $\mathscr{R}^m$. Of course if $\|\cdot\|$ is not strictly convex then the solution of (5.1) is not necessarily unique. Similarly, if $\|\cdot\|$ is not smooth then the dual problem (5.8) may have more than one solution. Nevertheless, if $\|\cdot\|$ is smooth and strictly convex then both problems have unique solutions. In this case the alignment relation enables us to retrieve a primal solution from a dual one and vice versa. (See [18] for detailed discussion of this issue.)

Let us turn now to consider the special case when $\mathscr{Z}$ happens to be a singleton. That is, $\mathscr{Z}$ contains only one point, $\mathbf{z}$ say. In this case the least norm problem (5.1) is reduced to the form

$$\begin{array}{ll} \text{minimize} & \|\mathbf{y}-\mathbf{z}\| \\ \text{subject to} & \mathbf{y} \in \overline{\mathscr{Y}}, \end{array} \tag{5.9}$$

while (5.8) takes the form

$$\begin{array}{ll} \text{maximize} & \eta(\mathbf{a}) = \mathbf{a}^{\mathrm{T}}\mathbf{z} - \alpha(\mathbf{a}) \\ \text{subject to} & \|\mathbf{a}\|' \leqslant 1, \end{array} \tag{5.10}$$

where, as before, $\alpha(\mathbf{a}) = \sup_{\mathbf{y}\in\mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y}$ is the support function of $\mathscr{Y}$. Recall that when $\|\mathbf{a}\|' = 1$ the objective function $\eta(\mathbf{a}) = \mathbf{a}^{\mathrm{T}}\mathbf{z} - \alpha(\mathbf{a})$ measures the distance between $\mathbf{z}$ and the hyperplane $\{\mathbf{x}|\mathbf{a}^{\mathrm{T}}\mathbf{x} = \alpha(\mathbf{a})\}$, which supports $\mathscr{Y}$ and separates $\mathscr{Y}$ and $\mathbf{z}$. Also, by following the arguments of Lemma 6, we see that (5.9) is always solvable. These observations result in the following conclusions.

Theorem 14 (The Nirenberg–Luenberger MND theorem). *The dual of the least norm problem* (5.9) *is the maximum distance problem* (5.10) *and both problems are solvable. Let* $\mathbf{y}^* \in \overline{\mathscr{Y}}$ *solve* (5.9) *and let* $\mathbf{a}^* \in \mathscr{R}^m$ *solve* (5.10). *Then*

$$\|\mathbf{a}^*\|' = 1 \quad \textit{and} \quad \eta(\mathbf{a}^*) = \|\mathbf{y}^* - \mathbf{z}\| = \mathrm{dist}(\mathbf{z}, \mathscr{Y}).$$

Furthermore, if $\mathbf{z} \notin \overline{\mathscr{Y}}$ *then* $-\mathbf{a}^*$ *and* $\mathbf{y}^* - \mathbf{z}$ *are aligned. That is,*

$$(-\mathbf{a}^*)^{\mathrm{T}}(\mathbf{y}^* - \mathbf{z}) = \|\mathbf{y}^* - \mathbf{z}\|. \tag{5.11}$$

The geometric interpretation of the last theorem is rather clear: The shortest distance from $\mathbf{z}$ to $\mathscr{Y}$ is equal to the maximum of the distances from $\mathbf{z}$ to any hyperplane separating $\mathbf{z}$ and $\mathscr{Y}$; and any pair of optimal solutions satisfies the alignment relation (5.11), see Fig. 1. Note that in contrast to the general case, here there is always an optimal point $\mathbf{y}^* \in \overline{\mathscr{Y}}$ such that

$$\|\mathbf{y}^* - \mathbf{z}\| = \mathrm{dist}(\mathbf{z}, \mathscr{Y}).$$

Part II: Examples of applications

6. Preliminary remarks

Perhaps the best way to learn is by example, so the second part of the paper brings several examples of applications. It is shown that the MND principle is a useful tool for studying the duality features of several least distance problems. Let $\mathscr{Y}$ and $\mathscr{Z}$ be two distinct closed convex sets in $\mathscr{R}^m$. Then the primal problem to be solved has the form

$$\text{minimize} \quad \{\|\mathbf{y}-\mathbf{z}\| \,|\, \mathbf{y} \in \mathscr{Y},\ \mathbf{z} \in \mathscr{Z}\}, \tag{6.1}$$

where $\|\cdot\|$ denotes an arbitrary norm on $\mathscr{R}^m$. As before $\|\cdot\|'$ denotes the corresponding dual norm. The least distance problems consider various types of convex sets, such as norm-balls, linear varieties, polyhedral cones, polyhedrons, polytopes, and norm-ellipsoids. The study of these problems illustrates how the general definition of the dual problem,

$$\begin{aligned} &\text{maximize} \quad \sigma(\mathbf{a}) = \inf_{\mathbf{z}\in\mathscr{Z}} \mathbf{a}^{\mathrm{T}}\mathbf{z} - \sup_{\mathbf{y}\in\mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y} \\ &\text{subject to} \quad \|\mathbf{a}\|' \leqslant 1, \end{aligned} \tag{6.2}$$

is casted into a specific maximization problem when $\mathscr{Y}$ and $\mathscr{Z}$ are specified.

The examples start by considering two distinct sets, where neither $\mathscr{Y}$ nor $\mathscr{Z}$ is a singleton. Yet, often the original problem (6.1) can be reformulated as a best approximation problem that seeks the shortest distance between a certain point and a convex set. This gives the problem a second geometric interpretation, and enables us to apply the Nirenberg–Luenberger MND theorem. The transformation of (6.1) into a best approximation problem is not surprising, as the new MND theorem considers the distance between the origin point and the difference set $\mathscr{Y} - \mathscr{Z}$. However, the new MND theorem has its own merits. Consider for example the distance between two polytopes, or the distance between two norm-ellipsoids.

7. The distance between two linear varieties

We start by considering the distance between two linear varieties in standard form,

$$\mathscr{Y} = \{\tilde{\mathbf{y}} + B\mathbf{u} | \mathbf{u} \in \mathscr{R}^{\ell_1}\} \quad \text{and} \quad \mathscr{Z} = \{\tilde{\mathbf{z}} - C\mathbf{v} | \mathbf{v} \in \mathscr{R}^{\ell_2}\},$$

where $\tilde{\mathbf{y}}$ and $\tilde{\mathbf{z}}$ are given vectors in $\mathscr{R}^m$, B is a real $m \times \ell_1$ matrix, and C is a real $m \times \ell_2$ matrix. In this case the least distance problem (6.1) takes the form

$$\text{minimize} \quad f(\mathbf{u}, \mathbf{v}) = \|\tilde{\mathbf{y}} + B\mathbf{u} - \tilde{\mathbf{z}} + C\mathbf{v}\|,$$

or, simply,

$$\text{minimize} \quad \|A\mathbf{x} - \mathbf{b}\|, \tag{7.1}$$

where $\mathbf{b} = \tilde{\mathbf{z}} - \tilde{\mathbf{y}}$, $A = [B, \ \ C]$ is an $m \times n$ matrix, $n = \ell_1 + \ell_2$, and $\mathbf{x} = (\mathbf{u}^{\mathrm{T}}, \mathbf{v}^{\mathrm{T}})^{\mathrm{T}} \in \mathscr{R}^n$ denotes the vector of unknowns. The last problem has a new geometric interpretation: It seeks **the shortest distance between a point b and the subspace**

$$\mathscr{W} = \left\{A\mathbf{x} | \mathbf{x} \in \mathscr{R}^n\right\} = \mathrm{Range}(A).$$

Let $\mathscr{W}^{\perp}$ denote the orthogonal complement of $\mathscr{W}$. That is,

$$\mathscr{W}^{\perp} = \{\mathbf{w} \in \mathscr{R}^m | A^{\mathrm{T}}\mathbf{w} = \mathbf{0}\} = \mathrm{Null}(A^{\mathrm{T}}).$$

Then

$$\sup_{\mathbf{w}\in\mathscr{W}} \mathbf{a}^{\mathrm{T}}\mathbf{w} = \begin{cases} 0 & \text{when } \mathbf{a} \in \mathscr{W}^{\perp}, \\ \infty & \text{otherwise.} \end{cases}$$

Hence the dual of (7.1) has the form

$$\begin{aligned} &\text{maximize} \quad \mathbf{b}^{\mathrm{T}}\mathbf{a} \\ &\text{subject to} \quad A^{\mathrm{T}}\mathbf{a} = 0 \quad \text{and} \quad \|\mathbf{a}\|' \leqslant 1. \end{aligned} \tag{7.2}$$

Let $\mathbf{x}^* \in \mathscr{R}^n$ solve (7.1) and let $\mathbf{a}^* \in \mathscr{R}^m$ solve (7.2). Then

$$\|A\mathbf{x}^* - \mathbf{b}\| = \mathbf{b}^{\mathrm{T}}\mathbf{a}^*,$$

since the two problems share the same optimal value. The last equality implies the following conclusion: Either the system $A\mathbf{x} = \mathbf{b}$ is solvable, or the system

$$A^{\mathrm{T}}\mathbf{a} = \mathbf{0} \quad \text{and} \quad \mathbf{b}^{\mathrm{T}}\mathbf{a} > 0$$

is solvable, but never both. This conclusion is essentially Gale's theorem of the alternatives for linear equalities, e.g., [13,18,32].

Let us consider for a moment the important case when both (7.1) and (7.2) are defined by the Euclidean norm. In this case the residual vector

$$\mathbf{r}^* = \mathbf{b} - A\mathbf{x}^*$$

satisfies $A^{\mathrm{T}}\mathbf{r}^* = \mathbf{0}$ and $\mathbf{a}^* = \mathbf{r}^*/\|\mathbf{r}^*\|_2$. Moreover, the vectors $\mathbf{q}^* = A\mathbf{x}^*$ and $\mathbf{r}^*$ constitute the Euclidean projections of $\mathbf{b}$ on Range(A) and Null(A^{T}), respectively. Thus, as Fig. 4 shows, every vector $\mathbf{b} \in \mathscr{R}^m$ has a unique **orthogonal decomposition** of the form

$$\mathbf{b} = \mathbf{q}^* + \mathbf{r}^*, \quad \mathbf{q}^* \in \text{Range}(A), \quad \mathbf{r}^* \in \text{Null}(A^{\mathrm{T}}). \tag{7.3}$$

Another important application occurs when solving the ℓ_1 problem

$$\text{minimize}\|A\mathbf{x} - \mathbf{b}\|_1. \tag{7.4}$$

The need for solving such problems arises in statistical regression, data fitting, and function approximation, e.g. [3,21,41,51]. The appeal of the ℓ_1 norm comes from the observation that ℓ_1 solutions are less sensitive to possibly large errors in the data. See [14] or [16]. The dual of (7.4) is obtained by writing (7.2) with the ℓ_∞ norm. This results in the problem

$$\begin{aligned} &\text{maximize} \quad \mathbf{b}^{\mathrm{T}}\mathbf{a} \\ &\text{subject to} \quad A^{\mathrm{T}}\mathbf{a} = \mathbf{0} \quad \text{and} \quad -\mathbf{e} \leqslant \mathbf{a} \leqslant \mathbf{e}, \end{aligned} \tag{7.5}$$

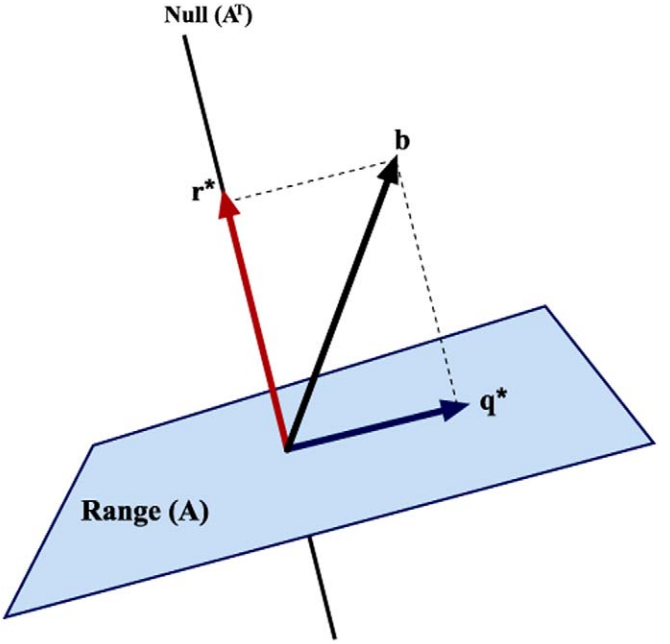

Fig. 4. The orthogonal decomposition.

where $\mathbf{e} = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathscr{R}^m$. The last problem is efficiently solved via the affine scaling method, e.g., [14,36].

The next example considers the ℓ_p problem

$$\text{maximize} \quad \|A\mathbf{x} - \mathbf{b}\|_p, \tag{7.6}$$

where $1 < p < \infty$. The dual of this problem has the form

$$\begin{aligned} &\text{maximize} && \mathbf{b}^{\mathrm{T}}\mathbf{a} \\ &\text{subject to} && A^{\mathrm{T}}\mathbf{a} = \mathbf{0} \quad \text{and} \quad \|\mathbf{a}\|_q \leqslant 1, \end{aligned} \tag{7.7}$$

where $q = p/(p-1)$. This gives rise to “dual methods” which are aimed at solving (7.7), e.g., [42,47,48]. It is also instructive to note that the dual of the problem

$$\text{minimize} \quad P(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|_p^p/p \tag{7.8}$$

has the form

$$\begin{aligned} &\text{maximize} && D(\mathbf{a}) = \mathbf{b}^{\mathrm{T}}\mathbf{a} - \|\mathbf{a}\|_q^q/q \\ &\text{subject to} && A^{\mathrm{T}}\mathbf{a} = \mathbf{0}, \end{aligned} \tag{7.9}$$

e.g., [12]. Thus, although (7.6) and (7.8) are essentially the same problem, the dual problems differ substantially.

Finally we consider the ℓ_∞ problem

$$\text{minimize} \quad \|A\mathbf{x} - \mathbf{b}\|_\infty, \tag{7.10}$$

whose dual has the form

$$\begin{aligned} &\text{maximize} && \mathbf{b}^{\mathrm{T}}\mathbf{a} \\ &\text{subject to} && A^{\mathrm{T}}\mathbf{a} = \mathbf{0} \quad \text{and} \quad \|\mathbf{a}\|_1 \leqslant 1. \end{aligned} \tag{7.11}$$

In this case both problems can be formulated and solved as linear programming problems, e.g., [41].

8. The distance between two shifted polyhedral convex cones

In this example both $\mathscr{Y}$ and $\mathscr{Z}$ are finitely generated cones of the form

$$\mathscr{Y} = \{\tilde{\mathbf{y}} + B\mathbf{u} | \mathbf{u} \in \mathscr{R}^{\ell_1} \text{ and } \mathbf{u} \geqslant \mathbf{0}\}$$

and

$$\mathscr{Z} = \{\tilde{\mathbf{z}} - C\mathbf{v} | \mathbf{v} \in \mathscr{R}^{\ell_2} \text{ and } \mathbf{v} \geqslant \mathbf{0}\},$$

where $\tilde{\mathbf{y}}, \tilde{\mathbf{z}} \in \mathscr{R}^m$, $B \in \mathscr{R}^{m \times \ell_1}$, and $C \in \mathscr{R}^{m \times \ell_2}$. Recall that a finitely generated convex cone in $\mathscr{R}^m$ is a closed convex set, e.g., [8] or [41]. To find the shortest distance between $\mathscr{Y}$ and $\mathscr{Z}$ one solves the problem

$$\begin{aligned} &\text{minimize} && f(\mathbf{u}, \mathbf{v}) = \|\tilde{\mathbf{y}} + B\mathbf{u} - (\tilde{\mathbf{z}} - C\mathbf{v})\| \\ &\text{subject to} && \mathbf{u} \geqslant \mathbf{0} \quad \text{and} \quad \mathbf{v} \geqslant \mathbf{0}, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\| \\ &\text{subject to} && \mathbf{x} \geqslant \mathbf{0}, \end{aligned} \tag{8.1}$$

where $\mathbf{b} = \tilde{\mathbf{z}} - \tilde{\mathbf{y}}$ and $A = [B, C]$ is a real $m \times n$ matrix, $n = \ell_1 + \ell_2$. The last problem has the following geometric interpretation: It seeks the shortest distance between $\mathbf{b}$ and the closed convex cone

$$\mathscr{K} = \{A\mathbf{x} | \mathbf{x} \in \mathscr{R}^n \text{ and } \mathbf{x} \geqslant \mathbf{0}\}.$$

Let

$$\mathscr{K}^* = \{\mathbf{a} | \mathbf{a}^{\mathrm{T}}\mathbf{k} \leqslant 0 \ \forall \mathbf{k} \in \mathscr{K}\} = \{\mathbf{a} | A^{\mathrm{T}}\mathbf{a} \leqslant \mathbf{0}\}$$

denote the **polar cone** of $\mathscr{K}$. Then, clearly,

$$\sup_{\mathbf{k}\in\mathscr{K}} \mathbf{a}^{\mathrm{T}}\mathbf{k} = \begin{cases} 0 & \text{when } \mathbf{a} \in \mathscr{K}^* \\ \infty & \text{otherwise.} \end{cases}$$

So the dual of (8.1) has the form

$$\begin{aligned} &\text{maximize} \quad \mathbf{b}^{\mathrm{T}}\mathbf{a} \\ &\text{subject to} \quad A^{\mathrm{T}}\mathbf{a} \leqslant \mathbf{0} \quad \text{and} \quad \|\mathbf{a}\|' \leqslant 1. \end{aligned} \tag{8.2}$$

Moreover, since $\mathscr{K}$ is a closed convex set in $\mathscr{R}^m$, there exists a point $\mathbf{x}^* \in \mathscr{R}^n$ that solves (8.1) and satisfies

$$\|A\mathbf{x}^* - \mathbf{b}\| = \mathbf{b}^{\mathrm{T}}\mathbf{a}^*, \tag{8.3}$$

where $\mathbf{a}^*$ solves (8.2). Let

$$\mathbf{r}^* = \mathbf{b} - A\mathbf{x}^*$$

denote the corresponding residual vector. Then, clearly, $\mathbf{r}^* \neq \mathbf{0}$ implies $\mathbf{b}^{\mathrm{T}}\mathbf{a}^* > 0$. This observation provides a simple proof of **Farkas' Lemma**: Either $\mathbf{b} \in \mathscr{K}$ or $\mathbf{b} \notin \mathscr{K}$ but never both. In the first case the system

$$A\mathbf{x} = \mathbf{b} \quad \text{and} \quad \mathbf{x} \geqslant \mathbf{0} \tag{8.4}$$

has a solution $\mathbf{x}^* \in \mathscr{R}^n$. In the second case, when $\mathbf{b} \notin \mathscr{K}$, the system

$$A^{\mathrm{T}}\mathbf{a} \leqslant \mathbf{0} \quad \text{and} \quad \mathbf{b}^{\mathrm{T}}\mathbf{a} > 0 \tag{8.5}$$

has a solution $\mathbf{a}^* \in \mathscr{R}^m$. Yet it is not possible that both systems are solvable.

A further consequence of the MND theorem is that $\mathbf{a}^*$ and $\mathbf{r}^*$ are aligned. This observation yields the relations

$$(\mathbf{a}^*)^{\mathrm{T}}(\mathbf{b} - A\mathbf{x}^*) = \|\mathbf{a}^*\|'\|\mathbf{b} - A\mathbf{x}^*\| = \|\mathbf{b} - A\mathbf{x}^*\| = \mathbf{b}^{\mathrm{T}}\mathbf{a}^*$$

and

$$(\mathbf{a}^*)^{\mathrm{T}}(A\mathbf{x}^*) = 0.$$

The last equality opens a simple way for establishing the polar decomposition which is associated with $\mathscr{K}$. For this purpose we consider the special case when both (8.1) and (8.2) are defined by the Euclidean norm. In this case $\mathbf{r}^*$ points at the same direction as $\mathbf{a}^*$, $(\mathbf{r}^*)^{\mathrm{T}}(A\mathbf{x}^*) = 0$, and $\mathbf{a}^* = \mathbf{r}^*/\|\mathbf{r}^*\|_2$. Consequently any vector $\mathbf{b} \in \mathscr{R}^m$ has a unique **polar decomposition** of the form

$$\mathbf{b} = \mathbf{k} + \mathbf{r}^*, \quad \mathbf{k} \in \mathscr{K}, \ \ \mathbf{r}^* \in \mathscr{K}^*, \quad \text{and} \quad (\mathbf{k})^{\mathrm{T}}\mathbf{r}^* = 0, \tag{8.6}$$

where $\mathbf{k} = A\mathbf{x}^*$ is the Euclidean projection of $\mathbf{b}$ on $\mathscr{K}$, and $\mathbf{r}^*$ is the Euclidean projection of $\mathbf{b}$ on $\mathscr{K}^*$ (see Fig. 5). For further discussions of polar cones and their properties see [24,29,38,39, 41,45].

9. The smallest deviation from consistency

In this example $\mathscr{Y}$ denotes the set of all the points $\mathbf{y} \in \mathscr{R}^m$ for which the linear system $B\mathbf{u} \geqslant \mathbf{y}$ is solvable. That is,

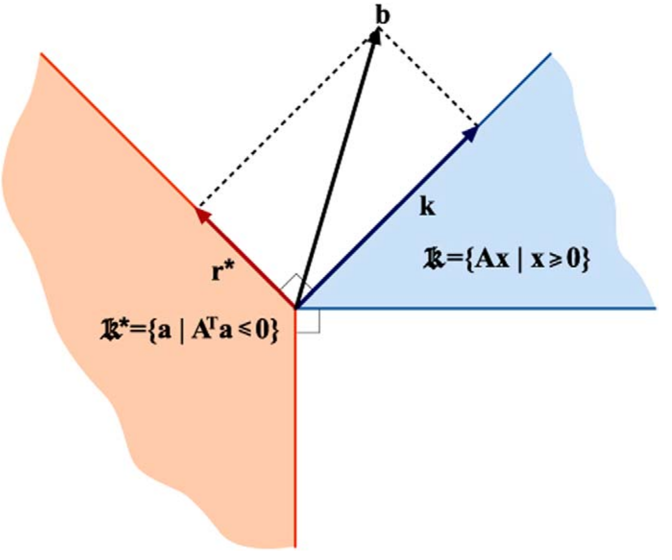

Fig. 5. The polar decomposition of Farkas' Lemma.

$$\mathscr{Y} = \left\{ B\mathbf{u} - \mathbf{p} | \mathbf{u} \in \mathscr{R}^{\ell_1}, \mathbf{p} \in \mathscr{R}^m, \text{ and } \mathbf{p} \geqslant \mathbf{0} \right\},$$

where B is a real $m \times \ell_1$ matrix. The other set is a linear variety in standard form,

$$\mathscr{Z} = \{\mathbf{b} - C\mathbf{v} | \mathbf{v} \in \mathscr{R}^{\ell_2}\},$$

where $\mathbf{b} \in \mathscr{R}^m$ and C is a real $m \times \ell_2$ matrix. With these definitions of $\mathscr{Y}$ and $\mathscr{Z}$ the least distance problem (6.1) takes the form

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{u}, \mathbf{v}, \mathbf{p}) = \|B\mathbf{u} - \mathbf{p} - (\mathbf{b} - C\mathbf{v})\| \\ &\text{subject to} \quad \mathbf{p} \geqslant \mathbf{0}. \end{aligned}$$

The last problem can be rewritten in the form

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{x}, \mathbf{p}) = \|A\mathbf{x} - \mathbf{p} - \mathbf{b}\| \\ &\text{subject to} \quad \mathbf{p} \geqslant \mathbf{0}, \end{aligned} \tag{9.1}$$

where $\mathbf{b} \in \mathscr{R}^m$, $A = [B, C]$ is a real $m \times n$ matrix, and $n = \ell_1 + \ell_2$. The unknowns here are $\mathbf{x} = (\mathbf{u}^{\mathrm{T}}, \mathbf{v}^{\mathrm{T}})^{\mathrm{T}} \in \mathscr{R}^n$ and $\mathbf{p} \in \mathscr{R}^m$. Observe that the set

$$\mathscr{F} = \{A\mathbf{x} - \mathbf{p} | \mathbf{x} \in \mathscr{R}^n, \mathbf{p} \in \mathscr{R}^m, \text{ and } \mathbf{p} \geqslant \mathbf{0}\}$$

consists of all the points $\mathbf{f} \in \mathscr{R}^m$ for which the linear system $A\mathbf{x} \geqslant \mathbf{f}$ is solvable. This observation gives (9.1) the following interpretation: Let $\mathbf{x}^*$ and $\mathbf{p}^*$ solve (9.1) and let

$$\mathbf{r}^* = \mathbf{b} - (A\mathbf{x}^* - \mathbf{p}^*)$$

denote the corresponding residual vector. If $\mathbf{r}^* = \mathbf{0}$ then $\mathbf{x}^*$ satisfies $A\mathbf{x}^* \geqslant \mathbf{b}$, which means that the system $A\mathbf{x} \geqslant \mathbf{b}$ is solvable. Otherwise, when $\mathbf{r}^* \neq \mathbf{0}$, the system $A\mathbf{x} \geqslant \mathbf{b}$ is inconsistent. Yet $\mathbf{x}^*$ satisfies $A\mathbf{x}^* \geqslant \mathbf{b} - \mathbf{r}^*$, which means that $\mathbf{r}^*$ provides the smallest correction vector. Note also that $\mathbf{r}^* \geqslant \mathbf{0}$, $\mathbf{p}^* \geqslant \mathbf{0}$ and $(\mathbf{r}^*)^{\mathrm{T}}\mathbf{p}^* = 0$.

Problem (9.1) can be viewed as an extension of the standard least norm problem (7.1) that handles inequalities instead of equalities: Solving (7.1) provides the smallest correction of $\mathbf{b}$ that makes the system $A\mathbf{x} = \mathbf{b}$ solvable. Similarly, solving (9.1) gives the smallest correction of $\mathbf{b}$ that makes the system $A\mathbf{x} \geqslant \mathbf{b}$ solvable.

An equivalent way to write $\mathscr{F}$ is

$$\mathscr{F} = \{H\mathbf{h} | \mathbf{h} \in \mathscr{R}^{m+2n} \text{ and } \mathbf{h} \geqslant \mathbf{0}\},$$

where $H = [A, -A, -I]$ is a real $m \times (m + 2n)$ matrix. This presentation implies that $\mathscr{F}$ is a finitely generated closed convex cone. Consequently

$$\sup_{\mathbf{f}\in\mathscr{F}} \mathbf{a}^{\mathrm{T}}\mathbf{f} = \begin{cases} 0 & \text{when } \mathbf{a} \in \mathscr{F}^* \\ \infty & \text{otherwise,} \end{cases}$$

where

$$\mathscr{F}^* = \{\mathbf{a} | \mathbf{a}^{\mathrm{T}}\mathbf{f} \leqslant 0 \ \forall \mathbf{f} \in \mathscr{F}\}$$

is the polar cone of $\mathscr{F}$. Moreover, as shown in [15],

$$\mathscr{F}^* = \{\mathbf{a} | A^{\mathrm{T}}\mathbf{a} = \mathbf{0} \text{ and } \mathbf{a} \geqslant \mathbf{0}\},$$

so the dual of (9.1) has the form

$$\begin{aligned} &\text{maximize} \quad \mathbf{b}^{\mathrm{T}}\mathbf{a} \\ &\text{subject to} \quad A^{\mathrm{T}}\mathbf{a} = \mathbf{0}, \quad \mathbf{a} \geqslant \mathbf{0}, \ \text{and } \|\mathbf{a}\|' \leqslant 1. \end{aligned} \tag{9.2}$$

Let $\mathbf{a}^*$ solve the dual. Here the primal–dual equality $\|\mathbf{r}^*\| = \mathbf{b}^{\mathrm{T}}\mathbf{a}^*$ provides a simple proof of **Gale's theorem of the alternative**: Either $\mathbf{r}^* = \mathbf{0}$, or $\mathbf{r}^* \neq \mathbf{0}$, but never both. In the first case the system $A\mathbf{x} \geqslant \mathbf{b}$ is solvable. In the second case the system

$$A^{\mathrm{T}}\mathbf{a} = \mathbf{0}, \quad \mathbf{a} \geqslant \mathbf{0}, \ \ \mathbf{b}^{\mathrm{T}}\mathbf{a} > 0,$$

is solvable. Yet it is not possible that both systems are solvable at the same time.

Combining the alignment relation

$$(\mathbf{a}^*)^{\mathrm{T}}(\mathbf{b} - A\mathbf{x}^* + \mathbf{p}^*) = \|\mathbf{a}^*\|'\|\mathbf{b} - A\mathbf{x}^* + \mathbf{p}^*\| = \|\mathbf{b} - A\mathbf{x}^* + \mathbf{p}^*\|$$

with the primal–dual equality gives

$$(\mathbf{a}^*)^{\mathrm{T}}(\mathbf{b} - A\mathbf{x}^* + \mathbf{p}^*) = \mathbf{b}^{\mathrm{T}}\mathbf{a}^*$$

and

$$(\mathbf{a}^*)^{\mathrm{T}}(A\mathbf{x}^* - \mathbf{p}^*) = 0.$$

These relations are sharpened when both (9.1) and (9.2) are defined by the Euclidean norm. In this case the alignment relation implies that $\mathbf{a}^* = \mathbf{r}^*/\|\mathbf{r}^*\|_2$. So $\mathbf{r}^*$ satisfies

$$A^{\mathrm{T}}\mathbf{r}^* = \mathbf{0}, \quad \mathbf{r}^* \geqslant \mathbf{0}, \ (\mathbf{r}^*)^{\mathrm{T}}(A\mathbf{x}^* - \mathbf{p}^*) = 0, \quad \text{and} \quad (\mathbf{r}^*)^{\mathrm{T}}\mathbf{p}^* = 0.$$

Consequently $\mathbf{b}$ has a unique "**polar decomposition**" of the form

$$\mathbf{b} = (A\mathbf{x}^* - \mathbf{p}^*) + \mathbf{r}^*, \quad A\mathbf{x}^* - \mathbf{p}^* \in \mathscr{F}, \ \ \mathbf{r}^* \in \mathscr{F}^*, \ \text{and } (\mathbf{r}^*)^{\mathrm{T}}(A\mathbf{x}^* - \mathbf{p}^*) = 0. \tag{9.3}$$

Moreover, as we have seen, $A\mathbf{x}^* - \mathbf{p}^*$ is the Euclidean projection of $\mathbf{b}$ on $\mathscr{F}$, while $\mathbf{r}^*$ is the Euclidean projection of $\mathbf{b}$ on $\mathscr{F}^*$ (see Fig. 6).

At this point it is instructive to compare the geometry of (7.1) with that of (9.1), when both problems are solved under the Euclidean norm. The traditional least squares approach has a simple geometric interpretation that is based on orthogonal decomposition of $\mathscr{R}^m$ into Range(A) and Null(A^{T}), see (7.3). Extending the least squares approach to handle an inconsistent system of linear inequalities, $A\mathbf{x} \geqslant \mathbf{b}$, changes the geometry of the problem: Now Range(A) is replaced by the closed convex cone $\mathscr{F}$, which consists of all the points $\mathbf{f} \in \mathscr{R}^m$ for which the system

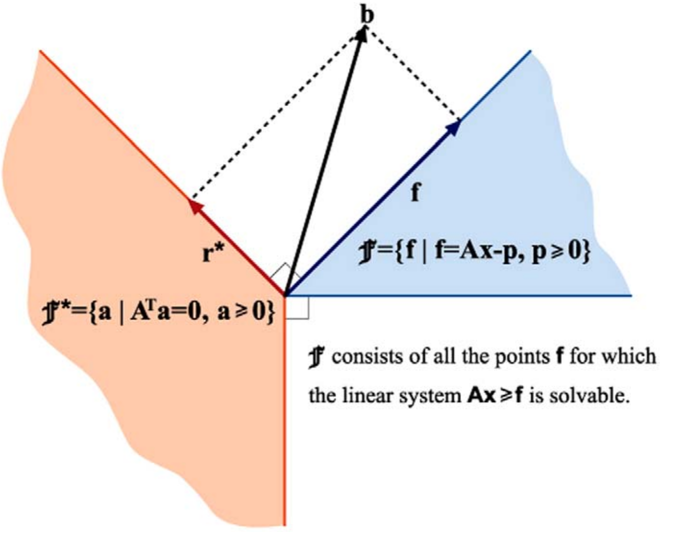

Fig. 6. The polar decomposition of the least deviation problem.

$A\mathbf{x} \geqslant \mathbf{f}$ is solvable. The subspace Null(A^{T}) is replaced by $\mathscr{F}^*$, the polar cone of $\mathscr{F}$, and the polar decomposition (9.3) replaces the orthogonal decomposition (7.3). In short, polarity replaces orthogonality. Compare Fig. 4 with Fig. 6.

Another important application occurs when solving the ℓ_1 problem

$$\text{minimize} \quad \|(\mathbf{b} - A\mathbf{x})_+\|_1. \tag{9.4}$$

Recall that $(\mathbf{b} - A\mathbf{x})_+$ is an m-vector whose ith component is $\max\{0, b_i - \mathbf{a}_i^{\mathrm{T}}\mathbf{x}\}$, where $\mathbf{a}_i^{\mathrm{T}}$ denotes the ith row of A. This problem is derived from (9.1) when using the ℓ_1 norm. The corresponding dual problem is obtained, therefore, by writing (9.2) with the ℓ_∞ norm. The resulting dual problem has the form

$$\begin{aligned} &\text{maximize} && \mathbf{b}^{\mathrm{T}}\mathbf{a} \\ &\text{subject to} && A^{\mathrm{T}}\mathbf{a} = \mathbf{0} \quad \text{and} \quad \mathbf{0} \leqslant \mathbf{a} \leqslant \mathbf{e}, \end{aligned} \tag{9.5}$$

where $\mathbf{e} = (1, 1, \cdots, 1)^{\mathrm{T}} \in \mathscr{R}^m$. The last problem is efficiently solved by the affine scaling method, see [16]. Note the similarity between (9.5) and (7.5).

As mentioned before, when using the ℓ_p norm, $1 < p < \infty$, there are two ways to pose the duality relations. The first one regards the problem

$$\text{minimize} \quad \|(\mathbf{b} - A\mathbf{x})_+\|_p, \tag{9.6}$$

whose dual has the form mm

$$\begin{aligned} &\text{maximize} && \mathbf{b}^{\mathrm{T}}\mathbf{a} \\ &\text{subject to} && A^{\mathrm{T}}\mathbf{a} = \mathbf{0}, \mathbf{a} \geqslant \mathbf{0}, \quad \text{and} \quad \|\mathbf{a}\|_q \leqslant 1, \end{aligned} \tag{9.7}$$

where $q = p/(p-1)$. The second option considers the problem

$$\text{minimize} \quad P(\mathbf{a}) = \|(\mathbf{b} - A\mathbf{x})_+\|_p^p/p, \tag{9.8}$$

whose dual has the form

$$\begin{aligned} &\text{maximize} \quad D(\mathbf{a}) = \mathbf{b}^T\mathbf{a} - \|\mathbf{a}\|_q^q/q \\ &\text{subject to} \quad A^T\mathbf{a} = \mathbf{0} \quad \text{and} \quad \mathbf{a} \geqslant \mathbf{0}, \end{aligned} \tag{9.9}$$

see [15]. Note the analogy between (9.6)–(9.9) and (7.6)–(7.9).

It is also interesting to consider the ℓ_∞ problem

$$\text{minimize} \quad \|(\mathbf{b} - A\mathbf{x})_+\|_\infty, \tag{9.10}$$

whose dual has the form

$$\begin{aligned} &\text{maximize} \quad \mathbf{b}^T\mathbf{a} \\ &\text{subject to} \quad A^T\mathbf{a} = \mathbf{0}, \quad \mathbf{a} \geqslant \mathbf{0}, \text{ and } \|\mathbf{a}\|_1 \leqslant 1. \end{aligned} \tag{9.11}$$

Recall that any solution of (9.11) must satisfy $\|\mathbf{a}\|_1 = 1$. Therefore, since $\mathbf{a} \geqslant \mathbf{0}$, the last constraint can be replaced with $\mathbf{e}^T\mathbf{a} = 1$. This enables us to write (9.11) in the form

$$\begin{aligned} &\text{maximize} \quad \mathbf{b}^T\mathbf{a} \\ &\text{subject to} \quad A^T\mathbf{a} = \mathbf{0}, \quad \mathbf{a} \geqslant \mathbf{0}, \text{ and } \mathbf{e}^T\mathbf{a} = 1. \end{aligned} \tag{9.12}$$

In other words, the dual of (9.10) is essentially a linear programming problem in standard form.

10. The nearest point in a polyhedron

The solution of (9.1) enables us to find a point in the polyhedron $\{\mathbf{x}|A\mathbf{x} \geqslant \mathbf{b}\}$ when this set is not empty. The problem considered in this example is to find a point in a polyhedron which is the closest to a given norm-ball. Let

$$\mathscr{Y} = \{\tilde{\mathbf{y}} + \mathbf{w} | \|\mathbf{w}\| \leqslant \rho\}$$

be a norm-ball of radius ρ in $\mathscr{R}^n$, centered at a point $\tilde{\mathbf{y}} \in \mathscr{R}^n$. Let

$$\mathscr{Z} = \{\mathbf{z}|A\mathbf{z} \geqslant \mathbf{b}\}$$

be a nonempty polyhedron in $\mathscr{R}^n$. As before, A is a real $m \times n$ matrix and $\mathbf{b} \in \mathscr{R}^m$. The distance between these sets is achieved by solving the least norm problem

$$\begin{aligned} &\text{minimize} \quad f(\mathbf{w}, \mathbf{z}) = \|\tilde{\mathbf{y}} + \mathbf{w} - \mathbf{z}\| \\ &\text{subject to} \quad A\mathbf{z} \geqslant \mathbf{b} \quad \text{and} \quad \|\mathbf{w}\| \leqslant \rho. \end{aligned} \tag{10.1}$$

The dual of (10.1) is derived by exploring the set

$$\mathscr{V} = \{\mathbf{v}|\mathbf{v} \in \mathscr{R}^m, A^T\mathbf{v} = \mathbf{a}, \mathbf{v} \geqslant \mathbf{0}\},$$

where $\mathbf{a}$ is a given point in $\mathscr{R}^m$. Assume first that this set is empty. Then, by Farkas' lemma, there exists a vector $\mathbf{u} \in \mathscr{R}^n$ such that $A\mathbf{u} \geqslant \mathbf{0}$ and $\mathbf{a}^T\mathbf{u} < 0$. Hence for any feasible point $\mathbf{z} \in \mathscr{Z}$ the ray $\{\mathbf{z} + \theta\mathbf{u}|\theta \geqslant 0\}$ is contained in $\mathscr{Z}$, which means that

$$\inf_{\mathbf{z}\in\mathscr{Z}} \mathbf{a}^T\mathbf{z} = -\infty.$$

Otherwise, when $\mathscr{V}$ is not empty,

$$\mathbf{a}^T\mathbf{z} = (A^T\mathbf{v})^T\mathbf{z} = (A\mathbf{z})^T\mathbf{v} \geqslant \mathbf{b}^T\mathbf{v} \quad \forall \mathbf{z} \in \mathscr{Z}, \mathbf{v} \in \mathscr{V}.$$

Hence the objective function of the **primal linear programming problem**

$$\begin{aligned} &\text{minimize} \quad \mathbf{a}^T\mathbf{z} \\ &\text{subject to} \quad A\mathbf{z} \geqslant \mathbf{b} \end{aligned} \tag{10.2}$$

is bounded from below, which means that this problem is solvable. (The existence of a point $\mathbf{z}^* \in \mathscr{Z}$ that solves (10.2) is easily proved by showing that the active set method terminates in a finite number of steps, e.g., [9].) A further use of Farkas' lemma shows that a point $\mathbf{z}^* \in \mathscr{Z}$ solves (10.2) if and only if there exists a vector $\mathbf{v}^* \in \mathscr{V}$ such that

$$(A\mathbf{z}^* - \mathbf{b})^{\mathrm{T}}\mathbf{v}^* = 0,$$

and

$$\mathbf{a}^{\mathrm{T}}\mathbf{z}^* = \mathbf{b}^{\mathrm{T}}\mathbf{v}^*.$$

(See Section 13.) In other words, $\mathbf{v}^*$ solves the **dual linear programming problem**

$$\begin{aligned} &\text{maximize} \quad \mathbf{b}^{\mathrm{T}}\mathbf{v} \\ &\text{subject to} \quad A^{\mathrm{T}}\mathbf{v} = \mathbf{a} \quad \text{and} \quad \mathbf{v} \geqslant \mathbf{0}, \end{aligned} \tag{10.3}$$

which is the dual problem of (10.2).

Summarizing the above two cases we conclude that

$$\inf_{\mathbf{z}\in\mathscr{Z}} \mathbf{a}^{\mathrm{T}}\mathbf{z} = \begin{cases} \sup\limits_{\mathbf{v}\in\mathscr{V}} \mathbf{b}^{\mathrm{T}}\mathbf{v} & \text{when } \mathscr{V} \neq \emptyset, \\ -\infty & \text{when } \mathscr{V} = \emptyset, \end{cases} \tag{10.4}$$

where $\emptyset$ denotes the empty set. The other part of the dual objective function satisfies

$$\sup_{\mathbf{y}\in\mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y} = \mathbf{a}^{\mathrm{T}}\tilde{\mathbf{y}} + \sup_{\|\mathbf{w}\|\leqslant\rho} \mathbf{a}^{\mathrm{T}}\mathbf{w} = \mathbf{a}^{\mathrm{T}}\tilde{\mathbf{y}} + \rho\|\mathbf{a}\|' = (A^{\mathrm{T}}\mathbf{v})^{\mathrm{T}}\tilde{\mathbf{y}} + \rho, \tag{10.5}$$

where the last equality relies on the relations $A^{\mathrm{T}}\mathbf{v} = \mathbf{a}$ and $\|\mathbf{a}\|' = 1$. Hence, by combining (10.4) with (10.5) we conclude that the dual of (10.1) has the form

$$\begin{aligned} &\text{maximize} \quad d(\mathbf{v}) = (\mathbf{b} - A\tilde{\mathbf{y}})^{\mathrm{T}}\mathbf{v} - \rho \\ &\text{subject to} \quad \|A^{\mathrm{T}}\mathbf{v}\|' \leqslant 1 \quad \text{and} \quad \mathbf{v} \geq \mathbf{0}. \end{aligned} \tag{10.6}$$

Since ρ is a nonnegative constant, the solution of (10.6) is not affected by the value of ρ as long as $\mathscr{Y}$ and $\mathscr{Z}$ are distinct sets. This observation indicates that (10.1) can be solved by considering the distance between $\tilde{\mathbf{y}}$ and $\mathscr{Z}$, which brings us to the problem of **finding the nearest point in a polyhedron**,

$$\begin{aligned} &\text{minimize} \quad \|\tilde{\mathbf{y}} - \mathbf{z}\| \\ &\text{subject to} \quad A\mathbf{z} \geqslant \mathbf{b}, \end{aligned} \tag{10.7}$$

whose dual has the form

$$\begin{aligned} &\text{maximize} \quad (\mathbf{b} - A\tilde{\mathbf{y}})^{\mathrm{T}}\mathbf{v} \\ &\text{subject to} \quad \|A^{\mathrm{T}}\mathbf{v}\|' \leqslant 1 \quad \text{and} \quad \mathbf{v} \geqslant \mathbf{0}. \end{aligned} \tag{10.8}$$

The least distance problems, (9.1) and (10.7) present different approaches for solving the basic **feasibility problem** of calculating a point $\mathbf{x} \in \mathscr{R}^n$ that satisfies $A\mathbf{x} \geqslant \mathbf{b}$. The solution of (9.1) guards against the possibility that the feasible region is empty. (In practice this often happens due to measurements errors in the data vector $\mathbf{b}$.) The other approach lacks this feature, but is capable of finding the point which is closest to $\tilde{\mathbf{y}}$. The need for solving feasibility problems arises in several applications. Perhaps the best known one is "Phase-1" of the Simplex method, e.g., [22]. Other important applications arise in medical image reconstruction from projections, and in inverse problems in radiation therapy, e.g., [5–7,23,28].

The duality relations (10.7)–(10.8) extend the results in [12], which concentrate on the ℓ_p norm, $1 < p < \infty$. It is also proved in [12] that the dual of the problem

$$\begin{array}{ll}\text{minimize} & F(\mathbf{z}) = \|\tilde{\mathbf{y}} - \mathbf{z}\|_p^p/p \\ \text{subject to} & A\mathbf{z} \geqslant \mathbf{b}\end{array} \tag{10.9}$$

has the form

$$\begin{array}{ll}\text{maximize} & D(\mathbf{v}) = (\mathbf{b} - A\tilde{\mathbf{y}})^{\mathrm{T}}\mathbf{v} - \|A^{\mathrm{T}}\mathbf{v}\|_q^q/q \\ \text{subject to} & \mathbf{v} \geqslant \mathbf{0},\end{array} \tag{10.10}$$

where $q = p/(p-1)$, and explicit rules are given for retrieving a primal solution from a dual one. A further simplification is gained by applying the Euclidean norm. In this case a primal solution is obtained from a dual one by the rule $\mathbf{z}^* = A^{\mathrm{T}}\mathbf{v}^*$. Maximizing the dual objective function by changing one variable at a time results in effective schemes for solving large sparse feasibility problems, e.g., [5,7,28].

11. The distance between two convex polytopes

In this example $\mathscr{Y}$ and $\mathscr{Z}$ are convex polytopes.

$$\mathscr{Y} = \{B\mathbf{w} | \mathbf{w} \in \mathscr{R}^{\ell},\ \mathbf{w} \geqslant \mathbf{0}, \mathbf{e}^{\mathrm{T}}\mathbf{w} = 1\}$$

and

$$\mathscr{Z} = \{C\mathbf{v} | \mathbf{v} \in \mathscr{R}^{n},\ \mathbf{v} \geqslant \mathbf{0},\quad \mathbf{e}^{\mathrm{T}}\mathbf{v} = 1\},$$

where $B \in \mathscr{R}^{m\times\ell}$, $C \in \mathscr{R}^{m\times n}$, and $\mathbf{e} = (1, 1, \ldots, 1)^{\mathrm{T}}$. The dimension of $\mathbf{0}$ and $\mathbf{e}$ depends on the context. The primal problem to be solved has, therefore, the form

$$\begin{array}{ll}\text{minimize} & f(\mathbf{w}, \mathbf{v}) = \|B\mathbf{w} - C\mathbf{v}\| \\ \text{subject to} & \mathbf{w} \geqslant \mathbf{0}, \quad \mathbf{v} \geqslant \mathbf{0}, \quad \mathbf{e}^{\mathrm{T}}\mathbf{w} = 1, \quad \mathbf{e}^{\mathrm{T}}\mathbf{v} = 1.\end{array} \tag{11.1}$$

Here it is easy to verify that

$$\sup_{\mathbf{y}\in\mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y} = \max\left\{\mathbf{b}_1^{\mathrm{T}}\mathbf{a}, \ldots, \mathbf{b}_\ell^{\mathrm{T}}\mathbf{a}\right\} \quad \text{and} \quad \inf_{\mathbf{z}\in\mathscr{Z}} \mathbf{a}^{\mathrm{T}}\mathbf{z} = \min\left\{\mathbf{c}_1^{\mathrm{T}}\mathbf{a}, \ldots, \mathbf{c}_n^{\mathrm{T}}\mathbf{a}\right\},$$

where $\mathbf{b}_1, \ldots, \mathbf{b}_\ell$, denote the columns of B, and $\mathbf{c}_1, \ldots, \mathbf{c}_n$, denote the columns of C. Hence the dual of (11.1) has the form

$$\begin{array}{ll}\text{maximize} & \sigma(\mathbf{a}) = \min\left\{\mathbf{c}_1^{\mathrm{T}}\mathbf{a}, \ldots, \mathbf{c}_n^{\mathrm{T}}\mathbf{a}\right\} - \max\{\mathbf{b}_1^{\mathrm{T}}\mathbf{a}, \ldots, \mathbf{b}_\ell^{\mathrm{T}}\mathbf{a}\} \\ \text{subject to} & \|\mathbf{a}\|' \leqslant 1.\end{array} \tag{11.2}$$

The need for calculating a hyperplane that separates between two given polytopes is a central problem in the fields of patterns recognition and machine learning, e.g., [2,31,33–35]. The fact that the maximal separation equals the smallest distance adds new insight into this problem. Note also that the solution of (11.2) remains meaningful even when the two polytopes are not disjoint.

Let us turn now to consider the case when $\mathscr{Y}$ happens to be a singleton, $\{\mathbf{b}\}$ say. In this case (11.1) is reduced to the least norm problem

$$\begin{array}{ll}\text{minimize} & \|A\mathbf{x}\| \\ \text{subject to} & \mathbf{e}^{\mathrm{T}}\mathbf{x} = 1 \quad \text{and} \quad \mathbf{x} \geqslant \mathbf{0},\end{array} \tag{11.3}$$

where A is a real $m \times n$ matrix whose columns are $\mathbf{a}_j = \mathbf{c}_j - \mathbf{b}$, $j = 1, \ldots, n$. Let $\mathbf{x}^* \in \mathscr{R}^n$ solve the last problem. Then, clearly, $C\mathbf{x}^*$ is a point of $\mathscr{Z}$ which is nearest to $\mathbf{b}$. Similarly, $\mathbf{r}^* = A\mathbf{x}^*$ is the smallest point of the polytope

$$\mathscr{P} = \{A\mathbf{x} | \mathbf{x} \in \mathscr{R}^n, \mathbf{x} \geqslant \mathbf{0}, \mathbf{e}^{\mathrm{T}}\mathbf{x} = 1\}.$$

Methods for calculating **the nearest (smallest) point in a polytope** were proposed by a number of authors, e.g., [10,37,52,53]. For further discussions of related issues see [4,30,46,50]. Adapting (11.2) to fit the new problem shows that the dual of (11.3) has the form

$$\begin{aligned}&\text{maximize} && \sigma(\mathbf{u}) = \min\left\{\mathbf{a}_1^{\mathrm{T}}\mathbf{u}, \ldots, \mathbf{a}_n^{\mathrm{T}}\mathbf{u}\right\}\\ &\text{subject to} && \|\mathbf{u}\|' \leqslant 1.\end{aligned} \tag{11.4}$$

Let $\mathbf{u}^*$ solve (11.4), then the primal–dual equality

$$\|\mathbf{r}^*\| = \min\left\{\mathbf{a}_1^{\mathrm{T}}\mathbf{u}^*, \ldots, \mathbf{a}_n^{\mathrm{T}}\mathbf{u}^*\right\}$$

provides a simple proof of **Gordan's theorem of the alternatives**: Either $\mathbf{r}^* = \mathbf{0}$, or $\mathbf{r}^* \neq \mathbf{0}$, but never both. In the first case the system

$$A\mathbf{x} = \mathbf{0}, \quad \mathbf{x} \geqslant \mathbf{0}, \quad \mathbf{e}^{\mathrm{T}}\mathbf{x} = 1, \tag{11.5}$$

has a solution $\mathbf{x}^* \in \mathscr{R}^n$. In the second case the system

$$\mathbf{a}_j^{\mathrm{T}}\mathbf{u} > 0, \quad j = 1, \ldots, n, \tag{11.6}$$

has a solution $\mathbf{u}^* \in \mathscr{R}^m$. Yet it is not possible that both systems are solvable.

The question which of the two systems is solvable can be answered in the following way. Let $\tilde{A}$ denote the $(m+1) \times n$ matrix whose first row is $\mathbf{e}^{\mathrm{T}}$ and the other rows are those of A. Let $\tilde{\mathbf{e}}_1$ denote the first column of the $(m+1) \times (m+1)$ identity matrix. Let $\tilde{\mathbf{x}} \in \mathscr{R}^n$ solve the least squares problem

$$\begin{aligned}&\text{minimize} && \|\tilde{A}\mathbf{x} - \tilde{\mathbf{e}}_1\|_2^2\\ &\text{subject to} && \mathbf{x} \geqslant \mathbf{0},\end{aligned} \tag{11.7}$$

and let $\tilde{\mathbf{r}} = A\tilde{\mathbf{x}}$ denote the resulting residual vector. Then, as shown in [10],

$$0 < \mathbf{e}^{\mathrm{T}}\tilde{\mathbf{x}} \leqslant 1 \tag{11.8}$$

and the vector $\mathbf{x}^* = \tilde{\mathbf{x}}/\mathbf{e}^{\mathrm{T}}\tilde{\mathbf{x}}$ solves the least distance problem

$$\begin{aligned}&\text{minimize} && \|A\mathbf{x}\|_2\\ &\text{subject to} && \mathbf{e}^{\mathrm{T}}\mathbf{x} = 1 \quad \text{and} \quad \mathbf{x} \geqslant \mathbf{0},\end{aligned} \tag{11.9}$$

whose dual has the form

$$\begin{aligned}&\text{maximize} && \sigma(\mathbf{u}) = \min\left\{\mathbf{a}_1^{\mathrm{T}}\mathbf{u}, \ldots, \mathbf{a}_n^{\mathrm{T}}\mathbf{u}\right\}\\ &\text{subject to} && \|\mathbf{u}\|_2 \leqslant 1.\end{aligned} \tag{11.10}$$

Moreover, if $\tilde{\mathbf{r}} = \mathbf{0}$ then $\mathbf{x}^*$ solves (11.5). Otherwise, when $\tilde{\mathbf{r}} \neq \mathbf{0}$, the vector $\mathbf{u}^* = \tilde{\mathbf{r}}/\|\tilde{\mathbf{r}}\|_2$ solves (11.10) and satisfies (11.6).

12. The distance between two norm-ellipsoids

The geometry of this problem is illustrated in Figs. 2 and 7. Here

$$\mathscr{Y} = \{\tilde{\mathbf{y}} + B\mathbf{w} \mid \|\mathbf{w}\|_* \leqslant 1\} \quad \text{and} \quad \mathscr{Z} = \{\tilde{\mathbf{z}} - C\mathbf{v} \mid \|\mathbf{v}\|_\dagger \leqslant 1\},$$

where $\tilde{\mathbf{y}} \in \mathscr{R}^m$, $B \in \mathscr{R}^{m\times n}$, $\|\cdot\|_*$ denotes an arbitrary norm on $\mathscr{R}^n$, $\tilde{\mathbf{z}} \in \mathscr{R}^m$, $C \in \mathscr{R}^{m\times \ell}$, and $\|\cdot\|_\dagger$ denotes an arbitrary norm on $\mathscr{R}^\ell$. The dual norms of $\|\cdot\|_*$ and $\|\cdot\|_\dagger$ are denoted by $\|\cdot\|_*'$ and $\|\cdot\|_\dagger'$, respectively. The usual definition of an ellipsoid in $\mathscr{R}^m$ refers to the case when $\|\cdot\|_*$ and$\|\cdot\|_\dagger$ denote the Euclidean norm, e.g., [1]. For this reason $\mathscr{Y}$ and $\mathscr{Z}$ are called "norm-ellipsoids". The distance between these sets is attained by solving the least norm problem

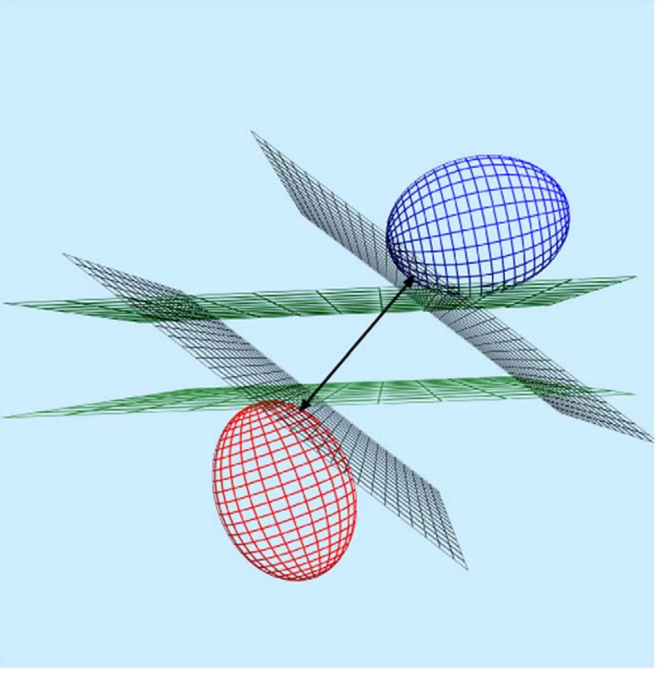

Fig. 7. The distance between two convex sets.

$$\begin{aligned} &\text{minimize} \quad \|\tilde{\mathbf{y}} - \tilde{\mathbf{z}} + B\mathbf{w} + C\mathbf{v}\| \\ &\text{subject to} \quad \|\mathbf{w}\|_* \leqslant 1 \quad \text{and} \quad \|\mathbf{v}\|_\dagger \leqslant 1. \end{aligned} \tag{12.1}$$

As before, $\|\cdot\|$ denotes an arbitrary norm on $\mathscr{R}^m$, while $\|\cdot\|'$ denotes the corresponding dual norm.

Here one can verify that

$$\inf_{\mathbf{z}\in\mathscr{Z}} \mathbf{a}^{\mathrm{T}}\mathbf{z} = \mathbf{a}^{\mathrm{T}}\tilde{\mathbf{z}} + \inf_{\|\mathbf{v}\|_\dagger \leqslant 1} \mathbf{a}^{\mathrm{T}}(-C\mathbf{v}) = \mathbf{a}^{\mathrm{T}}\tilde{\mathbf{z}} - \|C^{\mathrm{T}}\mathbf{a}\|'_\dagger$$

and

$$\sup_{\mathbf{y}\in\mathscr{Y}} \mathbf{a}^{\mathrm{T}}\mathbf{y} = \mathbf{a}^{\mathrm{T}}\tilde{\mathbf{y}} + \sup_{\|\mathbf{w}\|_* \leqslant 1} \mathbf{a}^{\mathrm{T}}B\mathbf{w} = \mathbf{a}^{\mathrm{T}}\tilde{\mathbf{y}} + \|B^{\mathrm{T}}\mathbf{a}\|'_*.$$

So the dual of (12.1) has the form

$$\begin{aligned} &\text{maximize} \quad \sigma(\mathbf{a}) = -(\tilde{\mathbf{y}} - \tilde{\mathbf{z}})^{\mathrm{T}}\mathbf{a} - \|B^{\mathrm{T}}\mathbf{a}\|'_* - \|C^{\mathrm{T}}\mathbf{a}\|'_\dagger \\ &\text{subject to} \quad \|\mathbf{a}\|' \leqslant 1. \end{aligned} \tag{12.2}$$

Let us turn now to consider the important case when $\mathscr{Z}$ happens to be a singleton, $\{\tilde{\mathbf{z}}\}$ say. In this case (12.1) is reduced to a **trust region** approximation problem

$$\begin{aligned} &\text{minimize} \quad \|B\mathbf{w} - \mathbf{g}\| \\ &\text{subject to} \quad \|\mathbf{w}\|_* \leqslant 1, \end{aligned} \tag{12.3}$$

where $\mathbf{g} = \tilde{\mathbf{y}} - \tilde{\mathbf{z}}$, while the dual of (12.3) has the form

$$\begin{aligned} &\text{maximize} \quad \sigma(\mathbf{a}) = -\mathbf{g}^{\mathrm{T}}\mathbf{a} - \|B^{\mathrm{T}}\mathbf{a}\|'_* \\ &\text{subject to} \quad \|\mathbf{a}\|' \leqslant 1. \end{aligned} \tag{12.4}$$

The trust region problem (12.3) has an independent geometric interpretation: Let $\mathbf{w}^*$ solve this problem then $B\mathbf{w}^*$ is a point of the norm-ellipsoid

$$\mathscr{E} = \{B\mathbf{w} \mid \|\mathbf{w}\|_* \leqslant 1\},$$

which is closest to $\mathbf{g}$. Let $\mathbf{r}^* = B\mathbf{w}^* - \mathbf{g}$ denote the corresponding residual vector. Then $\mathbf{r}^* = \mathbf{0}$ if and only if $\mathbf{g} \in \mathscr{E}$. Moreover, let $\mathbf{a}^*$ solve (12.4) then

$$\mathbf{g}^{\mathrm{T}}\mathbf{a}^* + \|B^{\mathrm{T}}\mathbf{a}^*\|_*' = -\|\mathbf{r}^*\| \tag{12.5}$$

and the vectors $\mathbf{a}^*$ and $\mathbf{r}^*$ are aligned. That is,

$$(\mathbf{a}^*)^{\mathrm{T}}\mathbf{r}^* = \|\mathbf{a}^*\|'\|\mathbf{r}^*\| = \|\mathbf{r}^*\|. \tag{12.6}$$

The primal–dual equality (12.5) provides a simple proof of **Dax' theorem of the alternatives**: Either the inequality

$$\mathbf{g}^{\mathrm{T}}\mathbf{a} + \|B^{\mathrm{T}}\mathbf{a}\|_*' < 0 \tag{12.7}$$

has a solution $\mathbf{a}^* \in \mathscr{R}^m$, or the system

$$B\mathbf{w} = \mathbf{g} \quad \text{and} \quad \|\mathbf{w}\|_* \leqslant 1, \tag{12.8}$$

has a solution $\mathbf{w}^* \in \mathscr{R}^n$, but never both, e.g., [11,13,18]. The last observation gives constructive optimality conditions for the ℓ_1 problem (7.4). See the next section.

13. Constructive optimality conditions

We have seen that the MND theorem connects theorems of the alternatives to least norm problems. In this section we illustrate how these relations are used to derive constructive optimality conditions and steepest descent directions. In all the coming examples A is a real $m \times n$ matrix, $\mathbf{b} = (b_1, \ldots, b_m)^{\mathrm{T}} \in \mathscr{R}^m$, and $\mathbf{x} \in \mathscr{R}^n$ denotes the vectors of unknowns. The rows of A are denoted as $\mathbf{a}_i^{\mathrm{T}}$, $i = 1, \ldots, m$.

The first example considers the linear programming problem

$$\begin{aligned} &\text{minimize} && \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ &\text{subject to} && A\mathbf{x} \geqslant \mathbf{b}, \end{aligned} \tag{13.1}$$

where $\mathbf{c} \in \mathscr{R}^n$. Let $\hat{\mathbf{x}} \in \mathscr{R}^n$ be a given feasible point. That is, $A\hat{\mathbf{x}} \geqslant \mathbf{b}$. Then there is no loss of generality in assuming that

$$\mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}} = b_i \quad \text{for } i = 1, \ldots, \ell,$$

and

$$\mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}} > b_i \quad \text{for } i = \ell + 1, \ldots, m.$$

Let $\hat{A}$ denote the $\ell \times n$ matrix whose rows are $\mathbf{a}_i^{\mathrm{T}}$, $i = 1, \ldots, \ell$. Then, clearly, a vector $\mathbf{u} \in \mathscr{R}^n$ is a feasible descent direction at $\hat{\mathbf{x}}$ if and only if it satisfies

$$\hat{A}\mathbf{u} \geqslant \mathbf{0} \quad \text{and} \quad \mathbf{c}^{\mathrm{T}}\mathbf{u} < 0. \tag{13.2}$$

It follows, therefore, that $\hat{\mathbf{x}}$ solves (13.1) if and only if (13.2) has no solution. This situation is characterized by **Farkas' lemma**, which says that either the system (13.2) is solvable, or the system

$$\hat{A}^{\mathrm{T}}\mathbf{y} = \mathbf{c} \quad \text{and} \quad \mathbf{y} \geqslant \mathbf{0} \tag{13.3}$$

is solvable, but never both. Moreover, let $\hat{\mathbf{y}} \in \mathscr{R}^\ell$ solve (13.3). Then the m-vector $\hat{\mathbf{v}} = \binom{\hat{\mathbf{y}}}{\mathbf{0}}$ satisfies $(A\hat{\mathbf{x}} - \mathbf{b})^{\mathrm{T}}\hat{\mathbf{v}} = 0$ and $\mathbf{c}^{\mathrm{T}}\hat{\mathbf{x}} = \mathbf{b}^{\mathrm{T}}\hat{\mathbf{v}}$, which means that $\hat{\mathbf{v}}$ solves the corresponding dual problem

$$\begin{aligned} &\text{maximize} \quad \mathbf{b}^{\mathrm{T}}\mathbf{v} \\ &\text{subject to} \quad A^{\mathrm{T}}\mathbf{v} = \mathbf{c} \quad \text{and} \quad \mathbf{v} \geqslant \mathbf{0}. \end{aligned} \tag{13.4}$$

As shown in Section 8, the question of whether (13.3) is solvable is answered by solving the least squares problem

$$\begin{aligned} &\text{minimize} \quad \|\hat{A}^{\mathrm{T}}\mathbf{y} - \mathbf{c}\|_2^2 \\ &\text{subject to} \quad \mathbf{y} \geqslant \mathbf{0}. \end{aligned} \tag{13.5}$$

Let $\hat{\mathbf{y}}$ solve (13.5) and let $\hat{\mathbf{r}} = \hat{A}^{\mathrm{T}}\hat{\mathbf{y}} - \mathbf{c}$ denote the corresponding residual vector. If $\hat{\mathbf{r}} = \mathbf{0}$ then, clearly, $\hat{\mathbf{y}}$ solves (13.3) and $\hat{\mathbf{x}}$ solves (13.1). Otherwise, when $\hat{\mathbf{r}} \neq \mathbf{0}$, the unit vector $\hat{\mathbf{u}} = \hat{\mathbf{r}}/\|\hat{\mathbf{r}}\|_2$ solves the **steepest descent problem**

$$\begin{aligned} &\text{minimize} \quad \mathbf{c}^{\mathrm{T}}\mathbf{u} \\ &\text{subject to} \quad \hat{A}\mathbf{u} \geqslant \mathbf{0} \quad \text{and} \quad \|\mathbf{u}\|_2 \leqslant 1, \end{aligned} \tag{13.6}$$

and satisfies $\mathbf{c}^{\mathrm{T}}\hat{\mathbf{u}} = -\|\hat{\mathbf{r}}\|_2$. In other words, $\hat{\mathbf{r}}$ points at the steepest descent direction!

The second example considers the ℓ_1 problem

$$\text{minimize} \quad f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|_1. \tag{13.7}$$

Let $\hat{\mathbf{x}}$ be a given point in $\mathscr{R}^n$. Here it is assumed for simplicity that

$$\mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}} = b_i \quad \text{for } i = 1, \ldots, \ell, \quad \text{and} \quad \mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}} \neq b_i \quad \text{for } i = \ell + 1, \ldots, m.$$

Let $\hat{A}$ denote the $\ell \times n$ matrix whose rows are $\mathbf{a}_i^{\mathrm{T}}$, $i = 1, \ldots, \ell$. The other rows of A are used to define the n-vector

$$\mathbf{g} = \sum_{i=\ell+1}^{m} \mathbf{a}_i \operatorname{sign}(\mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}} - b_i).$$

With these notations at hand one can verify that for any vector $\mathbf{u} \in \mathscr{R}^n$ there exists a positive constant, α say, such that

$$f(\hat{\mathbf{x}} + \theta\mathbf{u}) = f(\hat{\mathbf{x}}) + \theta\mathbf{g}^{\mathrm{T}}\mathbf{u} + \theta\|\hat{A}\mathbf{u}\|_1 \quad \forall 0 \leqslant \theta \leqslant \alpha.$$

It follows, therefore, that $\hat{\mathbf{x}}$ solves (13.7) if and only if there is no vector $\mathbf{u} \in \mathscr{R}^n$ that satisfies

$$\mathbf{g}^{\mathrm{T}}\mathbf{u} + \|\hat{A}\mathbf{u}\|_1 < 0. \tag{13.8}$$

The existence of a vector $\mathbf{u} \in \mathscr{R}^n$ that solves the last inequality is characterized by **Dax' theorem of the alternatives**, which says that either (13.8) has a solution, or the system

$$\hat{A}^{\mathrm{T}}\mathbf{y} = \mathbf{g} \quad \text{and} \quad \|\mathbf{y}\|_\infty \leqslant 1 \tag{13.9}$$

has a solution $\hat{\mathbf{y}} \in \mathscr{R}^\ell$, but never both.

The question of whether (13.9) is solvable is answered by solving the least squares problem

$$\begin{aligned} &\text{minimize} \quad \|\hat{A}^{\mathrm{T}}\mathbf{y} - \mathbf{g}\|_2^2 \\ &\text{subject to} \quad -\mathbf{e} \leqslant \mathbf{y} \leqslant \mathbf{e}, \end{aligned} \tag{13.10}$$

where $\mathbf{e} = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathscr{R}^\ell$. Let $\hat{\mathbf{y}}$ solve (13.10) and let $\hat{\mathbf{r}} = \hat{A}\hat{\mathbf{y}} - \mathbf{g}$ denote the resulting residual vector. If $\hat{\mathbf{r}} = \mathbf{0}$ then, clearly, $\hat{\mathbf{y}}$ solves (13.9), and $\hat{\mathbf{x}}$ solves (13.7). Otherwise, when $\hat{\mathbf{r}} \neq \mathbf{0}$, the unit vector $\hat{\mathbf{u}} = \hat{\mathbf{r}}/\|\hat{\mathbf{r}}\|_2$ satisfies

$$\mathbf{g}^{\mathrm{T}}\hat{\mathbf{u}} + \|\hat{A}\hat{\mathbf{u}}\|_1 = -\|\hat{\mathbf{r}}\|_2 < 0,$$

and solves the steepest descent problem

$$\begin{array}{ll} \text{minimize} & \mathbf{g}^{\mathrm{T}}\mathbf{u} + \|\hat{A}\mathbf{u}\|_1 \\ \text{subject to} & \|\mathbf{u}\|_2 \leqslant 1. \end{array} \tag{13.11}$$

That is, $\hat{\mathbf{u}}$ is the steepest descent direction at $\hat{\mathbf{x}}$.

The third example considers the ℓ_∞ problem

$$\text{minimize} \quad f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|_\infty. \tag{13.12}$$

Let $\hat{\mathbf{x}}$ be a given point in $\mathscr{R}^n$. If $\|A\hat{\mathbf{x}} - \mathbf{b}\|_\infty = 0$ then, clearly, $\hat{\mathbf{x}}$ solves (13.12). Hence there is no loss of generality in assuming that $\hat{\mathbf{x}}$ satisfies $\|A\hat{x} - \mathbf{b}\|_\infty > 0$,

$$\begin{array}{ll} |\mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}} - b_i| = \|A\mathbf{x} - \mathbf{b}\|_\infty & \text{for } i = 1, \ldots, \ell, \quad \text{and} \\ |\mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}} - b_i| < \|A\hat{\mathbf{x}} - \mathbf{b}\|_\infty & \text{for } i = \ell + 1, \ldots, m. \end{array}$$

Define $\hat{\mathbf{a}}_i = \mathbf{a}_i \operatorname{sign}(\mathbf{a}_i^{\mathrm{T}}\hat{\mathbf{x}} - b_i)$, $i = 1, \ldots, \ell$, and let $\hat{A}$ denote the $\ell \times n$ matrix whose rows are $\hat{\mathbf{a}}_i^{\mathrm{T}}$, $i = 1, \ldots, \ell$. Then for any vector $\mathbf{u} \in \mathscr{R}^n$ there exists a positive constant, α say, such that

$$f(\hat{\mathbf{x}} + \theta\mathbf{u}) = f(\hat{\mathbf{x}}) + \theta \max\{\hat{\mathbf{a}}_1^{\mathrm{T}}\mathbf{u}, \ldots, \hat{\mathbf{a}}_\ell^{\mathrm{T}}\mathbf{u}\} \quad \forall 0 \leqslant \theta \leqslant \alpha.$$

It follows, therefore, that $\hat{\mathbf{x}}$ solves (13.12) if and only if there is no vector $\mathbf{u} \in \mathscr{R}^n$ that satisfies

$$\hat{A}\mathbf{u} < \mathbf{e}, \tag{13.13}$$

where $\mathbf{e} = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathscr{R}^\ell$. The last condition is characterized by **Gordan's theorem of the alternatives**, which says that either (13.13) is solvable, or the system

$$\hat{A}^{\mathrm{T}}\mathbf{y} = \mathbf{0}, \quad \mathbf{y} \geqslant \mathbf{0}, \quad \mathbf{e}^{\mathrm{T}}\mathbf{y} = 1, \tag{13.14}$$

is solvable, but never both. See Section 11. The question which system is solvable is answered by solving the least squares problem

$$\begin{array}{ll} \text{minimize} & \|\tilde{A}^{\mathrm{T}}\mathbf{y} - \tilde{\mathbf{e}}_1\|_2^2 \\ \text{subject to} & \mathbf{y} \geqslant \mathbf{0}, \end{array} \tag{13.15}$$

where $\tilde{\mathbf{e}}_1 = (1, 0, 0, \ldots, 0) \in \mathscr{R}^{n+1}$ and $\tilde{A} = [\mathbf{e}, \hat{A}] \in \mathscr{R}^{\ell \times (n+1)}$. Let $\tilde{\mathbf{y}}$ solve (13.15) and let $\hat{\mathbf{r}} = \hat{A}^{\mathrm{T}}\tilde{\mathbf{y}}$ denote the resulting residual vector. If $\hat{\mathbf{r}} = \mathbf{0}$ then $\tilde{\mathbf{y}}$ solves (13.14). Otherwise, when $\hat{\mathbf{r}} \neq \mathbf{0}$, the unit vector $\hat{\mathbf{u}} = -\hat{\mathbf{r}}/\|\hat{\mathbf{r}}\|_2$ satisfies

$$\max\{\hat{\mathbf{a}}_1^{\mathrm{T}}\hat{\mathbf{u}}, \ldots, \hat{\mathbf{a}}_\ell^{\mathrm{T}}\hat{\mathbf{u}}\} = -\|\hat{\mathbf{r}}\| < 0,$$

and solves the steepest descent problem

$$\begin{array}{ll} \text{minimize} & \sigma(\mathbf{u}) = \max\{\hat{\mathbf{a}}_1^{\mathrm{T}}\mathbf{u}, \ldots, \hat{\mathbf{a}}_\ell^{\mathrm{T}}\mathbf{u}\} \\ \text{subject to} & \|\mathbf{u}\|_2 \leqslant 1. \end{array} \tag{13.16}$$

Similar constructive optimality conditions exist in several other problems. The ability to compute a steepest descent direction provides an effective way to resolve degeneracy (cycling) in active set methods and in algorithms for solving multifacility location problems. See [9,13,18] and the references therein.

14. Concluding remarks

The traditional "best approximation" theory concentrates on the distance between a point and a convex set. This is, perhaps, the reason that the new MND theorem has not been observed before. Extending the MND theorem to consider the distance between two convex sets enables it to handle a larger range of problems. Consider for example the distance between two polytopes. The need for calculating a hyperplane that separates two given polytopes is a central problem in the fields of patterns recognition and machine learning. The observation that the maximal separation equals the smallest distance adds new insight into this problem.

The examples described in this paper reveal the double role of duality in least norm problems. On one hand each least norm problem has its dual problem. For example, the dual of (7.4) is (7.5), and so forth. This paves the way for "dual methods" which are aimed at solving the corresponding dual problem. On the other hand, the question of whether a given point, $\hat{\mathbf{x}}$, solves the primal problem defines a second type of duality relation: The existence of a feasible descent direction at $\hat{\mathbf{x}}$ is connected to a certain theorem of the alternative, and a "secondary" least norm problem. To find a feasible descent direction we solve the secondary least norm problem and compute the resulting residual vector, $\hat{\mathbf{r}}$. If $\hat{\mathbf{r}} = \mathbf{0}$ then $\hat{\mathbf{x}}$ solves the primal problem. Otherwise $\hat{\mathbf{r}}$ points at the steepest descent direction at $\hat{\mathbf{x}}$.

It is true that the MND theorem is not as general as the duality theorems of Fenchel and Lagrange. Nevertheless, the MND theorem enjoys a number of pedagogical merits. First it has a simple geometric interpretation that visually illustrates the basic principles of duality. The representation of theorems by simple pictures makes them easier to understand and recall. Indeed, Figs. 1, 2 and 7 may convey the "duality principle" to a broad audience, including high school students and non-mathematicians. Second, as this paper shows, the MND theorem has a simple elementary proof which relies on pure geometric arguments. Third, in spite of its simplicity, the MND theorem applies to a large family of problems. Fourth, it is a useful tool for introducing and demonstrating several important concepts, such as polar decompositions, optimality conditions, steepest descent directions, theorems of the alternatives, etc.

The new MND theorem is valid in any finite dimensional real Hilbert space, $\mathscr{H}$, with inner product $\langle \mathbf{x}, \mathbf{y} \rangle$. The modification of the current proof to handle this setting is rather simple: $\mathscr{R}^m$ is replaced by $\mathscr{H}$, the Euclidean inner product $\mathbf{x}^{\mathrm{T}}\mathbf{y}$ is replaced by $\langle \mathbf{x}, \mathbf{y} \rangle$, the Euclidean norm is replaced by $\|\mathbf{x}\| = (\langle \mathbf{x}, \mathbf{x} \rangle)^{1/2}$, and so forth. Using the geometric Hahn–Banach theorem one can prove the Nirenberg–Luenberger MND theorem in any real (or complex) normed linear space, e.g., [19,20,27,29,40]. This observation suggests that the new MND theorem shares this property. One way to prove this assertion is by showing that the new MND theorem can be derived from the Nirenberg–Luenberger theorem in any normed linear space. (As before define $\mathscr{X} = \mathscr{Y} - \mathscr{Z}$ and consider the distance between $\mathbf{0}$ and $\mathscr{X}$.) However, this issue is beyond the scope of our paper.

References

[1] A. Ben-Tal, A. Nemirovski, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, SIAM, Philadelphia, 2001.
[2] K.P. Bennett, O.L. Mangasarian, Robust linear programming discrimination of two linearly inseparable sets, Optimiz. Methods Softw. 1 (1992) 23–34.
[3] P. Bloomfield, W.L. Steiger, Least Absolute Deviations, Birkhauser, Boston, 1983.
[4] W. Briec, Minimum distance to the complement of a convex set: a duality result, J. Optimiz. Theory Appl. 93 (1997) 301–319.

[5] Y. Censor, M.D. Altschuler, W.D. Powlis, A computational solution of the inverse problem in radiation-therapy treatment planning, Appl. Math. Comput. 25 (1988) 57–87.
[6] Y. Censor, T. Elfving, New methods for linear inequalities, Linear Algebra Appl. 42 (1982) 199–211.
[7] Y. Censor, S.A. Zenios, Parallel Optimization, Theory, Algorithms, and Applications, Oxford University Press, 1997.
[8] P.G. Ciarlet, Introduction to Numerical Linear Algebra and Optimization, Cambridge University Press, Cambridge, 1989.
[9] A. Dax, Linear programming via least squares, Linear Algebra Appl. 111 (1988) 313–324.
[10] A. Dax, The smallest point of a polytope, J. Optimiz. Theory Appl. 64 (1990) 419–422.
[11] A. Dax, A new theorem of the alternative, Math. Program. 47 (1990) 297–299.
[12] A. Dax, On minimum norm solutions, J. Optimiz. Theory Appl. 76 (1993) 183–193.
[13] A. Dax, The relationship between theorems of the alternative, least norm problems, steepest descent directions, and degeneracy: a review, Ann. Oper. Res. 46 (1993) 11–60.
[14] A. Dax, Loss and retention of accuracy in affine scaling methods, Optimiz. Methods Softw. 15 (2001) 121–151.
[15] A. Dax, The smallest correction of an inconsistent system of linear inequalities, Optimiz. Eng. 2 (2001) 349–359.
[16] A. Dax, The ℓ_1 solution of linear inequalities, Comput. Statist. Data Anal. 50 (2006) 40–60.
[17] A. Dax, The distance between two convex sets: the shortest distance equals the largest separation, Technical Report, Hydrological Service of Israel, 2004.
[18] A. Dax, V.P. Sreedharan, On theorems of the alternative and duality, J. Optimiz. Theory Appl. 94 (1997) 561–590.
[19] F. Deutsch, Best Approximation in Inner Product Spaces, Springer, 2001.
[20] F.R. Deutsch, P.H. Maserick, Applications of the Hahn–Banach theorem in approximation theory, SIAM Rev. 9 (1967) 516–530.
[21] B. Efron, Computer intensive methods in statistical regression, SIAM Rev. 30 (1988) 421–449.
[22] P.E. Gill, W. Murray, M.H. Wright, Numerical Linear Algebra and Optimization, Addison-Wesley, 1991.
[23] G.T. Herman, Image Reconstruction from Projections: The Fundamentals of Computerized Tomography, Academic Press, New York, 1980.
[24] J.-B. Hiriart-Urruty, C. Lemarechal, Convex Analysis and Minimization Algorithms I, Springer-Verlag, Berlin, Heidelberg, 1993.
[25] A.S. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell Publishing Company, New York, 1964.
[26] W. Kaplan, Maxima and Minima with Applications, Wiley, 1999.
[27] P.D. Lax, Functional Analysis, Wiley, 2002.
[28] A. Lent, Y. Censor, Extensions of Hildreth's row-action method for quadratic programming, SIAM J. Control Optimiz. 18 (1980) 444–454.
[29] D.G. Luenberger, Optimization by Vector Space Methods, Wiley, New York, 1969.
[30] D.G. Luenberger, Linear and Nonlinear Programming, Addision-Wesley, 1984.
[31] O.L. Mangasarian, Linear and nonlinear separation of patterns by linear programming, Oper. Res. 13 (1965) 444–452.
[32] O.L. Mangasarian, Nonlinear Programming, McGraw-Hill, New York, 1969.
[33] O.L. Mangasarian, Mathematical programming in machine learning, in: G. Di Pillo, F. Giannesi (Eds.), Nonlinear Optimization and Applications, Plenum Press, New York, 1996.
[34] O.L. Mangasarian, Arbitrary-norm separating plane, Oper. Res. Lett. 24 (1999) 15–23.
[35] O.L. Mangasarian, R. Setiono, W.H. Wolberg, Pattern recognition via linear programming: theory and application to medical diagnosis, in: T.F. Coleman, Y. Li (Eds.), Large-scale Numerical Optimization, SIAM, Philadelphia, 1990, pp. 22–31.
[36] P.H. Merz, An algorithm for percentile curves via constrained weighted L_1 regression, Technical Report, Chevron Research and Technology Co., 1998.
[37] B.F. Mitchell, V.F. Dem'yanov, V.N. Malozemov, Finding the point of a polyhedron closest to the origin, SIAM J. Control Optimiz. 12 (1974) 19–26.
[38] J.J. Moreau, Decomposition orthogonale d'un espace hilbertien selon deux cones mutuellement polaires, C.R. Acad. Sci. Paris 255 (1962) 238–240.
[39] J.J. Moreau, Convexity and duality, in: E.R. Caianiello (Ed.), Fundamental Analysis and Optimization, Academic Press, New York, 1966, pp. 145–169.
[40] L. Nirenberg, Functional analysis, Lectures given in 1960–61, notes by Lesley Sibner, New York University, 1961.
[41] M.R. Osborne, Finite Algorithms in Optimization and Data Analysis, Wiley, 1985.
[42] R.W. Owens, V.P. Sreedharan, Least squares methods to minimize errors in a smooth, strictly convex norm on $\mathscr{R}^m$, J. Approx. Theory 73 (1993) 180–193.
[43] M.J. Panik, Fundamentals of Convex Analysis, Kluwer, 1993.
[44] A.W. Roberts, D.E. Varberg, Convex Functions, Academic Press, New York, 1973.

[45] R.T. Rockafellar, Convex Analysis, Princeton University Press, 1970.
[46] I. Singer, Duality for Nonconvex Approximation and Optimization, Springer-Verlag, 2006.
[47] V.P. Sreedharan, Least squares algorithms for finding solutions of overdetermined linear equations which minimize error in an abstract norm, Numer. Math. 17 (1971) 387–401.
[48] V.P. Sreedharan, Least squares algorithms for finding solutions of overdetermined systems of linear equations which mininmize error in a smooth strictly convex norm, J. Approx. Theory 8 (1973) 46–61.
[49] J. Stoer, C. Witzgall, Convexity and Optimization in Finite Dimensions, Part I, Springer, Berlin, 1970.
[50] H.G. Tuenter, *L*1-distance projection onto convex sets: a simple characterization, J. Optimiz. Theory Appl. 112 (2002) 441–445.
[51] G.A. Watson, Choice of norms for data fitting and function approximation, Acta Numer. 7 (1998) 337–377.
[52] P. Wolfe, Algorithm for a least-distance programming problem, Math. Programm. Study 1 (1974) 190–205.
[53] P. Wolfe, Finding the nearest point in a polytope, Math. Programm. 11 (1976) 128–149.
[54] W.H. Yang, On generalized Hölder inequality, Nonlinear Anal. Theory Methods Appl. 16 (1991) 489–498.