To appear in JOURNAL OF CONVEX ANALYSIS, 2014.

Conic separation of finite sets

II. The non-homogeneous case

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

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

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

One has to select the aperture coefficient s, the axis y, and the apex z in such a way as to meet certain optimal separation criteria. The homogeneous case $z = 0$ has been treated in Part I of this work. We now discuss the more general case in which the apex of the cone is allowed to move in a certain region. The non-homogeneous case is structurally more involved and leads to challenging nonconvex nonsmooth optimization problems.

Mathematics Subject Classification: 90C26.
Key words: Conical separation, revolution cone, alternating minimization, DC programming, classification.

1 Introduction

Discriminating between a pair of finite sets in $\mathbb{R}^n$ by means of a separating hypersurface is a theme treated in several dozens of scientific publications. As separating hypersurface one may consider an hyperplane, a sphere, an ellipsoid, or another more sophisticated mathematical object. A brief introduction to this broad topic is given in our previous paper [2]. In this work we are interested in separating a pair

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

of mutually disjoint finite subsets of the Euclidean space $\mathbb{R}^n$ by means of a revolution cone

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

[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 (E-mail: astorino@icar.cnr.it)
[2]Dipartimento di Elettronica Informatica e Sistemistica, Università delle Calabria, 87036 Rende, Italy (E-mail: gaudioso@deis.unical.it)
[3]University of Avignon, Department of Mathematics, 33 rue Louis Pasteur, 84000 Avignon, France (E-mail: alberto.seeger@univ-avignon.fr)

The parameter z, called the apex of the cone, is to be selected from a given set $Z \subseteq \mathbb{R}^n$ of candidate apices. It is assumed once and for all that Z is nonempty and closed. The axis y is sought on the unit sphere $Y := \{y \in \mathbb{R}^n : \|y\| = 1\}$ and the aperture coefficient s on the interval $S := [0, 1]$.

For notational convenience we identify the revolution cone (1) with the parameter vector

$$w = (z, y, s) \in W := Z \times Y \times S \ \subseteq \ \mathbb{R}^{2n+1}.$$

Each term of the triplet w is recovered with an appropriate canonical projection: $z = \Pi_1 w$, $y = \Pi_2 w$, and $s = \Pi_3 w$. The precise formulation of the Conic Separation problem reads as follows:

$$(\mathrm{CS}) \quad \begin{cases} \text{find } w \text{ in } W \text{ such that} \\ \mathcal{A} \subseteq \{x \in \mathbb{R}^n : f(x, w) \leq 0\} \\ \mathcal{B} \subseteq \{x \in \mathbb{R}^n : f(x, w) \geq 0\}, \end{cases}$$

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

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

So, one needs to find a point $w \in W$ satisfying the nonlinear inequality system

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

Since W is a product of closed sets and f is continuous, one sees that

$$W_* = \{w \in W : \ w \text{ satisfies } (2)\}$$

is a closed set in $\mathbb{R}^{2n+1}$. Beware that (2) may be unsolvable, even if p and q are small integers.

Definition 1.1. *The pair $(\mathcal{A}, \mathcal{B})$ is* conically separable *if W_* is nonempty. If w belongs to W_*, then the corresponding revolution cone $\Gamma(w)$ is called a separator. If none of the inequalities in (2) is active, then the separator $\Gamma(w)$ is declared* strict *and one writes $w \in W_*^{\mathrm{st}}$.*

2 Existence of a separator

Sometimes it is convenient to handle the CS problem as a family of homogeneous conic separation problems indexed by the parameter $z \in Z$. From the very definition of the revolution cone (1) one sees that

$$(z, y, s) \in W_* \quad \Leftrightarrow \quad (y, s) \in M(z),$$

where $M(z)$ is the solution set to the Homogeneous Conic Separation problem

$$(\mathrm{HCS})_z \quad \begin{cases} \text{find } (y, s) \in Y \times S \text{ such that} \\ \mathcal{A}_z \subseteq \{x \in \mathbb{R}^n : \ f(x, 0, y, s) \leq 0\} \\ \mathcal{B}_z \subseteq \{x \in \mathbb{R}^n : \ f(x, 0, y, s) \geq 0\} \end{cases}$$

for the shifted datasets

$$\mathcal{A}_z := \{a - z : a \in \mathcal{A}\}, \ \ \mathcal{B}_z := \{b - z : b \in \mathcal{B}\}.$$

Thus, one can exploit the experience gained in our previous work [2] concerning the homogeneous case. The problem $(\mathrm{HCS})_z$ is about finding an homogeneous revolution cone

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

that separates the pair $(\mathcal{A}_z, \mathcal{B}_z)$.

Further comments on terminology are in order. In the parlance of set-valued analysis, W_* corresponds to the graph

$$\mathrm{gr}(M) = \{(z,y,s) \in W : (y,s) \in M(z)\}$$

of the set-valued map $M : Z \rightrightarrows Y \times S$. Since M may take empty values, it is helpful to distinguish between the set Z of candidate apices and the smaller set

$$\begin{aligned} Z_* &:= \{z \in Z : M(z) \neq \emptyset\} \\ &= \Pi_1(W_*) \\ &= \{\Pi_1 w : \Gamma(w) \text{ is a separator}\} \end{aligned} \tag{3}$$

of *admissible apices.*

Proposition 2.1. *The map $M : Z \rightrightarrows Y \times S$ is upper-semicontinuous. In particular, Z_* is a closed subset of Z.*

Proof. The proof is based on standard arguments of set-valued analysis. First of all, the graph of M is a closed subset of W. On the other hand,

$$\mathrm{Im}(M) := \bigcup_{z \in Z} M(z)$$

is contained in a compact set, namely $Y \times S$. It follows that M is upper-semicontinuous, i.e.,

$$M^-(U) := \{z \in Z : M(z) \cap U \neq \emptyset\}$$

is a closed subset of Z whenever U is a closed subset of $Y \times S$. In particular, $Z_* = M^-(Y \times S)$ is a closed subset of Z. □

By analogy with the characterization (3) of Z_*, each element of the set

$$Z_*^{\mathrm{st}} := \{\Pi_1 w : \Gamma(w) \text{ is a strict separator}\}$$

is called a *strictly admissible apex.* The following result comes without surprise.

Proposition 2.2. *The map $M : Z \rightrightarrows Y \times S$ is lower-semicontinuous at each point of Z_*^{st}. In particular, Z_*^{st} is open as subset of Z.*

Proof. Let $z_* \in Z_*^{\mathrm{st}}$. Let U be an open subset of $Y \times S$ such that $M(z_*) \cap U \neq \emptyset$. Pick any (y_*, s_*) in the above intersection. In particular, one has

$$\begin{cases} f(a_i, z_*, y_*, s_*) < 0 & \text{for all } i \in \mathbb{N}_p \\ f(b_j, z_*, y_*, s_*) > 0 & \text{for all } j \in \mathbb{N}_q. \end{cases} \tag{4}$$

Since f is continuous, there exists $\varepsilon > 0$ such that (4) remains true if z_* is changed by any point in the ball

$$B_\varepsilon^Z(z_*) := \{z \in Z : \|z - z_*\| < \varepsilon\}.$$

Hence, $M(z) \cap U \neq \emptyset$ for all $z \in B_\varepsilon^Z(z_*)$. This shows that M is the lower-semicontinuous at z_*. The above reasoning applied to the particular choice $U = Y \times S$ proves that z_* belongs to the interior of Z_*^{st} relative to the metric space Z. □

Clearly, the existence of a separator (respectively, a strict separator) is equivalent to the existence of an admissible apex (respectively, a strictly admissible apex).

Theorem 2.3. *Let $z \in Z$. Then*

(a) z is strictly admissible if and only if the polytopes

$$P_z := \operatorname{co}\left[\left\{\frac{a-z}{\|a-z\|} : a \in \mathcal{A}\backslash\{z\}\right\}\right], \tag{5}$$

$$Q_z := \operatorname{co}\left[\{0\} \bigcup \left\{\frac{b-z}{\|b-z\|} : b \in \mathcal{B}\backslash\{z\}\right\}\right] \tag{6}$$

do not intersect.

(b) z is admissible if and only if P_z and Q_z do not have a common interior point. In particular, $(\mathcal{A}, \mathcal{B})$ is conically separable if and only if there exists $z_* \in Z$ such that $\operatorname{int}(P_{z_*}) \cap \operatorname{int}(Q_{z_*}) \neq \emptyset$.

Proof. Statements (a) and (b) are obtained by applying [2, Theorem 2.3] to the homogeneous conic separation problem $(\mathrm{HCS})_z$. We omit writing down the details because the homogeneous case is discussed in length in [2]. □

Remark 2.4. If Z does not intersect the sample $\mathcal{A} \cup \mathcal{B}$, then the polytopes (5) and (6) take the simpler form

$$P_z = \operatorname{co}\{\mathbf{a_1}(z), \ldots, \mathbf{a_p}(z)\}, \tag{7}$$

$$Q_z = \operatorname{co}\{0, \mathbf{b_1}(z), \ldots, \mathbf{b_q}(z)\}. \tag{8}$$

Here the $\mathbf{a_i}$'s and $\mathbf{b_j}$'s are continuous functions on Z defined respectively by

$$\mathbf{a_i}(z) := \|a_i - z\|^{-1}(a_i - z),$$

$$\mathbf{b_j}(z) := \|b_j - z\|^{-1}(b_j - z).$$

Both (7) and (8) depend continuously (with respect to the Pompeiu-Hausdorff distance) on the parameter $z \in Z$.

3 Smallest angle separators

Geometrically speaking, the aperture coefficient s corresponds to the cosinus of the half-aperture angle of the revolution cone (1). Hence, a large aperture coefficient reflects a small half-aperture angle.

Definition 3.1. *Suppose that W_* is nonempty. If $\bar{w} = (\bar{z}, \bar{y}, \bar{s})$ solves*

$$\begin{cases} \text{maximize } s \\ (z, y, s) \in W_*, \end{cases} \tag{9}$$

then $\Gamma(\bar{w})$ is called a smallest angle separator (SAS). *The optimal value and solution set to the above maximization problem are denoted by $s_{\max}$ and W_{sas}, respectively.*

The set W_* is closed because Z is assumed to be closed. However, this fact alone does not guarantee the existence of a solution to (9). In other words, the set

$$W_{\rm sas} = \{w \in W_* : \Gamma(w) \text{ is a SAS}\}$$

could be empty. The question concerning the existence of a SAS can be settled for instance by asking Z to be compact and keeping of course the hypothesis that W_* is nonempty.

On the other hand, $W_{\rm sas}$ could have more than one element, i.e., the smallest angle criterion is not enough to single out a unique separator. The next proposition is a uniqueness result for smallest angle separators with a prescribed apex.

Proposition 3.2. *Assume the existence of a separator with positive aperture coefficient, i.e.,*

$$\exists\, \tilde{w} \in W_* \ \textit{such that}\ \Pi_3 \tilde{w} > 0. \tag{10}$$

Let $w_1 = (z_1, y_1, s_1)$ *and* $w_2 = (z_2, y_2, s_2)$ *be any two solutions to* (9)*. Then,*

$$s_1 = s_2 = s_{\max}\,, \tag{11}$$

$$z_1 = z_2 \ \textit{ implies } \ y_1 = y_2. \tag{12}$$

Proof. Since (9) is assumed to be solvable, the optimal value $s_{\max}$ is finite. Since w_1 is a solution to (9), one has $s_1 = s_{\max}$. For a similar reason, one has $s_2 = s_{\max}$. This takes care of (11). We now concentrate on (12). Let $\bar{z} := z_1 = z_2$. For proving $y_1 = y_2$, we start by observing that

$$s_{\max} \ = \ \sup_{z \in Z} v(z) \ = \ \sup_{z \in Z_*} v(z),$$

where $v(z)$ denotes the optimal value of the maximization problem

$$\begin{cases} \text{maximize } s \\ (y,s) \in M(z). \end{cases} \tag{13}$$

The interpretation of (13) is clear: one is searching for an homogeneous revolution cone $\Gamma_0(y,s)$ of smallest half-aperture angle that separates the pair $(\mathcal{A}_z, \mathcal{B}_z)$. Let us examine the above problem for the particular choice $z = \bar{z}$. Assumption (10) implies that $s_{\max} = v(\bar{z})$ is positive. Hence, one can apply [2, Theorem 3.1] and conclude that

$$M_{\rm sas}(\bar{z}) := \operatorname{argmax}\{s : (y,s) \in M(\bar{z})\}$$

is a singleton. Since $(y_1, s_{\max})$ and $(y_2, s_{\max})$ are both in $M_{\rm sas}(\bar{z})$, it follows that $y_1 = y_2$. □

As shown in the proof of Proposition 3.2, under assumption (10) one can construct a bijection between $W_{\rm sas}$ and the set

$$Z_{\rm sas} \ := \ \Pi_1(W_{\rm sas}) \ = \ \{\Pi_1 w : \Gamma(w) \text{ is a SAS}\}.$$

One gets in this way the following cardinality result.

Corollary 3.3. *If one assumes the existence of a separator with positive aperture coefficient, then*

$$\operatorname{card}(W_{\rm sas}) = \operatorname{card}(Z_{\rm sas}).$$

In particular, the number of smallest angle separators cannot exceed the number of admissible apices.

4 Largest angle separators

We now focuss the attention instead on separators that have a large half-aperture angle. When the aperture coefficient s is near 0, the revolution cone (1) is close to a half-space. In other words, conic separation with a small value of s resembles to classical separation by means of hyperplanes.

Definition 4.1. *Suppose that W_* is nonempty. If $\bar{w} = (\bar{z}, \bar{y}, \bar{s})$ solves*

$$\begin{cases} \text{minimize } s \\ (z, y, s) \in W_*, \end{cases} \tag{14}$$

then $\Gamma(\bar{w})$ is called a largest angle separator (LAS). *The optimal value and solution set to the above minimization problem are denoted by $s_{\min}$ and W_{las}, respectively.*

As done in the previous section, one can state a decomposition formula of the type

$$s_{\min} \;=\; \inf_{z \in Z} u(z) \;=\; \inf_{z \in Z_*} u(z),$$

where $u(z)$ denotes the optimal value of

$$\begin{cases} \text{minimize } s \\ (y, s) \in M(z). \end{cases} \tag{15}$$

The minimization problem (15) is about finding an homogeneous revolution cone $\Gamma_0(y, s)$ of largest half-aperture angle that separates the pair $(\mathcal{A}_z, \mathcal{B}_z)$.

4.1 Numerical computation of a LAS

Let us examine the problem (14) from the point of view of its numerical resolution. Finding a LAS amounts to solve the minimization problem

$$\begin{cases} \text{minimize } s \\ z \in Z, \; \|y\| = 1, \, 0 \leq s \leq 1 \\ s\,\|a_i - z\| - y^T(a_i - z) \leq 0 \quad \text{for all } i \in \mathbb{N}_p \\ s\,\|b_j - z\| - y^T(b_j - z) \geq 0 \quad \text{for all } j \in \mathbb{N}_q. \end{cases} \tag{16}$$

We have kept the inequality $s \leq 1$ even if it redundant. The reason is that, parallel to the homogeneous case, we resort to a relaxed version of problem (16) where such constraint plays a role. In fact we tackle the problem

$$h_* = \inf \{h(z, y, s) : \; z \in Z, \; \|y\| = 1, s \in [0, 1]\}, \tag{17}$$

where $h : \mathbb{R}^{2n+1} \to \mathbb{R}$ is given by

$$h(z, y, s) := \gamma s + E(z, y, s).$$

One views $\gamma > 0$ as a parameter expressing the tradeoff between the two objectives of maximizing the half-aperture angle and minimizing the classification error

$$E(z, y, s) := E_{\mathcal{A}}(z, y, s) + E_{\mathcal{B}}(z, y, s),$$

where

$$E_{\mathcal{A}}(z,y,s) := \sum_{i=1}^{p} \max\{0, -(a_i - z)^T y + \|a_i - z\| s\},$$

$$E_{\mathcal{B}}(z,y,s) := \sum_{j=1}^{q} \max\{0, (b_j - z)^T y - \|b_j - z\| s\}.$$

The function h is clearly continuous. The following example, which is merely academic, shows that the infimal value (17) may not be attained.

Example 4.2. Let $\mathcal{A} = \{(-2,0),(4,0)\}$ and $\mathcal{B} = \{(-1,0),(2,-1)\}$. As collection of candidate apices, consider the closed unbounded set

$$Z = \{(c,d) \in \mathbb{R}^2 : c \geq 1, cd \geq 1\}.$$

In such a case, the infimal value (17) is equal to 0. Indeed, for the sequence

$$(z_k, y_k, s_k) = ((k, 1/k), (0,1), 0)$$

one sees that $h(z_k, y_k, s_k) = 2/k$ goes to 0 as k goes to infinity. On the other hand, there is no $(z,y,s) \in Z \times Y \times [0,1]$ such that $h(z,y,s) = 0$.

A simple way to ensure that (17) admits a solution is to ask Z to be compact. This is what we do in the proposition below.

Proposition 4.3. *Let Z be compact. Then the minimization problem (17) admits a solution and its optimal value can be rewritten as*

$$h_* = \min\{h(z,y,s) : z \in Z, \|y\|^2 \geq 1, s \in [0,1]\}. \tag{18}$$

Furthermore, one of the following two situations occurs:

(a) $h_* = 0$. *This happens exactly when the datasets $\mathcal{A}$ and $\mathcal{B}$ are separable by a hyperplane. The optimal solutions to (18) are then of the form $(\bar{z}, \bar{y}, 0)$ with $\bar{y}$ being normal to a hyperplane which contains $\bar{z}$ and separates $\mathcal{A}$ and $\mathcal{B}$. An optimal solution to (17) is obtained simply by normalizing $\bar{y}$.*

(b) $h_* > 0$. *In this case the minimization problems (17) and (18) have the same solution set. In other words, any solution $(\bar{z}, \bar{y}, \bar{s})$ to (18) is such that $\|\bar{y}\| = 1$.*

Proof. One can write (17) as the concatenation of two minimization problems, namely

$$h_* = \min_{z \in Z} \left\{ \min_{\substack{\|y\|=1 \\ s \in [0,1]}} h(z,y,s) \right\}. \tag{19}$$

The inner minimization in (19) falls within the context of [2, Proposition 4.1]. Hence, the normalization constraint $\|y\| = 1$ can be changed by $\|y\|^2 \geq 1$. The second part of the proposition is also a matter of combining (19) and [2, Proposition 4.1]. The details are omitted. □

4.1.1 Alternating minimization: the first splitting scheme

In view of Proposition 4.3, we operate directly on problem (18). To numerically tackle this problem, we penalize first the nonconvex constraint $\|y\|^2 \geq 1$, that is to say, we consider the penalized version

$$\begin{cases} \text{minimize} \quad h(z,y,s) + \rho \max\{0, 1 - \|y\|^2\} \\ (z,y,s) \in Z \times \mathbb{R}^n \times [0,1], \end{cases} \tag{20}$$

where $\rho > 0$ is viewed as a penalty parameter. Note that (20) can be written in the equivalent form

$$\begin{cases} \text{minimize} \quad h(z,y,s) + \rho \max\{0, \|y\|^2 - 1\} - \rho(\|y\|^2 - 1) \\ (z,y,s) \in Z \times \mathbb{R}^n \times [0,1]. \end{cases} \tag{21}$$

The cost function in (21) is neither convex nor concave. However, one has:

Lemma 4.4. *For each fixed $s \in [0,1]$, the function*

$$(z,y) \mapsto \psi_s(z,y) := h(z,y,s) + \rho \max\{0, \|y\|^2 - 1\} - \rho(\|y\|^2 - 1)$$

can be written as a difference of two convex functions.

Proof. More than the result itself, what is important to us is to construct a particular pair of convex functions ψ'_s and ψ''_s such that $\psi_s = \psi'_s - \psi''_s$. To start with, we observe that a bilinear term like $g(z,y) := z^T y$ admits the representation

$$g(z,y) = \frac{\|z+y\|^2}{4} - \frac{\|z-y\|^2}{4}.$$

Parenthetically, the above representation constitutes an undominated DC decomposition of g in the sense of [3]. Taking also into account the general identity

$$\max\{0, r_1 - r_2\} = \max\{r_1, r_2\} - r_2$$

for all reals r_1 and r_2, one sees that the particular choice

$$\begin{aligned} \psi'_s(z,y) &= \gamma s + \sum_{i=1}^{p} \max\left\{-a_i^T y + \|a_i - z\| s + \frac{1}{4}\|z+y\|^2, \frac{1}{4}\|z-y\|^2\right\} \\ &\quad + \sum_{j=1}^{q} \max\left\{\frac{1}{4}\|z-y\|^2, -b_j^T y + \|b_j - z\| s + \frac{1}{4}\|z+y\|^2\right\} + \rho \max\left\{0, \|y\|^2 - 1\right\}, \\ \psi''_s(z,y) &= \frac{p}{4}\|z-y\|^2 + \frac{q}{4}\|z+y\|^2 + \sum_{j=1}^{q}(-b_j^T y + \|b_j - z\| s) + \rho(\|y\|^2 - 1) \end{aligned}$$

does the job of decomposing ψ_s as requested. □

Concerning the problem (21), the situation is then as follows:

- For a fixed $(z,y) \in Z \times \mathbb{R}^n$, the term $h(z,y,s)$ is polyhedral and convex as function of the scalar variable s. Hence, one can solve

$$\begin{cases} \text{minimize } h(z,y,s) \\ s \in [0,1] \end{cases}$$

  by a univariate minimization procedure of the same type as in [2].

- On the other hand, for a fixed $s \in [0,1]$, the cost term $\psi_s(z,y)$ can be put in a DC (Difference of Convex) format (cf. Lemma 4.4). Hence, one can solve

$$\begin{cases} \text{minimize} \ \ \psi_s(z,y) \\ (z,y) \in Z \times \mathbb{R}^n \end{cases}$$

  by means of the DCA method [1] for DC programming adopted also in [2].

Based on the above two observations, for solving the problem (21) we resort to a block coordinate descent method (cf. [6]) that performs a sequence of alternate minimizations with respect to (z,y) and to s. The details concerning the implementation of such an alternating minimization method are left to Section 4.2.

4.1.2 Alternating minimization: the second splitting scheme

Another way of solving (18) is to use an alternating minimization method coupled with a proximal point-linearized problem approach as described in [2, Section 4.2]. This time the minimization vector (z,y,s) is partitioned in a different way: one alternates between a minimization in the component (y,s) and one in the component z. The motivation behind this new splitting strategy is as follows:

- The machinery developed in [2, Section 4.2] for the homogeneous case is still valid for each fixed z. Hence, one can apply to such minimization phase the mentioned proximal point-linearized problem approach.
- On the other hand, for each fixed (y,s), the function $h(\cdot,y,s)$ can be easily put in a DC format and, consequently, the DCA method can be applied to such minimization phase too.

To see that $h(\cdot,y,s)$ can be put in a DC format it suffices to observe that $E_{\mathcal{A}}(\cdot,y,s)$ is already convex and that

$$E_{\mathcal{B}}(\cdot,y,s) = E'_{\mathcal{B}}(\cdot,y,s) \ - \ E''_{\mathcal{B}}(\cdot,y,s)$$

is a difference of two convex functions, namely

$$\begin{aligned} E'_{\mathcal{B}}(z,y,s) &= \sum_{j=1}^{q} \max\{(b_j - z)^T y, \|b_j - z\| s\} \\ E''_{\mathcal{B}}(z,y,s) &= s\sum_{j=1}^{q} \|b_j - z\|. \end{aligned}$$

4.2 Numerical tests with real-life data

In our numerical experiments we have focused on the problem of finding the largest angle homogeneous separator, which appears promising in view of application to classification area. We have proceeded along the guidelines described in Sections 4.1.1 and 4.1.2.

In particular, we have designed the DCA-II code for the first splitting scheme, where we have implemented the univariate minimization procedure as in [2, Section 4.2.2], combined with the DCA approach. As for the second splitting scheme, we have designed the Proximal-II code, which alternates the proximal point-linearized method [2, Section 4.2] for fixed z and the DCA method for fixed (y,s).

We recall that the DCA method [1] requires iterative solution of many convex optimization problems, which in our case are nonsmooth. To this aim we have embedded in our codes the package NCVX [4], which implements a bundle type approach enabling the resolution of nonsmooth optimization problems both convex and nonconvex.

We have considered the same test problems as in [2], whose list we report in Table 1.

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

As for the parameter setting, we have used a grid of possible values to preliminary tune the parameters γ, ρ in the program (21) and the parameters γ, δ in the proximal point-linearized problem [2, Section 4.2]. 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.

Since both the algorithms we have tested are of the local optimization type, we have implemented a multi-start approach. In particular, for each dataset we have considered several starting points $(z^{(0)}, y^{(0)}, s^{(0)})$, with $z^{(0)} = \mathbf{0}_n$, the zero vector in the sample space $\mathbb{R}^n$, $y^{(0)} = \pm e_i$ (where e_i is the ith unit vector, $i \in \{1, \ldots, n\}$) and $s^{(0)}$ uniformly set equal to 0.5. We have tested yet two other starting points: $(\mathbf{0}_n, \mathbf{1}_n/n, 0.5)$ where $\mathbf{1}_n$ is the vector of ones in $\mathbb{R}^n$ and a "warm starting" one, by letting $(y^{(0)}, s^{(0)})$ be the point obtained at the stop of the DCA and Proximal codes described in [2].

In Table 2 and 3 we consider, respectively, the performance of the codes DCA-II and Proximal-II in terms of average percentage of testing correctness only for the datasets where we have obtained a substantial improvement with respect to the homogeneous case (DCA and Proximal columns). We have selected, for each dataset, the best results obtained in our multistart approach. For useful references we report also the results obtained by using the LIBSVM package [5]. The best result for each dataset has been underlined.

Dataset	Linear-SEP	Conical-SEP	
	LIBSVM	DCA	DCA-II
Cancer	95.54	87.14	<u>95.86</u>
Heart	<u>85.19</u>	80.33	83.00
Pima	<u>76.30</u>	60.13	70.39

Table 2: DCA-II

It appears worth noting that the behavior of the non homogeneous conic separation approach is comparable, in terms of testing correctness, with linear separation implemented by LIBSVM.

Dataset	Linear-SEP	Conical-SEP	
	LIBSVM	Proximal	Proximal-II
Cancer	<u>95.54</u>	87.14	94.14
Diagnostic	95.95	96.84	<u>97.19</u>
Heart	<u>85.19</u>	77.00	82.33
Pima	<u>76.30</u>	64.16	69.22
Ionosphere	87.14	89.71	<u>90.00</u>
g10n	<u>98.91</u>	91.64	91.66

Table 3: Proximal-II

5 By way of conclusion

This paper is about separating a pair of finite sets in $\mathbb{R}^n$ by means of a suitable revolution cone. Below we briefly mention which are the main achievements of our work and also we point out what remains to be done in the near future:

i) In Section 4 we have introduced the concept of largest angle separator (LAS). We have explained in detail how to compute in practice such sort of separator. The numerical experiments reported in Section 4.2 show that a significant improvement in the quality of the separation is obtained with respect to the homogeneous case. From a theoretical point of view, such a conclusion is not so surprising after all: in the non-homogeneous case one has the possibility of placing the apex of the revolution cone in different loci, reducing in this way the classification error.

ii) In Section 3 we have introduced the concept of smallest angle separator (SAS). We have stated a few theoretical results associated to this concept, but we have not explained how to compute a SAS in practice. Of course, the numerical issue at hand is that of solving

$$\left\{ \begin{array}{l} \text{maximize } s \\ z \in Z,\ \|y\| = 1,\ 0 \le s \le 1 \\ s\,\|a_i - z\| - y^T(a_i - z) \le 0 \quad \text{for all } i \in \mathbb{N}_p \\ s\,\|b_j - z\| - y^T(b_j - z) \ge 0 \quad \text{for all } j \in \mathbb{N}_q. \end{array} \right. \tag{22}$$

We have not elaborated on the numerical resolution of this optimization problem, but it is clear that (22) can be treated in a similar way as (16).

iii) An interesting subject of future research is the robustness of the separation by means of a revolution cone. The homogeneous version of a most robust separator has been discussed in [2], but the analysis of the non-homogeneous case remains open.

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, M. Gaudioso, and A. Seeger. Conic separation of finite sets. I. The homogeneous case. *J. Convex Anal.*, 20 (2013).

[3] I.M. Bomze, and M. Locatelli. Undominated d.c. decompositions of quadratic functions and applications to Branch-and-Bound approaches. *Comput. Optim. Appl.*, 28, (2004), 227–245.

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

[5] 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.

[6] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.*, 109, (2001), 475–494.