Appeared in JOURNAL OF CONVEX ANALYSIS, Vol 21 (2014).

Conic separation of finite sets

I. The homogeneous case.

Annabella Astorino[1] - Manlio Gaudioso[2] - Alberto Seeger[3]

Abstract. This work addresses the issue of separating two finite sets in $\mathbb{R}^n$ by means of a suitable revolution cone

$$\Gamma(z, y, s) = \{x \in \mathbb{R}^n : s\,\|x - z\| - y^T(x - z) = 0\}.$$

The specific challenge at hand is to determine the aperture coefficient s, the axis y, and the apex z of the cone. These parameters have to be selected in such a way as to meet certain optimal separation criteria. Part I of this work focusses on the homogeneous case in which the apex of the revolution cone is the origin of the space. The homogeneous case deserves a separated treatment, not just because of its intrinsic interest, but also because it helps to built up the general theory. Part II of this work concerns the non-homogeneous case in which the apex of the cone can move in some admissible region. The non-homogeneous case is structurally more involved and leads to challenging nonconvex nonsmooth optimization problems.

Mathematics Subject Classification: 90C25, 90C26.
Key words: Conical separation, revolution cone, convex optimization, DC - optimization, proximal point techniques, classification.

1 Introduction

In ordinary parlance, distinguishing between two finite groups of elements sharing a common environment is usually called discrimination or separation. These words are employed also in a technical fashion in a number of areas of applied mathematics. There is a broad literature devoted to the problem of separating two finite sets $\mathcal{A} \subseteq \mathbb{R}^n$ and $\mathcal{B} \subseteq \mathbb{R}^n$ by means of a suitable hyperplane

$$\{x \in \mathbb{R}^n : y^T x - c = 0\}.$$

The affine separation problem has a long history that goes more than a century back. If a separating hyperplane does not exist, then one may resort to a nonlinear separating hypersurface. Spherical and ellipsoidal separating hypersurfaces have been studied in depth in the last decade [2, 4, 14, 16].

[1]Istituto di Calcolo e Reti ad Alte Prestazioni C.N.R., c/o Dipartimento di Elettronica Informatica e Sistemistica, Università delle Calabria, 87036 Rende, Italy (astorino@icar.cnr.it)
[2]Dipartimento di Elettronica Informatica e Sistemistica, Università delle Calabria, 87036 Rende, Italy (gaudioso@deis.unical.it)
[3]University of Avignon, Department of Mathematics, 33 rue Louis Pasteur, 84000 Avignon, France (alberto.seeger@univ-avignon.fr)

There is a half-century old work by Mangasarian [19] that briefly mentions the case of separation by means of a quadratic hypersurface of the form

$$\{x \in \mathbb{R}^n : x^T M x + y^T x - c = 0\}.$$

The theory of quadratic separation has been further developed by Falk and Karlov [10]. One can also consider polyhedral hypersurfaces [3, 12] or more sophisticated hypersurfaces as those mentioned in [18], but, of course, this burdens the analysis and the overall computational effort.

This paper addresses the issue of separating $\mathcal{A}$ and $\mathcal{B}$ by means of a revolution cone

$$\Gamma(z, y, s) = \{x \in \mathbb{R}^n : s\,\|x - z\| - y^T (x - z) = 0\}, \tag{1}$$

which is a manageable and easily visualizable mathematical object. Each parameter involved in the description of (1) has a clear geometric interpretation:

- The vector z, called the *apex* of the cone, is sought on some set $Z \subseteq \mathbb{R}^n$. The choice of Z depends on the concrete application under consideration.
- The vector y, called the *axis* of the cone, is sought on the unit sphere $\mathbb{S}_n$ of $\mathbb{R}^n$.
- The scalar $s \in [0, 1]$ is called *aperture coefficient*. Note that $\arccos s$ corresponds to the half-aperture angle of the cone.

We are using the expression “revolution cone” in a less restrictive manner than most authors: we are not requiring its convexity, nor having its apex at the origin. It helps to view (1) as the translation of a revolution cone with apex at the origin:

$$\Gamma(z, y, s) = z + \Gamma(0, y, s).$$

The convex envelope of $\Gamma(0, y, s)$ is a closed convex cone in the usual sense of convex analysis.

The exact formulation of the conic separation problem is described next. One is given a nonempty set $Z \subseteq \mathbb{R}^n$ and two mutually disjoint finite subsets of the Euclidean space $\mathbb{R}^n$, say

$$\mathcal{A} = \{a_1, \ldots, a_p\} \quad \text{and} \quad \mathcal{B} = \{b_1, \ldots, b_q\}.$$

For avoiding trivialities one assumes that $n \geq 2$, $\text{card}(\mathcal{A}) = p \geq 2$, and $\text{card}(\mathcal{B}) = q \geq 2$. The **C**onic **S**eparation problem reads as follows:

$$(\text{CS}) \quad \begin{cases} \text{Find a triplet } (z, y, s) \in Z \times \mathbb{S}_n \times [0, 1] \text{ such that} \\ \mathcal{A} \subseteq A(z, y, s) := \{x \in \mathbb{R}^n : f(x, z, y, s) \leq 0\}, \\ \mathcal{B} \subseteq B(z, y, s) := \{x \in \mathbb{R}^n : f(x, z, y, s) \geq 0\}, \end{cases}$$

where $f : \mathbb{R}^{3n+1} \to \mathbb{R}$ is the continuous function given by

$$f(x, z, y, s) = s\,\|x - z\| - y^T (x - z).$$

In other words, one must find a solution $(z, y, s) \in Z \times \mathbb{S}_n \times [0, 1]$ to the nonlinear inequality system

$$\begin{cases} f(a_i, z, y, s) \leq 0 & \text{for all } i \in \mathbb{N}_p := \{1, ..., p\}, \\ f(b_j, z, y, s) \geq 0 & \text{for all } j \in \mathbb{N}_q := \{1, ..., q\}. \end{cases} \tag{2}$$

The data sets $\mathcal{A}$ and $\mathcal{B}$ do not play a symmetric role. Indeed, the a_i's are to be captured by a convex set, whereas the b_j's are to be captured by a nonconvex set. Except for a recent paper by Kasimbeyli [17], separation through revolution cones has not been considered until today. The paper [17] takes place in an infinite dimensional normed space and it is so different in spirit from our work that is hard to extract a meaningful parallelism. Separation of two nonconvex sets by means of a general convex cone with apex at the origin is a theme that has been considered by recent scholarship; see, for instance, Henig [15] and Pappalardo and collaborators [9, 20, 23]. However, all these contributions differ substantially from the present work.

In general, one does not know in advance whether the CS problem is solvable or not. The data sets $\mathcal{A}$ and $\mathcal{B}$ are usually derived experimentally or through a certain random device.

Definition 1.1. *The pair* $(\mathcal{A}, \mathcal{B})$ *is* conically separable *if the solution set*

$$F(\mathcal{A}, \mathcal{B}) = \{(z, y, s) \in Z \times \mathbb{S}_n \times [0, 1] : (z, y, s) \text{ satisfies } (2)\}$$

is nonempty. If (z, y, s) *belongs to* $F(\mathcal{A}, \mathcal{B})$*, then* $\Gamma(z, y, s)$ *is called a* separator*. One adds the adjective* strict *if none of the inequalities in (2) is active.*

Affine separability, spherical separability, and quadratic separability are defined similarly by using the appropriate type of separation hypersurface. An hyperplane is a particular instance of a revolution cone. To see this, just take $s = 0$. Hence, affine separability implies conic separability. The converse statement is not true of course.

A situation of special interest occurs when $Z = \{0\}$, that is to say, when one forces the revolution cone to have its apex at the origin. This paper focusses on this particular situation and leaves to the companion paper [5] the analysis of the case in which the apex of the cone is allowed to move in a certain admissible region.

2 The homogeneous conic separation problem

As mentioned in the introductory section, this paper concentrates on the so-called **H**omogeneous **C**onic **S**eparation problem:

$$(\text{HCS}) \quad \begin{cases} \text{Find a pair } (y, s) \in \mathbb{S}_n \times [0, 1] \text{ such that} \\ \mathcal{A} \subseteq A_0(y, s) := A(0, y, s), \\ \mathcal{B} \subseteq B_0(y, s) := B(0, y, s). \end{cases}$$

So, the HCS problem is about finding a solution $(y, s) \in \mathbb{S}_n \times [0, 1]$ to the linear inequality system

$$\begin{cases} \|a_i\| s - a_i^T y \leq 0 & \text{for all } i \in \mathbb{N}_p \\ \|b_j\| s - b_j^T y \geq 0 & \text{for all } j \in \mathbb{N}_q. \end{cases} \tag{3}$$

If (y, s) is a solution to the HCS problem, then

$$\Gamma_0(y, s) := \Gamma(0, y, s) = \{x \in \mathbb{R}^n : s\,\|x\| - y^T x = 0\}$$

is called a *homogeneous separator*. One adds the adjective *strict* if none of the inequalities in (3) is active. The set

$$G(\mathcal{A}, \mathcal{B}) = \{(y, s) \in \mathbb{S}_n \times [0, 1] : (y, s) \text{ satisfies } (3)\}$$

is compact, but nonconvex in general.

Lemma 2.1. *If the total sample $\mathcal{A} \cup \mathcal{B}$ spans $\mathbb{R}^n$, then $G(\mathcal{A}, \mathcal{B})$ is path-connected.*

Proof. Let (y_0, s_0) and (y_1, s_1) be two different points in $G(\mathcal{A}, \mathcal{B})$. So, for each $k \in \{1, 2\}$, one has

$$\begin{cases} \|y_k\| = 1, \; s_k \geq 0, \\ \|a_i\| s_k \leq a_i^T y_k & \text{for all } i \in \mathbb{N}_p, \\ \|b_j\| s_k \geq b_j^T y_k & \text{for all } j \in \mathbb{N}_q. \end{cases}$$

We claim that y_0 and y_1 are not opposite vectors. Suppose to the contrary that $y_1 = -y_0$. In such a case, one gets

$$\begin{cases} 0 \leq \|a_i\| s_0 \leq a_i^T y_0 \\ 0 \leq \|a_i\| s_1 \leq a_i^T y_1 = -a_i^T y_0 \end{cases} \tag{4}$$

for all $i \in \mathbb{N}_p$, and

$$\begin{cases} \|b_j\| s_0 \geq b_j^T y_0 \\ \|b_j\| s_1 \geq b_j^T y_1 = -b_j^T y_0 \end{cases} \tag{5}$$

for all $j \in \mathbb{N}_q$. From (4) one sees that the a_i's are orthogonal to y_0 and that $s_0 = s_1 = 0$. Substituting the later information into (5) one deduces that also the b_j's are orthogonal to y_0. Hence, $\mathcal{A} \cup \mathcal{B}$ is contained in an homogeneous hyperplane, contradicting the assumption of the lemma. Since the claim is true, the line segment $[y_0, y_1]$ does not contains the vector 0, and, therefore, one can set

$$\mathbf{y}(t) = \frac{(1-t)y_0 + ty_1}{\|(1-t)y_0 + ty_1\|}, \quad \mathbf{s}(t) = \frac{(1-t)s_0 + ts_1}{\|(1-t)y_0 + ty_1\|}$$

for all $t \in [0, 1]$. One has constructed in this way a pair of continuous functions $\mathbf{y} : [0, 1] \to \mathbb{R}^n$ and $\mathbf{s} : [0, 1] \to \mathbb{R}$ such that

$$\begin{cases} (\mathbf{y}(0), \mathbf{s}(0)) = (y_0, s_0), \\ (\mathbf{y}(1), \mathbf{s}(1)) = (y_1, s_1), \\ (\mathbf{y}(t), \mathbf{s}(t)) \in G(\mathcal{A}, \mathcal{B}) \;\text{ for all } t \in [0, 1]. \end{cases} \tag{6}$$

This proves the path-connectedness of $G(\mathcal{A}, \mathcal{B})$. □

Remark 2.2. While dealing with the HCS problem, there is no loss of generality in assuming that

$$L := \operatorname{span}\{a_1, \ldots, a_p, b_1, \ldots, b_q\}$$

is equal to the whole space $\mathbb{R}^n$. Otherwise, one views $\mathcal{A}$ and $\mathcal{B}$ as subsets of the linear space L, and, accordingly, one tries to separate them by means of an homogeneous revolution cone in L.

Theorem 2.3. *Suppose that the a_i's and the b_j's are nonzero vectors. Then*

(a) A strict homogeneous separator exists if and only if the polytopes

$$\Xi_{\mathcal{A}} = \operatorname{co}\left\{\frac{a_1}{\|a_1\|}, \ldots, \frac{a_p}{\|a_p\|}\right\} \quad \text{and} \quad \Xi_{\mathcal{B}}^+ = \operatorname{co}\left\{0, \frac{b_1}{\|b_1\|}, \ldots, \frac{b_q}{\|b_q\|}\right\}$$

do not intersect.

(b) An homogeneous separator exists if and only if 0 does not belong to the interior of the Minkowski difference

$$\Xi_{\mathcal{A}} - \Xi_{\mathcal{B}}^+ = \left\{u - v : u \in \Xi_{\mathcal{A}}, v \in \Xi_{\mathcal{B}}^+\right\}. \tag{7}$$

Proof. We start by introducing the functions $\Psi_{\mathcal{A}} : \mathbb{R}^n \to \mathbb{R}$ and $\Phi_{\mathcal{B}} : \mathbb{R}^n \to \mathbb{R}$ given by

$$\Psi_{\mathcal{A}}(y) = \min_{1 \leq i \leq p} \frac{a_i^T y}{\|a_i\|} \qquad \text{and} \qquad \Phi_{\mathcal{B}}(y) = \max_{1 \leq j \leq q} \frac{b_j^T y}{\|b_j\|} \,.$$

We introduce also the nonnegative function $\Phi_{\mathcal{B}}^+ : \mathbb{R}^n \to \mathbb{R}$ given by

$$\Phi_{\mathcal{B}}^+(y) = \max\{0, \Phi_{\mathcal{B}}(y)\}.$$

A strict homogeneous separator exists if and only if the system

$$\begin{cases} \|a_i\| s - a_i^T y < 0 & \text{for all } i \in \mathbb{N}_p \\ \|b_j\| s - b_j^T y > 0 & \text{for all } j \in \mathbb{N}_q \end{cases}$$

holds for some $(y, s) \in \mathbb{S}_n \times [0, 1]$, that is to say,

$$\exists \, (y, s) \in \mathbb{S}_n \times [0, 1] \ \text{ s.t. } \ \Phi_{\mathcal{B}}(y) < s < \Psi_{\mathcal{A}}(y).$$

The above line can be rewritten as

$$\exists \, y \in \mathbb{S}_n \ \text{ s.t. } \ (\Psi_{\mathcal{A}} - \Phi_{\mathcal{B}}^+)(y) > 0. \tag{8}$$

But

$$(\Psi_{\mathcal{A}} - \Phi_{\mathcal{B}}^+)(y) = \min_{x \in W} x^T y \,,$$

where W is the convex compact set introduced in (7). Hence, (8) amounts to saying that $0 \notin W$, i.e., $\Xi_{\mathcal{A}}$ and $\Xi_{\mathcal{B}}^+$ do not intersect. Similarly, the existence of an homogeneous separator is equivalent to the condition

$$\exists \, y \in \mathbb{S}_n \ \text{ s.t. } \ (\Psi_{\mathcal{A}} - \Phi_{\mathcal{B}}^+)(y) \geq 0,$$

which in turn is equivalent to saying that 0 does not belong to the interior of W. □

Remark 2.4. Many authors have developed algorithms to compute the gap

$$\text{gap}(P_1, P_2) = \min\{\|u - v\| : u \in P_1, v \in P_2\}$$

between two arbitrary polytopes P_1 and P_2. As a by-product, these algorithms check whether the polytopes intersect or not. A detailed treatment of this topic can be found, for instance, in [13].

Remark 2.5. As a direct consequence of Theorem 2.3 one sees that the existence of a strict homogeneous separator implies the pointedness of the polyhedral cone

$$\text{cone}(\mathcal{A}) = \left\{ \sum_{i=1}^{p} \alpha_i a_i : \alpha_1 \geq 0, \ldots, \alpha_p \geq 0 \right\}$$

generated by the a_i's. Such observation is of course consistent with geometric intuition.

We next identify two special solutions to the HCS problem. The definition below focusses the attention on the angular size of the separator, but other selection criteria are also possible.

Definition 2.6. *If* (y_*, s_*) *solves the minimization problem*

$$\begin{cases} \text{minimize } s \\ (y,s) \in G(\mathcal{A},\mathcal{B}), \end{cases} \tag{9}$$

respectively the maximization problem

$$\begin{cases} \text{maximize } s \\ (y,s) \in G(\mathcal{A},\mathcal{B}), \end{cases} \tag{10}$$

then $\Gamma_0(y_*, s_*)$ *is called a* largest angle homogeneous separator*, respectively a* smallest angle homogeneous separator.

The above terminology has a clear geometric justification: to minimize (respectively, maximize) the coefficient s amounts to render the half-aperture angle of the homogeneous separator as large (respectively, small) as possible. A largest angle homogeneous separator, say $\Gamma_0(y_*, s_*)$, is better than an arbitrary homogeneous separator $\Gamma_0(y, s)$ in the following sense: if one adds to $\mathcal{A}$ a random point a_{p+1} with spherically symmetric probability distribution, then $\Gamma_0(y_*, s_*)$ has higher chances than $\Gamma_0(y, s)$ to remain a separator for the new pair $(\mathcal{A} \cup \{a_{p+1}\}, \mathcal{B})$. In other words, $\Gamma_0(y_*, s_*)$ is more stable than $\Gamma_0(y, s)$ with respect to a certain type of perturbation in $\mathcal{A}$. Similarly, a smallest angle homogeneous separator of $(\mathcal{A}, \mathcal{B})$ has better chances to remain a separator for a perturbed pair of the type $(\mathcal{A}, \mathcal{B} \cup \{b_{q+1}\})$.

Proposition 2.7. *Suppose that* $\mathcal{A} \cup \mathcal{B}$ *spans* $\mathbb{R}^n$ *and that* $G(\mathcal{A},\mathcal{B})$ *is nonempty. Then for any aperture coefficient* $\bar{s}$ *in the interval*

$$[s_{\min}, s_{\max}] := \left[\min_{(y,s)\in G(\mathcal{A},\mathcal{B})} s \, , \max_{(y,s)\in G(\mathcal{A},\mathcal{B})} s \right]$$

there exists an axis $\bar{y} \in \mathbb{S}_n$ *such that* $\Gamma_0(\bar{y}, \bar{s})$ *is an homogeneous separator.*

Proof. Since $G(\mathcal{A},\mathcal{B})$ is nonempty, the extremal problems (9) and (10) are both solvable. Let (y_0, s_0) be a solution to (9) and (y_1, s_1) be a solution to (10). Obviously, $s_0 = s_{\min}$ and $s_1 = s_{\max}$. Since $G(\mathcal{A},\mathcal{B})$ is path-connected by Lemma 2.1, there are continuous functions $\mathbf{y} : [0,1] \to \mathbb{R}^n$ and $\mathbf{s} : [0,1] \to \mathbb{R}$ satisfying (6). The intermediate value theorem applied to the function $\mathbf{s}(\cdot)$ ensures the existence of $\bar{t} \in [0,1]$ such that $\mathbf{s}(\bar{t}) = \bar{s}$. For completing the proof it suffices to take $\bar{y} = \mathbf{y}(\bar{t})$. □

As mentioned a few lines above, if $G(\mathcal{A},\mathcal{B})$ is nonempty, then (9) and (10) are both solvable. However, these optimization problems may have more than one solution.

Example 2.8. For the data sets

$$\mathcal{A} = \left\{ \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \end{bmatrix} \right\}, \quad \mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

one gets two largest angle homogeneous separators, namely

$$\left\{ x \in \mathbb{R}^2 : \left[x_1^2 + x_2^2 \right]^{1/2} - x_1 - x_2 = 0 \right\},$$
$$\left\{ x \in \mathbb{R}^2 : \left[x_1^2 + x_2^2 \right]^{1/2} - x_1 + x_2 = 0 \right\}.$$

These separators have of course the same half-aperture angle, but not the same axis.

Example 2.9. For the data sets

$$\mathcal{A} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}, \quad \mathcal{B} = \left\{ \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix} \right\}$$

one gets countable many smallest angle homogeneous separators, namely

$$\left\{ x \in \mathbb{R}^3 : (\cos t)x_1 + 0x_2 + (\sin t)x_3 = 0 \right\}$$

with $t \in [0, \pi/2]$. These hyperplanes are also largest angle homogeneous separators.

In general, a smallest angle homogeneous separator cannot be strict because it contains at least one the a_i's. By contrast, a largest angle homogeneous separator can be strict or not.

3 Computing a smallest angle homogeneous separator

If there is an homogeneous separator with positive aperture coefficient, then the smallest angle homogeneous separator is unique and can be found by solving a convex optimization problem. This is, in essence, what the next theorem asserts.

Theorem 3.1. *Suppose that $(\mathcal{A}, \mathcal{B})$ admits an homogeneous separator with positive aperture coefficient. Then*

(a) $(\mathcal{A}, \mathcal{B})$ has exactly one smallest angle homogeneous separator.

(b) Furthermore, the pair (y_*, s_*) that describes the smallest angle homogeneous separator is equal to the unique solution to the convex optimization problem

$$\left\{ \begin{array}{l} \textit{maximize } s \\ (y, s) \in \mathbb{R}^n \times \mathbb{R} \\ \|y\| \leq 1 \\ -a_i^T y + \|a_i\| s \leq 0 \quad \textit{for all } i \in \mathbb{N}_p \\ b_j^T y - \|b_j\| s \leq 0 \quad \textit{for all } j \in \mathbb{N}_q. \end{array} \right. \tag{11}$$

Proof. Written in full extent, the variational problem (10) reads

$$\left\{ \begin{array}{l} \text{maximize } s \\ (y, s) \in \mathbb{R}^n \times \mathbb{R} \\ \|y\| = 1 \\ 0 \leq s \leq 1 \\ -a_i^T y + \|a_i\| s \leq 0 \quad \text{for all } i \in \mathbb{N}_p \\ b_j^T y - \|b_j\| s \leq 0 \quad \text{for all } j \in \mathbb{N}_q. \end{array} \right. \tag{12}$$

Let $G_\diamond(\mathcal{A}, \mathcal{B})$ and $G_*(\mathcal{A}, \mathcal{B})$ be the solution sets to (11) and (12), respectively. We claim that

$$G_\diamond(\mathcal{A}, \mathcal{B}) \subseteq G_*(\mathcal{A}, \mathcal{B}). \tag{13}$$

We assume without loss of generality that the a_i's are nonzero vectors. One can easily check that $G_\diamond(\mathcal{A}, \mathcal{B})$ is nonempty, compact, and convex. Pick any $(y_\diamond, s_\diamond) \in G_\diamond(\mathcal{A}, \mathcal{B})$. In particular, $s_\diamond$ is the optimal value of (11). Since $(\mathcal{A},\mathcal{B})$ admits an homogeneous separator with positive aperture coefficient and the feasible set of (11) contains the feasible set of (12), one has

$$0 < s_* \le s_\diamond, \tag{14}$$

where s_* denotes the optimal value of (12). Since

$$\|a_1\| s_\diamond \le a_1^T y_\diamond \le \|a_1\|\, \|y_\diamond\| \le \|a_1\|,$$

one has $s_\diamond \le 1$. It is also clear that $y_\diamond \neq 0$. If $\|y_\diamond\| < 1$, then $\left(\|y_\diamond\|^{-1} y_\diamond, \|y_\diamond\|^{-1} s_\diamond\right)$ is a feasible pair for (11) with $\|y_\diamond\|^{-1} s_\diamond > s_\diamond$, contradicting the maximality of $s_\diamond$. Hence, $\|y_\diamond\| = 1$. We have shown in this way that $(y_\diamond, s_\diamond)$ is feasible for (12). In particular, $s_\diamond \le s_*$. In view of (14), one deduces that $s_\diamond = s_*$ and that $(y_\diamond, s_\diamond)$ is an optimal solution to (12). This completes the proof of (13). The reverse inclusion to (13) being obvious, one gets finally $G_\diamond(\mathcal{A},\mathcal{B}) = G_*(\mathcal{A},\mathcal{B})$. This common set must be a singleton. Indeed, if (y_1, s_1) and (y_2, s_2) were two distinct elements in $G_\diamond(\mathcal{A},\mathcal{B})$, then the midpoint $(1/2)(y_1, s_1) + (1/2)(y_2, s_2)$ would also be in $G_\diamond(\mathcal{A},\mathcal{B})$. In particular, one would have

$$\|(1/2)(y_1 + y_2)\| = \|y_1\| = \|y_2\| = 1,$$

a clear impossibility. □

Theorem 3.1 settles the question concerning the uniqueness and the practical computation of the smallest angle homogeneous separator. The numerical resolution of the convex optimization problem (11) offers no difficulty.

3.1 Dual interpretation of the smallest angle homogeneous separator

By applying the Fenchel-Rockafellar duality theorem to the convex optimization problem (11) one obtains a dual characterization for the smallest angle homogeneous separator. What the next theorem says is that (11) is nothing but a least norm problem over the convex polyhedron

$$C = \{u \in \mathbb{R}^n : (u, -1) \in Q\},$$

where Q is the polyhedral convex cone generated by the vectors

$$(a_1, -\|a_1\|), \ldots, (a_p, -\|a_p\|), (-b_1, \|b_1\|), \ldots, (-b_q, \|b_q\|). \tag{15}$$

Theorem 3.2. *Suppose that $(\mathcal{A},\mathcal{B})$ admits an homogeneous separator with positive aperture coefficient. Let $\Gamma(y_*, s_*)$ be the smallest angle homogeneous separator. Then*

$$s_* = \min_{u \in C} \|u\| \tag{16}$$

and $y_* = \|u_*\|^{-1} u_*$ with u_* denoting the unique solution to (16).

Proof. By Theorem 3.1 one knows that

$$s_* = \max_{(y,s) \in P} -c(y, s), \tag{17}$$

where P is the polyhedral convex cone in $\mathbb{R}^n \times \mathbb{R}$ given by (3) and $c : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is the lower-semicontinuous convex function given by

$$c(y, s) = \begin{cases} -s & \text{if } \|y\| \le 1 \\ \infty & \text{if } \|y\| > 1. \end{cases}$$

The effective domain of c is the set $\mathrm{dom}(c) = \{(y,s) \in \mathbb{R}^n \times \mathbb{R} : \|y\| \leq 1\}$. The intersection

$$P \cap \mathrm{int}[\mathrm{dom}(c)] = \{(y,s) \in P : \|y\| < 1\}$$

is nonempty because it contains the point $(y,s) = (0,0)$. Under such a constraint qualification condition one can apply [24, Theorem 31.4] and write

$$s_* = \min_{(u,r)\in Q} d(u,r), \tag{18}$$

where Q is the dual cone of P, i.e.,

$$Q = \{(u,r) \in \mathbb{R}^n \times \mathbb{R} : u^T y + rs \geq 0 \text{ for all } (y,s) \in P\},$$

and $d : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is the Legendre-Fenchel conjugate of c, i.e.,

$$d(u,r) = \sup_{(y,s)\in\mathbb{R}^n\times\mathbb{R}} \{u^T y + rs - c(y,s)\}.$$

Clearly, Q is the polyhedral convex cone generated by the vectors in (15) and

$$d(u,r) = \begin{cases} \|u\| & \text{if } r = -1 \\ \infty & \text{if } r \neq -1. \end{cases}$$

By substituting this information into (18) one gets

$$s_* = \min\{\|u\| : u \in \mathbb{R}^n \text{ s.t. } (u,-1) \in Q\},$$

proving in this way (16). Now, let u_* be the unique solution to (16) and let $r_* = -1$. Since (y_*, s_*) solves the primal problem (17) and (u_*, r_*) solves the dual problem (18), the orthogonality condition

$$u_*^T y_* + r_* s_* = 0$$

is in force (cf. [24, Theorem 31.4]). One gets in this way

$$u_*^T y_* = s_* = \|u_*\|. \tag{19}$$

But, on the other hand, one knows that

$$\|y_*\| = 1, \quad s_* > 0. \tag{20}$$

The combination of (19) and (20) yields $y_* = \|u_*\|^{-1} u_*$. □

We now briefly explain how to solve in practice the least norm problem (16). Observe that $u \in C$ if and only if there exist nonnegative scalars $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$ such that

$$(u,-1) = \sum_{i=1}^{p} \alpha_i \,(a_i, -\|a_i\|) + \sum_{j=1}^{q} \beta_j \,(-b_j, \|b_j\|).$$

Hence, finding the least norm element of C amounts to solve

$$\begin{cases} \text{minimize} \left\| \sum_{i=1}^{p} \alpha_i a_i - \sum_{j=1}^{q} \beta_j b_j \right\|^2 \\ \alpha_1 \geq 0, \ldots, \alpha_p \geq 0 \\ \beta_1 \geq 0, \ldots, \beta_q \geq 0 \\ \sum_{i=1}^{p} \|a_i\| \,\alpha_i - \sum_{j=1}^{q} \|b_j\| \,\beta_j \; = \; 1. \end{cases} \tag{21}$$

In others words, one must minimize a positive semidefinite quadratic form on the nonnegative orthant of $\mathbb{R}^{p+q}$ intersected with a non-homogenous hyperplane. The cost function of (21) is in fact the quadratic form associated to the Gramian matrix of the vectors $\{a_1, \ldots, a_p, -b_1, \ldots, -b_q\}$. The convex optimization problem (21) is quite simple, but one must be aware that p and q are usually large integers, and therefore the minimization process takes place in a high dimensional space. In order to reduce the dimension of the underlying space it is convenient to drop the data points in $\mathcal{A} \cup \mathcal{B}$ that are redundant. There are special techniques for detecting redundancy, but we shall not indulge on this matter.

4 Computing a largest angle homogeneous separator

Computing a largest angle homogeneous separator is a more difficult task because one has to solve a nonconvex optimization problem. The variational problem (9) reads as follows:

$$\begin{cases} \text{minimize } s \\ (y,s) \in \mathbb{R}^n \times \mathbb{R} \\ \|y\| = 1,\ 0 \le s \le 1 \\ -a_i^T y + \|a_i\| s \le 0 \ \text{ for all } i \in \mathbb{N}_p \\ \ b_j^T y - \|b_j\| s \ \le 0 \quad \text{for all } j \in \mathbb{N}_q. \end{cases} \tag{22}$$

One assumes without loss of generality that the a_i's and the b_j's are nonzero vectors. We keep the inequality $s \le 1$, which is implicit in the constraint involving the a_i's, because it is needed in a relaxed version of (22) that we are going to examine.

If one changes the normalization condition $\|y\| = 1$ by the convex constraint $\|y\| \le 1$, then one ends up with a convex optimization problem. Unfortunately, such convexification mechanism is here inappropriate because the optimal value of the convexified problem is always equal to 0 and the feasible point $(0,0)$ is optimal.

The problem (22) has a nonempty feasible region exactly when $(\mathcal{A}, \mathcal{B})$ is homogeneously conically separable, but this is not always the case in practice. Thus, in view to apply homogeneous conical separability to classification problems, it is useful to shift the attention to a "relaxed" optimization problem which is always feasible and such that an optimal solution provides

- either a good quality (large half-aperture angle) separator in case the pair $(\mathcal{A}, \mathcal{B})$ is homogeneously conically separable,

- or a large-angle-nearly-separating cone otherwise.

Following a rather standard way of reasoning in classification theory, such a goal can be achieved by introducing an appropriate classification error function and by constructing an optimization problem whose objective function accounts for both maximization of the half-aperture angle and minimization of the classification error. Thus one resorts to the formulation

$$h_* = \min \{h(y,s) : (y,s) \in \Omega,\ \|y\| = 1\}, \tag{23}$$

where $\Omega = \mathbb{R}^n \times [0,1]$ and

$$h(y,s) \ = \ \gamma s + \sum_{i=1}^{p} \max\{0, -a_i^T y + \|a_i\| s\} + \sum_{j=1}^{q} \max\{0, b_j^T y - \|b_j\| s\}.$$

Here $\gamma > 0$ is a parameter expressing the tradeoff between the two objectives previously mentioned.

Proposition 4.1. *The optimal value of (23) can be rewritten as*

$$h_* = \min\{h(y,s) : (y,s) \in \Omega,\ \|y\| \geq 1\}. \tag{24}$$

Furthermore, one of the following two situations occurs:

(a) $h_* = 0$*. This happens exactly when* $\mathcal{A}$ *and* $\mathcal{B}$ *can be separated by a homogeneous hyperplane. The optimal solutions to (24) are then of the form* $(\bar{y}, 0)$ *with* $\bar{y}$ *being orthogonal to a homogeneous hyperplane that separates* $\mathcal{A}$ *and* $\mathcal{B}$*. An optimal solution to (23) is obtained simply by normalizing* $\bar{y}$*.*

(b) $h_* > 0$*. In this case the minimization problems (23) and (24) have the same solution set. In other words, any solution* $(\bar{y}, \bar{s})$ *to (24) is such that* $\|\bar{y}\| = 1$*.*

Proof. The proof of the proposition is straightforward. It is essentially a matter of exploiting the fact that $h : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ is sublinear and nonnegative on Ω. The details are omitted. □

Proposition 4.1 allows us to operate on the minimization problem (24) instead of on (23). We describe next the two approaches we have adopted to numerically treat the problem (24). In the sequel we write the inequality constraint $\|y\| \geq 1$ in the equivalent form $\|y\|^2 \geq 1$.

4.1 The DC approach

The first approach is based on penalization of the nonconvex constraint $\|y\|^2 \geq 1$, that is to say, one considers the penalized problem

$$\begin{cases} \text{minimize} \quad h(y,s) + \rho \max\{0, 1 - \|y\|^2\} \\ (y,s) \in \Omega, \end{cases} \tag{25}$$

where $\rho > 0$ is a penalty parameter. In fact, (25) can be rewritten in the equivalent form

$$\begin{cases} \text{minimize} \quad h(y,s) + \rho \max\{0, \|y\|^2 - 1\} - \rho(\|y\|^2 - 1) \\ (y,s) \in \Omega, \end{cases} \tag{26}$$

where the cost function is of the DC (Difference of Convex) type. Indeed, the cost function of (26) is the difference $g_1 - g_2$ of the convex functions

$$\begin{aligned} g_1(y,s) &:= h(y,s) + \rho \max\{0, \|y\|^2 - 1\} \\ g_2(y,s) &:= \rho(\|y\|^2 - 1). \end{aligned}$$

Beware that g_1 is nonsmooth because h is polyhedral. Once the penalized problem has been put into the above form, algorithms for DC programming can be applied; see Section 4.3 for technical details.

4.2 The proximal point-linearized problem approach

The second approach we propose to deal with problem (24) is based on linearization of the norm constraint and introduction of both a proximal point and a feasibility restoration mechanism. We rewrite first problem (24) in the equivalent form of a differentiable nonlinear program by introducing auxiliary nonnegative variables v_i's and w_j's (grouped in vectors $v \in \mathbb{R}^p$ and $w \in \mathbb{R}^q$, respectively):

$$\left\{ \begin{array}{lll} h_* = & \min \ \gamma s + \sum_{i=1}^{p} v_i + \sum_{j=1}^{q} w_j & \\ & v_i \geq 0 & i \in \mathbb{N}_p \\ & v_i \geq -a_i^T y + \|a_i\| s & i \in \mathbb{N}_p \\ & w_j \geq 0 & j \in \mathbb{N}_q \\ & w_j \geq b_j^T y - \|b_j\| s & j \in \mathbb{N}_q \\ & \|y\|^2 \geq 1 & \\ & 0 \leq s \leq 1. & \end{array} \right. \tag{27}$$

Note that (27) would be a linear program were it not for the presence of the nonconvex constraint $\|y\|^2 \geq 1$. We open a parenthesis and say a few words on the Karush-Kuhn-Tucker (KKT) multiplier associated to this bothersome constraint. The KKT system for a (local) solution $(\bar{y}, \bar{v}, \bar{w}, \bar{s})$ to the problem (27) consists in stationarity conditions

$$\sigma \bar{y} + \sum_{i=1}^{p} \alpha_i a_i - \sum_{j=1}^{q} \beta_j b_j = 0 \tag{28}$$

$$\alpha_i + \mu_i = 1 \quad i \in \mathbb{N}_p \tag{29}$$

$$\beta_j + \xi_j = 1 \quad j \in \mathbb{N}_q \tag{30}$$

$$\gamma + \sum_{i=1}^{p} \alpha_i \|a_i\| - \sum_{j=1}^{q} \beta_j \|b_j\| - \lambda + \eta = 0, \tag{31}$$

together with complementary slackness conditions

$$\bar{v}_i \mu_i = 0 \quad i \in \mathbb{N}_p \tag{32}$$

$$\bar{w}_j \xi_j = 0 \quad j \in \mathbb{N}_q \tag{33}$$

$$(\bar{v}_i + a_i^T \bar{y} - \|a_i\| \bar{s}) \alpha_i = 0 \quad i \in \mathbb{N}_p \tag{34}$$

$$(\bar{w}_j - b_j^T \bar{y} + \|b_j\| \bar{s}) \beta_j = 0 \quad j \in \mathbb{N}_q \tag{35}$$

$$\sigma(\|\bar{y}\|^2 - 1) = 0$$

$$\bar{s} \lambda = 0 \tag{36}$$

$$\eta(-\bar{s} + 1) = 0, \tag{37}$$

for suitable nonnegativity multipliers

$$\left\{ \begin{array}{ll} \sigma \geq 0, \ \lambda \geq 0, \ \eta \geq 0 & \\ \alpha_i \geq 0, \ \mu_i \geq 0 & i \in \mathbb{N}_p \\ \beta_j \geq 0, \ \xi_j \geq 0 & j \in \mathbb{N}_q. \end{array} \right. \tag{38}$$

One has the following proposition.

Proposition 4.2. *Let $(\bar{y}, \bar{v}, \bar{w}, \bar{s})$ be a feasible point for (27) satisfying the KKT system (28)-(37) with multipliers as in (38). Then the cost term*

$$\bar{c} := \gamma \bar{s} + \sum_{i=1}^{p} \bar{v}_i + \sum_{j=1}^{q} \bar{w}_j$$

is equal to 0 if and only if $\sigma = 0$.

Proof. Suppose that $\bar{c} = 0$. Then $\bar{s} = 0$, $\bar{v} = 0$, and $\bar{w} = 0$. The combination of (34) and (35) yields then

$$\left(\sum_{i=1}^{p} \alpha_i a_i - \sum_{j=1}^{q} \beta_j b_j \right)^T \bar{y} = 0.$$

The above equality and (28) imply that $\sigma \|\bar{y}\|^2 = 0$. Hence, $\sigma = 0$. Conversely, suppose that $\sigma = 0$. From (28) one gets

$$\sum_{i=1}^{p} \alpha_i a_i - \sum_{j=1}^{q} \beta_j b_j = 0.$$

By combining the above equality with (34)-(35) one obtains

$$\kappa \bar{s} = \sum_{i=1}^{p} \alpha_i \bar{v}_i + \sum_{j=1}^{q} \beta_j \bar{w}_j \ \geq \ 0, \tag{39}$$

where

$$\kappa := \sum_{i=1}^{p} \alpha_i \|a_i\| - \sum_{j=1}^{q} \beta_j \|b_j\|.$$

If $\kappa \bar{s}$ were positive, then both κ and $\bar{s}$ would be positive. In such a case, the condition (45) yields $\lambda = 0$, contradicting the satisfaction of (31). Hence, $\kappa \bar{s} = 0$. This and (39) lead to

$$\sum_{i=1}^{p} \alpha_i \bar{v}_i + \sum_{j=1}^{q} \beta_j \bar{w}_j = 0.$$

The above equality, together with (29), (30), (32) and (33), imply that $\bar{v} = 0$ and $\bar{w} = 0$. To complete the proof one needs to show that $\bar{s} = 0$. If $\bar{s}$ were different from 0, then $\lambda = 0$ by (36). In such a case, (31) would imply that $\kappa < 0$, contradicting (39). □

Remark 4.3. If $\bar{c} = 0$, then also $h_* = 0$ and one is in the situation described by Proposition 4.1(a). In particular, $(\bar{y}/\|\bar{y}\|, 0)$ is a solution to the original problem (23) and $\bar{y}$ is orthogonal to a homogeneous hyperplane that separates $\mathcal{A}$ and $\mathcal{B}$.

We now return to the main flow of the presentation. To each feasible solution $(\bar{y}, \bar{s})$ to (24), with $h(\bar{y}, \bar{s}) > 0$, we associate a proximal point-linearized problem

$$(\mathcal{P}_{\bar{y}}) \quad \begin{cases} \text{minimize} \quad h(y, s) + \frac{1}{2}\varrho \|y - \bar{y}\|^2 \\ \bar{y}^T y = 1 \\ 0 \leq s \leq 1, \end{cases} \tag{40}$$

where ϱ is a positive "proximity" parameter. One can view the equality constraint in (40) as a linearized version of the normalization constraint $\|y\|^2 = 1$. The problem $\mathcal{P}_{\bar{y}}$ is not only convex, but it can be equivalently written as a special type of quadratic program:

$$(\mathcal{Q}_{\bar{y}}) \quad \begin{cases} \text{minimize} \ \ \gamma s + \sum_{i=1}^{p} v_i + \sum_{j=1}^{q} w_j + \frac{1}{2}\varrho \|y - \bar{y}\|^2 & \\ v_i \geq 0 & i \in \mathbb{N}_p \\ v_i \geq -a_i^T y + \|a_i\| s & i \in \mathbb{N}_p \\ w_j \geq 0 & j \in \mathbb{N}_q \\ w_j \geq b_j^T y - \|b_j\| s & j \in \mathbb{N}_q \\ \bar{y}^T y = 1 & \\ 0 \leq s \leq 1 & \end{cases}$$

The feasible set of $\mathcal{Q}_{\bar{y}}$ is a convex polyhedron and the cost function is a sum of a linear function and a nonnegative quadratic form. The KKT system for the quadratic program $\mathcal{Q}_{\bar{y}}$ consists of stationarity conditions

$$\begin{aligned} \varrho(y-\bar{y}) - \sigma\bar{y} - \sum_{i=1}^{p} \alpha_i a_i + \sum_{j=1}^{q} \beta_j b_j &= 0 \qquad (41) \\ \alpha_i + \mu_i &= 1 \quad i \in \mathbb{N}_p \\ \beta_j + \xi_j &= 1 \quad i \in \mathbb{N}_q \\ \gamma + \sum_{i=1}^{p} \alpha_i \|a_i\| - \sum_{j=1}^{q} \beta_j \|b_j\| - \lambda + \eta &= 0 \qquad (42) \end{aligned}$$

and complementarity slackness conditions

$$\begin{aligned} v_i \mu_i &= 0 \quad i \in \mathbb{N}_p \\ w_j\, \xi_j &= 0 \quad j \in \mathbb{N}_q \\ (v_i + a_i^T y - \|a_i\| s)\alpha_i &= 0 \quad i \in \mathbb{N}_p \qquad (43) \\ (w_j - b_j^T y + \|b_j\| s)\beta_j &= 0 \quad j \in \mathbb{N}_q \qquad (44) \\ s\lambda &= 0 \qquad (45) \\ \eta(-s+1) &= 0 \qquad (46) \end{aligned}$$

for suitable multipliers

$$\begin{cases} \sigma \in \mathbb{R},\ \lambda \geq 0,\ \eta \geq 0 & \\ \alpha_i \geq 0,\ \mu_i \geq 0 & i \in \mathbb{N}_p \\ \beta_j \geq 0,\ \xi_j \geq 0 & j \in \mathbb{N}_q. \end{cases} \qquad (47)$$

Note that the multiplier σ associated to the equality constraint $\bar{y}^T y = 1$ is unrestricted in sign. The following proposition holds:

Proposition 4.4. *Suppose that $(\bar{y}, \bar{v}, \bar{w}, \bar{s})$ is feasible point for $\mathcal{Q}_{\bar{y}}$. Then the following conditions are equivalent:*

- *(a) $(\bar{y}, \bar{v}, \bar{w}, \bar{s})$ is a solution to $\mathcal{Q}_{\bar{y}}$.*
- *(b) $(\bar{y}, \bar{v}, \bar{w}, \bar{s})$ satisfies the system (41)-(46) with multipliers as in (47).*
- *(c) $(\bar{y}, \bar{v}, \bar{w}, \bar{s})$ satisfies the system (28)-(37) with multipliers as in (38).*

Proof. The equivalence (a) $\Leftrightarrow$ (b) is a standard result concerning the minimization of a convex function on a convex polyhedron: satisfaction of the associated KKT system is necessary and sufficient for optimality. Clearly (c) $\Rightarrow$ (b), so we concentrate on the reverse implication. Suppose that (b) holds. By comparing the systems (41)-(46) and (28)-(37), one observes that all one needs to prove is nonnegativity of the multiplier σ. The satisfaction of (41) at $y = \bar{y}$, together with the fact that $\bar{y}^T \bar{y} = 1$, implies

$$\sigma + \sum_{i=1}^{p} \alpha_i a_i^T \bar{y} - \sum_{j=1}^{q} \beta_j b_j^T \bar{y} = 0.$$

By combining the above equality and the relations (43) and (44), one obtains

$$\sigma = \sum_{i=1}^{p} \alpha_i v_i + \sum_{j=1}^{q} \beta_j w_j - \bar{s}\kappa, \qquad (48)$$

where κ is as in the proof of Proposition 4.2. But, thanks to (42) and (45), one has

$$-\bar{s}\kappa \;=\; \bar{s}(\gamma - \lambda + \eta) \;=\; \bar{s}(\gamma + \eta) \;\geq\; 0.$$

This and (48) show that $\sigma \geq 0$. □

4.2.1 Proximal-feasibility restoration

We now explain our *proximal-feasibility restoration* approach to solve problem (24). The method is based on two distinct phases:

- *Phase I.* Given a current aperture coefficient $\bar{s} \in [0,1]$ and a current axis $\bar{y} \in \mathbb{S}_n$, we find a solution $(\tilde{y}, \tilde{s})$ to the convex program $\mathcal{P}_{\bar{y}}$.
- *Phase II.* We project $\tilde{y}$ onto $\mathbb{S}_n$ in order to produce a new axis $\bar{y}^{\,\mathrm{new}}$ and adjust $\tilde{s}$ in order to produce a new aperture coefficient $\bar{s}^{\,\mathrm{new}}$. To be more precise, we compute

  $$\bar{y}^{\,\mathrm{new}} = \tilde{y}/\|\tilde{y}\| \quad \text{and} \quad \bar{s}^{\,\mathrm{new}} = \mathbf{s}(\bar{y}^{\,\mathrm{new}}),$$

  where $\mathbf{s}(\cdot)$ is a function whose evaluation

  $$\mathbf{s}(y) := \arg\min_{0\leq s\leq 1} h(y,s) \tag{49}$$

  at an argument y requires to solve a univariate nonsmooth convex problem. The part is handled with the **U**nivariate **M**inimization **A**lgorithm (UMA) described in Section 4.2.2.

In practice, the convex program $\mathcal{P}_{\bar{y}}$ is solved by using its quadratic representation $\mathcal{Q}_{\bar{y}}$. However, in order to simplify the notation, we keep its original formulation. On the other hand, considering that 0 is a lower bound for the cost function of $\mathcal{P}_{\bar{y}}$, we assume that, for a given optimality threshold $\varepsilon > 0$, the initial cost function value is above such threshold.

The detailed presentation of the proximal-feasibility restoration algorithm is as follows:

- *Step 0. Initialization.* An optimality parameter $\varepsilon > 0$ is given, together with an initial point $(y^{(0)}, s^{(0)})$ such that $\|y^{(0)}\| = 1$, $s^{(0)} = \mathbf{s}(y^{(0)})$, and $h(y^{(0)}, s^{(0)}) > \varepsilon$. A distance parameter $\delta > 0$ is also given. Set $\varrho = 2\varepsilon/\delta^2$. Set the iteration counter $k = 0$.
- *Step 1. Solving the Proximal Point-Linearized Problem.* Solve the problem $\mathcal{P}_{y^{(k)}}$ and let $(\tilde{y}^{\,(k+1)}, \tilde{s}^{\,(k+1)})$ be the optimal solution. If

  $$\left\|\tilde{y}^{\,(k+1)} - y^{(k)}\right\| \leq \delta, \tag{50}$$

  then STOP.
- *Step 2. Projection and calculation of a new feasible point.* Set

  $$y^{(k+1)} = \tilde{y}^{\,(k+1)}/\|\tilde{y}^{\,(k+1)}\|$$

  and calculate $s^{(k+1)} = \mathbf{s}(y^{(k+1)})$ by calling the UMA. Set $k = k+1$ and return to Step 1.

In the following proposition we prove termination of the algorithm.

Proposition 4.5. *The termination criterion (50) is satisfied within a finite number of iterations.*

Proof. The point $(y^{(k)}, s^{(k)})$ is feasible for $\mathcal{P}_{y^{(k)}}$ and therefore

$$h\left(\tilde{y}^{(k+1)}, \tilde{s}^{(k+1)}\right) + \frac{1}{2}\varrho \left\|\tilde{y}^{(k+1)} - y^{(k)}\right\|^2 \leq h^{(k)},$$

where $h^{(k)} := h\left(y^{(k)}, s^{(k)}\right)$. So, if the stopping test (50) is not satisfied, then

$$h\left(\tilde{y}^{(k+1)}, \tilde{s}^{(k+1)}\right) \ < \ h^{(k)} - \frac{1}{2}\varrho\delta^2 \ = \ h^{(k)} - \varepsilon. \tag{51}$$

Note that $(y^{(k)})^T\tilde{y}^{(k+1)} = 1$ ensures $\|\tilde{y}^{(k+1)}\| \geq 1$. So, from (51) one gets

$$\begin{aligned} h^{(k+1)} \ &:= \ h\left(y^{(k+1)}, s^{(k+1)}\right) = h\left(y^{(k+1)}, \mathbf{s}(y^{(k+1)})\right) \\ &\leq \ h\left(\frac{\tilde{y}^{(k+1)}}{\|\tilde{y}^{(k+1)}\|}, \frac{\tilde{s}^{(k+1)}}{\|\tilde{y}^{(k+1)}\|}\right) \ = \ \|\tilde{y}^{(k+1)}\|^{-1} h\left(\tilde{y}^{(k+1)}, \tilde{s}^{(k+1)}\right) \\ &\leq \ h\left(\tilde{y}^{(k+1)}, \tilde{s}^{(k+1)}\right). \end{aligned}$$

The combination of (51) and the above inequality ensures that $h^{(k+1)} < h^{(k)} - \varepsilon$, that is, at each iteration the reduction in the objective function is at least ε. The thesis follows from the fact that 0 is a lower bound for the sequence $\{h^{(k)}\}_{k\in\mathbb{N}}$. □

Remark 4.6. The algorithm terminates when the stopping criterion at Step 1 is met. Observe that satisfaction of such test, taking into account condition (41), implies

$$\left\|\sigma^{(k)} y^{(k)} + \sum_{i=1}^{p} \alpha_i^{(k)} a_i - \sum_{j=1}^{q} \beta_j^{(k)} b_j\right\| \ = \ \varrho \left\|\tilde{y}^{(k+1)} - y^{(k)}\right\| \ \leq \ \varrho\delta, \tag{52}$$

where $\sigma^{(k)} \geq 0$, $\alpha_i^{(k)} \geq 0$, and $\beta_j^{(k)} \geq 0$ are multipliers associated to the solution to $\mathcal{P}_{y^{(k)}}$. Note that (52) can be viewed as an approximate satisfaction at $y^{(k)}$ of the condition (28) for problem (27).

4.2.2 The Univariate Minimization Algorithm

Now we describe how to solve the univariate minimization problem (49) once $y = \bar{y}$ has been fixed. The problem at hand is that of minimizing

$$s \in [0,1] \mapsto \varphi(s) := \gamma s + \sum_{i=1}^{p} \max\{0, -c_i + \|a_i\| s\} + \sum_{j=1}^{q} \max\{0, d_j - \|b_j\| s\},$$

where $c_i := a_i^T \bar{y}$ and $d_j := b_j^T \bar{y}$. The univariate function φ is convex and piecewise affine. A matter of computation shows that the subdifferential $\partial\varphi(s)$ of φ at s is the collection of points of the form

$$\gamma + \sum_{i\in I_+(s)} \|a_i\| + \sum_{i\in I_0(s)} \alpha_i \|a_i\| - \sum_{j\in J_+(s)} \|b_j\| - \sum_{j\in J_0(s)} \beta_j \|b_j\|$$

with

$$\begin{aligned} I_+(s) \ &= \ \{i \in \mathbb{N}_p : -c_i + \|a_i\| s > 0\}, \\ I_0(s) \ &= \ \{i \in \mathbb{N}_p : -c_i + \|a_i\| s = 0\}, \\ J_+(s) \ &= \ \{j \in \mathbb{N}_q : d_j - \|b_j\| s > 0\}, \\ J_0(s) \ &= \ \{j \in \mathbb{N}_q : d_j - \|b_j\| s = 0\}, \\ \alpha_i \ &\in \ [0,1] \quad \text{for all } i \in I_0(s), \\ \beta_j \ &\in \ [0,1] \quad \text{for all } j \in J_0(s). \end{aligned}$$

Hence, the right- and left-derivatives of φ at s are given respectively by

$$\begin{aligned} \varphi'_+(s) &:= \varphi'(s;1) = \gamma + \sum_{i\in I_+(s)\cup I_0(s)} \|a_i\| - \sum_{j\in J^+(s)} \|b_j\| \\ \varphi'_-(s) &:= -\varphi'(s;-1) = \gamma + \sum_{i\in I^+(s)} \|a_i\| - \sum_{j\in J_+(s)\cup J_0(s)} \|b_j\|. \end{aligned}$$

In particular one has

$$\begin{aligned} \varphi'_+(0) &:= \gamma + \sum_{i\in\mathbb{N}_p,\, c_i\le 0} \|a_i\| - \sum_{j\in\mathbb{N}_q,\ d_j>0} \|b_j\| \\ \varphi'_-(1) &:= \gamma + \sum_{i\in\mathbb{N}_p,\, c_i<\|a_i\|} \|a_i\| - \sum_{j\in\mathbb{N}_q,\, d_j\ge\|b_j\|} \|b_j\|. \end{aligned}$$

All these derivatives are easy to evaluate numerically. Parenthetically, observe that $\varphi'_+(0) \geq 0$ ensures $\mathbf{s}(\bar{y}) = 0$, while $\varphi'_-(1) \leq 0$ ensures $\mathbf{s}(\bar{y}) = 1$.

The Univariate Minimization Algorithm for calculating $\mathbf{s}(\bar{y})$ reads as follows:

- *Step 0. Testing the extreme points.* If either $\varphi'_+(0) \geq 0$ or $\varphi'_-(1) \leq 0$, then set $s(\bar{y}) = 0$ or 1, respectively, and terminate. Else select an accuracy parameter $\varepsilon > 0$ and set $l = 0$, $r = 1$.

- *Step 1. Locate the midpoint.* Select $s = (l + r)/2$ and calculate $\varphi'_+(s)$.

- *Step 2. Bisection.* If $\varphi'_+(s) \geq 0$, set $l = l$ and $r = s$. Else set $l = s$ and $r = r$. If $r - l \leq \varepsilon$, then set $\mathbf{s}(\bar{y}) = s$ and STOP. Else, return to Step 1.

4.3 Numerical experiments with real-life data

In our numerical experiments we have applied the largest angle homogeneous separator model (23), along the guidelines described in Sections 4.1 and 4.2. In particular, we have run two codes, implementing, respectively, the DC approach (DCA code) and the proximal point-linearized problem approach (Proximal code).

The DCA code is based on the DCA method [1] for solving the Difference of Convex problem stated in (26). DCA requires, at each iteration, to compute a solution to a convex program. To this aim we have used the subroutine NCVX (cf. [11]), which implements a bundle type approach enabling the resolution of nonsmooth optimization problems, be them convex or nonconvex. The Proximal code requires at each iteration solution of a convex quadratic problem. Again, we have used the subroutine NCVX to such purpose. Parenthetically, we mention that NCVX has been recently re-implemented in MATLAB and satisfactory results on machine learning problems are reported in [6].

As for the parameter setting, we have used a grid of possible values to preliminary tune the parameters γ, ρ in the DC program (26) and the parameters γ, δ in the proximal point-linearized problem. This has been done for each dataset.

Once the parameters have been fixed, we have adopted for each test problem the standard ten-fold cross validation protocol, which consists in splitting the dataset of interest into ten equally sized pieces. Nine of them are in turn used as training set and the remaining one as testing set. By correctness, as usual in classification literature, we intend the total percentage of well classified points (of both $\mathcal{A}$ and $\mathcal{B}$) when the algorithm stops. Of course, what we report is the average of such test correctness taken over the ten different experiments in the cross validation framework.

We note, however, that the correctness measure is not coincident with the adopted error function, even though, of course, a good correlation between them is expected.

We remark that both algorithms we have tested are of the local optimization type. Consequently, we have implemented a multi-start approach. In fact, for each dataset we have run the code considering the starting points $(y^{(0)}, s^{(0)}) = (\pm e_i, 0.5)$ (with $i \in \{1, \ldots, n\}$), where e_i is the ith unit vector and n is the dimension of the sample space. We have also tested the starting point $(\pm \mathbf{1}_n/n, 0.5)$ where $\mathbf{1}_n$ is the vector of ones in $\mathbb{R}^n$.

We have considered the following test problems drawn from the binary classification literature (cf. Table 1), where we indicate the dimension of the sample space and the total number of sample points.

Dataset	Dimension	Points
Cancer	9	699
Diagnostic	30	569
Heart	13	297
Pima	8	769
Ionosphere	34	351
Sonar	60	208
Galaxy	14	4192
g50c	50	550
g10n	10	550

Table 1: Datasets

The first six datasets are taken from the UCI Machine Learning Repository [21], Galaxy is the dataset used in galaxy discrimination with neural networks [22], while the last two test problems are described in [8].

In Tables 2 and 3 we report the numerical results in terms of average percentage of testing correctness. We have selected, for each dataset, the best results we have obtained corresponding to the different tested starting points. The best result for each dataset has been underlined.

In particular, in Table 2, to provide a useful reference, we compare our results with those obtained by using the LIBSVM package [7], a well established program library for Support Vector Machine (SVM) based classification. We have considered both linear kernel (Linear) and the best result obtained by applying two different kernels, polynomial or RBF (Kernel). Experimentations of all tested algorithms have been performed with no dataset preprocessing (normalization, scaling, etc).

Dataset	LIBSVM		Conical-SEP	
	Linear	Kernel	DCA	Proximal
Cancer	<u>95.54</u>	95.33	87.14	87.14
Diagnostic	95.95	96.48	<u>97.19</u>	96.84
Heart	<u>85.19</u>	82.82	77.33	77.00
Pima	<u>76.30</u>	75.78	60.13	64.16
Ionosphere	87.14	<u>94.86</u>	91.43	89.71
Sonar	78.81	<u>87.48</u>	77.14	77.62
Galaxy	94.78	<u>96.11</u>	93.10	87.28
g50c	95.27	94.36	96.08	<u>96.12</u>
g10n	<u>98.91</u>	93.82	90.52	91.64

Table 2: Homogeneous conical separation versus SVM with and without kernel

Moreover, in Table 3 we compare conical separation with spherical separation approaches: the

FC (Fixed Center) and UMC (Unconstrained Moving Center) codes [2], that implement, respectively, the fixed and the moving center spherical separation with no margin consideration.

Dataset	SPSEP		Conical-SEP	
	FC	UMC	DCA	Proximal
Cancer	97.00	95.71	87.14	87.14
Diagnostic	84.03	89.82	97.19	96.84
Heart	75.00	80.33	77.33	77.00
Pima	69.35	68.70	60.13	64.16
Ionosphere	51.14	72.00	91.43	89.71
Sonar	59.52	69.05	77.14	77.62
Galaxy	80.19	93.79	93.10	87.28
g50c	67.76	72.96	96.08	96.12
g10n	54.02	81.04	90.52	91.64

Table 3: Homogeneous conical separation versus spherical separation

Our experimentation shows that conical separability provides interesting results on some of the tested datasets. It can be considered yet another useful tool for approaching practical classification problems.

5 The most robust homogeneous separator

A classical method for separating the sets $\mathcal{A}$ and $\mathcal{B}$ in a robust way is to find the widest strip of "no man's land" which can be placed between both sets. One defines the most robust separating hyperplane as the hyperplane that is equidistant from both sides of that strip, see Figure 1.

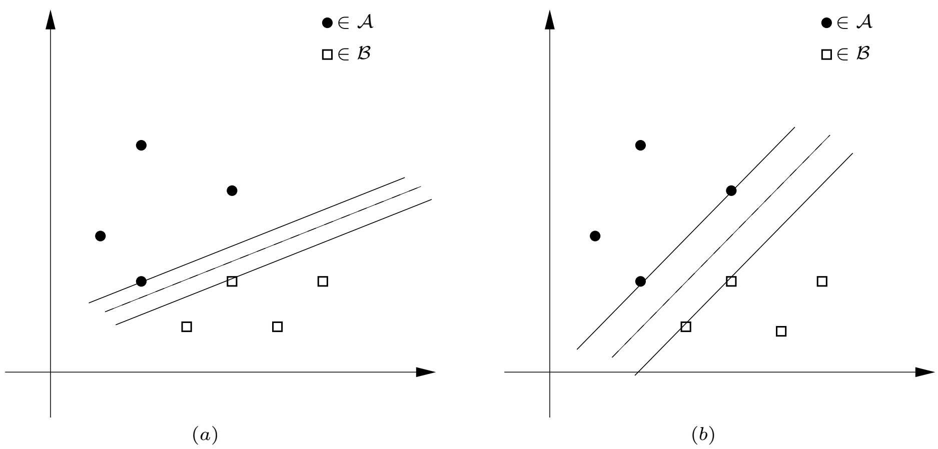

Figure 1: (a) A separating hyperplane. (b) A robust separating hyperplane.

Inspired by this classical idea, we introduce a robustness concept for homogeneous separators and explain how to compute a most robust homogeneous separator. We assume that $G(\mathcal{A},\mathcal{B})$ is nonempty and consider the multivalued map $\mathcal{G}:\mathbb{R}^n \rightrightarrows \mathbb{R}$ given by

$$\mathcal{G}(y) = \{s \in \mathbb{R} : (y,s) \in G(\mathcal{A},\mathcal{B})\}.$$

The graph of $\mathcal{G}$ is equal to the compact set $G(\mathcal{A},\mathcal{B})$. Hence, the effective domain

$$D(\mathcal{G}) = \{y \in \mathbb{R}^n : \mathcal{G}(y) \neq \emptyset\}$$

is a compact set. The conic no man's land associated to the axis $y \in D(\mathcal{G})$ is the region

$$R(y) = \{x \in \mathbb{R}^n : \, s_{\min}(y)\, \|x\| \leq y^T x \leq s_{\max}(y)\, \|x\|\},$$

where

$$s_{\min}(y) = \min\{s : s \in \mathcal{G}(y)\} \quad \text{and} \quad s_{\max}(y) = \max\{s : s \in \mathcal{G}(y)\}.$$

Since $s_{\min} : D(\mathcal{G}) \to \mathbb{R}$ and $s_{\max} : D(\mathcal{G}) \to \mathbb{R}$ are continuous functions on a compact set, the maximization problem

$$\begin{cases} \text{maximize } (s_{\max} - s_{\min})(y) \\ y \in D(\mathcal{G}) \end{cases} \tag{53}$$

admits at least one solution. The problem (53) is about finding a conic no man's land with largest possible "angular width".

Definition 5.1. *A solution $\bar{y}$ to (53) is called a* robust axis*. One says that $\Gamma_0(\bar{y},\bar{s})$ is a* most robust homogeneous separator *if the axis $\bar{y}$ is robust and*

$$\bar{s} = \cos\left(\frac{\arccos[s_{\min}(\bar{y})] \, + \, \arccos[s_{\max}(\bar{y})]}{2}\right). \tag{54}$$

The special choice (54) ensures that $\Gamma_0(\bar{y},\bar{s})$ divides the region $R(\bar{y})$ into two conic portions of equal aperture. Figure 2 illustrates the distinction between robustness and non-robustness.

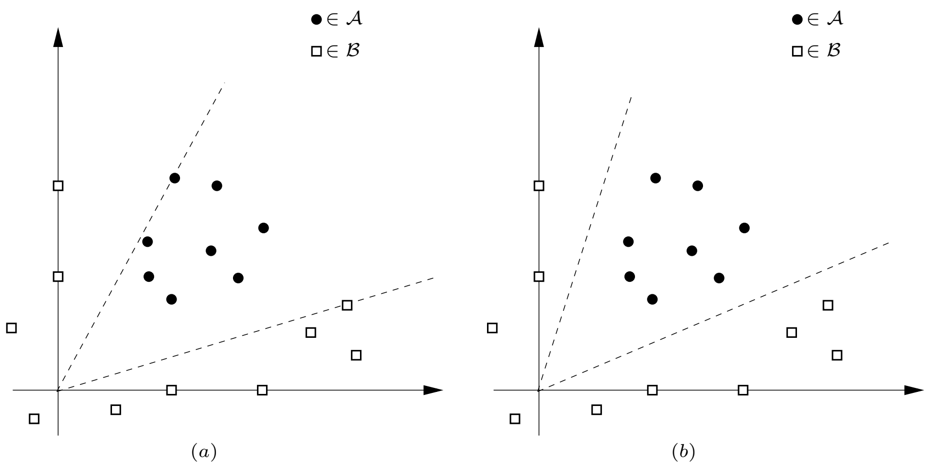

Figure 2: (a) An homogeneous separator. (b) The most robust homogeneous separator.

With the helps of the functions $\Psi_{\mathcal{A}}$ and $\Phi^+_{\mathcal{B}}$ one can can be reformulated (53) in a simpler manner. In fact, one has:

Theorem 5.2. *Suppose that the a_i's and the b_j's are nonzero vectors and that $G(\mathcal{A},\mathcal{B})$ is nonempty. Then* *(a) The problem (53) has the same optimal value and solution set as*

$$\begin{cases} \textit{maximize } (\Psi_{\mathcal{A}} - \Phi^+_{\mathcal{B}})(y) \\ \|y\| = 1. \end{cases} \tag{55}$$

(b) If $(\mathcal{A}, \mathcal{B})$ admits a strict homogeneous separator, then the most robust homogeneous separator is unique. In fact, its axis is the unique solution to the convex optimization problem

$$\begin{cases} \textit{maximize } (\Psi_{\mathcal{A}} - \Phi^+_{\mathcal{B}})(y) \\ \|y\| \leq 1. \end{cases} \tag{56}$$

Proof. The problem (53) is clearly equivalent to

$$\begin{cases} \text{maximize } s_1 - s_0 \\ (y, s_0, s_1) \in \mathbb{S}_n \times [0,1] \times [0,1] \\ \Phi_{\mathcal{B}}(y) \leq s_0 \leq \Psi_{\mathcal{A}}(y) \\ \Phi_{\mathcal{B}}(y) \leq s_1 \leq \Psi_{\mathcal{A}}(y). \end{cases} \tag{57}$$

Since $s_1 - s_0$ is to be maximized, one can add the constraint $s_1 \geq s_0$ and rewrite (57) as

$$\begin{cases} \text{maximize } s_1 - s_0 \\ \|y\| = 1 \\ 0 \leq s_0 \leq s_1 \leq 1 \\ \Phi_{\mathcal{B}}(y) \leq s_0 \\ s_1 \leq \Psi_{\mathcal{A}}(y). \end{cases}$$

The constraint $s_1 \leq 1$ is superfluous because $\Psi_{\mathcal{A}}$ is majorized by 1 on the unit sphere $\mathbb{S}_n$. So, after simplification one gets

$$\begin{cases} \text{maximize } s_1 - s_0 \\ \|y\| = 1 \\ \Phi^+_{\mathcal{B}}(y) \leq s_0 \leq s_1 \leq \Psi_{\mathcal{A}}(y). \end{cases}$$

This leads to (55) and completes the proof of (a). The existence of a strict homogeneous separator implies that $(\Psi_{\mathcal{A}} - \Phi^+_{\mathcal{B}})(\tilde{y}) > 0$ for some $\tilde{y} \in \mathbb{S}_n$. We claim that (55) and (56) have the same solution set. By a convexity argument, the common solution set is then a singleton. Let y_* be a solution to (56). Hence,

$$(\Psi_{\mathcal{A}} - \Phi^+_{\mathcal{B}})(y_*) \ \geq \ (\Psi_{\mathcal{A}} - \Phi^+_{\mathcal{B}})(\tilde{y}) \ > \ 0,$$

and therefore $y_* \neq 0$. If $\|y_*\| < 1$, then $\hat{y}_* = \|y_*\|^{-1} y_*$ is a unit vector such that

$$(\Psi_{\mathcal{A}} - \Phi^+_{\mathcal{B}})(\hat{y}_*) \ = \ \|y_*\|^{-1} (\Psi_{\mathcal{A}} - \Phi^+_{\mathcal{B}})(y_*) \ > \ (\Psi_{\mathcal{A}} - \Phi^+_{\mathcal{B}})(y_*),$$

contradicting the optimality of y_*. Hence, $\|y_*\| = 1$. This proves our claim and completes the proof of the theorem. □

The next theorem provides a dual characterization for the most robust homogeneous separator. In essence, what this result says is that the axis of the most robust homogeneous separator can be found by solving a least norm problem on a certain polytope.

Theorem 5.3. *Suppose that the a_i's and the b_j's are nonzero vectors and that $(\mathcal{A}, \mathcal{B})$ admits a strict homogeneous separator. Then*

$$\max_{y \in D(\mathcal{G})} (s_{\max} - s_{\min})(y) \;=\; \min\{\|x\| : x \in \Xi_{\mathcal{A}} - \Xi_{\mathcal{B}}^{+}\} \tag{58}$$

Furthermore, the axis of the most robust homogeneous separator is given by $\bar{y} = \|\bar{x}\|^{-1}\bar{x}$ with $\bar{x}$ denoting the least norm element of the polytope $\Xi_{\mathcal{A}} - \Xi_{\mathcal{B}}^{+}$.

Proof. Let δ be the maximum on the right-hand side of (58). By Theorem 5.2(b) one has

$$\delta \;=\; \max_{\|y\| \le 1} (\Psi_{\mathcal{A}} - \Phi_{\mathcal{B}}^{+})(y) \;=\; \max_{\|y\| \le 1} \min_{x \in W} x^T y, \tag{59}$$

where $W = \Xi_{\mathcal{A}} - \Xi_{\mathcal{B}}^{+}$. Thanks to the Kneser minimax theorem one can exchange the order of the maximum and the minimum in (59). Hence,

$$\delta \;=\; \min_{x \in W} \max_{\|y\| \le 1} x^T y \;=\; \min_{x \in W} \|x\|.$$

For proving the second part of the theorem we exploit the fact that

$$(\Psi_{\mathcal{A}} - \Phi_{\mathcal{B}}^{+})(\bar{y}) = \|\bar{x}\| \tag{60}$$

$$\|\bar{y}\| = 1, \;\; \bar{x} \in W. \tag{61}$$

The equality (60) can be rewritten as

$$\left(\bar{x}^T \bar{y} - \min_{x \in W} x^T \bar{y}\right) + \left(\|\bar{x}\| - \bar{x}^T \bar{y}\right) = 0.$$

Due to (61), both terms in the above sum are nonnegative. Since their sum is equal to zero, each term must be equal to zero. In particular, $\bar{x}^T \bar{y} = \|\bar{x}\|$ and therefore $\bar{y} = \|\bar{x}\|^{-1}\bar{x}$. □

As mentioned in Remark 2.4, there is an extensive literature dealing the problem of estimating the gap between two polytopes. In fact, finding the least norm element of $\Xi_{\mathcal{A}} - \Xi_{\mathcal{B}}^{+}$ is not a difficult matter. One just need to solve the quadratic minimization problem

$$\begin{cases} \text{minimize} \left\| \sum_{i=1}^{p} \alpha_i \frac{a_i}{\|a_i\|} - \left(\beta_0 0 + \sum_{j=1}^{q} \beta_j \frac{b_j}{\|b_j\|}\right) \right\|^2 \\ \alpha_1 \ge 0, \ldots, \alpha_p \ge 0, \beta_0 \ge 0, \beta_1 \ge 0, \ldots, \beta_q \ge 0 \\ \sum_{i=1}^{p} \alpha_i = 1, \;\; \sum_{j=0}^{q} \beta_j = 1. \end{cases} \tag{62}$$

By getting rid of the variable β_0 one gets a reduced problem

$$\begin{cases} \text{minimize} \left\| \sum_{i=1}^{p} \alpha_i \frac{a_i}{\|a_i\|} - \sum_{j=1}^{q} \beta_j \frac{b_j}{\|b_j\|} \right\|^2 \\ \alpha_1 \ge 0, \ldots, \alpha_p \ge 0, \beta_1 \ge 0, \ldots, \beta_q \ge 0 \\ \sum_{i=1}^{p} \alpha_i = 1, \;\; \sum_{j=1}^{q} \beta_j \le 1, \end{cases}$$

but somehow it is simpler to work with the original model (62). Note that (62) concerns the minimization of a positive semidefinite quadratic form over the Cartesian product of two standard simplices.

References

[1] L.T.H. An and P.D. Tao. A D.C. optimization algorithm for solving the trust-region subproblem. *SIAM J. Optim.*, 8 (1998), 476–505.

[2] A. Astorino, A. Fuduli, and M. Gaudioso. DC models for spherical separation. *J. Global Optim.*, 48 (2010), 657-669.

[3] A. Astorino and M. Gaudioso. Polyhedral separability through successive LP. J. Optim. Theory Appl., 112 (2002), 265-293.

[4] A. Astorino and M. Gaudioso. Ellipsoidal separation for classification problems. *Optim. Methods Softw.*, 20 (2005), 267-276.

[5] A. Astorino, M. Gaudioso, and A. Seeger. Conic separation of finite sets. II. The non-homogeneous case. July 2012, submitted.

[6] C. Bergeron, G. Moore, J. Zaretzki, C.M. Breneman and K.P. Bennett. Fast bundle algorithm for multiple instance learning. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011. In http://doi.ieeecomputersociety.org/10.1109/TPAMI.2011.194.

[7] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2 (2011), pp. 27:1–27:27. Software available at www.csie.ntu.edu.tw/∼cjlin/libsvm.

[8] O. Chapelle and A. Zien. Semi–supervised classification by low density separation. In *Proc. 10th Intern. Workshop on Artificial Intelligence and Statistics*, 2005, pp. 57–64.

[9] P.H. Dien, G. Mastroeni, M. Pappalardo, and P.H. Quang. Regularity conditions for constrained extremum problems via image space: the nonlinear case. *J. Optim. Theory Appl.*, 80 (1994), 19-37.

[10] J.E. Falk and V.E. Karlov. Robust separation of finite sets via quadratics. *Comput. Oper. Res.*, 28 (2001), 6, 537-561.

[11] A. Fuduli, M. Gaudioso and G. Giallombardo. Minimizing nonconvex nonsmooth functions via cutting planes and proximity control. *SIAM J. Optim.*, 14 (2004), 743–756.

[12] R.N. Gasimov and G. Ozturk. Separation via polyhedral conic functions. *Optim. Methods Softw.*, 21 (2006), 527-540.

[13] E.G. Gilbert, D.W. Johnson, and S.S. Keerthi. A fast procedure for computing the distance between complex objects. *IEEE J. Robotics Automation*, 4 (1988), 193–203.

[14] Y-J. Gotoh, H. Konno, S. Uryasev, and A. Yuki. Failure discrimination by semidefinite programming. In *Financial Engineering, E-Commerce and Supply Chain*, P.M. Pardalos and V.K. Tsitsiringos (Eds.), Kluwer Acad. Press, Dordrecht, 2002.

[15] M.I. Henig. A cone separation theorem. *J. Optim. Theory Appl.*, 36 (1982), 451-455.

[16] M.L. Hoai, A.L.T. Hoai, P.D. Tao, and V.N. Ngai. An efficient DCA for spherical separation. In *Intelligent Information and Database Systems.* Lecture Notes in Computer Science, Vol. 6592, 2011, 421–431.

[17] R. Kasimbeyli. A nonlinear cone separation theorem and scalarization in nonconvex vector optimization. *SIAM J. Optim.*, 20 (2009), 1591-1619.

[18] K. Kim and H.S. Ryoo. Data separation via a finite number of discriminant functions: a global optimization approach. *Appl. Math. Comput.*, 190 (2007), 476-489.

[19] O.L. Mangasarian. Linear and nonlinear separation of patterns by linear programming. *Operations Res.* 13 (1965), 444-452.

[20] G. Mastroeni and M. Pappalardo. Separation and regularity in the image space. In *New Trends in Mathematical Programming*, F. Giannessi, S. Komlosi, and T. Rapcsak (Eds.), Kluwer Acad. Publ., pp. 181-190, 1998.

[21] P.M. Murphy and D.W. Aha, UCI *repository of machine learning databases.* In www.ics.uci.edu/∼mlearn/MLRepository.html, 1992.

[22] S. Odewahn, E. Stockwell, R. Pennington, R. Humphreys, and W. Zumach. Automated star/galaxy discrimination with neural networks. *Astrophysical Journal*, 103 (1992), 318–331.

[23] M. Pappalardo. On the image regularity condition. *J. Optim. Theory Appl.*, 121 (2004), 673-678.

[24] R.T. Rockafellar. *Convex Analysis.* Princeton Univ. Press, Princeton, 1970.