

# Minmax-distance approximation and separation problems: geometrical properties

Frank Plastria · Emilio Carrizosa

Received: 12 March 2007 / Accepted: 11 May 2010 / Published online: 4 June 2010  
© Springer and Mathematical Programming Society 2010

**Abstract** A center hyperplane in the  $d$ -dimensional space minimizes the maximum of its distances from a finite set of points  $A$  with respect to possibly different gauges. In this note it is shown that a center hyperplane exists which is at (equal) maximum distance from at least  $d + 1$  points of  $A$ . Moreover the projections of the points among these which lie above the center hyperplane cannot be separated by another hyperplane from the projections of those that are below it. When all gauges involved are smooth, all center hyperplanes satisfy these properties. This geometric property allows us to improve and generalize previously existing results, which were only known for the case in which all distances are measured using a common norm. The results also extend to the constrained case where for some points it is prespecified on which side of the hyperplane (above, below or on) they must lie. In this case the number of points lying on the hyperplane plus those at maximum distance is at least  $d + 1$ . It follows that solving such global optimization problems reduces to inspecting a finite set of candidate solutions. Extensions of these results to a separation problem are outlined.

**Keywords** Gauge-distance to hyperplane · Center hyperplane · Linear separation

**Mathematics Subject Classification (2000)** 90C47 · 62H30

---

The research of the second author is partially supported by grants MTM2008-3032, Spain, and FQM-329 of Junta de Andalucía, Spain.

---

F. Plastria (✉)  
MOSI, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium  
e-mail: Frank.Plastria@vub.ac.be  
URL: <http://homepages.vub.ac.be/~faplastr>

E. Carrizosa  
Fac. de Matemáticas, Universidad de Sevilla, Tarfia s/n, 41012 Sevilla, Spain  
e-mail: [ecarrizosa@us.es](mailto:ecarrizosa@us.es)

## 1 Introduction

In continuous location models, [31], one has a set of users, identified with a set  $A$  in  $\mathbb{R}^d$ , and the location for a facility of given shape is sought so that some function of the user-facility distances (e.g. the weighted sum of the distances, or the largest distance) is optimized. Although the facility is typically assumed to be a point, a good number of papers address the case in which the facility to be located has a different shape, (e.g. sphere, hyperplane or polygon in the plane), leading to location models for *extensive facilities*. See [2, 11, 36] and the references therein.

These location problems not only have evident applications in Operations Research, see e.g. [3, 4, 11, 12, 20, 25, 27, 37, 38, 41], but are also considered in other fields. In Approximation Theory and Statistics orthogonal regression calls for finding a hyperplane minimizing some function of its distances to given data-points [14, 15, 18, 19, 24, 29, 32, 39]. In Computational Geometry [16] obtains the width of a finite set as the smallest possible distance between two parallel hyperplanes enclosing the set, which is equivalent to minimizing the maximal distance of any point to the hyperplane in the middle between these. In Machine Learning [1, 5], given two linearly separable sets of data-points, the support vector machine is a hyperplane maximizing the minimum distance to the data-points while separating the two sets. For sets that are not linearly separable, for any hyperplane one associates each set to one of the two corresponding halfspaces, and seeks a hyperplane minimizing the sum of distances of the points to their halfspace [6, 22, 33].

In general, the resulting optimization problems are multimodal, and thus can only be solved to optimality using expensive global optimization tools such as continuous branch and bound. However, as happens also in other location models, for particular shapes of the facility it is often possible to identify a so-called *finite dominating set*, i.e. a finite set of candidates known to contain an optimal solution. Hence, the search of an optimum may be reduced to exhaustive enumeration of this finite candidate set.

For problems in higher dimension, neither branch and bound schemes nor the inspection of finite sets of (sets of) candidates are realistic alternatives, and one is forced to use heuristics. Taking into account the geometrical properties adds a combinatorial structure in the continuous optimization problem which can be successfully exploited to develop more powerful heuristics [34].

As examples of finite dominating sets the following basic cases are known. In the plane with Euclidean distances, some median line, i.e. a line minimizing the sum of weighted distances to a set of points, is known [29] to pass through at least two of these points. Similarly, some center line, i.e. a line minimizing the maximum weighted distance to a finite set of points, exists at maximum weighted distance from three of these points [27]. More generally, for any norm distance in any dimension  $d$ , some median hyperplane will pass through at least  $d$  affinely independent points, and some center hyperplane will have maximal weighted distance to at least  $d + 1$  points [36, p153]. Other examples can be found in location of segments [17], obnoxious planes, [10], as well as for separation problems by hyperplanes [6, 33, 40].

In this paper we study such geometric properties for center hyperplanes when distances are measured by norms or gauges, thereby sharpening and generalizing known results on finite dominating sets. After reviewing in Sect. 2 basic properties of gauges,

distances and center problems, we give in Sect. 3 a series of geometrical properties of the center hyperplanes, enabling us to answer a number of open problems found in the literature, as summarized in Sect. 3.5. Moreover, center problems with constraints are also considered, allowing us to also address in Sect. 4, as a byproduct, a problem of approximate separation by a hyperplane of two sets of points that are not linearly separable.

### 2 Gauge distances and center hyperplanes

Given  $u \in \mathbb{R}^d, u \neq 0$  and  $\beta \in \mathbb{R}$ , let  $H(u, \beta)$  denote the hyperplane of equation  $\langle u ; x \rangle = \beta$ , and let  $H(u, \beta)^\geq$  and  $H(u, \beta)^\leq$  denote the two closed halfspaces with common boundary  $H(u, \beta)$ ,

$$\begin{aligned}
 H(u, \beta)^\geq &\stackrel{\text{def}}{=} \left\{ x \in \mathbb{R}^d : \langle u ; x \rangle \geq \beta \right\} \\
 H(u, \beta)^\leq &\stackrel{\text{def}}{=} \left\{ x \in \mathbb{R}^d : \langle u ; x \rangle \leq \beta \right\}
 \end{aligned}$$

The hyperplane  $H(u, \beta)$  separates (respectively strictly separates) the subsets  $S_1$  and  $S_2$  of  $\mathbb{R}^d$ , if  $S_1 \subset H(u, \beta)^\geq$  and  $S_2 \subset H(u, \beta)^\leq$ , or inversely (resp.  $S_1 \cap H(u, \beta)^\leq = \emptyset$  and  $S_2 \cap H(u, \beta)^\geq = \emptyset$ , or inversely).

Distances will be measured through finite gauges, [13,26,32]: Given a compact convex set  $B$  in  $\mathbb{R}^d$  containing the origin in its interior, its gauge  $\gamma$  is defined as

$$\gamma(x) \stackrel{\text{def}}{=} \min\{t \geq 0 \mid x \in tB\}, \tag{1}$$

and its dual gauge  $\gamma^\circ$  as

$$\gamma^\circ(v) \stackrel{\text{def}}{=} \max\{\langle v ; y \rangle \mid \gamma(y) \leq 1\}.$$

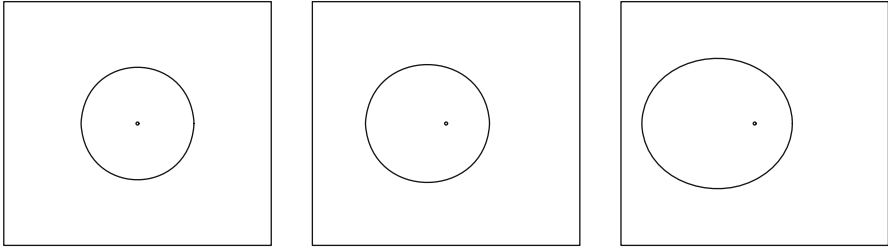
The class of gauges includes the class of norms, since norms are those gauges whose unit ball  $B$  is symmetric with respect to the origin. Observe that, by definition, if  $\gamma$  is a norm, then its dual  $\gamma^\circ$  is also a norm, and thus  $\gamma^\circ(-v) = \gamma^\circ(v)$  for all  $v$ . Any positive multiple  $\lambda\gamma (\lambda > 0)$  of a gauge  $\gamma$  is a gauge with dual gauge  $(\lambda\gamma)^\circ = \frac{\gamma^\circ}{\lambda}$ .

Interesting instances of gauges which are not norms are, for instance, the *skewed gauges*, used in Statistics to define a multivariate version of quantiles, [8], and also suggested in Location Theory to model transportation on an inclined plane, in presence of steady wind, [30], or to a moving target [7]. Skewed gauges have the form

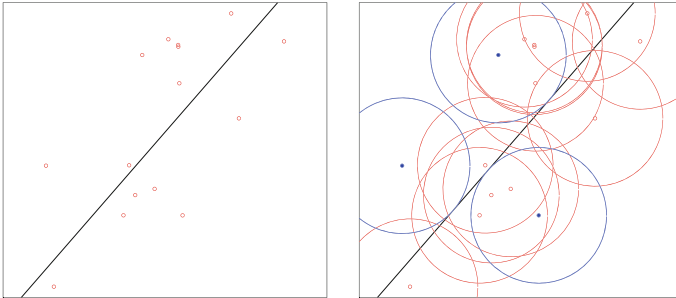
$$\gamma(x) = \|x\| + p^\top x,$$

where  $\|\cdot\|$  is a norm and  $p$  is a vector with  $\|p\|^\circ < 1$ . See Fig. 1 for the unit ball of the gauge  $\gamma$  for  $\|\cdot\|$  being the Euclidean norm and different values of  $p$ , namely  $p = (0, 0)$  (and thus  $\gamma$  is the Euclidean norm),  $p = (0.3, 0)$ , and  $p = (0.5, 0)$  respectively.

See also [30] for further properties on skewed gauges.



**Fig. 1** Unit ball of skewed gauges



**Fig. 2** Center hyperplane (Euclidean norm) and balls of smallest common radius.

Gauges are convex functions in  $\mathbb{R}^d$ . The subdifferential  $\partial\gamma(x)$  of the gauge  $\gamma$  at  $x \in \mathbb{R}^d$  is never empty, and can be expressed as

$$\partial\gamma(x) = \{v \in \mathbb{R}^d, \gamma^\circ(v) \leq 1 \mid \langle x; v \rangle = \gamma(x)\}.$$

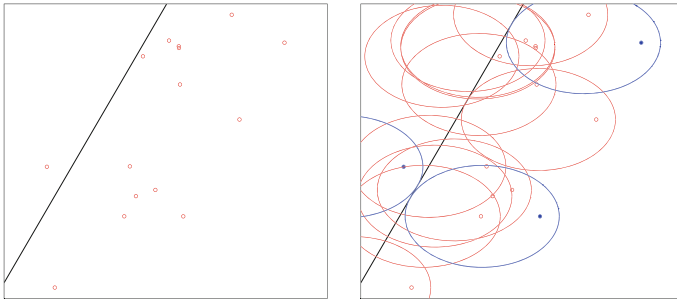
Note that when the gauge  $\gamma$  is smooth at  $x$ , the subdifferential is reduced to a single subgradient: the gradient  $\nabla\gamma(x)$  of  $\gamma$  at  $x$ .

Given a hyperplane  $H$ , the  $\gamma$ -distance of a point  $a \in \mathbb{R}^d$  to  $H$  is defined as

$$d_\gamma(a, H) \stackrel{\text{def}}{=} \min\{\gamma(x - a) \mid x \in H\}.$$

Given a non-empty finite set  $A \subset \mathbb{R}^d$ , together with corresponding gauges  $\gamma_a$  on  $\mathbb{R}^d$  with unit ball  $B_a(a \in A)$ , and two finite (not necessarily disjoint, possibly empty) sets  $A^\geq, A^\leq \subset \mathbb{R}^d$ , a *constrained center hyperplane* is a hyperplane  $H(u, \beta)$  such that the two closed halfspaces  $H(u, \beta)^\geq$  and  $H(u, \beta)^\leq$  contain respectively  $A^\geq$  and  $A^\leq$ , and it minimizes the highest  $\gamma_a$ -distance to  $a \in A$ , i.e., it minimizes the common radius  $r$  needed to guarantee that all balls  $a + rB_a$  intersect  $H(u, \beta)$ .

As an illustration, the left side of Fig. 2 (respectively Fig. 3) represents the unconstrained ( $A^\geq = A^\leq = \emptyset$ ) center hyperplane for the points given, with all points measuring distances with a common gauge, namely, the Euclidean norm (respectively, a skewed gauge). Such a hyperplane  $H(u, \beta)$  has been obtained using the results derived in Sect. 3 below.



**Fig. 3** Center hyperplane (skewed gauge) and balls of smallest common radius

For the center hyperplane  $H(u, \beta)$  obtained, the balls  $a + rB_a$  of smallest common radius  $r$  needed to guarantee nonempty intersection with  $H(u, \beta)$  are plotted in the right side of Figs. 2 and 3. Three balls, with centers marked with solid circles, are tangent to  $H(u, \beta)$ , a property stated in generality in Sect. 3.

Center hyperplane problems have already been extensively considered in the literature in the unconstrained case, i.e., for  $A^\geq = A^\leq = \emptyset$ , e.g. [11, 14, 15, 19, 23, 24, 36, 37, 39], first for Euclidean distances, later for a general norm distance. Some geometrical properties have been obtained for these cases and will be generalized here, both in form and in application range.

The inclusion of the constraints given by  $A^\geq, A^\leq \subset \mathbb{R}^d$ , allows to make the link to a problem typical in classification: finding the hyperplane optimizing the margin needed to separate two sets which are not linearly separable. As detailed in Sect. 4, this margin optimization problem is strongly related to a center hyperplane problem for adequate choices of the sets  $A^\geq, A^\leq$ .

On the other hand, our formulation enables us also to accommodate within the same framework *anchored* center hyperplane problems, as introduced in [38]: a hyperplane  $H(u, \beta)$  is sought minimizing the largest distance to a set of points  $A$ , and constrained to pass through all points of some  $A_0 \subset A$ . It then follows that such a problem is a particular instance of (4) for  $A^\geq = A^\leq = A_0$ .

To express center hyperplane problems as optimization problems, we make use of the following closed-form formula for  $\gamma$ -distances derived in [32]:

$$d_\gamma(a, H(u, \beta)) = \max \left\{ \frac{\beta - \langle u ; a \rangle}{\gamma^\circ(u)}, \frac{\langle u ; a \rangle - \beta}{\gamma^\circ(-u)} \right\}, \tag{2}$$

which simplifies to

$$d_\gamma(a, H(u, \beta)) = \frac{|\langle u ; a \rangle - \beta|}{\gamma^\circ(u)} \tag{3}$$

when  $\gamma$  (and thus  $\gamma^\circ$ ) is a norm. Note that, for each point  $a$  and gauge  $\gamma$ , the distance function

$$\mathcal{D}_a : (\mathbb{R}^d \setminus \{0\}) \times \mathbb{R} \rightarrow \mathbb{R}^+ : (u, \beta) \mapsto d_\gamma(a, H(u, \beta))$$

is continuous.

For  $(u, \beta) \in (\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}$ , let  $\Psi(u, \beta)$  denote the highest  $\gamma_a$ -distance from  $a \in A$  to the hyperplane  $H(u, \beta)$ ,

$$\Psi(u, \beta) \stackrel{\text{def}}{=} \max_{a \in A} d_{\gamma_a}(a, H(u, \beta)),$$

where the  $\gamma_a$  distances are given by (2).

Finding a constrained center hyperplane  $H(u, \beta)$  amounts to solving the optimization problem

$$\begin{aligned} \min \quad & \Psi(u, \beta) \\ \text{s.t.} \quad & A^{\geq} \subset H(u, \beta)^{\geq}, \quad A^{\leq} \subset H(u, \beta)^{\leq} \\ & u \in (\mathbb{R}^d \setminus \{0\}), \beta \in \mathbb{R}. \end{aligned} \tag{4}$$

Note that since the objective and the constraints of this optimization problem are all invariant for multiplication by a positive constant, the constraint  $u \neq 0$  may be replaced at our convenience by  $\|u\| = 1$ , where  $\|\cdot\|$  is any norm (or gauge).

**Assumption 1** Throughout the paper, we assume that (4) is non-degenerate, in the sense that it is feasible (i.e.  $A^{\geq}$  and  $A^{\leq}$  can be separated by a hyperplane), and has strictly positive optimal value (i.e.  $A$  is not contained in a hyperplane separating  $A^{\geq}$  and  $A^{\leq}$ ).

This is the case for example in the ‘classical’ situation when no constraints are present and  $A$  has dimension  $d$ .

### 3 Geometrical properties of center hyperplanes

In what follows we first prove existence of optimal solutions to (4), and then derive several structural and geometrical properties of optimal solutions. These properties reduce the search for a center hyperplane to the inspection of a finite set of candidates, polynomial in the number of data points for fixed dimension  $d$ , but exponential in  $d$ .

Somewhat stronger results are obtained for the case in which every gauge  $\gamma_a$  is smooth, and/or when all gauges are weighted versions of a same norm or gauge.

#### 3.1 Existence of center hyperplanes

Let us first consider Problem (4) for  $u \neq 0$  fixed. Define

$$\begin{aligned} \delta^{\leq}(u) &= \max_{a \in A^{\leq}} \langle u ; a \rangle \\ \delta^{\geq}(u) &= \min_{a \in A^{\geq}} \langle u ; a \rangle \\ I(u) &= \{ \beta : \delta^{\leq}(u) \leq \beta \leq \delta^{\geq}(u) \}, \end{aligned}$$

and the problem

$$\min_{\beta \in I(u)} \Psi(u, \beta) \tag{5}$$

Note that the feasibility assumption of (5) is equivalent to

$$\delta^{\leq}(u) \leq \delta^{\geq}(u). \tag{6}$$

**Lemma 1** For fixed  $u \in \mathbb{R}^d \setminus \{0\}$  satisfying (6), one has

1. there exists a unique  $\beta^*(u)$  optimal solution to (5)
2.  $-\max_{a \in A \leq \cup A} \|a\| \|u\| \leq \beta^*(u) \leq \max_{a \in A \geq \cup A} \|a\| \|u\|$
3. there exists at least one active  $a \in A$ , i.e. such that  $d_{\gamma_a}(a, H(u, \beta^*(u))) = \Psi(u, \beta^*(u))$ . Moreover, if  $\delta^{\leq}(u) < \beta^*(u) < \delta^{\geq}(u)$ , then there exist at least two active points.

In addition,  $\beta^*$  is a continuous function of  $u$ , and the intersection of the unit Euclidean ball with its graph

$$\{(u, \beta^*(u)) \mid \|u\| = 1, \delta^{\leq}(u) \leq \delta^{\geq}(u)\}$$

is compact.

*Proof* By the expression (2) for the distance (with respect to a gauge) to a hyperplane, the objective  $\Psi(u, \beta)$  is seen to be the maximum of a finite number of affine functions with strictly positive or strictly negative slope, tending to  $+\infty$  when  $\beta$  tends to  $\pm\infty$ . It follows that the problem obtained when the constraints on  $\beta$  are dropped,

$$\min_{\beta \in \mathbb{R}} \max_{a \in A} \left\{ \frac{\beta - \langle u ; a \rangle}{\gamma_a^\circ(u)}, \frac{\langle u ; a \rangle - \beta}{\gamma_a^\circ(-u)} \right\} \tag{7}$$

has just one optimal solution  $\beta(u)$ .

By convexity,  $\beta(u)$  must be such that the right derivative (respectively left derivative) of  $\Psi(u, \beta)$  at  $\beta = \beta(u)$  is non-negative (respectively non-positive). This happens iff  $\beta(u)$  satisfies

$$\frac{\beta(u) - \langle u ; a \rangle}{\gamma_a^\circ(u)} = \frac{\langle u ; b \rangle - \beta(u)}{\gamma_b^\circ(-u)} \tag{8}$$

for some  $a, b \in A$ . Also, by Assumption 1 we have  $\Psi(u, \beta(u)) \neq 0$ , so we must have  $a \neq b$ . In other words,  $\beta(u)$  is always of the form  $\beta_{ab}(u)$ ,

$$\begin{aligned} \beta_{ab}(u) &= \frac{\gamma_a^\circ(u)}{\gamma_a^\circ(u) + \gamma_b^\circ(-u)} \langle u ; b \rangle + \frac{\gamma_b^\circ(-u)}{\gamma_a^\circ(u) + \gamma_b^\circ(-u)} \langle u ; a \rangle \\ &\in [\langle u ; b \rangle, \langle u ; a \rangle] \\ &\subset \left[ -\max_{a \in A} \|a\| \|u\|, \max_{a \in A} \|a\| \|u\| \right] \end{aligned} \tag{9}$$

for some  $a \neq b \in A$ .

Now the constrained problem (5) is a one-dimensional convex problem, whose unconstrained version (7) has as unique optimal solution  $\beta(u)$ . Therefore the constrained optimal solution is also unique and obtained by

$$\beta^*(u) = \max \left( \delta^{\leq}(u), \min \left( \beta(u), \delta^{\geq}(u) \right) \right) \in \left[ - \max_{a \in A^{\leq} \cup A} \|a\| \|u\|, \max_{a \in A^{\geq} \cup A} \|a\| \|u\| \right] \tag{10}$$

Existence of an active  $a \in A$  is a direct consequence of the definition of the objective  $\Psi$ . When  $\delta^{\leq}(u) < \beta^*(u) < \delta^{\geq}(u)$  we have  $\beta^*(u) = \beta(u) = \beta_{ab}(u)$  for some  $a \neq b \in A$ , which are then both active points.

When  $\|u\| = 1$ , it was shown above that  $\beta(u)$  takes values in the compact interval  $[- \max_{a \in A} \|a\|, \max_{a \in A} \|a\|]$ . Since  $\beta$  is an *argmin*-function of the continuous maximal distance function to the points of  $A$ , this implies its continuity on  $\|u\| = 1$ . Equation (10) and the continuity of  $\delta^{\leq}$  and  $\delta^{\geq}$ , imply that  $\beta^*$  is continuous in  $u$  on the set  $\{u \in \mathbb{R}^d \mid \|u\| = 1, \delta^{\leq}(u) \leq \delta^{\geq}(u)\}$ . The announced compactness then also directly follows.  $\square$

**Theorem 2** *There always exists a constrained center hyperplane.*

*Proof* By the previous lemma, the center hyperplane problem is reduced to the following problem

$$\begin{aligned} \min \quad & \Psi(u, \beta^*(u)) \\ \text{s.t.} \quad & \delta^{\leq}(u) \leq \delta^{\geq}(u) \\ & \|u\| = 1, \end{aligned} \tag{11}$$

which is the minimization of a continuous function on a compact domain. This proves existence of an optimal solution to (11), and hence, by equivalence, of an optimal solution  $(u^*, \beta^*)$  to Problem (4). Hence,  $H(u^*, \beta^*)$  is a constrained center hyperplane.  $\square$

### 3.2 Structural properties

For any solution  $(u, \beta) \in (\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}$ , define the following sets of points

Upper active distance points

$$D^{\geq}(u, \beta) \stackrel{\text{def}}{=} \{a \in A \cap H^{\geq}(u, \beta) \mid d_{\gamma_a}(a, H(u, \beta)) = \Psi(u, \beta)\} \tag{12}$$

Lower active distance points

$$D^{\leq}(u, \beta) \stackrel{\text{def}}{=} \{a \in A \cap H^{\leq}(u, \beta) \mid d_{\gamma_a}(a, H(u, \beta)) = \Psi(u, \beta)\} \tag{13}$$

Active distance points

$$D(u, \beta) \stackrel{\text{def}}{=} D^{\geq}(u, \beta) \cup D^{\leq}(u, \beta) \tag{14}$$



Active constraint points

$$C(u, \beta) \stackrel{\text{def}}{=} (A^{\geq} \cup A^{\leq}) \cap H(u, \beta)$$

Active points

$$Act(u, \beta) \stackrel{\text{def}}{=} D(u, \beta) \cup C(u, \beta)$$

*Remark 3* By the definition of  $\Psi$  the sets  $D^{\geq}(u, \beta)$  and  $D^{\leq}(u, \beta)$  cannot be both empty, i.e.  $D(u, \beta) \neq \emptyset$ . By the nondegeneracy assumption 1, for any feasible  $u$  these sets are also disjoint. Additionally, Lemma 1 and its proof show that for any optimal solution  $(u^*, \beta^*)$  with  $C(u^*, \beta^*) = \emptyset$ , they are both non-empty.

A point  $p_a \in H(u, \beta)$  is called a  $\gamma$  – projection of  $a$  on the hyperplane  $H(u, \beta)$  when  $d_{\gamma}(a, H(u, \beta)) = \gamma(p_a - a)$ , which happens (see [32]) if and only if

$$\begin{aligned} \frac{u}{\gamma^{\circ}(u)} &\in \partial\gamma(p_a - a) \quad \text{when } a \in H(u, \beta)^{\leq} \\ \frac{-u}{\gamma^{\circ}(-u)} &\in \partial\gamma(p_a - a) \quad \text{when } a \in H(u, \beta)^{\geq}. \end{aligned} \tag{15}$$

Equivalently,  $H(u, \beta)$  is a supporting hyperplane at  $p_a$  of the  $\gamma$ -ball  $a + d_{\gamma}(a, H(u, \beta))B$  of center  $a$  and radius  $d_{\gamma}(a, H(u, \beta))$ , where  $B$  denotes the unit ball of  $\gamma$ .

**Theorem 4** *Let  $(u^*, \beta^*)$  be an optimal solution to (4), whose objective value  $z^*$  is strictly positive by assumption. For each  $a \in D^{\geq}(u^*, \beta^*)$  (resp.  $D^{\leq}(u^*, \beta^*)$ ) let  $p_a$  be a  $\gamma_a$ -projection of  $a$  on  $H(u^*, \beta^*)$ , and denote the set of these projections  $p_a$  by  $P^{\geq}$  (resp.  $P^{\leq}$ ). One has*

1. *The sets  $P^{\geq}$  and  $P^{\leq}$  cannot be strictly separated by a hyperplane containing  $C(u^*, \beta^*)$ .*
2. *If each  $\gamma_a$  is a smooth gauge, then  $H(u^*, \beta^*)$  is the only hyperplane containing  $C(u^*, \beta^*)$  that separates  $P^{\geq}$  and  $P^{\leq}$*

*Proof* We prove both properties by contradiction.

1. Suppose some hyperplane  $H(u, \beta)$  containing  $C(u^*, \beta^*)$  strictly separates  $P^{\geq}$  and  $P^{\leq}$ . We can assume without loss of generality that  $\langle u ; p_a \rangle < \beta$  for all  $p_a \in P^{\geq}$  and  $\langle u ; p_a \rangle > \beta$  for all  $p_a \in P^{\leq}$ . Defining  $u_k = u^* + ku$  and  $\beta_k = \beta^* + k\beta$ , consider the hyperplanes  $H(u_k, \beta_k)$  for small values of  $k > 0$ . One then has:
  - For any  $a \in A^{\geq} \subset H^{\geq}(u^*, \beta^*)$ , either  $a \in C(u^*, \beta^*) \subset H(u_k, \beta_k)$ , or  $\langle u^* ; a \rangle - \beta^* > 0$ . Now  $\langle u_k ; a \rangle - \beta_k = \langle u^* ; a \rangle - \beta^* + k(\langle u ; a \rangle - \beta)$  and for  $k > 0$  sufficiently small, this quantity is strictly positive, implying that  $a \in H^{\geq}(u_k, \beta_k)$ . Hence, for any  $k > 0$  sufficiently small, we have  $A^{\geq} \subset H^{\geq}(u_k, \beta_k)$ .
  - Similarly for any  $k > 0$  sufficiently small, we have  $A^{\leq} \subset H^{\leq}(u_k, \beta_k)$ .

- Also for any  $a \in A \setminus D(u^*, \beta^*)$  we have

$$D_a(u^*, \beta^*) = d_{\gamma_a}(a, H(u^*, \beta^*)) < z^*,$$

and, by continuity of  $D_a$ ,  $D_a(u_k, \beta_k) < z^*$  as soon as  $k > 0$  is sufficiently small.

- Consider now any  $a \in D^{\geq}(u^*, \beta^*)$  and its projection point  $p_a \in P^{\geq}$ . Since  $\gamma_a(p_a - a) = d_{\gamma_a}(a, H(u^*, \beta^*)) = z^* > 0$ ,  $a$  cannot lie on  $H(u^*, \beta^*)$ , so we must have  $\langle u^* ; a \rangle > \beta^*$ . It follows then as above that for any  $k > 0$  sufficiently small  $\langle u_k ; a \rangle > \beta_k$ .  
 On the other hand,  $p_a \in H(u^*, \beta^*)$ , so  $\langle u^* ; p_a \rangle = \beta^*$ , while, by assumption,  $\langle u ; p_a \rangle < \beta$ , and therefore  $\langle u_k ; p_a \rangle < \beta_k$  as soon as  $k > 0$ . Therefore, for sufficiently small  $k > 0$ , there exists a (unique) point  $a_k \in ]a, p_a[$  such that  $\langle u_k ; a_k \rangle = \beta_k$ , i.e.  $a_k \in H(u_k, \beta_k)$ . Hence,

$$d_{\gamma_a}(a, H(u_k, \beta_k)) \leq \gamma_a(a_k - a) < \gamma_a(p_a - a) = z^*$$

- Analogously we obtain that  $d_{\gamma_a}(a, H(u_k, \beta_k)) < z^*$  for sufficiently small  $k > 0$  and for all  $a \in D^{\leq}(u^*, \beta^*)$ .

In other words, for sufficiently small  $k > 0$ ,  $(u_k, \beta_k)$  would be a feasible solution to problem (4) with  $\Psi(u_k, \beta_k) < z^*$ , contradicting the optimality of  $(u^*, \beta^*)$ .

2. Assume now that each  $\gamma_a$  is a smooth gauge, and that some hyperplane  $H(u, \beta)$  different from  $H(u^*, \beta^*)$  contains  $C(u^*, \beta^*)$  and separates  $P^{\geq}$  and  $P^{\leq}$ . Without loss of generality we assume further that  $\langle u ; p_a \rangle \leq \beta$  for all  $p_a \in P^{\geq}$ , while  $\langle u ; p_a \rangle \geq \beta$  for all  $p_a \in P^{\leq}$ .

All arguments used above may be used again unchanged, except for those  $a \in D(u^*, \beta^*)$  for which  $\langle u ; p_a \rangle = \beta$ . But for any such  $a$  all hyperplanes  $H(u_k, \beta_k) (k > 0)$  pass through  $p_a$  and are different from  $H(u^*, \beta^*)$ . And this latter is a hyperplane of support to the  $\gamma_a$ -ball  $a + z^* B_a$  at  $p_a$ , which is unique, by smoothness of  $\gamma_a$ . Therefore no  $H(u_k, \beta_k)$  can be a hyperplane of support at  $p_a$  to this  $\gamma_a$ -ball, and therefore they must all contain some interior points of this ball. But this means that  $d_{\gamma_a}(a, H(u_k, \beta_k)) < z^*$  for any  $k > 0$ . Therefore we may conclude again that for sufficiently small  $k > 0$ ,  $(u_k, \beta_k)$  would be a feasible solution to problem (4) with  $\Psi(u_k, \beta_k) < z^*$ , contradicting the optimality of  $(u^*, \beta^*)$ .

□

**Theorem 5** *Suppose that each  $\gamma_a$  is a smooth gauge. Let  $(u^*, \beta^*)$  be an optimal solution to (4). One has*

1. *The set  $Act(u^*, \beta^*)$  contains at least  $d + 1$  points.*
2. *Let  $v$  be a smooth norm. If for all  $a \in A$ ,  $\gamma_a = \omega_a v$  for some positive  $\omega_a$ , then  $Act(u^*, \beta^*)$  cannot be contained in a hyperplane, i.e. contains  $d + 1$  affinely independent points.*
3. *When no constraints are present ( $A^{\geq} = A^{\leq} = \emptyset$ ) and all  $\gamma_a$  are equal to a same smooth gauge  $\gamma$  (i.e. unweighted), then  $Act(u^*, \beta^*)$  cannot be contained in a hyperplane.*

*Proof 1.* Consider any feasible solution  $(u, \beta)$  with no more than  $d$  active points, and we will show that  $(u, \beta)$  is not optimal to (4). For each  $a \in D^{\geq}(u, \beta)$  (resp.  $D^{\leq}(u, \beta)$ ) let  $p_a$  be a  $\gamma_a$ -projection of  $a$  on  $H(u, \beta)$ , and denote the set of these projections by  $P^{\geq}$  (resp.  $P^{\leq}$ ).

By Remark 3,  $D(u, \beta) \neq \emptyset$ . Take  $\bar{a} \in D(u, \beta)$ . Then  $((P^{\geq} \cup P^{\leq}) \setminus \{p_{\bar{a}}\}) \cup C(u, \beta)$  consists of at most  $d - 1$  points, and hence there exist hyperplanes  $H(u', \beta')$  different from  $H(u, \beta)$  containing this set. By construction any such hyperplane contains  $C(u, \beta)$  and separates (non strictly!)  $P^{\geq}$  and  $P^{\leq}$ . The second part of Theorem 4 implies that  $(u, \beta)$  cannot be an optimal solution.

- Let now  $\gamma_a = \omega_a v$  with  $\omega_a > 0$  ( $a \in A$ ), and consider any solution  $(u, \beta)$ , for which  $Act(u, \beta)$  is fully contained in some hyperplane  $H(u', \beta')$ . Note that, since  $D(u, \beta) \neq \emptyset$ , there exists an  $a \in D(u, \beta) \subset Act(u, \beta) \subset H(u', \beta')$  for which, by non-degeneracy,  $d_{\gamma_a}(a, H(u, \beta)) = \Psi(u, \beta) > 0$ , so  $a \notin H(u, \beta)$ , and hence  $H(u', \beta') \neq H(u, \beta)$ . We may assume without loss of generality that  $v^\circ(u) = v^\circ(-u) = 1$ . Let then  $q \in \mathbb{R}^d$  be some vector with  $v(q) = 1$  and  $\langle u; q \rangle = 1$  (this always exists: take any  $q \in \partial v^\circ(u)$ ), and let us show that for every  $a \in D^{\geq}(u, \beta)$  the point  $p_a = a - \frac{\Psi(u, \beta)}{\omega_a} q$  is then a  $\gamma_a$ -projection of  $a$  on  $H(u, \beta)$ : First, we have  $\Psi(u, \beta) = d_{\gamma_a}(a, H(u, \beta)) = \frac{\langle u; a \rangle - \beta}{\gamma_a^\circ(-u)} = \frac{\langle u; a \rangle - \beta}{v^\circ(-u)/\omega_a} = \omega_a(\langle u; a \rangle - \beta)$ , so  $\langle u; p_a \rangle = \langle u; a \rangle - \frac{\Psi(u, \beta)}{\omega_a} = \beta$ , i.e.  $p_a \in H(u, \beta)$ . Secondly,  $\gamma_a(p_a - a) = \omega_a v(p_a - a) = v(\Psi(u, \beta)q) = \Psi(u, \beta) = d_{\gamma_a}(a, H(u, \beta))$ . Similarly, for every  $a \in D^{\leq}(u, \beta)$ , the point  $p_a = a + \frac{\Psi(u, \beta)}{\omega_a} q$  is a  $\gamma_a$ -projection of  $a$  on  $H(u, \beta)$ .

Consider now the sets  $P^{\geq}$  and  $P^{\leq}$  so defined, and assume  $\langle u'; q \rangle \geq 0$  (otherwise change the signs of  $u'$  and  $\beta'$ ). For each  $p_a \in P^{\geq}$  we have  $\langle u'; p_a \rangle = \langle u'; a \rangle - \frac{\Psi(u, \beta)}{\omega_a} \langle u'; q \rangle \leq \beta'$ , while for each  $p_a \in P^{\leq}$  we have  $\langle u'; p_a \rangle = \langle u'; a \rangle + \frac{\Psi(u, \beta)}{\omega_a} \langle u'; q \rangle \geq \beta'$ , or, in other words,  $H(u', \beta')$  separates  $P^{\geq}$  and  $P^{\leq}$ . Since also  $C(u, \beta) \subset Act(u, \beta) \subset H(u', \beta') \neq H(u, \beta)$ , Theorem 4 shows that  $(u, \beta)$  cannot be optimal.

- Finally, let  $A^{\geq} = A^{\leq} = \emptyset$  and all  $\gamma_a = \gamma$  ( $a \in A$ ), and consider any solution  $(u, \beta)$ , for which  $Act(u, \beta)$  is fully contained in some hyperplane  $H(u', \beta')$ . Note that as above  $H(u', \beta') \neq H(u, \beta)$ . Moreover, by Remark 3, both  $D^{\geq}(u, \beta)$  and  $D^{\leq}(u, \beta)$  are nonvoid, which implies that  $H(u', \beta')$  is not parallel to  $H(u, \beta)$ . Let  $q^{\geq} \in \mathbb{R}^d$  be some vector with  $\gamma(q^{\geq}) = 1$  and  $\langle -u; q^{\geq} \rangle = \gamma^\circ(-u)$ , and  $q^{\leq} \in \mathbb{R}^d$  be some vector with  $\gamma(q^{\leq}) = 1$  and  $\langle u; q^{\leq} \rangle = \gamma^\circ(u)$ . Similarly to above, for every  $a \in D^{\geq}(u, \beta)$  (resp.  $D^{\leq}(u, \beta)$ ) the point  $p_a = a + \Psi(u, \beta)q^{\geq}$  (resp.  $p_a = a + \Psi(u, \beta)q^{\leq}$ ) is then a  $\gamma$ -projection of  $a$  on  $H(u, \beta)$ .

Note that  $q^{\geq} \neq q^{\leq}$ , since otherwise  $\gamma^\circ(u) = \langle u; q^{\leq} \rangle = -\langle -u; q^{\geq} \rangle = -\gamma^\circ(-u)$  which would imply the ruled out  $u = 0$ . Therefore  $\frac{\Psi(u, \beta)}{2}(q^{\geq} - q^{\leq}) \stackrel{\text{def}}{=} \bar{q} \neq 0$ , and since the hyperplane  $H = H(u', \beta') + \frac{\Psi(u, \beta)}{2}(q^{\geq} + q^{\leq})$  satisfies  $P^{\geq} \subset H + \bar{q}$ , while  $P^{\leq} \subset H - \bar{q}$ , it strictly separates  $P^{\geq}$  and  $P^{\leq}$ . Theorem 4 then shows that  $(u, \beta)$  cannot be optimal.

□

For the case of possibly nonsmooth gauges  $\gamma$ , the following slightly weaker results may now be obtained:

**Theorem 6** For arbitrary gauges we have:

1. There always exists some  $(u^*, \beta^*)$ , optimal solution to (4), with at least  $d + 1$  active points.
2. Let  $v$  be a norm. If for all  $a \in A$ ,  $\gamma_a = \omega_a v$  for some positive  $\omega_a$ , then for this optimal solution  $Act(u^*, \beta^*)$  cannot be contained in a hyperplane.
3. If the problem is unconstrained and all gauges are equal to a fixed gauge  $\gamma$ , then for this optimal solution  $Act(u^*, \beta^*)$  cannot be contained in a hyperplane.

*Proof* 1. Take a sequence  $(c_k)_k$  of strictly positive scalars converging to zero; for each  $a \in A$ , define the gauge  $\gamma_{a,k}$  whose dual gauge is

$$\gamma_{a,k}^\circ(x) \stackrel{\text{def}}{=} \gamma_a^\circ(x) + c_k \|x\|,$$

where  $\|\cdot\|$  stands for the Euclidean norm. The gauges  $\gamma_{a,k}$  are smooth, since their dual gauge is strictly convex. For each  $k$ , Theorem 2 implies the existence of some  $(u_k^*, \beta_k^*)$  defining a center hyperplane when the gauges considered are  $\gamma_{a,k}$ . Moreover, Theorem 5 asserts that  $|Act(u_k^*, \beta_k^*)| \geq d + 1$ .

By homogeneity, we can assume without loss of generality that each  $u_k^*$  is taken with unit Euclidean norm. By Lemma 1 we also have  $\beta_k^* = \beta^*(u_k^*)$  and the sequence  $\{(u_k^*, \beta_k^*)\}_k$  is contained in a compact set. Thus it contains some subsequence converging to some  $(\bar{u}, \bar{\beta})$ . Out of this subsequence and due to the finiteness of  $A$ ,  $A^\geq$ ,  $A^\leq$ , we may select a subsequence for which the sets  $C(u_k^*, \beta_k^*)$ ,  $D^\geq(u_k^*, \beta_k^*)$ ,  $D^\leq(u_k^*, \beta_k^*)$ , are constant. For this subsequence, the union of these sets  $Act(u_k^*, \beta_k^*)$  is therefore also constant and, by convergence, is evidently contained in  $Act(\bar{u}, \bar{\beta})$ , which therefore contains at least  $d + 1$  elements. By construction,  $(\bar{u}, \bar{\beta})$  will be optimal for (4) when the gauges  $\gamma_a$  are considered.

2. For the second part, when each  $\gamma_a$  is proportional to the same norm  $v$ ,  $\gamma_a = \omega_a v$ , we may define the auxiliary norms  $\gamma_{a,k}$  by their dual

$$\gamma_{a,k}^\circ(x) \stackrel{\text{def}}{=} \gamma_a^\circ(x) + \frac{c_k}{\omega_a} \|x\| = \frac{1}{\omega_a} (v^\circ(x) + c_k \|x\|)$$

for which, for each fixed  $k$ , we have  $\gamma_{a,k} = \omega_a (v^\circ + c_k \|\cdot\|)^\circ$  so, for all  $a \in A$ ,  $\gamma_{a,k}$  is proportional to a same smooth norm. By the second part of Theorem 5, the limit solution  $(\bar{u}, \bar{\beta})$  constructed as above has an active point set  $Act(\bar{u}, \bar{\beta})$  which cannot be contained in a hyperplane.

3. Previous reasoning applies also in the unconstrained case with a single gauge  $\gamma$ , now using for each  $k$  the fixed auxiliary gauge  $\gamma_k$  with dual  $\gamma^\circ + c_k \|\cdot\|$ , and the third part of Theorem 5.

□

### 3.3 Algorithm in special case

Theorem 6 leads to a brute-force finite enumerative procedure to find an *unconstrained* center hyperplane provided all distances are measured with one common gauge  $\gamma$ ,

and the points in  $A$  are in *general position*, i.e., for any  $k = 1, \dots, d + 1$ , any subset of  $A$  with cardinality  $k$  has affine dimension  $k - 1$ . Indeed, by Theorem 6, we can restrict the search of an optimal solution to (4) to those  $(u, \beta)$  whose set  $D(u, \beta) = D^{\leq}(u, \beta) \cup D^{\geq}(u, \beta)$  of active distance points, as defined in (12)–(14) has at least  $d + 1$  active points. For any such  $(u^*, \beta^*)$ ,

$$\begin{aligned} d_{\gamma}(a, H(u, \beta)) &= \frac{\beta - \langle u ; a \rangle}{\gamma^{\circ}(u)} = z \quad \forall a \in D^{\leq}(u^*, \beta^*) \\ d_{\gamma}(a, H(u, \beta)) &= \frac{\langle u ; a \rangle - \beta}{\gamma^{\circ}(-u)} = z \quad \forall a \in D^{\geq}(u^*, \beta^*) \end{aligned} \tag{16}$$

where  $z$  denotes the optimal value of the center problem.

Choose arbitrary  $a_{\leq} \in D^{\leq}(u^*, \beta^*)$ ,  $a_{\geq} \in D^{\geq}(u^*, \beta^*)$ . This is possible, since, by Remark 3, both are non-empty. Hence (16) can also be written as

$$\begin{aligned} \frac{\beta - \langle u ; a \rangle}{\gamma^{\circ}(u)} &= \frac{\beta - \langle u ; a_{\leq} \rangle}{\gamma^{\circ}(u)} \quad \forall a \in D^{\leq}(u^*, \beta^*) \\ \frac{\langle u ; a \rangle - \beta}{\gamma^{\circ}(-u)} &= \frac{\langle u ; a_{\geq} \rangle - \beta}{\gamma^{\circ}(-u)} \quad \forall a \in D^{\geq}(u^*, \beta^*) \\ \frac{\beta - \langle u ; a_{\leq} \rangle}{\gamma^{\circ}(u)} &= \frac{\langle u ; a_{\geq} \rangle - \beta}{\gamma^{\circ}(-u)}, \end{aligned} \tag{17}$$

or equivalently as

$$\begin{aligned} \langle u ; a \rangle &= \langle u ; a_{\leq} \rangle \quad \forall a \in D^{\leq}(u^*, \beta^*) \\ \langle u ; a \rangle &= \langle u ; a_{\geq} \rangle \quad \forall a \in D^{\geq}(u^*, \beta^*) \\ \frac{\beta - \langle u ; a_{\leq} \rangle}{\gamma^{\circ}(u)} &= \frac{\langle u ; a_{\geq} \rangle - \beta}{\gamma^{\circ}(-u)}. \end{aligned} \tag{18}$$

Since we are assuming that the points in  $A$  are in general position, the homogeneous system of linear equations

$$\begin{aligned} \langle u ; a - a_{\leq} \rangle &= 0 \quad \forall a \in D^{\leq}(u^*, \beta^*) \\ \langle u ; a - a_{\geq} \rangle &= 0 \quad \forall a \in D^{\geq}(u^*, \beta^*) \end{aligned} \tag{19}$$

has as solution a line, which must thus be of the form  $\{\vartheta u^* : \vartheta \in \mathbb{R}\}$ . Hence, defining  $\beta(\cdot)$  as

$$\beta(\omega) = \frac{\gamma^{\circ}(\omega)}{\gamma^{\circ}(\omega) + \gamma^{\circ}(-\omega)} \langle \omega ; a_{\geq} \rangle + \frac{\gamma^{\circ}(-\omega)}{\gamma^{\circ}(\omega) + \gamma^{\circ}(-\omega)} \langle \omega ; a_{\leq} \rangle, \tag{20}$$

taking an arbitrary solution  $u \neq 0$  of (19), one has that either  $(u, \beta(u))$  or  $(-u, \beta(-u))$  is a positive multiple of  $(u^*, \beta^*)$ , and thus it defines the same hyperplane and half-spaces than  $(u^*, \beta^*)$ .

In other words, in order to find an optimal unconstrained center hyperplane it suffices to

1. construct all possible non-empty  $D^{\leq}, D^{\geq} \subset A$ , with  $D^{\leq} \cup D^{\geq}$  of cardinality  $d + 1$ ,
2. for any such  $D^{\leq}, D^{\geq} \subset A$ , choose an arbitrary  $a_{\leq} \in D^{\leq}, a_{\geq} \in D^{\geq}$  and construct a solution  $u$  of the homogeneous system of linear equations

$$\begin{aligned} \langle u ; a - a_{\leq} \rangle &= 0 & \forall a \in D^{\leq} \\ \langle u ; a - a_{\geq} \rangle &= 0 & \forall a \in D^{\geq} \end{aligned} \tag{21}$$

3. using (20) evaluate the objective function at candidates  $(u, \beta(u)), (-u, \beta(-u))$ .

This strategy was used to solve the instances given in Figs. 2–3.

### 3.4 Necessity of the assumptions

The following lemma allows to easily show non-optimality of certain configurations.

**Lemma 7** *Let  $z$  be a known feasible objective value, i.e.  $z = \Psi(u, \beta)$  for some feasible  $(u, \beta)$ . Consider two disjoint subsets  $D^{\geq}$  and  $D^{\leq}$  of  $A$ , not both empty, and a subset  $C$  of  $A^{\geq} \cup A^{\leq}$ .*

1. *If the two unions of  $\gamma_a$ -balls of radius  $z$ ,*

$$\bigcup_{a \in D^{\geq}} (a + zB_{\gamma_a}) \quad \text{and} \quad \bigcup_{a \in D^{\leq}} (a + zB_{\gamma_a}),$$

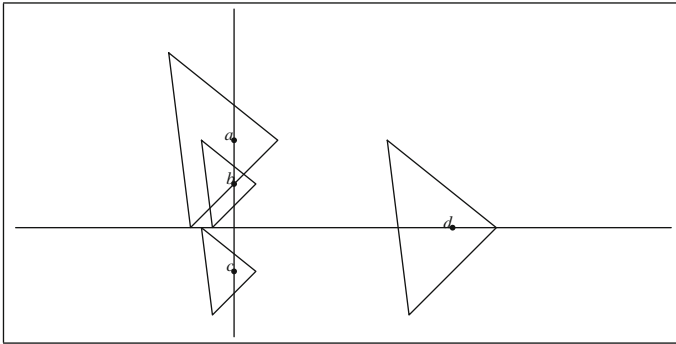
*can be strictly separated by some hyperplane containing  $C$ , then no optimal solution  $(u^*, \beta^*)$  can have  $D^{\geq}(u^*, \beta^*) \subset D^{\geq}, D^{\leq}(u^*, \beta^*) \subset D^{\leq}$  and  $C(u^*, \beta^*) \subset C$ .*

2. *Let  $\gamma_a$  be smooth for each  $a \in A$ . If the two unions of  $\gamma_a$ -balls of radius  $z$  can be separated by at least two different hyperplanes containing  $C$ , then no optimal solution  $(u^*, \beta^*)$  can have  $D^{\geq}(u^*, \beta^*) \subset D^{\geq}, D^{\leq}(u^*, \beta^*) \subset D^{\leq}$  and  $C(u^*, \beta^*) \subset C$ .*

*Proof* Consider any feasible solution  $(u, \beta)$  with  $D^{\geq}(u, \beta) \subset D^{\geq}, D^{\leq}(u, \beta) \subset D^{\leq}, C(u, \beta) \subset C$ , and  $\Psi(u, \beta) \leq z$  (if no such exists, the result is true), and we will show that  $(u, \beta)$  is not optimal. By assumption, for any  $a \in D(u, \beta)$  and any  $\gamma_a$ -projection  $p_a$  of  $a$  on  $H(u, \beta)$ , we have  $\gamma_a(p_a - a) = d_{\gamma_a}(a, H(u, \beta)) \leq z$ , so  $p_a \in a + zB_{\gamma_a}$ . Hence, for any choice of  $P^{\geq}$  and  $P^{\leq}$  we have

$$\begin{aligned} P^{\geq} &\subset \bigcup_{a \in D^{\geq}} (a + zB_{\gamma_a}) \\ P^{\leq} &\subset \bigcup_{a \in D^{\leq}} (a + zB_{\gamma_a}). \end{aligned}$$

Moreover, under the assumption of part 1, these sets can be strictly separated by a hyperplane containing  $C$ , which, by Theorem 4, proves that  $(u, \beta)$  is not optimal.



**Fig. 4** A counterexample in  $\mathbb{R}^2$  with a single weighted gauge

On the other hand, under the assumptions of part 2,  $P^\geq$  and  $P^\leq$  can be separated by two different hyperplanes containing  $C$ , which, again by Theorem 4, proves that  $(u, \beta)$  is not optimal.  $\square$

This lemma allows us to build examples showing that the assumptions yielding the full dimensionality of the set of active points in Theorem 6 cannot be relaxed to either a single, but unequally weighted, gauge, nor to mixed norms, even in  $\mathbb{R}^2$ .

*Example 8* Take as  $\gamma$  the triangular gauge in  $\mathbb{R}^2$  with extreme points of its unit ball  $(-1.5, 2)$ ,  $(-1, -2)$  and  $(1, 0)$ . Consider the set  $A = \{a, b, c, d\}$  where  $a=(0, 4)$ ,  $b=(0, 2)$ ,  $c=(0, -2)$  and  $d = (10, 0)$ , and respective weights  $\omega_a = \omega_d = 1$  and  $\omega_b = \omega_c = 2$ .

We want to solve the unconstrained minmax hyperplane problem with these data. Figure 4 shows the balls of weighted distance 2 around each point of  $A$ . It is evident that the line  $x_2 = 0$  satisfies the necessary conditions for optimality, with value  $z = 2$ . Here  $D^\geq(u, \beta) = \{a, b\}$  and  $D^\leq(u, \beta) = \{c\}$ , and they are all on a same line.

Let us show that no better solution can be found. Note that any better solution would have to be a line cutting all interiors of these balls.

We consider all possible choices for  $D^\geq(u, \beta)$  and  $D^\leq(u, \beta)$  in turn and rule each one out.

- $\{a\}, \{b, c\}$  :Impossible : to be at the same side and at equal distance to  $b$  and  $c$  the line should be vertical, so would not cut  $B_d$ .
- $\{b\}, \{a, c\}$  : Impossible : any line with  $a$  and  $c$  at the same side, has  $b$  also on this same side.
- $\{d\}, \{a, b\}$  : Impossible by Lemma 7: the vertical line  $x_1 = 5$  separates the balls strictly.
- $\{d\}, \{a, c\}$  idem.
- $\{d\}, \{b, c\}$  idem.
- $\{a\}, \{d, c\}$  : Impossible by Lemma 7: the line through  $(-1.75, 0)$  and  $(0, 1.75)$  separates the balls strictly.
- $\{c\}, \{d, b\}$  : Impossible by Lemma 7: the line through  $(-1.25, 0)$  and  $(0, -1.25)$  separates the balls strictly.
- $\{b\}, \{a, d\}$  : Impossible : the line should be parallel to  $ad$ , and intersect  $[a, b]$ , so cannot intersect  $B_c$ .

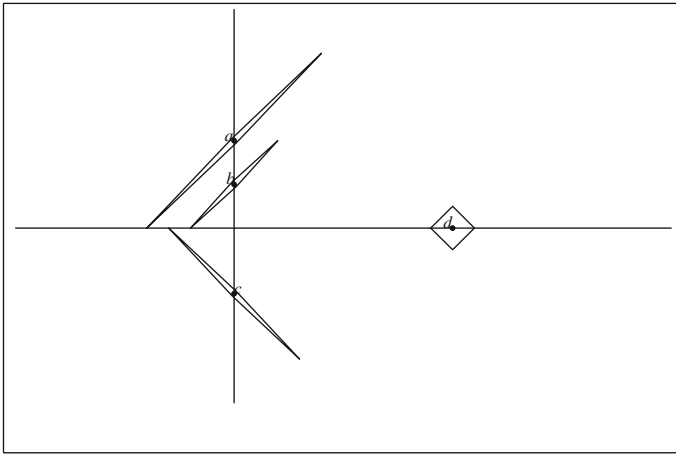


Fig. 5 A counterexample in  $\mathbb{R}^2$  with mixed norms

- $\{c\}, \{a, d\}$  : Impossible : the line should be parallel to  $ad$  and intersect  $B_c$ , but then it cannot intersect  $B_b$ .
- $\{a\}, \{b, d\}$  : Impossible : the line should intersect  $B_d$  and  $[a, b]$ , so cannot intersect  $B_c$ .
- $\{b\}, \{c, d\}$  : Impossible : any line at equal  $\gamma$ -distance smaller than 2, separating  $b$  and  $c$ , and intersecting  $B_d$ , cannot intersect  $B_a$ ; indeed, any such line would be increasing with slope less than 1, intersecting the first axis between point  $(-1.5, 0)$  and the origin.

Hence, part 2 of Theorem 6 does not hold if  $v$  is an arbitrary (non-symmetric) gauge □

*Example 9* Consider the set  $A = \{a, b, c, d\}$  where  $a = (0, 4), b = (0, 2), c = (0, -3)$  and  $d = (10, 0)$ . Each norm  $v_a, v_b$  and  $v_c$  has quadrangular unit ball with extreme points  $\pm\alpha(1, 1)$  and  $\pm\beta(1, -1)$ , with  $(\alpha, \beta) = (4, 0.1), (2, 0.1), (0.1, 3)$  respectively, whereas  $v_d = \ell_1$ .

We want to solve the unconstrained minmax hyperplane problem with these data. Figure 5 shows the balls  $B_s$  of  $v_s$ -distance 1 around each point  $s \in A$ . It is evident that the line  $x_2 = 0$  satisfies the necessary conditions for optimality, with value  $z = 1$ . Here  $D^{\geq}(u, \beta) = \{a, b\}$  and  $D^{\leq}(u, \beta) = \{c\}$ , and they are all on a same line.

Any better solution would have to be a line cutting all interiors of these balls. Let us show again that no better solution can be found, by considering all possible choices for  $D^{\geq}(u, \beta)$  and  $D^{\leq}(u, \beta)$  in turn and ruling each one out.

- $\{b\}, \{a, c\}$  : Impossible : any line with  $a$  and  $c$  at the same side, has  $b \in ]a, c[$  also on this same side.
- $\{d\}, \{a, b\}$  : Impossible by Lemma 7: the vertical line  $x_1 = 8$  separates the balls strictly.
- $\{d\}, \{a, c\}$  idem.
- $\{d\}, \{b, c\}$  idem.



- $\{a\}, \{b, c\}$  : Impossible by Lemma 7: the line through  $(-3.5, 0)$  and  $(0, 3.5)$  separates the balls strictly.
- $\{a\}, \{d, c\}$  : idem
- $\{a\}, \{b, d\}$  : idem
- $\{c\}, \{d, b\}$  : Impossible by Lemma 7: the line through  $(-2.5, 0)$  and  $(0, 2.5)$  separates the balls strictly.
- $\{b\}, \{a, d\}$  : Impossible : no line intersecting  $B_d$  and  $[a, b]$  can also intersect  $B_c$ .
- $\{c\}, \{a, d\}$  : Impossible : any line at equal  $\gamma_a$  and  $\gamma_c$ -distance smaller than 1, separating  $a$  and  $c$ , and intersecting  $B_d$ , cannot intersect  $B_b$ .
- $\{b\}, \{c, d\}$  : Impossible : no line intersecting  $B_b, B_c$  and  $B_d$  can also intersect  $B_a$ .

Hence, part 2 of Theorem 6 does not hold if  $\nu$  is an arbitrary (non-symmetric) gauge

□

*Remark 10* In both previous examples the center hyperplane is unique. In other words, any other hyperplane with at least  $d + 1$  active points yields a strictly larger objective value. Therefore, similarly to the approximation arguments used in the proof of Theorem 5 one may adapt these examples to involve a single weighted smooth gauge or mixed smooth norms sufficiently close to the polyhedral ones used above, so that the same choice of  $D^{\geq}$  and  $D^{\leq}$  yields the unique central hyperplane, thereby showing that the assumptions of Theorem 5 also cannot be relaxed to either a single weighted smooth gauge, or to mixed smooth norms, even in  $\mathbb{R}^2$ .

### 3.5 Connection with known results

When  $A^{\geq} = A^{\leq}$ , the constrained center hyperplane problem becomes an anchored hyperplane problem, as defined in [38], in which the sought hyperplane must pass through all points of  $\mathcal{P} = A^{\geq} = A^{\leq}$ . Setting  $k$  the affine dimension of  $\mathcal{P}$ , we may equivalently reduce  $\mathcal{P}$  to  $k$  affinely independent points. Since the hyperplane must contain these  $k$  points of  $\mathcal{P}$ , Theorem 6 states the existence of an anchored center hyperplane with at least  $d + 1 - k$  distance active points. When all gauges are smooth, by Theorem 5, this property holds for all anchored center hyperplanes. In the case of a single weighted norm, the second part of Theorem 6 shows that all  $d + 1$  active points may be affinely independent, generalising Theorems 2 and 3 of [38] which state affine independence of the distance active points for at least one, respectively all, anchored center hyperplanes. Note that the non-separability result of Theorem 4 states the even stronger property that the upper distance active points cannot be strictly separated (simply separated in the smooth case) from the lower active distance points by another anchored hyperplane containing  $\mathcal{P}$ .

The table on p182 of [36] states a number of conjectures concerning unconstrained center hyperplane problems to which our previous results provide answers.

- For (possibly mixed) gauges the first part of Theorem 6 confirms the conjectured property Cen1, i.e. existence of  $d + 1$  distance active points.
- For a single (weighted) gauge the same table conjectures a property called Cen1', asserting existence of  $d + 1$  affinely independent distance active points. For an

unweighted gauge the third part of Theorem 6 shows this is correct, and when this gauge is smooth Theorem 5 confirms the property for all (unweighted) center hyperplanes. However, for a weighted gauge affine independence cannot be guaranteed, even in the plane, as shown in Example 8.

- The corresponding negative conjecture about property Cen1' in case of mixed gauges turns out to be correct, and may even be strengthened to mixed norms, already in the plane, as shown in Example 9.

#### 4 Distance to linear separation

In Sect. 3 geometrical properties of center hyperplane problems have been derived. These geometrical properties have enabled us to give an answer to a number of conjectures posed in [36], as detailed in Sect. 3.5. Moreover, as stated in Sect. 3.3, the existence of finite dominating sets for the single-gauge case reduces the search of an unconstrained center hyperplane to inspecting a finite candidate list. Now we go one step further by showing how these results are also of direct use in a quite different problem of linear separation, as posed in what follows.

Consider two non-empty finite sets  $\mathcal{S}^{\geq}, \mathcal{S}^{\leq} \subset \mathbb{R}^d$ . Such sets are said to be linearly separable if  $(u, \beta) \in (\mathbb{R}^d \setminus \{0\}) \times \mathbb{R}$  exists such that

$$\begin{aligned} \mathcal{S}^{\geq} &\subset H(u, \beta)^{\geq} \\ \mathcal{S}^{\leq} &\subset H(u, \beta)^{\leq} \end{aligned} \quad (22)$$

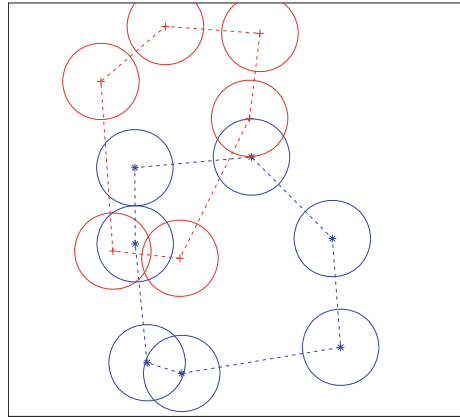
Procedures to check the feasibility of (22) and to construct particular feasible solutions have a very long tradition in the literature, e.g. [9, 21, 28, 35], and applications, among other fields, in Machine Learning.

When linear separation is not possible, i.e., when (22) is unfeasible, we may try to perturb the points in  $\mathcal{S}^{\geq}$  and  $\mathcal{S}^{\leq}$  so that the perturbed sets are linearly separable. The smallest perturbation needed to make the sets separable will be referred to as the *distance to linear separability*  $\tau(\mathcal{S}^{\geq}, \mathcal{S}^{\leq})$  of the sets  $\mathcal{S}^{\geq}, \mathcal{S}^{\leq}$ .

Let us formally introduce the concept of distance to linear separability. For each point  $a \in \mathcal{S}^{\geq} \cup \mathcal{S}^{\leq}$ , let  $B_a$  be a compact convex set in  $\mathbb{R}^d$  containing the origin in its interior, which will represent the perturbations (translations) of unit strength on point  $a$ . Define the gauge  $\gamma_a$  following (1), thus if point  $a$  is perturbed to  $a + z_a$ , we will say that the strength of the perturbation at  $a$  is  $\gamma_a(z_a)$ . For instance in Fig. 6 the points in  $\mathcal{S}^{\geq}, \mathcal{S}^{\leq}$ , represented respectively by crosses and stars, are the extreme points of two non-disjoint (thus not linearly separable) polytopes. The balls  $B_a$  around each point are derived from the Euclidean distances. It should be stressed that, although the very same gauge  $\gamma$  may be used for all data points, we also allow different points to have different associated gauges, reflecting possibly different types of sources of perturbations for each datapoint.

The distance to linear separability  $\tau(\mathcal{S}^{\geq}, \mathcal{S}^{\leq})$  of the nonseparable sets  $\mathcal{S}^{\geq}, \mathcal{S}^{\leq}$  is defined as the smallest perturbation strength required to make the sets separable. In

**Fig. 6** Perturbations with a common gauge (Euclidean)



other words,  $\tau(\mathcal{S}^{\geq}, \mathcal{S}^{\leq})$  is the optimal value of the optimization problem

$$\begin{aligned}
 & \min \lambda \\
 & \text{s.t. } \langle u ; a + z_a \rangle - \beta \geq 0 \quad \forall a \in \mathcal{S}^{\geq} \\
 & \quad \langle u ; a + z_a \rangle - \beta \leq 0 \quad \forall a \in \mathcal{S}^{\leq} \\
 & \quad \gamma_a(z_a) \leq \lambda \quad \forall a \in \mathcal{S}^{\geq} \cup \mathcal{S}^{\leq} \\
 & \quad u \neq 0 \\
 & \quad \beta, \lambda \in \mathbb{R}.
 \end{aligned} \tag{23}$$

Since, by assumption, the sets  $\mathcal{S}^{\geq}, \mathcal{S}^{\leq}$  are not linearly separable, the optimal value of (23) is strictly positive. By definition of dual gauge, we have for  $u, \lambda$  fixed, that

$$\begin{aligned}
 \max \{ \langle u ; z_a \rangle : \gamma_a(z_a) \leq \lambda \} &= \lambda \gamma_a^\circ(u) \\
 \min \{ \langle u ; z_a \rangle : \gamma_a(z_a) \leq \lambda \} &= -\lambda \gamma_a^\circ(-u),
 \end{aligned} \tag{24}$$

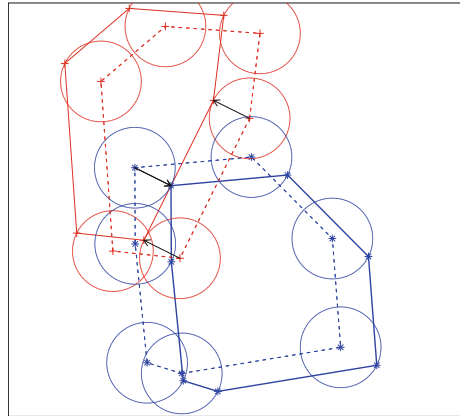
and the maximum (respectively the minimum) is attained at any  $z_a \in \lambda \partial \gamma_a^\circ(u)$  (respectively at any  $z_a \in \lambda \partial \gamma_a^\circ(-u)$ ). Hence, Problem (23) can be rephrased as

$$\begin{aligned}
 & \min \lambda \\
 & \text{s.t. } \langle u ; a \rangle + \lambda \gamma_a^\circ(u) - \beta \geq 0 \quad \forall a \in \mathcal{S}^{\geq} \\
 & \quad \langle u ; a \rangle - \lambda \gamma_a^\circ(-u) - \beta \leq 0 \quad \forall a \in \mathcal{S}^{\leq} \\
 & \quad u \neq 0 \\
 & \quad \beta, \lambda \in \mathbb{R},
 \end{aligned} \tag{25}$$

and, due to the positive homogeneity of the gauges, also as

$$\begin{aligned}
 & \min \max \left\{ \max_{a \in \mathcal{S}^{\geq}} \frac{\beta - \langle u ; a \rangle}{\gamma_a^\circ(u)}, \max_{a \in \mathcal{S}^{\leq}} \frac{\langle u ; a \rangle - \beta}{\gamma_a^\circ(-u)} \right\} \\
 & \text{s.t. } u \neq 0 \\
 & \quad \beta \in \mathbb{R}.
 \end{aligned} \tag{26}$$

**Fig. 7** Separable sets after minimal perturbation



*Remark 11* If all datapoints in  $S^{\geq}$  measure all distances with the same common gauge  $\gamma_{S^{\geq}}$ , then all problems in the first line of (24) attain their maxima at a common vector  $p_{S^{\geq}}$ . In other words, all points in  $S^{\geq}$  need a common perturbation vector  $p_{S^{\geq}}$ . Similarly, if all datapoints in  $S^{\leq}$  measure all distances with the same common gauge  $\gamma_{S^{\leq}}$ , a common perturbation  $p_{S^{\leq}}$  is obtained for all points in  $S^{\leq}$ , yielding the sets  $S^{\geq} + p_{S^{\geq}}$  and  $S^{\leq} + p_{S^{\leq}}$  linearly separable. If  $\gamma_{S^{\geq}} = \gamma_{S^{\leq}}$  and the common gauge is a norm, then  $p_{S^{\leq}} = -p_{S^{\geq}}$ , and thus we are moving  $S^{\geq}$  and  $S^{\leq}$  the same length in opposite directions. This is shown in Fig. 7, which represents the optimal perturbation for the sets in Fig. 6.

*Remark 12* We have from (26) that, for a given nonempty finite set  $S$ , the distance  $\tau(S, S)$  to linear separability of  $S$  with itself is given by the optimal value of

$$\begin{aligned} \min \max & \left\{ \max_{a \in S} \frac{\beta - \langle u, a \rangle}{\gamma_a^\circ(u)}, \max_{a \in S} \frac{\langle u, a \rangle - \beta}{\gamma_a^\circ(-u)} \right\} \\ \text{s.t. } & u \neq 0 \\ & \beta \in \mathbb{R}, \end{aligned} \tag{27}$$

which coincides with (4). In other words,  $\tau(S, S)$  is the optimal value of the center problem, and the  $(u, \beta)$  yielding  $\tau(S, S)$ , a center hyperplane.

Hence, the center problem can be seen as a particular case of the problem of finding the distance to linear separability.

This remark shows that distance to linear separability of two sets is an extension of the (unconstrained) center problem, this latter boils down to calculating the distance to linear separability of two copies of the given set. In fact, as shown below, the two problems are actually equivalent: distance to linear separability of two sets can be seen as a constrained center hyperplane problem. Hence, the geometrical properties of the approximation problem studied in previous sections are directly applicable to the separation problem in this section.

**Theorem 13** Let  $S^{\geq}$  and  $S^{\leq}$  be two finite subsets of  $\mathbb{R}^d$  which are nonseparable, then

1. There exists an optimal solution  $(u^*, \beta^*, \lambda^*)$  to Problem (25). Moreover, any such  $(u^*, \beta^*, \lambda^*)$  satisfies the following:

(a) The hyperplane  $H(u^*, \beta^*)$  is a constrained center hyperplane for the sets  $A^{\geq}(u^*, \beta^*), A^{\leq}(u^*, \beta^*), A(u^*, \beta^*)$  defined as

$$\begin{aligned} A^{\geq}(u^*, \beta^*) &\stackrel{\text{def}}{=} \{a \in \mathcal{S}^{\geq} \mid \langle u^* ; a \rangle > \beta^*\} \\ A^{\leq}(u^*, \beta^*) &\stackrel{\text{def}}{=} \{a \in \mathcal{S}^{\leq} \mid \langle u^* ; a \rangle < \beta^*\} \\ A(u^*, \beta^*) &\stackrel{\text{def}}{=} (\mathcal{S}^{\geq} \setminus A^{\geq}(u^*, \beta^*)) \cup (\mathcal{S}^{\leq} \setminus A^{\leq}(u^*, \beta^*)) \end{aligned} \tag{28}$$

(b) Define the sets

$$\begin{aligned} I^{\geq}(u^*, \beta^*, \lambda^*) &\stackrel{\text{def}}{=} \{a \in \mathcal{S}^{\geq} : \frac{\beta^* - \langle u^* ; a \rangle}{\gamma_a^{\circ}(u^*)} = \lambda^*\} \\ I^{\leq}(u^*, \beta^*, \lambda^*) &\stackrel{\text{def}}{=} \{a \in \mathcal{S}^{\leq} : \frac{\langle u^* ; a \rangle - \beta^*}{\gamma_a^{\circ}(-u^*)} = \lambda^*\}. \end{aligned} \tag{29}$$

Then,  $I^{\geq}(u^*, \beta^*, \lambda^*)$  and  $I^{\leq}(u^*, \beta^*, \lambda^*)$  are both nonempty, disjoint, and cannot be strictly separated by a hyperplane parallel to some  $P \in \bigcap_{a \in I^{\geq}(u^*, \beta^*, \lambda^*) \cup I^{\leq}(u^*, \beta^*, \lambda^*)} \partial \gamma_a^{\circ}(u^*)$ . Moreover, if all gauges  $\gamma_a$  are smooth, then separation by such a hyperplane is not possible.

2. There exists some  $(u^*, \beta^*, \lambda^*)$ , optimal for (25), such that  $I^{\geq}(u^*, \beta^*, \lambda^*) \cup I^{\leq}(u^*, \beta^*, \lambda^*)$  has cardinality at least  $d + 1$ . When all gauges  $\gamma_a$  are smooth, this property holds for any optimal solution.

*Proof* Existence of optimal solutions for Problem (26), and thus also to Problem (25), is easily shown along the same lines as in Theorem 2 for center hyperplanes. Let  $(u^*, \beta^*, \lambda^*)$  be any such optimum. Let us first show Part 1a. Given any  $(u, \beta)$  with  $u \neq 0, A^{\geq}(u^*, \beta^*) \subset H(u, \beta)^{\geq}$  and  $A^{\leq}(u^*, \beta^*) \subset H(u, \beta)^{\leq}$ , define the objective function of the corresponding center hyperplane problem (4) as

$$\Psi(u, \beta) = \max_{a \in A(u^*, \beta^*)} \max \left\{ \frac{\beta - \langle u ; a \rangle}{\gamma_a^{\circ}(u)}, \frac{\langle u ; a \rangle - \beta}{\gamma_a^{\circ}(-u)} \right\}. \tag{30}$$

First observe that  $\Psi(u, \beta) > 0$ , since otherwise  $H(u, \beta)$  would separate  $\mathcal{S}^{\geq}$  and  $\mathcal{S}^{\leq}$ , which was assumed not to be possible. One has

$$\begin{aligned} \Psi(u, \beta) &= \max \left\{ \max_{a \in A(u^*, \beta^*)} \frac{\beta - \langle u ; a \rangle}{\gamma_a^{\circ}(u)}, \max_{a \in A(u^*, \beta^*)} \frac{\langle u ; a \rangle - \beta}{\gamma_a^{\circ}(-u)} \right\} \\ &= \max \left\{ \max_{a \in \mathcal{S}^{\geq} \setminus A^{\geq}(u^*, \beta^*)} \frac{\beta - \langle u ; a \rangle}{\gamma_a^{\circ}(u)}, \max_{a \in \mathcal{S}^{\leq} \setminus A^{\leq}(u^*, \beta^*)} \frac{\beta - \langle u ; a \rangle}{\gamma_a^{\circ}(u)}, \right. \\ &\quad \left. \max_{a \in \mathcal{S}^{\geq} \setminus A^{\geq}(u^*, \beta^*)} \frac{\langle u ; a \rangle - \beta}{\gamma_a^{\circ}(-u)}, \max_{a \in \mathcal{S}^{\leq} \setminus A^{\leq}(u^*, \beta^*)} \frac{\langle u ; a \rangle - \beta}{\gamma_a^{\circ}(-u)} \right\} \end{aligned}$$

$$\begin{aligned} &\geq \max \left\{ \max_{a \in \mathcal{S}^{\geq} \setminus A^{\geq}(u^*, \beta^*)} \frac{\beta - \langle u; a \rangle}{\gamma_a^\circ(u)}, \max_{a \in \mathcal{S}^{\leq} \setminus A^{\leq}(u^*, \beta^*)} \frac{\langle u; a \rangle - \beta}{\gamma_a^\circ(-u)} \right\} \\ &= \max \left\{ \max_{a \in \mathcal{S}^{\geq}} \frac{\beta - \langle u; a \rangle}{\gamma_a^\circ(u)}, \max_{a \in \mathcal{S}^{\leq}} \frac{\langle u; a \rangle - \beta}{\gamma_a^\circ(-u)} \right\}, \end{aligned}$$

where the last equality follows from the fact that, by assumption,  $A^{\geq}(u^*, \beta^*) \subset H(u, \beta)^{\geq}$  and  $A^{\leq}(u^*, \beta^*) \subset H(u, \beta)^{\leq}$ . By definition of the sets  $A^{\geq}(u^*, \beta^*)$  and  $A^{\leq}(u^*, \beta^*)$ , one has that

$$\lambda^* = \Psi(u^*, \beta^*) = \max \left\{ \max_{a \in \mathcal{S}^{\geq}} \frac{\beta^* - \langle u^*; a \rangle}{\gamma_a^\circ(u^*)}, \max_{a \in \mathcal{S}^{\leq}} \frac{\langle u^*; a \rangle - \beta^*}{\gamma_a^\circ(-u^*)} \right\}. \tag{31}$$

Moreover, by assumption,  $(u^*, \beta^*, \lambda^*)$  is optimal for Problem (26). This implies for an arbitrary feasible  $(u, \beta)$  that

$$\begin{aligned} \Psi(u, \beta) &\geq \max \left\{ \max_{a \in \mathcal{S}^{\geq}} \frac{\beta - \langle u; a \rangle}{\gamma_a^\circ(u)}, \max_{a \in \mathcal{S}^{\leq}} \frac{\langle u; a \rangle - \beta}{\gamma_a^\circ(-u)} \right\} \\ &\geq \max \left\{ \max_{a \in \mathcal{S}^{\geq}} \frac{\beta^* - \langle u^*; a \rangle}{\gamma_a^\circ(u^*)}, \max_{a \in \mathcal{S}^{\leq}} \frac{\langle u^*; a \rangle - \beta^*}{\gamma_a^\circ(-u^*)} \right\} = \Psi(u^*, \beta^*), \end{aligned}$$

showing that  $H(u^*, \beta^*)$  is a constrained center hyperplane, as asserted.

To show Part 1b, observe that since, by assumption,  $\mathcal{S}^{\geq}$  and  $\mathcal{S}^{\leq}$  cannot be separated, we have that  $\Psi(u^*, \beta^*) > 0$ . Moreover, the set of active constraint points  $C(u^*, \beta^*)$ , defined as  $(A^{\geq}(u^*, \beta^*) \cup A^{\leq}(u^*, \beta^*)) \cap H(u^*, \beta^*)$ , is, by construction, empty. Hence, by Remark 3, the sets of active distance points  $D^{\geq}(u^*, \beta^*)$  and  $D^{\leq}(u^*, \beta^*)$  of the center hyperplane  $H(u^*, \beta^*)$  are nonempty and disjoint. But

$$\begin{aligned} D^{\geq}(u^*, \beta^*) &= \{a \in A(u^*, \beta^*) \cap H(u^*, \beta^*)^{\geq} : d_{\gamma_a}(a, H(u^*, \beta^*)) = \Psi(u^*, \beta^*)\} \\ &= \{a \in ((\mathcal{S}^{\geq} \setminus A^{\geq}(u^*, \beta^*)) \cup (\mathcal{S}^{\leq} \setminus A^{\leq}(u^*, \beta^*))) \cap H(u^*, \beta^*)^{\geq} : \\ &\quad d_{\gamma_a}(a, H(u^*, \beta^*)) = \Psi(u^*, \beta^*)\} \\ &= \{a \in (\mathcal{S}^{\leq} \cap H(u^*, \beta^*)^{\geq}) : d_{\gamma_a}(a, H(u^*, \beta^*)) = \Psi(u^*, \beta^*)\} \\ &= \{a \in \mathcal{S}^{\leq} : \frac{\langle u^*; a \rangle - \beta^*}{\gamma_a^\circ(-u^*)} = \Psi(u^*, \beta^*) = \lambda^*\} \\ &= I^{\leq}(u^*, \beta^*, \lambda^*), \end{aligned} \tag{32}$$

and similarly

$$D^{\leq}(u^*, \beta^*) = I^{\geq}(u^*, \beta^*, \lambda^*). \tag{33}$$

Hence,  $I^{\geq}(u^*, \beta^*, \lambda^*)$  and  $I^{\leq}(u^*, \beta^*, \lambda^*)$  are both nonempty and disjoint.

Assume  $p \in \bigcap_{a \in I^{\geq}(u^*, \beta^*, \lambda^*) \cup I^{\leq}(u^*, \beta^*, \lambda^*)} \partial \gamma_a^\circ(u^*)$  exists. We will show that no hyperplane parallel to  $p$  can strictly separate the sets  $I^{\geq}(u^*, \beta^*, \lambda^*)$  and  $I^{\leq}(u^*, \beta^*, \lambda^*)$ .

Define the sets

$$\begin{aligned}
 P^{\geq} &\stackrel{\text{def}}{=} I^{\geq}(u^*, \beta^*, \lambda^*) + \Psi(u^*, \beta^*)p \\
 P^{\leq} &\stackrel{\text{def}}{=} I^{\leq}(u^*, \beta^*, \lambda^*) - \Psi(u^*, \beta^*)p.
 \end{aligned}$$

It follows that  $P^{\geq}$  and  $P^{\leq}$  are projections on  $H(u^*, \beta^*)$  of the sets  $I^{\geq}(u^*, \beta^*, \lambda^*)$  and  $I^{\leq}(u^*, \beta^*, \lambda^*)$  respectively. Indeed, for any  $a \in I^{\geq}(u^*, \beta^*, \lambda^*)$ , one has that  $p \in \partial\gamma_a^{\circ}(u^*)$ , i.e.  $\langle p; u^* \rangle = \gamma_a^{\circ}(u^*)$  and  $(\gamma_a^{\circ})^{\circ}(p) = \gamma_a(p) = 1$ , so first

$$\langle u^*; a + \Psi(u^*, \beta^*)p \rangle = \langle u^*; a + \frac{\beta^* - \langle u^*; a \rangle}{\gamma_a^{\circ}(u^*)}p \rangle = \beta^*,$$

thus  $a + \Psi(u^*, \beta^*)p \in H(u^*, \beta^*)$ , and secondly  $\gamma_a(a + \Psi(u^*, \beta^*)p - a) = \Psi(u^*, \beta^*)\gamma_a(p) = \Psi(u^*, \beta^*) = d_{\gamma_a}(a, H(u^*, \beta^*))$  so  $a + \Psi(u^*, \beta^*)p$  is a  $\gamma_a$ -projection of  $a$  on  $H(u^*, \beta^*)$ . Hence,  $P^{\geq}$  is a projection of  $I^{\geq}(u^*, \beta^*, \lambda^*)$ , and, by a similar argument,  $P^{\leq}$  is a projection of  $I^{\leq}(u^*, \beta^*, \lambda^*)$  on  $H(u^*, \beta^*)$ .

Assuming then that  $I^{\geq}(u^*, \beta^*, \lambda^*)$  and  $I^{\leq}(u^*, \beta^*, \lambda^*)$  can be strictly separated by some  $H$  parallel to  $p$  would mean that  $H$  strictly separates the sets  $P^{\geq}$  and  $P^{\leq}$ , contradicting Theorem 4.

Assume now that each  $\gamma_a$  is smooth, and we will show that separation is not possible.  $H$  parallel to  $p \neq 0$ , would mean  $y = x + p \in H$  for any  $x \in H \cap H(u^*, \beta^*)$ . But, since  $p \in \partial\gamma_a^{\circ}(u^*)$  for all  $a$ , we have that

$$\begin{aligned}
 d_{\gamma_a}(y, H(u^*, \beta^*)) &= \max \left\{ \frac{\langle x + p; u^* \rangle - \beta^*}{\gamma_a^{\circ}(u^*)}, \frac{\beta^* - \langle x + p; u^* \rangle}{\gamma_a^{\circ}(-u^*)} \right\} \\
 &= \max \left\{ \frac{\langle u^*; p \rangle}{\gamma_a^{\circ}(u^*)}, \frac{\langle -u^*; p \rangle}{\gamma_a^{\circ}(-u^*)} \right\} > 0.
 \end{aligned}$$

Hence  $y \notin H(u^*, \beta^*)$ , and so  $H \neq H(u^*, \beta^*)$ .

If  $H$  separates  $I^{\geq}(u^*, \beta^*, \lambda^*)$  and  $I^{\leq}(u^*, \beta^*, \lambda^*)$ , it also separates  $I^{\geq}(u^*, \beta^*, \lambda^*) + \Psi(u^*, \beta^*)p$  and  $I^{\leq}(u^*, \beta^*, \lambda^*) - \Psi(u^*, \beta^*)p$ . Hence  $H$  is a hyperplane different from  $H(u^*, \beta^*)$  which separates  $P^{\geq}$  and  $P^{\leq}$ , which contradicts Theorem 4 when the gauges are smooth.

Now let us show Part 2, first assuming that all gauges  $\gamma_a$  are smooth. Let  $(u^*, \beta^*, \lambda^*)$  be an optimal solution to (25). Then,  $(u^*, \beta^*)$  is also an optimal solution to

$$\begin{aligned}
 \min \max_{a \in A} \max &\left\{ \frac{\langle u; a \rangle - \beta}{\gamma_a^{\circ}(u)}, \frac{\beta - \langle u; a \rangle}{\gamma_a^{\circ}(-u)} \right\} \\
 \text{s.t. } &\langle u; a \rangle - \beta \geq 0 \quad \forall a \in A^{\geq}(u^*, \beta^*) \\
 &\langle u; a \rangle - \beta \leq 0 \quad \forall a \in A^{\leq}(u^*, \beta^*).
 \end{aligned} \tag{34}$$

By Theorem 5,  $Act(u^*, \beta^*)$  contains at least  $d + 1$  points. Since, by construction, none of the constraints is active at  $(u^*, \beta^*)$ , we have that  $I^{\geq}(u^*, \beta^*, \lambda^*) \cup I^{\leq}(u^*, \beta^*, \lambda^*)$  contains at least  $d + 1$  points. This shows the result for smooth gauges.

For the general case, take, as in the proof of Theorem 6, a sequence of positive scalars  $c_k$  converging to zero, and the sequence of smooth gauges  $\gamma_{a,k} = (\gamma_a^{\circ} + c_k \|\cdot\|)^{\circ}$ .

Applying the result for the optimal solutions  $(u_k, \beta_k, \lambda_k)$  of the problems with smooth gauges, it follows that the sets  $I^{\geq}(u_k, \beta_k, \lambda_k) \cup I^{\leq}(u_k, \beta_k, \lambda_k)$  have cardinality at least  $d + 1$ .

Since such sets can take only a finite set of values, a subsequence, converging to some  $(u^*, \beta^*, \lambda^*)$  can be extracted with constant  $I^{\geq}(u_k, \beta_k, \lambda_k) \cup I^{\leq}(u_k, \beta_k, \lambda_k)$ . For such  $(u^*, \beta^*, \lambda^*)$ , the result follows.  $\square$

*Remark 14* Using the same arguments as in Sect. 3.3 for center hyperplanes, one can derive from Theorem 13 a procedure to calculate the distance to linear separation by complete enumeration of a finite candidate set when a common gauge  $\gamma$  is used to measure distances, since the list of candidate hyperplanes can be obtained by solving linear systems of equations.

Applying this strategy to the problem in Fig. 6 one obtains the optimal solution represented in Fig. 7.

## 5 Concluding remarks

In this paper we have obtained geometrical properties of the center hyperplanes when distances are measured by arbitrary norms, or, more generally, gauges. We have extended known results, and have given proofs or counterexamples to several open problems, stated as conjectures in [36].

In Sect. 4 the results have been extended to a separation problem. The results obtained for center hyperplanes immediately lead to similar results for separation.

For both types of problems these results lead to a brute force solution method by solving a finite set of candidate subproblems. How these properties may lead to efficient algorithms and quick heuristics in large-dimensional problems, as suggested in [34] for a related problem, deserves further attention.

## References

1. Bennett, K.P., Mangasarian, O.L.: Robust linear programming discrimination of two linearly inseparable sets. *Optim. Methods Softw.* **1**, 23–34 (1992)
2. Blanquero, R., Carrizosa, E., Hansen, P.: Locating objects in the plane using Global Optimization techniques. *Math. Oper. Res.* **34**, 837–858 (2009)
3. Brimberg, J., Juel, J., Schöbel, A.: Linear facility location in three dimensions—models and solution methods. *Oper. Res.* **50**, 1050–1057 (2002)
4. Brimberg, J., Juel, J., Schöbel, A.: Locating a minisum circle in the plane. *Discrete Appl. Math.* **157**, 901–912 (2009)
5. Burges, C.: A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **2**, 121–167 (1998)
6. Carrizosa, E., Plastria, F.: Optimal expected-distance separating halfspace. *Math. Oper. Res.* **33**, 662–677 (2008)
7. Cera, M., Mesa, J.A., Ortega, F.A., Plastria, F.: Locating a central hunter on the plane. *J. Optim. Theory Appl.* **136**, 155–166 (2008)
8. Chaudhuri, P.: On a geometric notion of quantiles for multivariate data. *J. Am. Stat. Assoc.* **91**, 862–872 (1996)
9. Cortes, C., Vapnik, V.: Support vector networks. *Mach. Learn.* **20**, 273–297 (1995)
10. Díaz-Báñez, J.M., López, M.A., Sellarès, J.A.: Locating an obnoxious plane. *Eur. J. Oper. Res.* **173**, 556–564 (2006)



11. Díaz-Báñez, J.M., Mesa, J.A., Schöbel, A.: Continuous location of dimensional structures. *Eur. J. Oper. Res.* **152**, 22–44 (2004)
12. Drezner, Z., Steiner, G., Wesolowsky, G.O.: On the circle closest to a set of points. *Comput. Oper. Res.* **29**, 637–650 (2002)
13. Durier, R., Michelot, C.: Geometrical properties of the Fermat–Weber problem. *Eur. J. Oper. Res.* **20**, 332–343 (1985)
14. Houle, M.E., Imai, H., Imai, K., Robert, J.M.: Weighted orthogonal  $L_\infty$ -approximation and applications. *Lect. Notes Comput. Sci.* **382**, 183–191 (1989)
15. Houle, M.E., Imai, H., Imai, K., Robert, J.M., Yamamoto, P.: Orthogonal weighted linear  $L_1$  and  $L_\infty$  approximation and applications. *Discrete Appl. Math.* **43**, 217–232 (1993)
16. Houle, M.E., Toussaint, G.: Computing the width of a set. In: *Proceedings of the 1st ACM Symposium of Computational Geometry*, pp. 1–7 (1985)
17. Imai, H., Lee, D.T., Yang, C.D.: 1-segment center problems. *ORSA J. Comput.* **4**, 426–434 (1992)
18. Korneenko, N.M., Martini, H.: Approximating finite weighted point sets by hyperplanes. *Lect. Notes Comput. Sci.* **447**, 276–286 (1990)
19. Korneenko, N.M., Martini, H.: Hyperplane approximation and related topics. In: Pach, J. (ed.) *New trends in discrete and computational geometry*, pp. 135–162. Springer, Berlin (1993)
20. Lee, D.T., Wu, Y.F.: Geometric complexity of some location problems. *Algorithmica* **1**, 193–211 (1986)
21. Mangasarian, O.L.: Linear and nonlinear separation of patterns by linear programming. *Oper. Res.* **13**, 444–452 (1965)
22. Mangasarian, O.L.: Arbitrary-norm separating plane. *Oper. Res. Lett.* **24**, 15–23 (1999)
23. Martini, H., Schöbel, A.: Two characterizations of smooth norms. *Geometriae Dedicata* **77**, 173–183 (1999)
24. Martini, H., Schöbel, A.: Median and center hyperplanes in Minkowski spaces—a unified approach. *Discrete Math.* **241**, 407–426 (2001)
25. Megiddo, N., Tamir, A.: On the complexity of locating linear facilities in the plane. *Oper. Res. Lett.* **1**, 194–197 (1982)
26. Michelot, C.: The mathematics of continuous location. *Stud. Locat. Anal.* **5**, 59–83 (1993)
27. Morris, J.G., Norback, J.P.: Linear facility location, Solving extensions of the basic problem. *Eur. J. Oper. Res.* **12**, 90–94 (1983)
28. Nilsson, N.: *Learning machines: foundations of trainable pattern classifying systems*. McGraw-Hill, New York (1965)
29. Norback, J.P., Morris, J.G.: Fitting hyperplanes by minimizing orthogonal deviations. *Math. Program.* **19**, 102–105 (1980)
30. Plastria, F.: On destination optimality in asymmetric distance Fermat–Weber problems. *Ann. Oper. Res.* **40**, 355–369 (1992)
31. Plastria, F.: Continuous location problems: research, results and questions. In: Drezner, Z. (ed.) *Facility location: a survey of applications and methods*, pp. 85–127. Springer, Berlin (1995)
32. Plastria, F., Carrizosa, E.: Gauge-distances and median hyperplanes. *J. Optim. Theory Appl.* **110**, 173–182 (2001)
33. Plastria, F., Carrizosa, E.: Optimal distance separating halfspace. Working Paper: Report BEIF/124, Vrije Universiteit Brussel, 2002. [http://www.optimization-online.org/DB\\_FILE/2004/10/970.pdf](http://www.optimization-online.org/DB_FILE/2004/10/970.pdf) (2002)
34. Plastria, F., De Bruyne, S., Carrizosa, E.: Alternating local search based VNS for linear classification. *Ann. Oper. Res.* **174**, 121–134 (2010)
35. Rosenblatt, F.: *Principles of neurodynamics: perceptrons and the theory of brain mechanisms*. Spartan Books, New York (1962)
36. Schöbel, A.: *Locating lines and hyperplanes*. Kluwer Academic Press, Dordrecht (1998)
37. Schöbel, A.: Locating least-distant lines in the plane. *Eur. J. Oper. Res.* **106**, 152–159 (1998)
38. Schöbel, A.: Anchored hyperplane location problems. *Discrete Computat. Geom.* **29**, 229–238 (2003)
39. Still, G., Streng, M.: The Chebyshev hyperplane optimization problem. *J. Global Optim.* **11**, 361–376 (1997)
40. Vapnik, V.N.: *Statistical learning theory*. Wiley, New York (1998)
41. Wesolowsky, G.O.: Location of the median line for weighted points. *Environ. Plan. A* **7**, 163–170 (1975)