Visualizing data as objects by DC (difference of convex) optimization

Emilio Carrizosa[1], Vanesa Guerrero[1], and Dolores Romero Morales[2]

[1]Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Seville, Spain
{ecarrizosa, vguerrero}@us.es
[2]Copenhagen Business School, Frederiksberg, Denmark
drm.eco@cbs.dk

Abstract

In this paper we address the problem of visualizing in a bounded region a set of individuals, which has attached a dissimilarity measure and a statistical value. This problem, which extends the standard Multidimensional Scaling Analysis, is written as a global optimization problem whose objective is the difference of two convex functions (DC). Suitable DC decompositions allow us to use the DCA algorithm in a very efficient way. Our algorithmic approach is used to visualize two real-world datasets.

Keywords: Data Visualization, DC functions, DC algorithm, Multidimensional Scaling Analysis

1 Introduction

In the Big Data era, Data Visualization is an area of interest to specialists from a wide variety of disciplines, [14, 15, 26, 27]. The information managed must be processed and, what is even more important, understood. Data Visualization techniques arise to respond to this requirement by developing specific frameworks to depict complex data structures as easy-to-interpret graphics, [40, 50].

Mathematical Optimization has contributed significantly to the development of this area during recent years, see [13, 30, 42] and the references therein. Nowadays, complex datasets pose new challenges in order to visualize the data in such a way that patterns are captured and useful information is extracted. Special attention is paid to represent the underlying dissimilarity relationships that data may have. Classical dimensionality reduction techniques, such as Principal Component Analysis, [43], or Multidimensional Scaling (MDS), [29, 34, 52], have been customized to deal with more complex data structures, [1, 5, 16], and to make the interpretability of the results easier via, for instance, sparse models, [9, 8, 18].

Apart from adapting existing methods, specific problems may call also for new approaches. For instance, in addition to the dissimilarity measure, the data may have attached a statistical variable, to be related with the size of each object in the graphical representation of the dataset, [20]. This is the case for geographical data, to be visualized on a map in which countries are resized according to, for instance, population rates, but maintaining the neighboring relationships of countries. This type of representations, known as cartograms, [51], leads to plots in which countries are replaced by geometrical objects, frequently circles or rectangles, while the neighborhood relationships and the size of the objects are sought to be well represented. A key issue is how such problems are expressed as optimization programs, and which optimization tools are available to cope with them. For uses of optimization applied to cartograms construction and related visualization frameworks we refer the reader to [6, 10, 11, 20, 21, 28, 31, 47, 49] and references therein.

In this paper we present a new mathematical programming framework to build a visualization map, in which a set of N individuals are depicted as convex objects in a bounded region $\Omega \subset \mathbb{R}^n$, usually $n \leq 3$. These objects must have a volume proportional to a given statistical value associated with the individuals, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_N)$, and they should be placed accordingly to a dissimilarity measure attached to the individuals, $\boldsymbol{\delta} = (\delta_{ij})_{i,j=1,\ldots,N}$. In order to locate the objects in Ω, a reference object $\mathcal{B}$ is used, to be translated and expanded. However, since our final goal is to obtain a visualization map which allows the analysts to understand the data they are working with, a criterion which somehow controls the appearance of the plot needs to be also considered. We will deal with this paradigm by focusing on how the objects are spread out over Ω.

Leaving aside the statistical values $\boldsymbol{\omega}$, the purpose of representing dissimilarities between individuals reminds to MDS, [5, 16, 18, 29, 34, 35, 39, 52], which aims to represent the dissimilarity between individuals as empirical distances between points in an unbounded space of lower dimension. Although our visualization model may seem very close to MDS, it has the special feature of representing in the bounded region Ω not only dissimilarities as distances between objects, but also the statistical measure $\boldsymbol{\omega}$ through the volumes of the objects in Ω. Our visualization tool is able to rescale the dissimilarities between the individuals and the statistical values associated to them to fit in Ω. Observe that fitting the objects into Ω may yield representations in which the objects intersect if their sizes are not small enough, but, on the other hand, too small objects obstruct the visualization of the statistical measure. Ideally the objects should be spread out across the visualization map. This aim will be also taken into account when modeling the problem.

The methodology proposed in this paper has applications in fields others than Data Visualization, such as for instance, Location Analysis or Distance Geometry. In location problems, the facilities to be located are usually considered as points. However, a natural extension is to consider facilities as dimensional structures, see [19], and DC techniques have been specifically applied to this generalization, [3, 12]. Ours can also be seen as a problem in Distance Geometry optimization, as carefully reviewed in [39]. In Distance Geometry, a graph realization problem consists of finding a configuration of points such that their (Euclidean) distances fit a given dissimilarity matrix. Among them is the Sensor Network Location problem, [46, 48, 54, 58], in which one assumes that some individuals are anchors (their location is known) and the remaining ones are sensors, whose location is to be obtained so that their Eculidean distances fit the dissimilarities. Thus, our method can also be applied to the Sensor Network Location problem, in which sensors and anchors have a nonnegligible area.

In this paper, the construction of a visualization map with the three characteristics mentioned above is written as a global biobjective optimization problem with convex constraints. We show that the objective function of the aggregate problem can be expressed as a difference of convex (DC) function, and thus DC optimization tools can be used to solve the optimization program.

The rest of the paper is organized as follows. In Section 2 the biobjective optimization program to build the visualization map is formalized. In Section 3, structural properties of the optimization problem are analyzed. In Section 4, we present our algorithmic approach. Numerical results for two datasets of different size and nature are included in Section 5. Some conclusions and extensions are presented in Section 6.

2 The visualization model

In our model we have a reference object $\mathcal{B}$, which is a compact convex subset of $\mathbb{R}^n$, symmetric with respect to the origin, interior to $\mathcal{B}$. Each individual i is associated with a set of the form $\boldsymbol{c}_i + \tau r_i \mathcal{B}$, where $r_i \geq 0$ is chosen so that the volume of $r_i\mathcal{B}$ is proportional to the statistical value $\omega_i \geq 0$, $\boldsymbol{c}_i$ is a translation vector and τ is a common positive rescaling for all objects. We seek the values of the variables $\boldsymbol{c}_i$, $i = 1, \ldots, N$, and τ so that objects $\boldsymbol{c}_i + \tau r_i \mathcal{B}$ are contained in Ω. The previously described representation is illustrated in Figure 1.

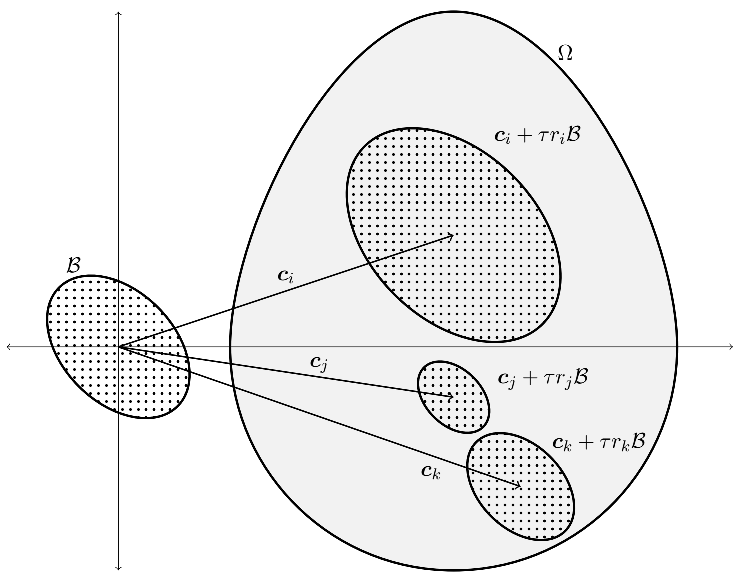

Figure 1: Example in $\mathbb{R}^2$ of a visualization region Ω, a reference object $\mathcal{B}$ and three individuals i, j and k defined through the translation vectors $\boldsymbol{c}_i$, $\boldsymbol{c}_j$ and $\boldsymbol{c}_k$, which are scaled via τr_i, τr_j and τr_k.

Henceforth, we deal with a biobjective optimization problem: the distances between the objects representing the individuals i and j must resemble the dissimilarities δ_{ij} between such individuals, and the objects must be spread out in Ω to make the visualization easier. The two criteria are formalized in what follows.

2.1 First objective: distances resemble dissimilarities

Regarding the first objective, a function d, which gives us a strictly positive distance between two non-intersecting objects representing individuals i and j and zero otherwise, needs to be considered. Thus, we define the function g, which assigns such distance to two individuals i and

j, as follows

$$
\begin{array}{llll}
g: & \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
 & (\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) & \longmapsto & d(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B}).
\end{array} \tag{1}
$$

Then, to quantify the resemblance between the distances in the visualization map and the dissimilarities, the summation over all the individuals of the squared differences between the distances and the rescaled dissimilarities through a positive variable κ will be minimized. Thus, we consider as first objective the function F_1 defined as

$$
\begin{array}{llll}
F_1: & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
 & (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau, \kappa) & \longmapsto & \displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \left[g(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) - \kappa \delta_{ij} \right]^2 .
\end{array}
$$

Observe that for simplicity all pairs (i, j) are considered in the summation in F_1, but our analysis remains valid if only some (i, j) pairs of objects, as done e.g. in [53].

2.2 Second objective: spread

To avoid that the objects collapse in a small subregion of Ω, we encourage objects to be spread out all over Ω. There are several ways to model spread. For instance, we could use the overall volume occupied by the objects, the amount of intersections between them, or the distances between the objects. This last option is the one analyzed in detail in this paper, and therefore, our aim is to maximize the sum over all the individuals of the distances between the objects representing them. Let F_2 be a function which, given the translation vectors $\boldsymbol{c}_i$, and the rescaling parameter, τ, computes the spread of the visualization map in such way. Then, written in minimization form, one has

$$
\begin{array}{llll}
F_2: & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
 & (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau) & \longmapsto & -\displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau).
\end{array}
$$

Note that F_2 does not distinguish between how much the objects intersect, since it penalizes in the same way two objects one on top of the other as two tangent objects. A possible way to quantify the amount of intersection between two objects is by measuring the minimum-norm translation of such objects which makes them not to intersect. This leads to the concept of penetration depth, [23, 56].

Let $\|\cdot\|$ be a norm in $\mathbb{R}^n$. Given two convex compact sets, $A_1, A_2 \in \mathbb{R}^n$, the penetration depth of A_1, A_2 is defined as

$$
\pi(A_1, A_2) = \min_{\boldsymbol{p}} \left\{ \|\boldsymbol{p}\| : \, int(\boldsymbol{p} + A_1) \cap A_2 = \emptyset \right\},
$$

where *int* denotes the interior of a set.

Thus, the amount of intersection between the objects in the visualization map can be quantified as the sum over all the individuals of the squared penetration depth between pairs of them, yielding the function F_2^{Π} defined as

$$
\begin{array}{rlll}
F_2^{\Pi}: & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
& (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau) & \longmapsto & \displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \pi^2 \left(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B} \right).
\end{array}
$$

However, the penetration depth does not measure how separated the objects are. Then, an alternative to the two previous spread criteria, namely F_2 and F_2^{Π}, which does take into account both the amount of intersection and the separation of the objects, consists of measuring the distance between the centers of the objects. Maximizing the sum over all the individuals of the squared distances between the centers gives an alternative spread criterion, namely

$$
\begin{array}{rlll}
F_2^{\boldsymbol{c}}: & \mathbb{R}^n \times \ldots \times \mathbb{R}^n \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\
& (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau) & \longmapsto & -\displaystyle\sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2.
\end{array}
$$

2.3 Problem statement

The problem of building a visualization map in which a set of convex objects in the form $\boldsymbol{c}_i + \tau r_i \mathcal{B}$ are represented in a region Ω, satisfying that the distances between the objects resemble the dissimilarities between the individuals and the map is spread enough, can be stated as a biobjective optimization problem. By proceeding in the usual way, we consider the convex combination of the objectives and solve the aggregate problem, see [22]. Thus, given $\lambda \in [0,1]$, the Visualization Map problem, (VM), is stated as follows

$$
\begin{array}{cl}
\displaystyle\min_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N,\tau,\kappa} & \lambda F_1(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau, \kappa) + (1-\lambda) F_2(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \tau) \\
\text{s.t.} & \boldsymbol{c}_i + \tau r_i \mathcal{B} \subseteq \Omega, \ i = 1, \ldots, N \\
& \tau \in T \\
& \kappa \in K,
\end{array} \tag{VM}
$$

where $K, T \subset \mathbb{R}^+$.

3 Properties

In this section we study the structure of problem (VM). We will prove that its objective function is DC, by considering distance functions d, defined in the space of compact convex sets of $\mathbb{R}^n$, which satisfy the following:

Assumption 1. *The function d, defined on pairs of compact convex sets of $\mathbb{R}^n$, satisfies for any A_1, A_2*

(i) $d \geq 0$ *and* d *is symmetric*

(ii) $d(A_1, A_2) = d(A_1 + z, A_2 + z)$, $\forall z \in \mathbb{R}^n$

(iii) The function $d_z : z \in \mathbb{R}^n \longmapsto d(z + A_1, A_2)$ *is convex and satisfies for all* $\theta > 0$ *that* $d_z(\theta A_1, \theta A_2) = \theta d_{\frac{1}{\theta} z}(A_1, A_2)$.

Typical instances of d satisfying (i)-(iii) are

1. The infimum distance, defined as

$$d(A_1, A_2) = \inf\{\|a_1 - a_2\| : a_1 \in A_1, a_2 \in A_2\} \tag{d1}$$

2. The supremum distance, defined as

$$d(A_1, A_2) = \sup\{\|a_1 - a_2\| : a_1 \in A_1, a_2 \in A_2\} \tag{d2}$$

3. The average distance, defined as

$$d(A_1, A_2) = \frac{1}{vol_n(A_1) vol_n(A_2)} \int \|a_1 - a_2\| d\mu_1 d\mu_2, \tag{d3}$$

where $vol_n(\cdot)$ denotes the volume of a set in $\mathbb{R}^n$ and μ_1, μ_2 are probability distributions with support A_1 and A_2.

Note that the functions d defined in (*d1*)–(*d3*). correspond with the well-known single linkage, complete linkage and the average distances in Cluster Analysis, [30].

Observe that, thanks to the Assumption 1, the distance between two objects representing individuals i and j, given by the function g in (1), can be expressed as

$$g(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) = \tau d_{\frac{1}{\tau}(\boldsymbol{c}_i - \boldsymbol{c}_j)}(r_i \mathcal{B}, r_j \mathcal{B}), \tag{2}$$

and thus g is the perspective of the convex function $f(\boldsymbol{c}_i, \boldsymbol{c}_j) = d_{\boldsymbol{c}_i - \boldsymbol{c}_j}(\omega_i \mathcal{B}, \omega_j \mathcal{B})$. Hence, g is convex as well, see e.g. [32] for the proof.

Elementary tools of DC optimization enable us to show that objective function in (VM), namely $\lambda F_1 + (1-\lambda) F_2$, is DC, and a DC decomposition can be given. The result is presented in Proposition 1 and the proof is included in the Appendix for the sake of completeness.

Proposition 1. *One has that $\lambda F_1 + (1-\lambda) F_2$ is DC, and a decomposition is given by*

$$\lambda F_1 + (1-\lambda) F_2 = u - (\lambda F_1 + (1-\lambda) F_2),$$

where

$$u = \sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \max\{3\lambda - 1, 0\} g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) + 2\lambda(\kappa \delta_{ij})^2$$

The two alternative functions for the spread presented in Section 2, namely F_2^{Π} and $F_2^{\boldsymbol{c}}$, are also DC functions, as stated in the following results.

Proposition 2. *Let h_{ij} be defined as the penetration depth between $\boldsymbol{c}_i + \tau r_i \mathcal{B}$ and $\boldsymbol{c}_j + \tau r_j \mathcal{B}$, namely*

$$\begin{array}{rlll} h_{ij}: & \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^+ & \longrightarrow & \mathbb{R}^+ \\ & (\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) & \longmapsto & \pi\left(\boldsymbol{c}_i + \tau r_i \mathcal{B}, \boldsymbol{c}_j + \tau r_j \mathcal{B}\right). \end{array}$$

Denoting as $\gamma^\circ_{\mathcal{B}}$ the dual norm with unit ball $\mathcal{B}$, one has that h_{ij} is DC, and it has a decomposition in $h_{ij} = u - (u - h_{ij})$, where

$$u \;=\; \max\left\{ \max_{\substack{\boldsymbol{\xi} \in \mathbb{R}^n \\ \|\boldsymbol{\xi}\|=1}} \left\{ \boldsymbol{\xi}^\top (\boldsymbol{c}_j - \boldsymbol{c}_i) - \tau(r_i + r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi}), 0 \right\} \right\},$$

Proof. See Appendix. □

Corollary 1. *One has that the function $\lambda F_1 + (1-\lambda)F_2^{\Pi}$ is DC.*

Proof. The function F_1 is DC. Indeed, it is sufficient to take $\lambda = 1$ in Proposition 1. F_2^{Π} is also DC by using Proposition 2 and Proposition 3.7 in [55]. Then, since the summation of DC function is also DC, the result holds. □

Corollary 2. *One has that the function $\lambda F_1 + (1-\lambda)F_2^{\boldsymbol{c}}$ is DC.*

Proof. Since the function F_1 is DC (take $\lambda = 1$ in Proposition 1) and $F_2^{\boldsymbol{c}}$ is concave, since it is minus the summation of squares of a nonnegative convex function, the result holds. □

Corollaries 1–2 state that the functions $\lambda F_1 + (1-\lambda)F_2^{\Pi}$ and $\lambda F_1 + (1-\lambda)F_2^{\boldsymbol{c}}$ are DC. DC decompositions for them are readily available from the DC decomposition of F_1 in Proposition 1 ($\lambda = 1$), Proposition 2 and the concavity of $F_2^{\boldsymbol{c}}$.

Showing that a function is DC and giving explicitly a DC decomposition enables us to use DC optimization algorithms. It is well known that the performance of the procedures may strongly depend on the choice of the DC decomposition, [2, 4, 24]. We give now an alternative DC decomposition of the form of those addressed in [37, 45], namely, a DC decomposition involving a quadratic convex separable function. In the rest of the paper, we work with the expression of d given by (*d1*), namely the infimum distance. We will show in Section 4 that such alternative decomposition yields a simple DCA algorithm, whose convergence follows from the general convergence results of DCA, [36, 38, 44].

Proposition 3. *The function $\lambda F_1 + (1-\lambda)F_2$, where d is the infimum distance (d1), can be expressed as a DC function, $\lambda F_1 + (1-\lambda)F_2) = u - (u - \lambda F_1 + (1-\lambda)F_2)$, where the quadratic separable convex function u is given by*

$$u = \max\{3\lambda - 1, 0\} \cdot \left[\sum_{i=1,\dots,N} 8\|\boldsymbol{c}_i\|^2 + \tau^2 \sum_{\substack{i,j=1,\dots,N \\ i\neq j}} \beta_{ij} \right] + 2\lambda\kappa^2 \sum_{\substack{i,j=1,\dots,N \\ i\neq j}} \delta_{ij}^2,$$

where β_{ij} satisfies $\beta_{ij} \geq 2\|r_i b_i - r_j b_j\|^2$ for all $b_i, b_j \in \mathcal{B}$.

Proof. See Appendix. □

4 The algorithmic approach

Propositions 1-3 and Corollaries 1-2 show that (VM), as well as its variants, is an optimization problem with a DC objective function, with a DC decomposition available, and simple constraints. Then, DC optimization tools can be used, either of exact nature for very low dimensional problems, [2, 3], or heuristics, as the DCA, [36, 38, 44]. This is the approach we are following in this paper, and we refer the reader to [17, 35] for alternative mathematical optimization approaches to MDS.

Roughly speaking, DCA consists of an iterative process in which a sequence of convex programs are solved. Given a DC program of the form $\min\{f(x) = u(x) - v(x) : x \in \mathbb{R}^n\}$, at each iteration, the concave part $(-v(x))$ is replaced by its affine majorization at a certain $x_0 \in \mathbb{R}^n$, and the resulting convex problem is then solved. However, running times would be dramatically reduced if a DC decomposition of the objective were available so that the convex optimization problems to be solved at each stage were trivial, in the sense that an explicit expression for the optimal solution is available. This idea has been studied in [37, 45] and it will be customized to our problem in what follows.

When the DCA scheme is applied to problem (VM) with the DC decomposition given in Proposition 3, we see that the convex subproblems to be solved at each stage have the form

$$\begin{array}{ll} \min\limits_{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N,\tau,\kappa} & \left\{ \sum\limits_{i=1,\ldots,N} M_i^{\boldsymbol{c}} \|\boldsymbol{c}_i\|^2 + M^\kappa \kappa^2 + M^\tau \tau^2 + \sum\limits_{i=1,\ldots,N} {\boldsymbol{c}_i}^\top \boldsymbol{q}_i^{\boldsymbol{c}} + p^\kappa \kappa + p^\tau \tau \right\} \\ \text{s.t.} & \quad \boldsymbol{c}_i + \tau r_i \mathcal{B} \subseteq \Omega, \ i = 1, \ldots, N \\ & \quad \tau \in T \\ & \quad \kappa \in K, \end{array}$$

for scalars $M_i^{\boldsymbol{c}}$, M^κ, $M^\tau \geq 0$, vectors $\boldsymbol{q}_i^{\boldsymbol{c}}$ and scalars p^κ and p^τ.

Such problem is written as a two separate problems,

$$\min_{\kappa \in K} \ \left\{ M^\kappa \kappa^2 + p^\kappa \kappa \right\} + \min_{\substack{\boldsymbol{c}_i + \tau r_i \mathcal{B} \subseteq \Omega \\ \tau \in T}} \left\{ \sum_{i=1,\ldots,N} M_i^{\boldsymbol{c}_i} \|\boldsymbol{c}\|_i^2 + {\boldsymbol{c}_i}^\top \boldsymbol{q}_i^{\boldsymbol{c}} + M^\tau \tau^2 + p^\tau \tau^2 \right\} \tag{3}$$

The first problem in (3) is a convex problem in one variable, for which a closed form can be given for its optimal value. The second problem in (3) is separable in the variables $\boldsymbol{c}_i$ if the linking variable τ were fixed at τ_0. For this reason, an alternating strategy seems to be plausible, in which one alternates the optimization of τ for $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N$ fixed, (and this is a one dimensional quadratic problem and thus a closed formula for the optimal solution is readily obtained), and then for τ fixed, the centers $\boldsymbol{c}_i$ are to be optimized. But this is done by solving separately N optimization problems of the form

$$\begin{array}{ll} \min\limits_{\boldsymbol{c}_i} & M_i^{\boldsymbol{c}_i} \|\boldsymbol{c}_i\|^2 + {\boldsymbol{c}_i}^\top \boldsymbol{q}_i^{\boldsymbol{c}} \\ \text{s.t.} & \boldsymbol{c}_i \in \Omega - \tau r_i \mathcal{B}. \end{array} \tag{4}$$

Two particular cases of (4) have an amenable structure, yielding a closed formula for the optimal solution, and thus avoiding any call to numerical optimization routines. Indeed, suppose Ω is a rectangle, for simplicity taken as $[0, 1]^n$, and $\mathcal{B}$ is the disc centered at the origin with radius r_0. Then, the constraint in (4) can be rewritten as

$$\tau_0 r_0 r_i \leq c_{ij} \leq 1 - \tau_0 r_0 r_i, \ j = 1, \ldots, n,$$

and thus (4) is expressed as

$$\sum_{j=1,\ldots,n} \min_{c_{ij}} \left\{ M_i^{\boldsymbol{c}_i} c_{ij}^2 + q_{ij}^{\boldsymbol{c}} c_{ij} : \ \tau_0 r_0 r_i \leq \boldsymbol{c}_{ij} \leq 1 - \tau_0 r_0 r_i \right\} \tag{5}$$

In other words, (4) is decomposed into n one dimensional quadratic problems on an interval, and thus a closed formula is readily obtained for the optimal solution of each problem of the form (5), and thus also for (4).

Similarly, suppose Ω and $\mathcal{B}$ are discs centered at the origin, and radius 1 and r_0 respectively. Then, (4) is rewritten as

$$\begin{array}{ll} \min_{\boldsymbol{c}_i} & M_i^{\boldsymbol{c}_i}\|\boldsymbol{c}_i\|^2 + \boldsymbol{c}_i{}^\top \boldsymbol{q}_i^{\boldsymbol{c}} \\ \text{s.t.} & \|\boldsymbol{c}_i\| \le 1 - \tau_0 r_0 r_i. \end{array} \tag{6}$$

Karush-Kuhn-Tucker conditions immediately yield an expression for the optimal solution of (6).

Summarizing, while DCA could be applied to solve (VM) for an arbitrary DC decomposition of the objective function, we see that the DC decomposition of Proposition 3 is particularly attractive, since, for some convenient choices of Ω (a rectangle or a disc) and $\mathcal{B}$ (a disc) yield a closed formula for the optimal solution of the subproblems to be addressed at each stage of the DCA, thus avoiding the need of using numerical optimization routines. See [37] also for other problems in which this strategy has been successful.

5 Numerical illustrations

The methodology in Section 4 is illustrated using two real-world datasets of diverse nature, to be plotted in two different visualization regions $\Omega \subseteq \mathbb{R}^2$. The DCA algorithm has been coded in C and the experiments have been carried out in a Windows 8.1 PC Intel® Core™ i7-4500U, 16GB of RAM. The first dataset consists of $N = 11$ financial markets across Europe and Asia. The statistical value ω_i relates to the importance of market i relative to the world market portfolio, [25], and the dissimilarity δ_{ij} is based on the correlation between markets i and j, [5]. The second dataset is a social network of $N = 200$ musicians, modeled as a graph, where there is an arc connecting two nodes if one musician was influential on the other, [20]. The statistical value ω_i represents the out degree of node i and the dissimilarity between musicians i and j is based on the shortest distance from node i to j.

Throughout this section, we set $\lambda = 0.9$ and $\mathcal{B}$ equal to the circle centered at $(0, 0)$ with radius equal to one. Since (VM) is a multimodal problem and the DCA may get stuck at a local optimum, 100 runs of a multistart are executed. At each run, 3 steps of an alternating procedure are performed, where each step executes 50 iterations of the DCA to optimize $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \kappa$ for τ fixed, and then τ is solved analytically for the so-obtained $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_N, \kappa$.

Figure 2 plots the financial markets dataset on the visualization region $\Omega = [0, 1] \times [0, 1]$, with the scaling parameters ranging in the intervals $K = T = [0.4, 0.6]$. Observe that, the European markets are clustered above the Asian ones, covering the upper half rectangle. These two clusters are represented with different colours. Figure 3 plots the musicians' social network taking a circular visualization region, namely $\Omega = \mathcal{B}$, with the scaling parameters ranging in the intervals $K = [0.075, 0.100]$ and $T = [0.015, 0.030]$, respectively. In the plot at the top, we find all musicians. In the plot at the bottom, we have highlighted one of the most influential nodes, the Rolling Stones, and the connected nodes: musicians influencing the Rolling Stones (respectively, those influenced by them) can be found in a lighter (respectively darker) colour.

6 Concluding remarks and extensions

In this paper we have addressed the problem of representing, in a so-called visualization region Ω, a set of individuals by means of convex objects so that the distance between the objects fits

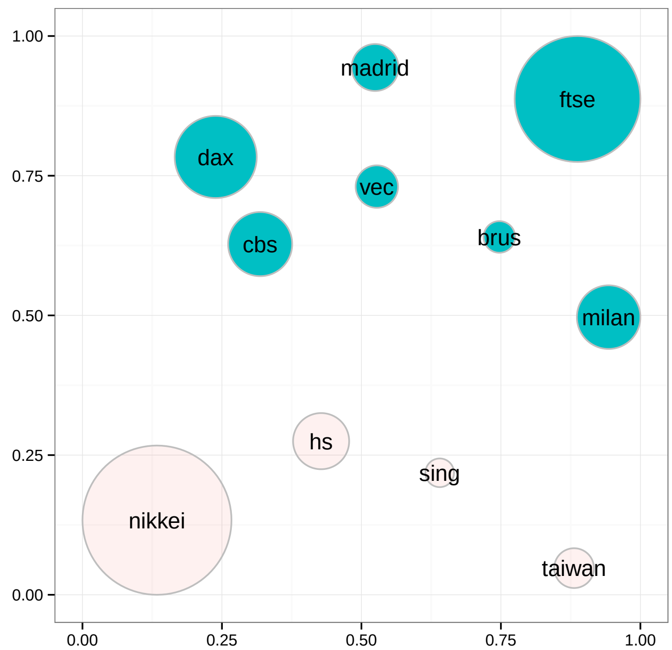

Figure 2: Visualizing financial markets

as close as possible a given dissimilarity matrix, the volume of the objects represents a statistical variable, and, at the same time, the spread of the objects within Ω is maximized.

The problem has been formulated as a DC optimization problem, and the powerful heuristic DCA has been proposed as solution approach. For particular choices of the visualization region Ω (a rectangle and a disc), the reference object (a disc) and the function d (the infimum distance), closed formulas for the optimal solutions of the DCA subproblems are obtained, thus avoiding the need to use numerical optimization routines. The examples presented demonstrate the usefulness of our approach.

Several extensions deserve further analysis.

In the algorithmic section, we have considered the infimum distance ($d1$). Instead, one can consider the supremum distance ($d2$) or the average one ($d3$). It should be observed that the average distance between two convex sets may not have an easy expression, and thus approximations may be needed, [33, 57].

We have assumed the reference object $\mathcal{B}$ to be convex, to guarantee the convexity of the function giving the infimum distance and thus allowing us to express (VM) as a DC optimization problem. For arbitrary sets $\mathcal{B}$ the infimum distance function may not be DC, see [2]. However, as discussed e.g. in [3], important classes of nonconvex sets (e.g. finite union of convex sets) make the infimum distance function DC, and thus the analysis in this paper extends gracefully to such cases. It should be observed that if the supremum distance or the average distance are used instead, then the distance function is convex for arbitrary reference objects, and thus the objective function is DC regardless of the shape of $\mathcal{B}$.

Another promising extension to be modeled is the case in which objects have associated not a dissimilarity $\boldsymbol{\delta}$, but a time series of dissimilarities $\{\boldsymbol{\delta}^s : \ s = 1, \ldots, S\}$. In this case, we

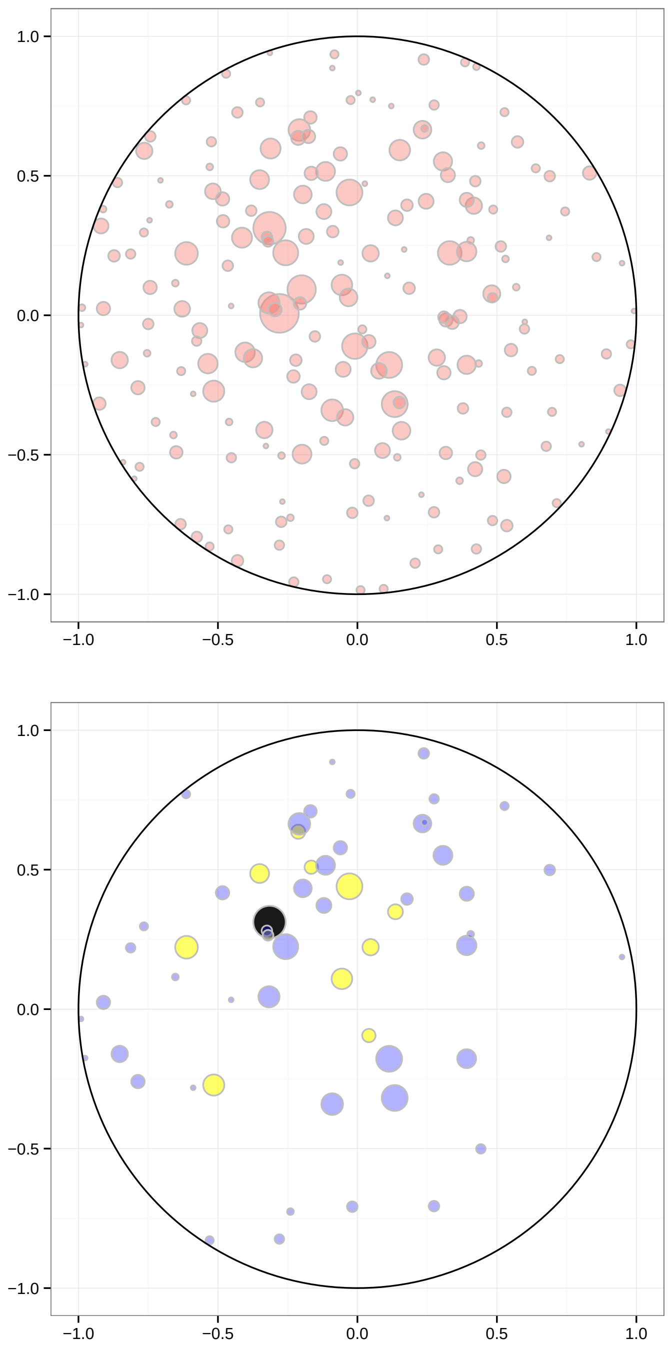

Figure 3: Visualizing the musicians' social network

seek each individual to be represented at each time instant $s = 1, \dots, S$ by an object so that distances between objects are as close as possible to those in $\boldsymbol{\delta}^s$, but, at the same time, smooth transitions take place between the representation at time s and $s+1$, $s = 1, \dots, S-1$. The approach developed in this paper can be adapted to include such smoothness criterion too.

Regarding the optimization, we have proposed DCA as a plausible approach, which can quickly handle problems of non-negligible size since, for convenient choices of Ω and $\mathcal{B}$, (costly) numerical routines are not needed to solve the subproblems at each stage of the DCA.

Convergence to the algorithm to the global optimum is not guaranteed, and thus DCA may get stuck at a local optima. A better performance can be obtained if instead of a uniform multistart, a more guided strategy is used, or if DCA is plugged, as a local search routine, within a strategy which avoids local optima, such as (continuous) Variable Neighborhood Search, [7, 41]. This extension calls for further analysis and testing.

References

[1] H. Abdi, L. J. Williams, D. Valentin, and M. Bennani-Dosse. STATIS and DISTATIS: optimum multitable principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(2):124–167, 2012.

[2] R. Blanquero and E. Carrizosa. Continuous location problems and big triangle small triangle: constructing better bounds. *Journal of Global Optimization*, 45(3):389–402, 2009.

[3] R. Blanquero, E. Carrizosa, and P. Hansen. Locating objects in the plane using global optimization techniques. *Mathematics of Operations Research*, 34(4):837–858, 2009.

[4] I. M. Bomze, M. Locatelli, and F. Tardella. New and old bounds for standard quadratic optimization: dominance, equivalence and incomparability. *Mathematical Programming*, 115(1):31–64, 2008.

[5] I. Borg and P.J.F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005.

[6] K. Buchin, B. Speckmann, and S. Verdonschot. Evolution strategies for optimizing rectangular cartograms. In N. Xiao, M.-P. Kwan, M.F. Goodchild, and S. Shekhar, editors, *Geographic Information Science*, volume 7478 of *Lecture Notes in Computer Science*, pages 29–42. Springer, 2012.

[7] E. Carrizosa, M. Dražić, Z. Dražić, and N. Mladenović. Gaussian variable neighborhood search for continuous optimization. *Computers & Operations Research*, 39(9):2206–2213, 2012.

[8] E. Carrizosa and V. Guerrero. Biobjective sparse principal component analysis. *Journal of Multivariate Analysis*, 132:151–159, 2014.

[9] E. Carrizosa and V. Guerrero. rs-Sparse principal component analysis: A mixed integer nonlinear programming approach with VNS. *Computers & Operations Research*, 52:349–354, 2014.

[10] E. Carrizosa, V. Guerrero, and D. Romero Morales. A multi-objective approach to visualize adjacencies in weighted graphs by rectangular maps. Technical report, IMUS, Sevilla, Spain, 2015.

[11] E. Carrizosa, V. Guerrero, and D. Romero Morales. Piecewise rectangular visualization maps: A large neighborhood seach approach. Technical report, IMUS, Sevilla, Spain, 2015.

[12] E. Carrizosa, M. Muñoz-Márquez, and J. Puerto. Location and shape of a rectangular facility in $\mathbb{R}^n$. Convexity properties. *Mathematical Programming*, 83(1-3):277–290, 1998.

[13] E. Carrizosa and D. Romero Morales. Supervised classification and mathematical optimization. *Computers & Operations Research*, 40(1):150–165, 2013.

[14] C. P. Chen and C.-Y. Zhang. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Information Sciences*, 275:314–347, 2014.

[15] J. Choo and H. Park. Customizing computational methods for visual analytics with big data. *IEEE Computer Graphics and Applications*, 33(4):22–28, 2013.

[16] T. F. Cox and M. A. A. Cox. *Multidimensional scaling*. CRC Press, 2000.

[17] J. De Leeuw and W.J. Heiser. Convergence of correction matrix algorithms for multidimensional scaling. In J.C. Lingoes, E.E. Roskam, and I. Borg, editors, *Geometric Representations of Relational Data*, pages 735–752. Mathesis Press, Ann Arbor, MI, 1977.

[18] V. De Silva and J. B. Tenenbaum. Sparse multidimensional scaling using landmark points. Technical report, Stanford University, 2004.

[19] J.M. Díaz-Báñez, J. A. Mesa, and A. Schöbel. Continuous location of dimensional structures. *European Journal of Operational Research*, 152(1):22–44, 2004.

[20] M. Dörk, S. Carpendale, and C. Williamson. Visualizing explicit and implicit relations of complex information spaces. *Information Visualization*, 11(1):5–21, 2012.

[21] D. Dorling. Area cartograms: their use and creation. In *Concepts and Techniques in Modern Geography series no. 59*. University of East Anglia: Environmental Publications, 1996.

[22] M. Ehrgott. A discussion of scalarization techniques for multiple objective integer programming. *Annals of Operations Research*, 147(1):343–360, 2006.

[23] A. Elkeran. A new approach for sheet nesting problem using guided cuckoo search and pairwise clustering. *European Journal of Operational Research*, 231(3):757–769, 2013.

[24] A. Ferrer and J. E. Martínez-Legaz. Improving the efficiency of DC global optimization methods by improving the DC representation of the objective function. *Journal of Global Optimization*, 43(4):513–531, 2009.

[25] T. Flavin, M. Hurley, and F. Rousseau. Explaining stock market correlation: A gravity model approach. *The Manchester School*, 70:87–106, 2002.

[26] K. Fountoulakis and J. Gondzio. Performance of first- and second-order methods for big data optimization. Technical Report ERGO-15-005, 2015.

[27] K. Fountoulakis and J. Gondzio. A second-order method for strongly convex ℓ_1-regularization problems. Forthcoming in *Mathematical Programming*, 2015.

[28] E. Gomez-Nieto, F. San Roman, P. Pagliosa, W. Casaca, E. S. Helou, M. C. F. de Oliveira, and L. G. Nonato. Similarity preserving snippet-based visualization of web search results. *IEEE Transactions on Visualization and Computer Graphics*, 20(3):457–470, 2014.

[29] J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–338, 1966.

[30] P. Hansen and B. Jaumard. Cluster analysis and mathematical programming. *Mathematical Programming*, 79(1-3):191–215, 1997.

[31] R. Heilmann, D. A. Keim, C. Panse, and M. Sips. Recmap: Rectangular map approximations. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 33–40. IEEE Computer Society, 2004.

[32] J.B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms*. Springer, 1993.

[33] T. Koshizuka and O. Kurita. Approximate formulas of average distances associated with regions and their applications to location problems. *Mathematical Programming*, 52(1-3):99–123, 1991.

[34] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[35] H. A. Le Thi and T. Pham Dinh. D.C. Programming Approach to the Multidimensional Scaling Problem. In A. Migdalas, P.M. Pardalos, and P. Värbrand, editors, *From Local to Global Optimization*, volume 53 of *Nonconvex Optimizations and Its Applications*, pages 231–276. Springer, 2001.

[36] H. A. Le Thi and T. Pham Dinh. DC programming approaches for distance geometry problems. In A. Mucherino, C. Lavor, L. Liberti, and N. Maculan, editors, *Distance Geometry*, pages 225–290. Springer, 2013.

[37] H.A. Le Thi. An efficient algorithm for globally minimizing a quadratic function under convex quadratic constraints. *Mathematical programming*, 87:401–426, 2000.

[38] H.A. Le Thi and T. Pham Dinh. The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems. *Annals of Operations Research*, 133(1-4):23–46, 2005.

[39] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino. Euclidean distance geometry and applications. *SIAM Review*, 56(1):3–69, 2014.

[40] S. Liu, W. Cui, Y. Wu, and M. Liu. A survey on information visualization: recent advances and challenges. *The Visual Computer*, 30(12):1373–1393, 2014.

[41] N. Mladenović, M. Dražić, V. Kovačevic-Vujčić, and M. Čangalović. General variable neighborhood search for the continuous optimization. *European Journal of Operational Research*, 191(3):753–770, 2008.

[42] S. Olafsson, X. Li, and S. Wu. Operations research and data mining. *European Journal of Operational Research*, 187(3):1429–1448, 2008.

[43] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2:559–572, 1901.

[44] T. Pham Dinh and H.A. Le Thi. Convex analysis approach to d.c. programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica*, 22(1):289–355, 1997.

[45] T. Pham Dinh and H.A. Le Thi. A branch-and-bound method via DC optimization algorithm and ellipsoidal technique for box constrained nonconvex quadratic programming problems. *Journal of Global Optimization*, 13:171–206, 1998.

[46] T. K. Pong and P. Tseng. (Robust) edge-based semidefinite programming relaxation of sensor network localization. *Mathematical programming*, 130(2):321–358, 2011.

[47] R. L. Rabello, G. R. Mauri, G. M. Ribeiro, and L. A. N. Lorena. A clustering search metaheuristic for the point-feature cartographic label placement problem. *European Journal of Operational Research*, 234(3):802–808, 2014.

[48] A. M.-C. So and Y. Ye. Theory of semidefinite programming for sensor network localization. *Mathematical Programming*, 109(2-3):367–384, 2007.

[49] B. Speckmann, M. van Kreveld, and S. Florisson. A linear programming approach to rectangular cartograms. In *Proceedings of the 12th International Symposium on Spatial Data Handling*, pages 527–546. Springer, 2006.

[50] J. Thomas and P.C. Wong. Visual analytics. *IEEE Computer Graphics and Applications*, 24(5):20–21, 2004.

[51] W. Tobler. Thirty five years of computer cartograms. *Annals of the Association of American Geographers*, 94(1):58–73, 2004.

[52] W.S. Torgerson. *Theory and Methods of Scaling.* Wiley, 1958.

[53] M.W. Trosset. Extensions of classical multidimensional scaling via variable reduction. *Computational Statistics*, 17:147–163, 2002.

[54] P. Tseng. Second-order cone programming relaxation of sensor network localization. *SIAM Journal on Optimization*, 18(1):156–185, 2007.

[55] H. Tuy. *Convex Analysis and Global Optimization.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[56] S. Umetani, M. Yagiura, S. Imahori, T. Imamichi, K. Nonobe, and T. Ibaraki. Solving the irregular strip packing problem via guided local search for overlap minimization. *International Transactions in Operational Research*, 16(6):661–683, 2009.

[57] R. Vaughan. Approximate formulas for average distances associated with zones. *Transportation Science*, 18(3):231–244, 1984.

[58] Z. Wang, S. Zheng, Y. Ye, and S. Boyd. Further relaxations of the semidefinite programming approach to sensor network localization. *SIAM Journal on Optimization*, 19(2):655–673, 2008.

Appendix

Proof of Proposition 1

One has

$$\begin{aligned}
&\lambda F_1 + (1-\lambda)F_2 = \\
&\quad = \sum_{\substack{i,j=1,\dots,N \\ i\neq j}} \left\{ \lambda \left[g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) - \kappa \delta_{ij} \right]^2 - (1-\lambda) g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) \right\} \\
&\quad = \sum_{\substack{i,j=1,\dots,N \\ i\neq j}} \left\{ (3\lambda - 1) g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) + 2\lambda\kappa^2\delta_{ij}^2 - \lambda (g(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) + \kappa\delta_{ij})^2 \right\}
\end{aligned}$$

In Section 2.1, the convexity of the function g was stated. Moreover, since g, λ, $\delta_{ij} \geq 0$, then $g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau)$, $2\lambda\kappa^2\delta_{ij}^2$ and $(g(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) + K\delta_{ij})^2$ are convex. Finally, $(3\lambda - 1)g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau)$ is convex for $3\lambda - 1 \geq 0$ and concave otherwise. $\square$

Proof of Proposition 2

For convex sets A_1, A_2, the condition in Definition 2.2 is equivalent to the existence of a separating hyperplane between the sets $\boldsymbol{p} + A_1$ and A_2, i.e., of some $\boldsymbol{\xi} \neq 0$, such that

$$\boldsymbol{\xi}^\top(\boldsymbol{p} + \boldsymbol{a}_1) \leq \boldsymbol{\xi}^\top \boldsymbol{a}_2 \qquad \forall \boldsymbol{a}_1 \in A_1, \boldsymbol{a}_2 \in A_2.$$

Without loss of generality, we can consider $\|\boldsymbol{\xi}\| = 1$ and thus we have

$$\begin{aligned}
\pi(A_1, A_2) \quad = \quad & \min_{\boldsymbol{p},\boldsymbol{\xi}\in\mathbb{R}^n} && \|\boldsymbol{p}\| \\
& \text{s.t.} && \boldsymbol{\xi}^\top(\boldsymbol{p} + \boldsymbol{a}_1) \leq \boldsymbol{\xi}^\top \boldsymbol{a}_2 \qquad \forall \boldsymbol{a}_1 \in A_1, \boldsymbol{a}_2 \in A_2 \\
& && \|\boldsymbol{\xi}\| = 1.
\end{aligned}$$

Thus, h_{ij} can be written as follows

$$\begin{aligned}
h_{ij}(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) \quad = \quad & \min_{\boldsymbol{p},\boldsymbol{\xi}\in\mathbb{R}^n} && \|\boldsymbol{p}\| \\
& \text{s.t.} && \boldsymbol{\xi}^\top(\boldsymbol{p} + \boldsymbol{c}_i + \tau r_i \boldsymbol{x}_i) \leq \boldsymbol{\xi}^\top(\boldsymbol{c}_j + \tau r_j \boldsymbol{x}_j) \qquad \forall \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{B} \\
& && \|\boldsymbol{\xi}\| = 1.
\end{aligned}$$

Equivalently, the first constraint, i.e.,

$$\boldsymbol{\xi}^\top(\boldsymbol{p} + \boldsymbol{c}_i + \tau r_i \boldsymbol{x}_i) \leq \boldsymbol{\xi}^\top(\boldsymbol{c}_j + \tau r_j \boldsymbol{x}_j) \quad \forall \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathcal{B},$$

can be written as follows,

$$\boldsymbol{\xi}^\top(\boldsymbol{p}+\boldsymbol{c}_i)+\tau r_i \max_{\boldsymbol{x}\in\mathcal{B}} \boldsymbol{\xi}^\top \boldsymbol{x} \le \boldsymbol{\xi}^\top \boldsymbol{c}_j + \tau r_j \min_{\boldsymbol{x}\in\mathcal{B}} \boldsymbol{\xi}^\top \boldsymbol{x}.$$

Let $\gamma^\circ_{\mathcal{B}}$ denote the dual of the norm with unit ball $\mathcal{B}$, i.e.,

$$\gamma^\circ_{\mathcal{B}}(z) = \max_y \{y^\top z : \, y \in \mathcal{B}\}$$

Since $\mathcal{B}$ is assumed to be symmetric with respect to the origin, we have

$$\begin{aligned}\max_{x\in\mathcal{B}} \boldsymbol{\xi}^\top x &= \gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})\\ \min_{x\in\mathcal{B}} \boldsymbol{\xi}^\top x &= -\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi}).\end{aligned}$$

Hence, by replacing the expression of the dual norm in the constraint above, one has

$$\begin{aligned} h_{ij}(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau) \;=\; & \min_{\boldsymbol{p},\boldsymbol{\xi}\in\mathbb{R}^n} && \|\boldsymbol{p}\|\\ & \text{s.t.} && \boldsymbol{\xi}^\top \boldsymbol{p} \le \boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) - \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})\\ &&& \|\boldsymbol{\xi}\| = 1.\end{aligned}$$

For $\boldsymbol{\xi}$ fixed with $\|\boldsymbol{\xi}\| = 1$, let $\eta(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) - \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})$. It follows that the inner minimum in $h_{ij}(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau)$, i.e., for $\boldsymbol{\xi}$ fixed, is the distance from the origin to the halfspace $\boldsymbol{\xi}^\top p \le \eta(\boldsymbol{\xi})$, and such distance equals 0, if 0 belongs to the halfspace, i.e., if $0 \le \boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) - \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})$, and $-\eta(\boldsymbol{\xi})$ else. Hence

$$\begin{aligned} h_{ij}(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau) &= \min_{\substack{\boldsymbol{\xi}\in\mathbb{R}^n\\ \|\boldsymbol{\xi}\|=1}} \max\left\{0, -\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) + \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})\right\}\\ &= \max\left\{0, \min_{\substack{\boldsymbol{\xi}\in\mathbb{R}^n\\ \|\boldsymbol{\xi}\|=1}} -\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) + \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})\right\}\end{aligned}$$

But, for $\boldsymbol{\xi}$ fixed, the function $(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau) \longmapsto -\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) + \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})$ is affine, and thus the function $(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau) \longmapsto \min_{\substack{\boldsymbol{\xi}\in\mathbb{R}^n\\ \|\boldsymbol{\xi}\|=1}} -\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) + \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})$ is the minimum of affine functions, and is thus concave. Hence, h_{ij} is the maximum between 0 and a concave function, which is DC, whose decomposition is

$$\begin{aligned} & h_{ij}(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau) =\\ &\quad = \max\left\{0, \min_{\substack{\boldsymbol{\xi}\in\mathbb{R}^n\\ \|\boldsymbol{\xi}\|=1}} \left\{-\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) + \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})\right\}\right\}\\ &\quad = \max\left\{-\min_{\substack{\boldsymbol{\xi}\in\mathbb{R}^n\\ \|\boldsymbol{\xi}\|=1}} \left\{-\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) + \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})\right\}, 0\right\} + \min_{\substack{\boldsymbol{\xi}\in\mathbb{R}^n\\ \|\boldsymbol{\xi}\|=1}} \left\{-\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) + \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})\right\}\\ &\quad = \max\left\{\max_{\substack{\boldsymbol{\xi}\in\mathbb{R}^n\\ \|\boldsymbol{\xi}\|=1}} \left\{\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) - \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi}), 0\right\}\right\} - \max_{\substack{\boldsymbol{\xi}\in\mathbb{R}^n\\ \|\boldsymbol{\xi}\|=1}} \left\{\boldsymbol{\xi}^\top(\boldsymbol{c}_j-\boldsymbol{c}_i) - \tau(r_i+r_j)\gamma^\circ_{\mathcal{B}}(\boldsymbol{\xi})\right\}\end{aligned}$$

$$= u(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) - (u(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) - h_{ij}(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau)).$$

□

Proof of Proposition 3

Before giving the proof of Proposition 3, the following technical result is needed.

Lemma 1. *Let $\beta_{ij} \in \mathbb{R}$ be such that $\beta_{ij} \geq 2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2$, $\forall \boldsymbol{b}_i, \boldsymbol{b}_j \in \mathcal{B}$. Then, g^2 can be expressed as a DC function, $g^2 = u - (u - g^2)$, where*

$$u = 2\|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 + \beta_{ij}\tau^2.$$

Proof.

$$\begin{aligned}
&g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) = \\
&\quad = \min_{\boldsymbol{b}_i, \boldsymbol{b}_j \in \mathcal{B}} \|\boldsymbol{c}_i - \boldsymbol{c}_j + \tau(r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j)\|^2 \\
&\quad = \min_{\boldsymbol{b}_i, \boldsymbol{b}_j \in \mathcal{B}} \left\{\|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 + \tau^2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2 + 2\tau(\boldsymbol{c}_i - \boldsymbol{c}_j)^\top(r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j)\right\} \\
&\quad = \min_{\boldsymbol{b}_i, \boldsymbol{b}_j \in \mathcal{B}} \{\|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 + \tau^2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2 \\
&\qquad\quad + \|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 + \tau^2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2 - \|\boldsymbol{c}_i - \boldsymbol{c}_j - \tau(r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j)\|^2\} \\
&\quad = 2\|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 + \beta_{ij}\tau^2 + \min_{\boldsymbol{b}_i, \boldsymbol{b}_j \in \mathcal{B}} \left\{-\beta_{ij}\tau^2 + 2\tau^2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2 - \|\boldsymbol{c}_i - \boldsymbol{c}_j - \tau\left(r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\right)\|^2\right\} \\
&\quad = 2\|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 + \beta_{ij}\tau^2 + \min_{\boldsymbol{b}_i, \boldsymbol{b}_j \in \mathcal{B}} \left\{\tau^2\left(2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2 - \beta_{ij}\right) - \|\boldsymbol{c}_i - \boldsymbol{c}_j - \tau\left(r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\right)\|^2\right\} \\
&\quad = 2\|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 + \beta_{ij}\tau^2 - \max_{\boldsymbol{b}_i, \boldsymbol{b}_j \in \mathcal{B}} \left\{\|\boldsymbol{c}_i - \boldsymbol{c}_j - \tau\left(r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\right)\|^2 - \tau^2\left(2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2 - \beta_{ij}\right)\right\}
\end{aligned}$$

Observe that taking $\beta_{ij} \in \mathbb{R}$ such that

$$2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2 - \beta_{ij} \leq 0 \qquad \forall \boldsymbol{b}_i, \boldsymbol{b}_j \in \mathcal{B},$$

the function

$$(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau) \longmapsto \|\boldsymbol{c}_i - \boldsymbol{c}_j - \tau\left(r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\right)\|^2 - \tau^2\left(2\|r_i\boldsymbol{b}_i - r_j\boldsymbol{b}_j\|^2 - \beta_{ij}\right)$$

is convex. Since the maximum of convex functions is convex, hence taking $u = 2\|\boldsymbol{c}_i - \boldsymbol{c}_j\|^2 + \beta_{ij}\tau^2$, we have obtained a DC decomposition for g^2 as in the statement. □

We proof now Proposition 3:

If $\lambda < \dfrac{1}{3}$, considering Proposition 1, one has

$$\lambda F_1 + (1-\lambda)F_2 = \sum_{\substack{i,j=1,\ldots,N \\ i \neq j}} \left\{2\lambda\kappa^2\delta_{ij}^2 - \left[\lambda(g + \kappa\delta_{ij})^2 - (3\lambda - 1)g^2(\boldsymbol{c}_i, \boldsymbol{c}_j, \tau)\right]\right\},$$

and thus $u = \sum_{\substack{i,j=1,\dots,N\\ i\neq j}} 2\lambda\kappa^2\delta_{ij}^2$ holds.

If $\lambda \geq \dfrac{1}{3}$, by using the DC decomposition for g^2 obtained in Lemma 1 and Proposition 1, one has

$$
\begin{aligned}
&\lambda F_1 + (1-\lambda)F_2 = \\
&\quad = \sum_{\substack{i,j=1,\dots,N\\ i\neq j}} \left\{(3\lambda-1)g^2(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau) + 2\lambda\kappa^2\delta_{ij}^2 - \lambda(g(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau)+\kappa\delta_{ij})^2\right\} \\
&\quad = \sum_{\substack{i,j=1,\dots,N\\ i\neq j}} \Big\{2(3\lambda-1)\|\boldsymbol{c}_i-\boldsymbol{c}_j\|^2 + (3\lambda-1)\beta_{ij}\tau^2 + 2\lambda\kappa^2\delta_{ij}^2 - \big[\lambda(g(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau)+\kappa\delta_{ij})^2 \\
&\qquad\quad + (3\lambda-1)\max_{\boldsymbol{b}_i,\boldsymbol{b}_j\in\mathcal{B}}\left\{\|\boldsymbol{c}_i-\boldsymbol{c}_j-\tau\left(r_i\boldsymbol{b}_i-r_j\boldsymbol{b}_j\right)\|^2 - \tau^2\left(2\|r_i\boldsymbol{b}_i-r_j\boldsymbol{b}_j\|^2-\beta_{ij}\right)\right\}\big]\Big\} \\
&\quad = \sum_{i=1,\dots,N} 8(3\lambda-1)(N-1)\|\boldsymbol{c}_i\|^2 + (3\lambda-1)\tau^2 \sum_{\substack{i,j=1,\dots,N\\ i\neq j}} \beta_{ij} + 2\lambda\kappa^2 \sum_{\substack{i,j=1,\dots,N\\ i\neq j}} \delta_{ij}^2 \\
&\qquad\quad - \big[(3\lambda-1)\|\boldsymbol{c}_i+\boldsymbol{c}_j\|^2 + \lambda(g(\boldsymbol{c}_i,\boldsymbol{c}_j,\tau)+\kappa\delta_{ij})^2 \\
&\qquad\quad + (3\lambda-1)\max_{\boldsymbol{b}_i,\boldsymbol{b}_j\in\mathcal{B}}\left\{\|\boldsymbol{c}_i-\boldsymbol{c}_j-\tau\left(r_i\boldsymbol{b}_i-r_j\boldsymbol{b}_j\right)\|^2 - \tau^2\left(2\|r_i\boldsymbol{b}_i-r_j\boldsymbol{b}_j\|^2-\beta_{ij}\right)\right\}\big]
\end{aligned}
$$

□