



**HAL**  
open science

## Projection under pairwise distance controls

Hiba Alawieh, Nicolas Wicker, Christophe Biernacki

► **To cite this version:**

Hiba Alawieh, Nicolas Wicker, Christophe Biernacki. Projection under pairwise distance controls. 2016. hal-01420662v1

**HAL Id: hal-01420662**

**<https://hal.science/hal-01420662v1>**

Preprint submitted on 20 Dec 2016 (v1), last revised 23 Dec 2020 (v5)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Projection under pairwise distance controls

Alawieh Hiba\*, Nicolas Wicker, Christophe Biernacki

Université Lille 1 - UFR de Mathématiques, cité scientifique  
59655 Villeneuve d'Ascq, France

August 29, 2016

## Abstract

Visualization of high-dimensional and possibly complex (non continuous for instance) data onto a low-dimensional space may be difficult. Several projection methods have been already proposed for displaying such high-dimensional structures on a lower-dimensional space, but the information lost is not always easy to use. Here, a new projection paradigm is presented to describe a non-linear projection method that takes into account the projection quality of each projected point in the reduced space, this quality being directly available in the same scale as this reduced space. More specifically, this novel method allows a straightforward visualization data in  $\mathbb{R}^2$  with a simple reading of the approximation quality, and provides then a novel variant of dimensionality reduction.

*Keywords:* Data visualization, dimension reduction, multidimensional scaling, principal component analysis.

---

\*Corresponding author

# 1 Introduction

Several domains in science use data with large number of variables. This data can be viewed as a large matrix and extracting results from this type of matrix is often hard and complicate. In such cases, it is desirable to reduce the number of dimensions of data by conserving as much information as possible from the given initial matrix. Many methods for performing dimension reduction exist. To quote a few, principal component analysis (PCA), multidimensional scaling (MDS) and self-organizing map (Mardia *et al.*, 1979) are perhaps the most known. Usually, a quality of projection is assessed by the percentage of variance which is conserved or by a stress factor but this quality takes only into account what happens globally. In some projection methods, like PCA, a local measure is defined to indicate the projection quality of each projected point taken individually. This local measure is evaluated by the squared cosine of angle between the principal space and the vector of the point. A good representation in the projected space is hinted by high squared cosine values. This measure is useful in cases of linear projection as happens in PCA but cannot be applied to the case of non-linear projection.

In this paper, we propose a new non-linear projection method that projects the points in a reduced space by using the pairwise distance between pairs of points and by taking into account the projection quality of each point taken individually. This projection leads to a representation of the points as circles with a different radius associated to each point. Henceforth, this method will be called "Projection under pairwise distance control". The main contributions of this study are to give a simple data visualization in  $\mathbb{R}^2$  with a straightforward interpretation and provide a new variant of dimensionality reduction. First, the new projection method is presented in the section 2. Then, in section 3, the algorithms used in the resolution of optimization problems related to this method are illustrated. Next, section 4 shows the application of this method to various real data sets. Finally, section 5 concludes this work.

## 2 Projection under pairwise distance control

Let us consider  $n$  points given by their pairwise distance noted  $d_{ij}$  for  $i, j \in \{1, \dots, n\}$ . The task here is to project these points using distances into a reduced space  $\mathbb{R}^k$  by introducing additional variables, called hereafter radii, that indicate to which extent the projection of each point is accurate. The local quality is then given by the values of the radii. A good quality projection of point  $i$  is indicated by a small radius value noted  $r_i$ . It will be important to note that both units of  $d_{ij}$ 's and  $r_i$ 's are identical, allowing direct comparison. Before developing our method, an overview of principal component analysis (PCA) is presented to highlight the interest of our method.

### 2.1 Principal Component Analysis (PCA)

PCA method is the most used method in the data visualization and dimensionality reduction. This method is a linear projection technique applied when the data is linearly separable. PCA problem can be stated as an optimization problem involving the squared Euclidean distances (Mardia *et al.*, 1979). This optimization problem is the following:

$$\mathcal{P}_{\text{PCA}} : \begin{cases} \min_A \sum_{1 \leq i < j \leq n} |d_{ij}^2 - \|Ay_i - Ay_j\|^2| \\ \text{s.t. } \text{rank}(A) = k \\ AA^T = I_p \end{cases}$$

where  $y_i \in \mathbb{R}^p$  is the original coordinates vector of point  $i$ ,  $d_{ij}^2$  is the squared distance for couple  $(i, j)$  given by  $\|y_i - y_j\|^2$  and  $A$  is the projection matrix of dimension  $p \times q$ . Let us consider now  $n$  variables  $R_1, \dots, R_n \in \mathbb{R}$ , the sum of which bounds the objective function. So, the PCA optimization problem  $\mathcal{P}_{\text{PCA}}$  can be rewritten as:

$$\mathcal{P}_{\text{PCA}} : \begin{cases} \min_{R_1, \dots, R_n, A} \sum_{i=1}^n R_i \\ \text{s.t. } \sum_{i=1}^n R_i \geq \frac{1}{n-1} \sum_{1 \leq i < j \leq n} |d_{ij}^2 - \|Ay_i - Ay_j\|^2| \\ \text{rank}(A) = k \\ AA^T = I_p \end{cases}$$

Moreover, a new optimization problem, denoted as  $\mathcal{P}_{R,A}$ , is defined below. The constraints number of this latest problem is equal to  $\frac{n(n-1)}{2} + 2$ . So, by summing all the constraint of type  $|d_{ij}^2 - \|Ay_i - Ay_j\|^2| \leq R_i + R_j$  for all  $1 \leq i < j \leq n$ , we obtain the PCA problem. thus, it is clear that  $\mathcal{P}_{PCA}$  is a relaxed problem of problem  $\mathcal{P}_{R,A}$ .

$$\mathcal{P}_{R,A} : \begin{cases} \min_{R_1, \dots, R_n, A} \sum_{i=1}^n R_i \\ s.t. \quad |d_{ij}^2 - \|Ay_i - Ay_j\|^2| \leq R_i + R_j, \text{ for } 1 \leq i < j \leq n \\ \text{rank}(A) = k \\ AA^T = I_p \end{cases}$$

Problem  $\mathcal{P}_{R,A}$  is hard to be solved as it is needed to deal simultaneously with the orthogonality constraint and with the constraints on the variables  $R_i$ .

Besides, PCA cannot take into account non-linear structures, since it describes the data in terms of a linear subspace. Furthermore, the only measures used to evaluate the projection quality of points are the squared cosines values which can only be used in the case of linear projection. Thus, the individual control of projection is not more guaranteed using non-linear projection method.

So, there is the requirement of a non-linear projection method that controls individually the projection of points, hence the interest of our proposed method.

## 2.2 Our proposed method

Let  $x_1, \dots, x_n$  be the coordinates of the projected points in  $\mathbb{R}^k$ . Radii are an important element of the paper introduced to assess how much the distance between two projected points  $(i, j)$  given by  $\|x_i - x_j\|$  is far from given distance  $d_{ij}$ . Indeed, radii  $(r_i, r_j)$  for couple  $(i, j)$  are small when  $\|x_i - x_j\|$  is close to  $d_{ij}$ . Figure 1 depicts this idea: for all points  $i \in \{1, \dots, n\}$  the projected point of each point  $i$  belongs to a sphere with center  $x_i$  and radius  $r_i$  such that  $\|x_i - x_j\| + r_i + r_j \leq d_{ij} \leq \|x_i - x_j\| + r_i + r_j$ .

This idea can be expressed by finding the value of radii that satisfy these two constraints:

- $\sum_{i=1}^n r_i$  is minimum.
- $d_{ij} \in \{\|x_i - x_j\| - r_i - r_j, \|x_i - x_j\| + r_i + r_j\}, \forall 1 \leq i < j \leq n$ .

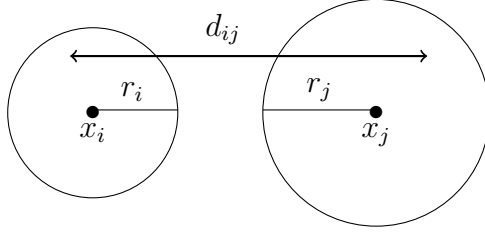


Figure 1: Examples of radii for bounding the original distance  $d_{ij}$

The projection under pairwise distance control problem can be written as the following optimization problem:

$$\mathcal{P}_{r,x} : \begin{cases} \min_{r_1, \dots, r_n \in \mathbb{R}, x_1, \dots, x_n \in \mathbb{R}^k} \sum_{i=1}^n r_i \\ \text{s.t. } |d_{ij} - \|x_i - x_j\|| \leq r_i + r_j, \text{ for } 1 \leq i < j \leq n \end{cases}$$

Of course, by fixing the coordinates vectors  $x_i$  for all  $i \in 1, \dots, n$ , using principal component analysis or any other projection method, the problem can easily be solved in  $(r_1, \dots, r_n)$  using linear programming. This problem can be written as follows:

$$\mathcal{P}_r : \begin{cases} \min_{r_1, \dots, r_n \in \mathbb{R}} \sum_{i=1}^n r_i \\ \text{s.t. } |d_{ij} - \|x_i - x_j\|| \leq r_i + r_j, \text{ for } 1 \leq i < j \leq n \\ x_i \in \mathbb{R}^k, \text{ for } i = 1, \dots, n \end{cases}$$

We can remark that a solution of  $\mathcal{P}_r$  always exists. Indeed, to satisfy the constraints it is enough to increase all  $r_i$ . Besides, solving  $\mathcal{P}_r$  with fixed coordinates  $(x_1, \dots, x_n)$  does not lead in general to the optimum of problem  $\mathcal{P}_{r,x}$ .

### 2.3 Visualization example

Let us apply our projection method to a simple example by taking a tetrahedron with all pairwise distance is equal to 1. For problem  $\mathcal{P}_r$ , the coordinates  $x_i$  for  $i = 1, \dots, 4$  are obtained using multidimensional scaling. Using linear and non-linear optimization packages in Matlab respectively for problems  $\mathcal{P}_r$  and  $\mathcal{P}_{r,x}$  give a value of  $\sum_{i=1}^n r_i$  equal to 0.7935 for problem  $\mathcal{P}_r$  and 0.4226 for  $\mathcal{P}_{r,x}$ . Figure 2 corresponds to the first solution and Figure 3 corresponds to the second one. In Figures 2 and 3, we depict circles with different radii.

The circle color is related to the radius values, the shades of gray lie between white and black in the descending direction of the radius values; the smaller the radius, the darker the circle. The points that have circles with small radii are considered as well projected points. Note that the points that are represented as points and not circles are very well projected, having radii almost equal to zero. In Figure 2, half of the points is well projected whereas the other half have large radii indicating that they are not well projected. In Figure 3 one circle appears marking that the projection quality using problem  $\mathcal{P}_{r,x}$  is better than  $\mathcal{P}_r$ .

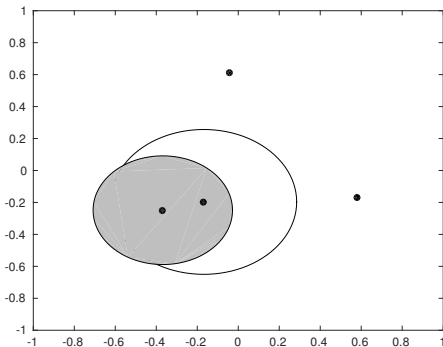


Figure 2: Solution of  $\mathcal{P}_r$  using MDS. The objective value is equal to 0.7935.

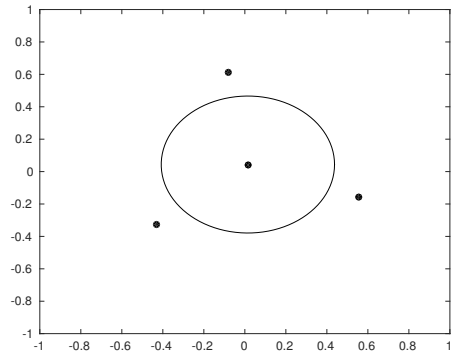


Figure 3: Solution of  $\mathcal{P}_{r,x}$ . The objective value is equal to 0.4226.

## 2.4 Link with other methods

Multidimensional fitting (MDF) (Berge et al., 2010) is a method that modifies the coordinates of a set of points in order to make the distances calculated on the modified coordinates similar to given distances on the same set of points. We call “target matrix” the matrix that contains the points coordinates and “reference matrix” the matrix that contains the given distances.

Let us note  $X = \{x_1 | \dots | x_n\}$  the target matrix and  $D = \{d_{ij}\}$  the reference matrix. The objective function of MDF problem is given by:

$$\sum_{1 \leq i < j \leq n} |d_{ij} - \|x_i - x_j\||.$$

**Property 1.** Problem  $\mathcal{P}_{r,x}$  is bounded below by  $\frac{1}{n-1} \sum_{1 \leq i < j \leq n} |d_{ij} - \|x_i - x_j\||$  where  $x_1, \dots, x_n$  is the optimum of the associated MDF problem.

*Proof.* By summing all the constraints of problem  $\mathcal{P}_{r,x}$  we obtain:

$$\begin{aligned} \sum_{1 \leq i < j \leq n} |d_{ij} - \|x_i - x_j\|| &\leq \sum_{1 \leq i < j \leq n} r_i + r_j = (n-1) \sum_{i=1}^n r_i \\ \Rightarrow \sum_{i=1}^n r_i &\geq \frac{1}{n-1} \sum_{1 \leq i < j \leq n} |d_{ij} - \|x_i - x_j\||, \text{ which concludes the proof.} \quad \square \end{aligned}$$

## 3 Optimization tools

### 3.1 Lower Bound of problem $\mathcal{P}_{r,x}$

Minimization problem  $\mathcal{P}_{r,x}$  is too hard to be solved exactly. A way to assess how good a solution, is to provide a lower bound on the objective function. Then, if the bound is close to the best solution found, we can conclude that this solution is fix. Thus, in this section we want to present a lower bound of problem  $\mathcal{P}_{r,x}$  given by property 2.

**Property 2.** The lower bound of problem  $\mathcal{P}_{r,x}$  is given by:

$$\sum_{i=1}^n r_i^{opt} \geq \min_M \max\{f(M); g(M); h(M)\}.$$

with  $f, g, h$  three functions defined in the following proof.

*Proof.* Let  $x_1, \dots, x_n; r_1, \dots, r_n$  a feasible solution of  $\mathcal{P}_{r,x}$  then

$$\text{for all } 1 \leq i < j \leq n, \exists M \text{ such that } \|x_i - x_j\| \leq M.$$

We consider three functions noted  $f, g, h$  depending on  $M$  as follows:

- $f(M) = \sqrt{\left(1 - \frac{n}{3}\right) M^2 + \frac{1}{n-1} \sum_{i < j} d_{ij}^2} - M.$
- $g(M) = |D - d_{max}|.$
- $h(M) = \min_{i < j} \max_{k < l; k, l \neq i, j} \min\{L_{ijkl}^1; L_{ijkl}^2\}.$



where:

- $d_{max} = \max_{1 \leq i < j \leq n} \{d_{ij}\},$
- $L_{ijkl}^1 = \max \left\{ \frac{d_{jk} - d_{kl}}{2}, \frac{d_{jl} - d_{kl}}{2}, |d_{ij} - M| \right\},$
- $L_{ijkl}^2 = \max \left\{ \frac{d_{ik} - d_{kl}}{2}, \frac{d_{il} - d_{kl}}{2}, |d_{ij} - M| \right\}.$

Using results presented in appendix *B*, we can write:

$$\left\{ \begin{array}{l} \sum_{i=1}^n r_i \geq f(M^*) \\ \sum_{i=1}^n r_i \geq g(M^*) \\ \sum_{i=1}^n r_i \geq h(M^*). \end{array} \right.$$

Consequently,

$$\sum_{i=1}^n r_i \geq \max\{f(M); g(M); h(M)\}. \quad (1)$$

The inequality (1) is true for all solutions of  $\mathcal{P}_{r,x}$  particularly for the optimal solution.

Thus:

$$\sum_{i=1}^n r_i^{opt} \geq \max\{f(M^{opt}); g(M^{opt}); h(M^{opt})\}. \quad (2)$$

Hence, the lower bound is given by:

$$\sum_{i=1}^n r_i^{opt} \geq \min_M \max\{f(M); g(M); h(M)\} \text{ for all feasible solutions.}$$

Given  $M$ , a lower bound of problem  $\mathcal{P}_{r,x}$  is derived. Afterwards, a bound free of  $M$  is given by minimizing the bounds depending on  $M$ .  $\square$

By applying this bound to the tetrahedron example, the three functions are plotted. The result is shown in Figure 4. The lower bound is equal to 0.1276 for  $M = 1.1276$  and as we have seen the minimum obtained by solving  $\mathcal{P}_{r,x}$  is equal to 0.42.

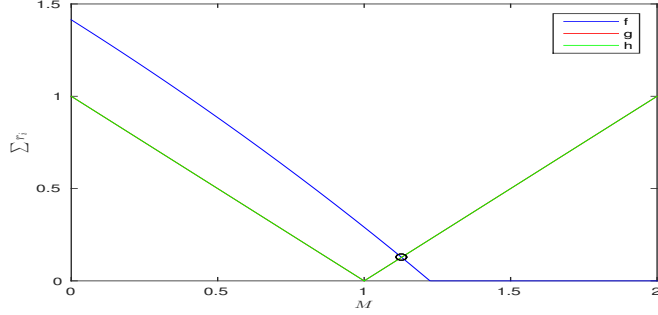


Figure 4: The curves of the three functions  $f, g$  and  $h$ . Functions  $g$  and  $h$  are equal due to the fact that all distances are equal to 1. The minimal intersection is given by the black circle for  $M = 1.1276$  and  $\sum_{i=1}^n r_i > 0.1276$

### 3.2 Initialization point of problem $\mathcal{P}_{r,x}$

Different resolutions of problem  $\mathcal{P}_{r,x}$  can be obtained using different initial values of matrix  $X$ . Three possible initial values can be used. The first of them is the matrix obtained by PCA or another projection method. In what follows, we present two other possibilities.

**Initial point using squared distances** The optimization problem  $\mathcal{P}_{r,x}$  can be changed by taking the squared distances between points instead of the distances. Rewriting  $r_i^2$  as  $R_i$ , the problem is changed into

$$\mathcal{P}_{R,x} : \begin{cases} \min_{R_1, \dots, R_n \in \mathbb{R}, x_1, \dots, x_n \in \mathbb{R}^k} \sum_{i=1}^n R_i \\ \text{s.t. } |d_{ij}^2 - \|x_i - x_j\|^2| \leq R_i + R_j, \text{ for } 1 \leq i < j \leq n. \end{cases}$$

The transformation is interesting as if the constraints of problem  $\mathcal{P}_{R,x}$  are satisfied, the constraints of problem  $\mathcal{P}_{r,x}$  will also be satisfied. Indeed:

$$\begin{aligned} & |d_{ij}^2 - \|x_i - x_j\|^2| \leq R_i + R_j = r_i^2 + r_j^2 \\ \Rightarrow & (d_{ij} - \|x_i - x_j\|)(d_{ij} + \|x_i - x_j\|) \leq r_i^2 + r_j^2 \leq (r_i + r_j)^2 \\ \Rightarrow & |d_{ij} - \|x_i - x_j\||^2 \leq (r_i + r_j)^2 \\ \Rightarrow & |d_{ij} - \|x_i - x_j\|| \leq (r_i + r_j). \end{aligned}$$

That way problem  $\mathcal{P}_{R,x}$  can serve as an initial step for solving problem  $\mathcal{P}_{r,x}$ .

**Initial point using improved solution of problem  $\mathcal{P}_r$**  First, we give two properties which provide a way to improve the optimization results of problem  $\mathcal{P}_{r,x}$ .

**Property 3.** *Let us consider a point  $x_i$  such that for an index  $j$ , the following inequality is saturated :*

$$|d_{ij} - \|x_i - x_j\|| \leq r_i + r_j,$$

*and the other inequalities involving  $i$  are not saturated. Then, the corresponding solution can be improved by moving  $x_i$  along the direction  $x_j - x_i$  to decrease  $r_i$  and  $|d_{ij} - \|x_i - x_j\||$ .*

*Proof.* The above condition means that  $x_i$  is rewritten  $x_i + a(x_j - x_i)$  with  $a \in \mathbb{R}$  and we look for  $a$  such that  $|d_{ij} - \|x_i + a(x_j - x_i) - x_j\|| < r_i + r_j$ . In particular  $a \leq 0$  if  $d_{ij} - \|x_i - x_j\| \geq 0$  and  $a > 0$  otherwise. Let us now consider the other inequalities corresponding to index pairs  $(i, k)$  with  $k \neq j$ . For each of them, either  $\exists a \in [a'_k, a''_k]$  with  $a'_k < 0$  and  $a''_k > 0$  such that

$$|d_{ij} - \|x_i + a(x_j - x_i) - x_j\|| \leq r_i + r_j,$$

as these constraints are unsaturated. Finally, if we take  $a$  different from 0 in  $[a', a'']$  with  $a' = \max_k a'_k$  and  $a'' = \min_k a''_k$ , all constraints involving  $i$  get unsaturated so that  $r_i$  can be decreased, decreasing so the objective function. Depending on whether  $a$  must be negative or positive, we take  $a = a'$  or  $a = a''$  respectively.

□

Another manner to improve the resolution of problem  $\mathcal{P}_{r,x}$  is to effectuate a scale change by multiplying the coordinates  $x_i$ , for  $i = 1, \dots, n$ , by a constant  $a \in \mathbb{R}$ . Thus, the new optimization problem is given by:

$$\mathcal{P}_{r,a} : \begin{cases} \min_{r_1, \dots, r_n, a \in \mathbb{R}} \sum_{i=1}^n r_i \\ \text{s.t. } |d_{ij} - a\|x_i - x_j\|| \leq r_i + r_j \end{cases}$$

**Property 4.** *Let  $r_1, \dots, r_n; x_1, \dots, x_n$  be a feasible solution of  $\mathcal{P}_{r,x}$ , if  $\exists a$  such that  $\eta(a) < \sum_{i=1}^n r_i$  with  $\eta(a) = \sum_{1 \leq i < j \leq n} |d_{ij} - a\|x_i - x_j\||$ , then  $\exists \tilde{r}_1, \dots, \tilde{r}_n$  a solution of  $\mathcal{P}_{r,a}$  such that  $\sum_{i=1}^n \tilde{r}_i < \sum_{i=1}^n r_i$ .*

*Proof.* Let us consider  $r_1, \dots, r_n; x_1, \dots, x_n$  a feasible solution of problem  $\mathcal{P}_{r,x}$  and  $a, \tilde{r}_1, \tilde{r}_2, \dots, \tilde{r}_n$  the optimal solution of  $\mathcal{P}_{r,a}$ . For the solution of  $\mathcal{P}_{r,a}$ , for each point  $i$ , we have a certain saturated constraint associated to point  $k$  noted  $C_{ik(i)}$ , otherwise it would not be an optimum. So, we have:

$$\begin{aligned}
|d_{i1} - a\|x_i - x_1|| &\leq \tilde{r}_i + \tilde{r}_1 \\
&\vdots \\
|d_{ik(i)} - a\|x_i - x_{k(i)}|| &= \tilde{r}_i + \tilde{r}_{k(i)} \\
&\vdots \\
|d_{ij} - a\|x_i - x_j|| &\leq \tilde{r}_i + \tilde{r}_j \\
&\vdots \\
|d_{in} - a\|x_i - x_n|| &\leq \tilde{r}_i + \tilde{r}_n.
\end{aligned}$$

Then,  $|d_{ik(i)} - a\|x_i - x_{k(i)}|| = \tilde{r}_i + \tilde{r}_{k(i)} \geq \tilde{r}_i$ . By summing all points  $i$ , for  $i = 1, \dots, n$ , we obtain:

$$\sum_{i=1}^n |d_{ik(i)} - a\|x_i - x_{k(i)}|| \geq \sum_{i=1}^n \tilde{r}_i.$$

Thus

$$\sum_{1 \leq i < j \leq n} |d_{ij} - a\|x_i - x_j|| \geq \sum_{i=1}^n |d_{ik(i)} - a\|x_i - x_{k(i)}|| \geq \sum_{i=1}^n \tilde{r}_i.$$

Note  $\eta(a) = \sum_{1 \leq i < j \leq n} |d_{ij} - a\|x_i - x_j||$ , then if  $\eta(a) < \sum_{i=1}^n r_i$  there is a solution of  $\mathcal{P}_{r,a}$  such that  $\sum_{i=1}^n \tilde{r}_i < \sum_{i=1}^n r_i$ . □

The new initial point is then given by using these two properties as follows:

- firstly, improve the solution of  $\mathcal{P}_r$  using property 2 by solving  $\mathcal{P}_{r,a}$ .
- secondly, improve the solution of  $\mathcal{P}_{r,a}$  using property 1.

### 3.3 Algorithm 1

Using the different initial values of matrix  $X$  presented above, we solve now problem  $\mathcal{P}_{r,x}$ . For this task, we introduce a new algorithm denoted algorithm 1 which gives the best solution that can be obtained using the different initial values cited above. This algorithm

is consisted of two steps: initialization step and optimization step and it is presented as follows:

---

**Algorithm 1**

---

Input:  $D$ : distance matrix,  $N$ : number of iteration.

***Initialization step***

Project the points using PCA or MDS.

Solve  $\mathcal{P}_r$  using an interior-point method. Obtained solution:  $(X_{\mathcal{P}_r}, r_{\mathcal{P}_r})$ .

Solve  $\mathcal{P}_{R,x}$  using an active-set method and starting from the solution of  $\mathcal{P}_r$  obtained at the previous step. Obtained solution:  $(X_{\mathcal{P}_{R,x}}, R_{\mathcal{P}_{R,x}})$ .

$X_0 \leftarrow X_{\mathcal{P}_{R,x}}$ .

**for**  $t = 1$  to  $N$  **do**

Solve  $\mathcal{P}_{r,a}$  starting from  $X_0$  using an interior-point method. .

Improve the solution of  $\mathcal{P}_{r,a}$  using property 1. Obtained solution:  $(X_{\mathcal{P}_{r,a}}^I, r_{\mathcal{P}_{r,a}}^I)$ .

$X_0 \leftarrow X_{\mathcal{P}_{r,a}}^I$ .

**end for**

***Optimization step***

Optimize  $\mathcal{P}_{r,x}$  using an active-set method and starting from  $X_0$ ,  $X_{\mathcal{P}_r}$  and  $X_{\mathcal{P}_{R,x}}$ .

Choose the minimal solution obtained by these three different starting points.

---

### 3.4 Algorithm 2

Problem  $\mathcal{P}_{r,x}$  is a hard problem, so it is natural to resort to stochastic optimization methods. In the present case, Metropolis-Hastings algorithm (Johansen, 2007) allows us to build a Markov chain with a desired stationary distribution. The only delicate part is the choice of the proposal distribution and the necessity to solve a  $\mathcal{P}_r$  problem at each iteration. In details, this Metropolis-Hastings algorithm requires:

1- *A target distribution:*

The target distribution is related with the objective function of problem  $\mathcal{P}_{r,x}$  and it is given by:

$$\pi(s) \propto \exp\left(\frac{-E(x)}{T}\right),$$

with  $E$  an application given by:

$$E : \quad \mathbb{R}^n \quad \longmapsto \quad \mathbb{R}$$

$$x = (x_1, \dots, x_n) \quad \longmapsto \quad E(x) = \text{Solution of problem } \mathcal{P}_r \text{ with } x \text{ fix.}$$

The variable  $T$  is the temperature parameter, to be fixed according to the value range of  $E$ .

2- *A proposal distribution:*

The choice of the proposal distribution is very important to obtain interesting results. It should be chosen in such a way that the proposal distribution gets close to the target distribution. The proposal distribution  $q(X \rightarrow \cdot)$  has been constructed as follows, giving priority to the selection of points involved in saturated constraints:

- For each point  $i$ , choose a point  $j^{(i)}$  with probability equal to:

$$P_{j^{(i)}} = \frac{\lambda \exp(-\lambda(r_i + r_{j^{(i)}} - |d_{ij^{(i)}} - \|x_i - x_{j^{(i)}}\||))}{\sum_{k=1, k \neq i}^n \lambda \exp(-\lambda(r_i + r_k - |d_{ik} - \|x_i - x_k\||))}$$

- Choose a constant  $c_{ij^{(i)}}$  using Gaussian distribution  $\mathcal{N}_k(0, \sigma)$ .
- Generate a matrix  $X^*$  by moving each vector  $x_i$  of matrix  $X^{t-1}$  as follows:
  - \* If  $d_{ij^{(i)}} - \|x_i - x_{j^{(i)}}\| > 0$  then  $x_i^* = x_i + |c_{ij^{(i)}}|L_i$ .
  - \* else  $x_i^* = x_i - |c_{ij^{(i)}}|L_i$ .

$$\text{with } L_i = \frac{x_i - x_j}{\|x_i - x_j\|}.$$

3- *A linear optimization problem:*

For the matrix  $X$  generated in each iteration, we solve the linear optimization problem  $\mathcal{P}_r$ .

## 4 Numerical application

The presented projection method has been applied to different types of real data sets so as to illustrate its generality.

## 4.1 The data

Four real data sets are used and divided into three categories:

- Quantitative data: Iris and cars data sets.
- Categorical data: Soybean data set.
- Functional data: Coffee data set.

The Iris data set (Anderson, 1935) is a famous data set and is presented to show that the projection is as expected. This data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. The four variables studied in this data set are: sepal length, sepal width, petal length and petal width (in *cm*). Cars data set (Saporta, 2006) is a data set studied in the book of Saporta (Table 17.1, page 428). This data set describes 18 cars according to various variables (cylinders, power, length, width, weight, speed).

The soybean data set (Stepp, 1984) from *UCI Machine Learning Repository* characterizes 47 soybean disease case histories defined over 35 attributes. Each observation is identified by one of the 4 diseases: Diaporthe Stem Canker (D1), charcoal Rot (D2), Rhizoctonia Root Rot (D3) and Phytophthora Rot (D3).

The coffee data set is a time series data set used in chemometrics to classify food types. This kind of time series is seen in many applications in food safety and quality insurance. This data set is taken from *UCR time Series Classification and Clustering* website (Chen et al., 2015). *Coffea Arabica* and *Coffea Canephora* variant Robusta are the two species of coffee bean which have acquired a worldwide economic importance and many methods have been developed to discriminate between these two species by chemical analysis (Briandet et al., 2000).

## 4.2 Experimental setup

In practice, we have tested our method on the different data sets by solving the optimization problem  $\mathcal{P}_{r,x}$  using algorithm 1 and also the proposed Metropolis-Hastings algorithm (algorithm 2). Each time, a distance matrix is required. For the quantitative data, we compute the Euclidean distance between points  $y_i$ , for  $i = 1, \dots, n$ , by the known formula

$d_{ij} = \sqrt{\sum_{k=1}^p (y_{ik} - y_{jk})^2}$ . For categorical data, the distance between two soybean diseases  $(i, j)$  is given through Eskin dissimilarity (or proximity) measure (Boriah *et al.* , 2008) computed by the formula  $p_{ij} = \sum_{t=1}^Q w_t p_{ij}^t$  where  $p_{ij}^t = \begin{cases} 1 & \text{if } i^t = j^t \\ \frac{n_k^2}{n_k^2 + 2} & \text{else} \end{cases}$ ,  $p_{ij}^t$  is the per-attribute Eskin dissimilarity between two values for the categorical attribute indexed by  $t$ ,  $w_t$  is the weight assigned to the attribute  $t$ ,  $Q$  is the number of attributes and  $n_t$  is the number of values taken by each attribute. Then, using the formula which transforms the dissimilarity into similarity:  $p_{ij} = 1 - s_{ij}$ , the distances can be given by the standard transformation formula from similarity to distance:  $d_{ij} = \sqrt{s_{ii} - 2s_{ij} + s_{jj}}$ . On top of that, to compute the distances between the curves of functional data, we have chosen a measure of proximity similar to that studied in Ieva (2012). In this article, the authors develop a proper classification designed to distinguish the grouping structure by using a functional k-means clustering procedure with three sorts of distances. So, in our work we choose one of these three proximity measures forasmuch their results are similar. Thus, the proximity measure chosen between two curves  $F_i$  and  $F_j$  is the following:  $d_0(F_i, F_j) = \sqrt{\int_{\mathcal{T}} (F_i^0(t) - F_j^0(t))^2 dt}$ . This measure is calculated using the function `metric.lp()` of the `fda.usc` package for the **R** software.

To solve the different optimization problems, we have used the optimization toolbox in MATLAB. For problem  $\mathcal{P}_r$  and  $\mathcal{P}_{r,a}$ , we apply firstly PCA – for quantitative data – or MDS – for categorical and functional data – and then a linear programming package is used to solve the optimization problems using an interior-point algorithm. Problems  $\mathcal{P}_{r,x}$  and  $\mathcal{P}_{R,x}$  are non-linear optimization problems, therefore we use a non-linear programming package to solve it selecting the active-set algorithm to obtain the best values of  $(x_1, \dots, x_n)$  and  $(r_1, \dots, r_n)$ . This iterative algorithm is composed of two phases. In the first phase (the feasibility phase), the objective function is ignored while a feasible point is found for the constraints, in the second phase (the optimality phase), the objective function is minimized while feasibility is maintained (Wong , 2011).

Our proposed Metropolis-Hastings algorithm can provide a good solution if parameters  $\lambda$ ,  $\sigma$  and  $T$  are chosen adequately. For instance,  $\lambda$  should be such that the points belonging



Table 1: Optimization solution of problem  $\mathcal{P}_{r,x}$  for different data sets.

	$\sum r_i^{\text{Algo 1}}$	$\sum r_i^{\text{MH}}$	Lower Bound
Iris	16.19	17.2	1.07
Cars	3.27	3.35	1.21
Soybean	3.98	3.93	0.29
Coffee	21.68	21.97	0.89

to unsaturated constraints are chosen with small probabilities. Therefore, we take it equal to 100. For the other parameters  $\sigma$  and  $T$ , we take their values respectively in a range from 0.01 and 100.

As we have mentioned in the section of visualization, the visualization of the projection of each point  $i$  in  $\mathbb{R}^2$  is presented as a circle having  $x_i$  as center and  $r_i$  as radius so as the projected point belongs to this circle and this is the specificity of our method. For each data set, we show the circles obtained for each point after resolution of optimization problem  $\mathcal{P}_{r,x}$ . To compare the projection quality of our representation with that obtained by PCA, we use the squared cosine values as PCA projection quality.

Furthermore, the lower bound defined in section 3.1 is each time computed.

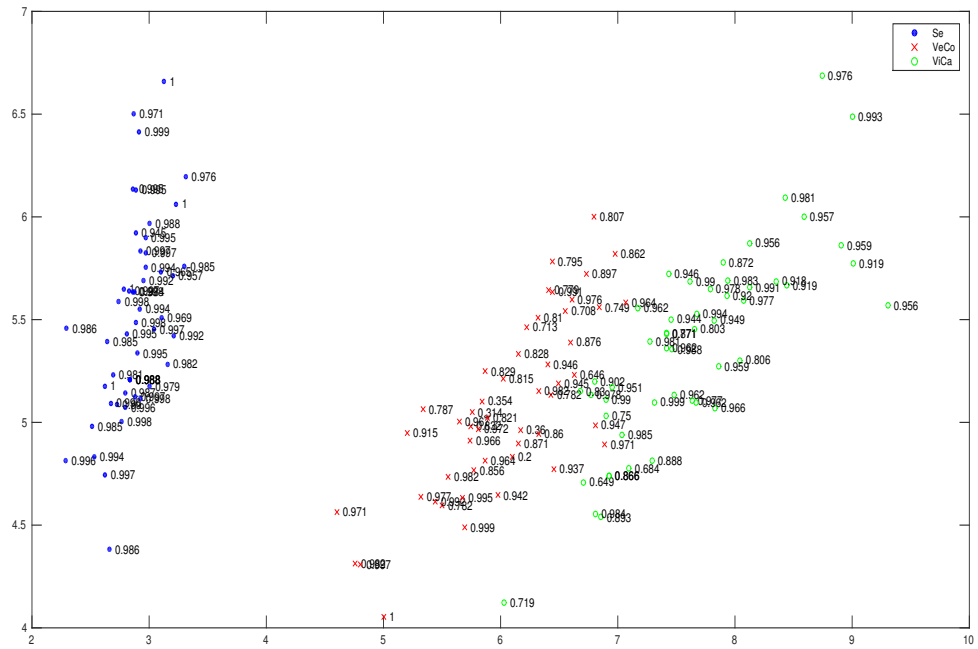
## 4.3 Results

### 4.3.1 Visualization data in $\mathbb{R}^2$

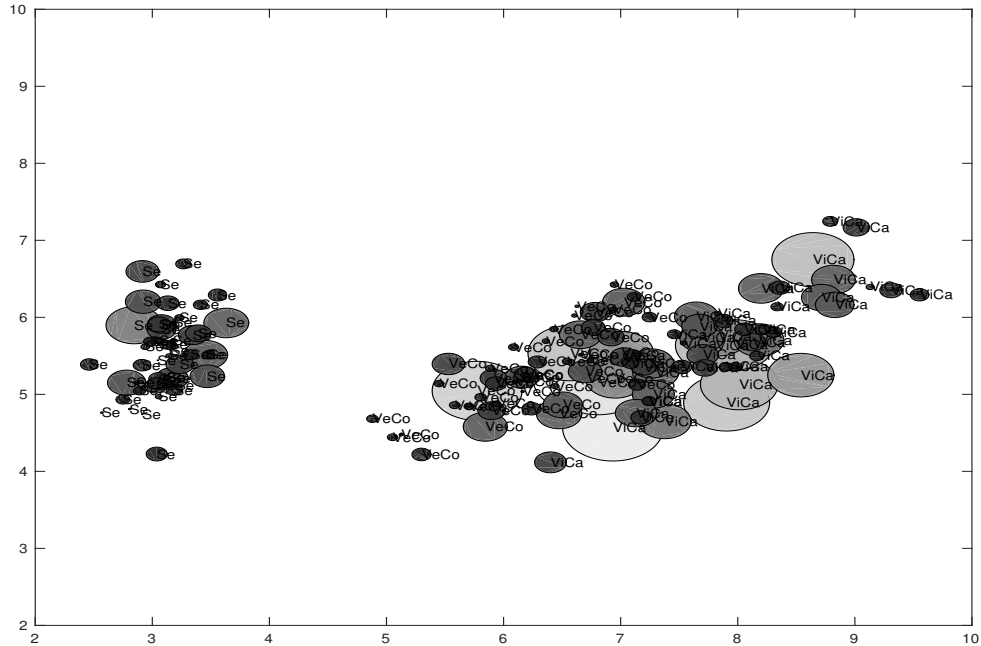
The optimization results for these four data sets are given in Table 1. For each data, we give the algorithm 1 and Metropolis-Hastings results with which initial starting point is used in algorithm 1. The lower bound value for each data set is also given in this table.

Figures 5 and 6 depict the results of projection under pairwise distance control for quantitative data. This projection is compared with the projection given by PCA by plotting the projection of the points indexed by their squared cosine values.

In the projection of Iris data set showed in Figure 5b, it is interesting to remark that appealingly two areas are well separated. This corresponds to the well-known fact that Iris

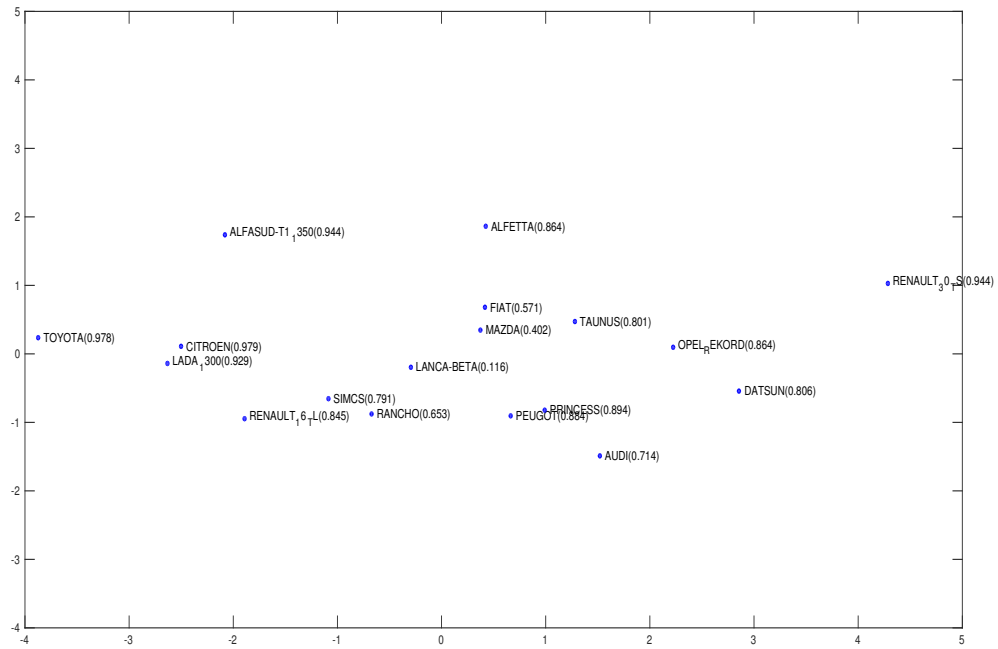


(a)

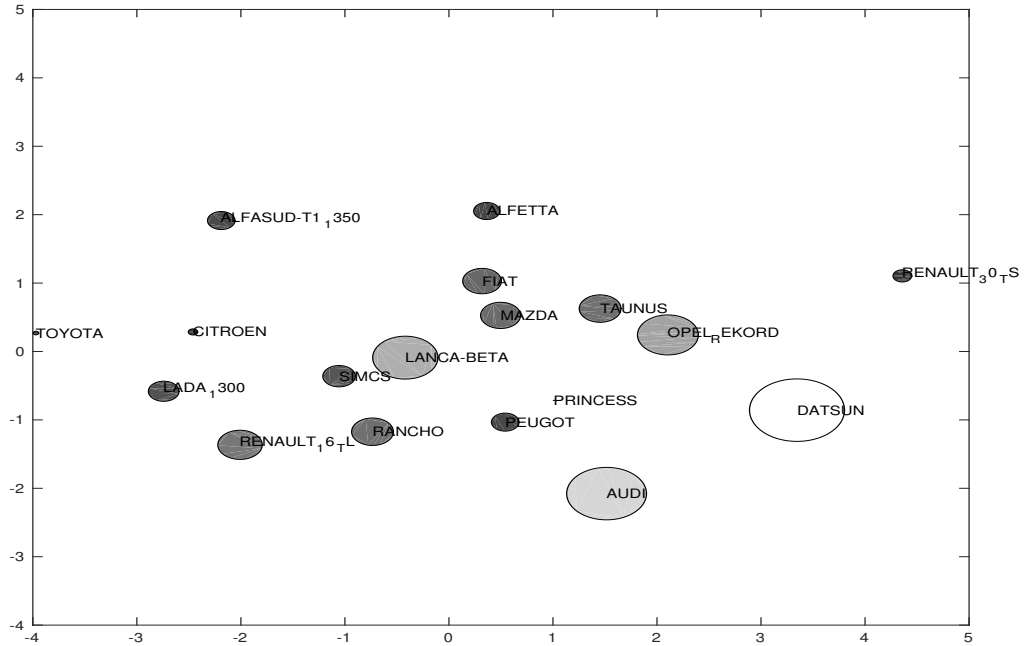


(b)

Figure 5: Projection of Iris data set. (a) and (b) show the projection quality using PCA and projection under pairwise distance control methods respectively. Two well separated groups can be observed.



(a)



(b)

Figure 6: Projection of cars data set. (a) and (b) show the projection quality using PCA and projection under pairwise distance control methods respectively. For PCA, the values of the quality are given between parentheses near each cars.

versicolor and virginica are close whereas the species *Iris setosa* are more distant.

Concerning cars data set, the projection of points using projection under pairwise distance control is given in Figure 6*b*. The expensive cars as "Audi 100", "Alfetta-1.66", "Dastun-200L", "Renault 30" are well-separated from the low-standard cars as "Lada-1300", "Toyota Corolla", "Citroen GS Club", "Simca 1300". We remark that the expensive cars are located on the right and low-standard one are located on the left.

By comparing the projection quality for each method presented in Figures 5 and 6 for these two data sets, we can say that our method projected the points without giving any importance to any group. Indeed, Figure 5*a* depicts a group with small values of quality measure and a group with high values of quality measure whereas the radii obtained by projection under pairwise distance control method are distributed in an equivalent way. Additionally, from Figure 6*b*, we can assert that the projected points obtained using projection under pairwise distance control method are well separated as there is no intersection between the circle. Moreover, the pairwise distances are significant in our method and give an interpretation on the position between points whereas the distances between the projected points using PCA are not interpretable as the cosine values can not be interpreted as distances. This is the particular strength of our method. Hence, projection under pairwise distance control suggests an absolute interpretation whereas PCA gives a relative one. From this, we can conclude from Figure 6*b* that there is a big difference between the two cars "Toyota" and "Renault 3" as the distances between this two cars is very important. Conversely, the distance between "Lada1300" and "Citroen" cars are small indicating then the closeness of these two cars. Note here that these two cars are very well projected leading to a very good interpretation.

For the qualitative and functional data sets, it is necessary to verify that the matrix  $B$  obtained by MDS method is semi-definite positive to use the quality measure  $\cos^2$  because the starting point of optimization is obtained from MDS. After that, in case of positiveness of matrix  $B$ , we can calculate the quality measure. In the projection of the soybean data set, four classes have been shown in Figure 7 and each class contains the diseases number of the class. But basically, the whole set of points can be divided in two large classes. Indeed, It is clear that class 2 is well separated from the others classes as there is no intersection

between the circles of class 2 and the circles of others classes. Moreover, class 1 can be considered as well separated class from classes 3 and 4 if we do not take into account the point  $D_3^*$ . Classes 3 and 4 are not at all well separated as we can exhibit that there are different intersections between the circles of these two classes. This result is figured in Stepp

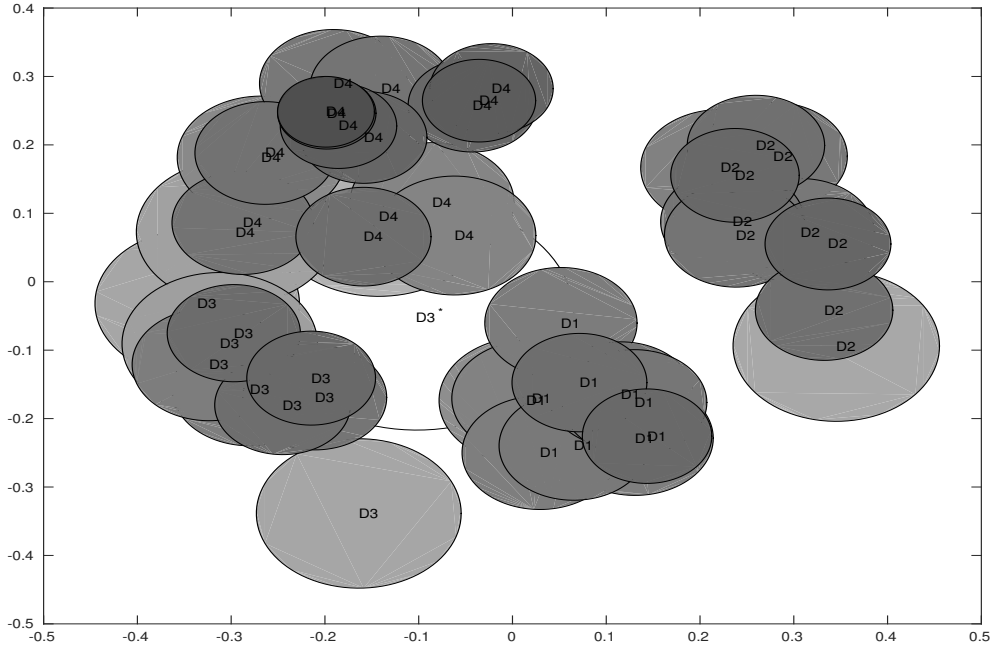


Figure 7: Projection under pairwise distance control for soybean dat set. Four groups are presented, indexed by D1, D2, D3 and D4.

(1984) which lists the value "normal" for the first two classes and "irrelevant" for the later two classes. The comparison of projection under pairwise distance control result with PCA is not possible for this data set because the matrix  $B$  is not semi-definite positive.

The coffee data set has been studied in several articles (Briandet et al. (2000); Bagnall et al. (2012)) and different classification methods have shown the different groups contained in this data set using our method and PCA. We can see clearly in Figure 8 the grouping structure that is obtained. In Figure 8*b*, we show that we have succeeded in differentiating the Arabica from Robusta coffee. These two classes are clearly presented, the first class indexed by 1 corresponding to Arabica coffee and the second one indexed by 2 corresponding to Robusta coffee. These classes are not well separated by comparing with the results of



quantitative data, since there are many intersections. Therefore, the representation of the points as circles and not as coordinates points gives more information about the real class of points and shows the points who have the possibility to be misplaced in a class.

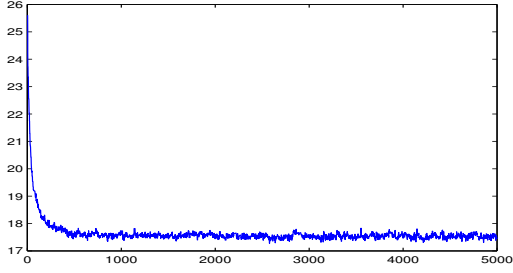
Figure 8a shows the projection quality using PCA. As all the eigenvalues of matrix  $B$  are positive, so we can compute the quality measure given by PCA. Comparing the projection quality of PCA and projection under pairwise distance control given respectively by Figures 8a and 8b, we can observe that the quality of projection of the set of points is pretty steady. Additionally, Metropolis-Hastings has been applied to these data sets. The trace plots of the optimization problem  $\mathcal{P}_{r,x}$  are shown in Figure 9 after 5000 iterations. Returning to Table 1, we can exhibit that Metropolis-Hastings algorithm solutions are very close to those obtained using the optimization package of Matlab and reciprocally. Thus, the obtained radii should be close to the optimum.

Finally, we present the lower bound computed from the three functions described in Section 3. The lower bound is taken by the minimal intersection of these functions. Returning to Table 1, we observe that in on case (cars), this lower bound indicates that the found solution is not far from the optimum. In the other cases, it seems that the lower bound while providing a good starting point can be improved. Note that this bound for tetrahedron example gives also good results forasmuch as algorithm 1 provides a solution three time smaller than the bound.

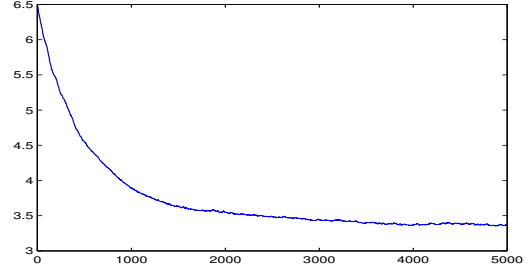
### 4.3.2 Dimensionality reduction results

One of high-dimensional data studies objectives is to choose from a large number of variables those which are important for understanding the underlying phenomena of study. So, the aim will be to reduce the dimension rather than to visualize data in  $\mathbb{R}^2$ . So, our method can also serve to reduce the number of variables by taking into account the minimal value of  $\sum_{i=1}^n r_i$ .

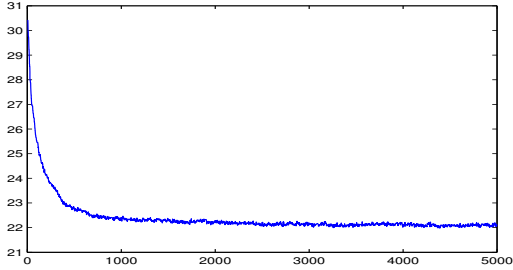
Here, we have solved the problem  $\mathcal{P}_{r,x}$  using the different possible dimension values. We have plotted in Figure 10 the values of  $\sum_{i=1}^n r_i$  as a guide for choosing the reduced number of variables. This figure shows the values of  $\sum_{i=1}^n r_i$  for the different data sets using different dimensions. It is clear to see that the value of  $\sum_{i=1}^n r_i$  decreases when the dimension



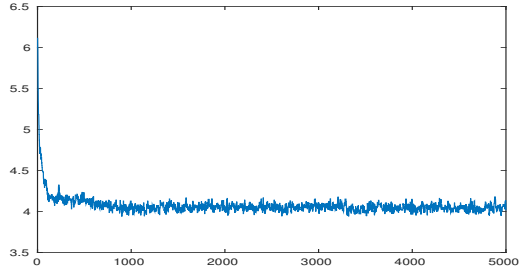
(a) Iris data set



(b) cars data set



(c) Coffee data set



(d) Soybean data set

Figure 9: Trace plots of Metropolis Hastings for different data sets. The x-axis corresponds to the iteration number and y-axis to the value of  $\sum_{i=1}^n r_i$ .

increases.

The main problem which is widely posed in dimension reduction methods is the determination of the number of components that are needed to be retained. Many methods have been discussed in the literature (Jolliffe (1986); Besse (1992),...) to determine the dimension of reduced space relying on different strategies related to the good explanation or the good prediction. So, with our method the choice of the reduced space dimension is related to the locally projection quality of points and how much the user is interested by the projection quality of points.

Concerning the quantitative data sets (Iris and cars), if the main objective of the user is to obtain a very good projection quality then a choice of three components against 4 for iris and 6 for cars can be a good choice as the value of  $\sum_{i=1}^n r_i$  is small and there is not a big difference between this value for this dimension and the values of the higher dimensions. For coffee data set, a dimensionality reduction from 56 sample time series down to 6 simple



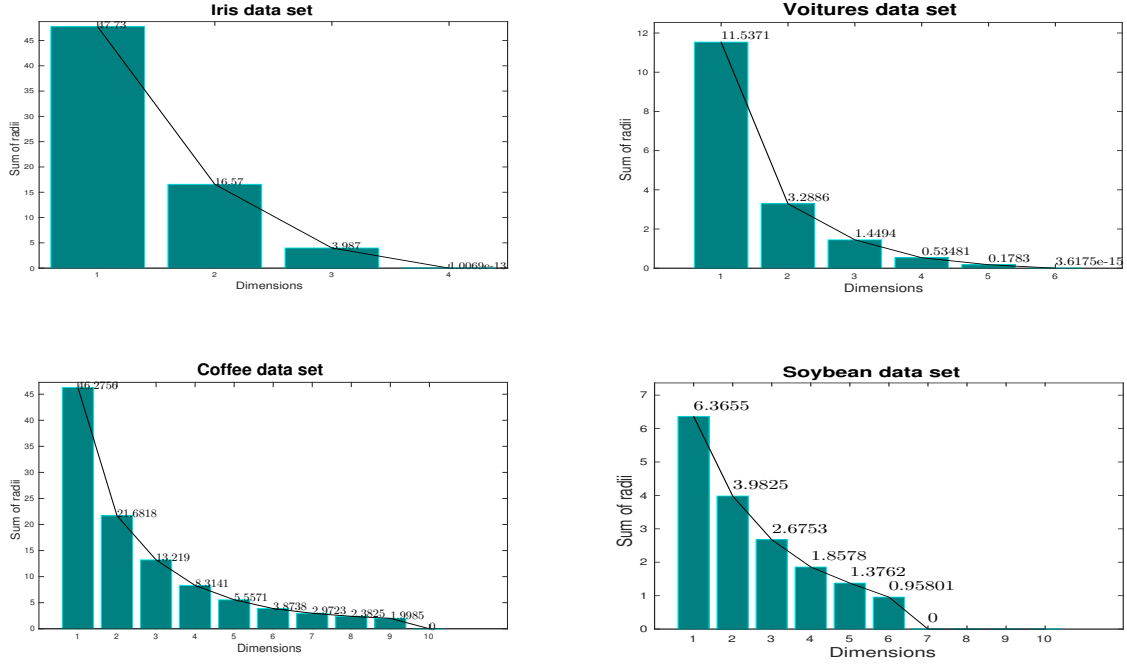


Figure 10: The scree plot of  $\sum_{i=1}^n r_i$  for different dimensions for the four data sets.

extracted features is considered as a good choice. The same idea can be seen for soybean data set, a reduced space dimension equal to 4 can be considered as efficient reduced space. Moreover, a comparison of our results with the existent results shows a coherence between them. For Iris data set, Chiu (1996) and Liu and Setino (1995) conclude that the number of variables can be reduced to 2 as the petal length and petal width variables are the most important variables among all variables. Similarly, this result can be seen for cars data set, Saporta (2006) in his book (Table 7.4.1 page 178) notices that the conservation of two dimensions leads to the explanation of 88% of inertia. So, these results seem very similar to our results, the important decrease is located between dimensions 1 and 2. The other decreases are negligible for these two data sets. Selection variables is studied on time series coffee data set in Andrews and McNicholas (2014). Using several analysis methods, the number of selected variables ranges between 2 and 13. This result is also seen using our method, a number of reduced variables taken between 2 and 9 gives a good quality projection of the points. Concerning soybean data set, Dela Cruz shows in his paper Dela Cruz (2015) that the 35 attributes can be reduced to 15 and here with our method, we

have succeeded to reduce the attributes to 6 by having a very good projection.

Hence, the presented results confirm that we can reduce the dimension non-linearly and still keep a way of assessing as reasonable number of dimensions. and that is efficient as a dimensionality reduction method.

#### 4.4 Advantages of projection under pairwise distance control method

As we have seen, our presented method has several advantages. To summarize:

firstly, it is a non-linear projection method which takes into account the projection quality of each point individually. Secondly, the distances between projected points are related to the initial distances between points offering a way to interpret easily the distances observed in the projection plane. Thirdly, the quality distribution between the points seems to be evenly distributed.

## 5 Conclusion

The purpose of this article was to outline a new non-linear projection method based on a new local measure of projection quality. Of course, in some projection methods a local measure is given but this measure cannot be applied unless in cases of linear projections, and even then it is not suitable for graphical representation.

The quality of projection is given here by additional variables called radii, which enable to give a bound on the original distances. We have shown that the idea can be written as an optimization problem in order to minimize the sum of the radii under some constraints. As the solution of this problem cannot be obtained exactly, we have developed different algorithms and proposed a lower bound for the objective function. As such, the method described here needs further research to improve the lower bound in order to assess how close the algorithms are from the minimum.

## References

Anderson E (1935), The Irises of the Gaspé Peninsula. Bull. Am. Iris Soc. 59: p. 2–5.

- Andrews, J. L. and McNicholas, P.D. (2014), Variable Selection for Clustering and Classification. *Journal of Classification*, 31, 136-153, 2014.
- Bagnall, A., Davis, L., Hills, J., and Lines, J. (2012), Transformation Based Ensembles for Time Series Classification. *Proceedings of the 12th SIAM International Conference on Data Mining (SDM 2012)*, 307–319, 2012
- Berge, C., Froloff, N., Kalathur, R.K., Maumy, M., Poch, O., Raffelsberger, W. and Wicker, N. (2010), Multidimensional fitting for multivariate data analysis. *Journal of Computational Biology*, 17, 723–732, 2010.
- Besse, P. (1992), PCA stability and choice of dimensionality. *Statistics & Probability Letters*, 13, 405-410, 1992.
- Boriah, S., Chandola, V., and Kumar, V. (2008), Similarity Measures for Categorical Data: A Comparative Evaluation, *Proceedings of the SIAM International Conference on Data Mining, SDM 2008, April 24-26, 2008, Atlanta, Georgia, USA.*
- Briandet, R., Kemsley, E. K., and Wilson, R. H. (1996), Discrimination of arabica and robusta in instant coffee by fourier transform infrared spectroscopy and chemometrics. *J. Agric. Food Chem.* 44, 1996.
- Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A. and Batista, G. (2015), The UCR Time Series Classification Archive, 2015, [www.cs.ucr.edu/~eamonn/time\\_series\\_data/](http://www.cs.ucr.edu/~eamonn/time_series_data/)
- Chiu, S. L. (1996), Method and Software for Extracting Fuzzy Classification Rules by Subtractive Clustering. In *Proceedings of North American Fuzzy Information Processing Society Conf. (NAFIPS'96)*, 1996.
- Ieva, F., Paganoni, A.M., Pigoli, D., and Vitelli, V. (2012), Multivariate functional clustering for the analysis of ECG curves morphology. *Journal of the Royal Statistical Society, series C. Applied Statistics*, in press, 2012.
- Johansen, A. M. and Evers, L. (2007) *Monte Carlo Methods*. Department of Mathematics, University of Bristol, 2007.

- Jolliffe, I.T. (1986), *Principal Component Analysis* (Springer, NewYork), 1986.
- Dela Cruz, G. B. (2015), Comparative Study of Data Mining Classification Techniques over Soybean Disease by Implementing PCA-GA. *International Journal of Engineering Research and General Science*, 3, 2015.
- Liu, H., Setiono, R. (1995), Chi2: feature selection and discretization of numeric attributes. In *Proceedings., Seventh International Conference on Tools with Artificial Intelligence (TAT'95)*, 1995.
- Mardia, K.V., Kent, J.T. and Bibby, J.M., (1979). *Multivariate analysis*, Academic Press, London, p. 406–407, 1979.
- Saporta, G. (2006), *Probabilités, analyse des données et statistique*. Technip, 2006, p. 428.
- Stepp, R. (1984), *Conjunctive conceptual clustering*, Doctoral dissertation, department of computer science, university of Illinois, Urbana-Champaign, IL, 1984.
- Wong, E. (2011), *Active-Set Methods for Quadratic Programming*, Ph.D. thesis, university of California, San Diego, 2011.

## Appendix A: Three lemmas

Let  $(S)$  be a sphere with center  $O$  and radius  $r$ ,  $x_1, \dots, x_n$  be  $n$  points such that  $\|x_i - O\| \leq r$  and  $g$  the center of gravity of points  $x_1, \dots, x_n$ . Using this hypothesis, we have developed three lemmas.

### Appendix A.1:

**Lemma A.1.** *For all points  $x_1, \dots, x_n$ , we have:*

$$\|x_i - O\| = r \text{ when } \sum_{i=1}^n \|x_i - g\|^2 \text{ is maximum .}$$

*Proof.* Let  $x$  a point inside the sphere  $(S)$  and  $y$  a point belonging  $(S)$ . The point  $x$  and  $y$  have as coordinates  $(x_1, x_2)$  and  $(r, 0)$  respectively. We want to show that by moving  $y$  by small movements along the sphere, we can approach the point  $x$  to the sphere border increasing thus  $\sum_{i=1}^n \|x_i - g\|^2$ .

We note  $y', x'$  the new positions after movements of  $y$  and  $x$  respectively. Let  $\theta$  be the angle between  $(Oy)$  and  $(Oy')$  and  $\vec{u}_\theta$  the displacement of  $y$ .

Approaching  $x$  to the sphere border requires the opposite movements of  $x$  and  $y$  with equal length. This constraint is necessary to keep the center in the same position.

We distinguish two cases:

1-  $x$  having  $x_2 < 0$ .

2-  $x$  having  $x_2 > 0$ .

The two cases are illustrated in Figure 11.

For case 1,  $x$  in the lower half of the sphere requires that  $y$  moves positively with angle  $\theta \in [0; \frac{\pi}{2}]$ . In this case, the vector  $u_\theta$  is given by:  $u_\theta^+ = \begin{pmatrix} \cos \theta - 1 \\ \sin \theta \end{pmatrix}$  and the inner product  $\langle u_\theta^+, xx' \rangle$  is given by:

$$\langle u_\theta^+, xx' \rangle = (x'_1 - x_1)(\cos \theta - 1) + (x'_2 - x_2) \sin \theta \quad (3)$$

Here, we have  $x'_1 \geq x_1$  and  $x'_2 \leq x_2$  that imply  $(x'_1 - x_1) \geq 0$  and  $(x'_2 - x_2) \leq 0$ . Moreover, we have  $0 \leq \cos \theta \leq 1$  and  $\sin \theta \geq 0$  that give  $\langle u_\theta^+, xx' \rangle < 0$ .

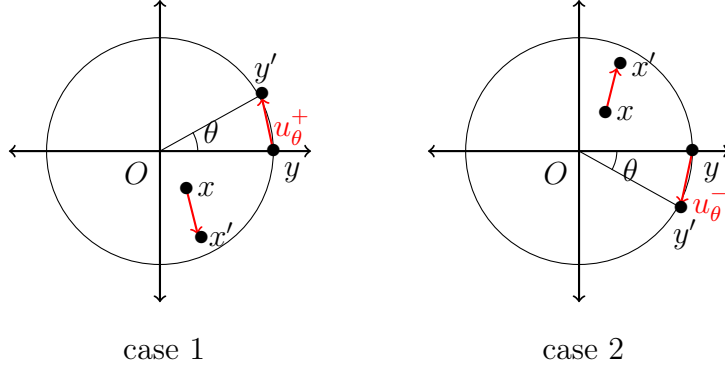


Figure 11: Representation of movements of points  $x$  and  $y$  in cases 1 and 2.

For case 2,  $x$  in the upper half of the sphere requires that  $y$  moves negatively with angle  $\theta \in [3\frac{\pi}{2}, 2\pi]$ . In this case, the vector  $u_\theta$  is given by:  $u_\theta^- = \begin{pmatrix} \cos \theta - 1 \\ -\sin \theta \end{pmatrix}$  and the inner product is:

$$\langle u_\theta^-, xx' \rangle = (x'_1 - x_1)(\cos \theta - 1) - (x'_2 - x_2) \sin \theta \quad (4)$$

Here, we have  $x'_1 \geq x_1$  and  $x'_2 \leq x_2$  that imply  $(x'_1 - x_1) \geq 0$  and  $x'_2 - x_2 \leq 0$ . Moreover, we have  $0 \leq \cos \theta \leq 1$  and  $\sin \theta \leq 0$  that give  $\langle u_\theta^-, xx' \rangle < 0$ .  $\square$

## Appendix A.2:

**Lemma A.2.** *The center of gravity  $g$  of  $x_1, \dots, x_n$  is the center of sphere ( $S$ ) i.e.  $O = g$ .*

*Proof.* We have:

$$\begin{aligned} \sum_{i=1}^n \|x_i - g\|^2 &= \sum_{i=1}^n \|x_i - O + O - g\|^2 \\ &= \sum_{i=1}^n \|x_i - O\|^2 + \sum_{i=1}^n \|O - g\|^2 + 2 \sum_{i=1}^n (x_i - O)'(O - g) \\ &= \sum_{i=1}^n \|x_i - O\|^2 + \sum_{i=1}^n \|O - g\|^2 + 2n(g - O)'(O - g) \\ &= \sum_{i=1}^n \|x_i - O\|^2 + n\|O - g\|^2 - 2n\|O - g\|^2 \\ &= \sum_{i=1}^n \|x_i - O\|^2 - n\|O - g\|^2 \end{aligned}$$

All the points belongs the sphere ( $S$ ) as a result of lemma A.1. So,  $\|x_i - O\|^2$  is fixed and equal to  $r^2$ . Thus, maximizing inertia  $\sum_{i=1}^n \|x_i - g\|^2$  amounts to minimizing  $\|O - g\|^2$ . Then, the minimum of  $\|O - g\|^2$  is zero so that  $O = g$ .  $\square$

### Appendix A.3:

**Lemma A.3.** *If  $\sum_{i=1}^n \|x_i - g\|^2$  is maximum for points  $x_1, \dots, x_n$  under constraint  $\|x_i - x_j\| \leq M$  with  $M = \max_{(i,j)} \{\|x_i - x_j\|\}$  then an upper bound of  $\sum_{i=1}^n \|x_i - g\|^2$  is given by:  $\frac{nM^2}{3}$*

*Proof.* Let  $x_1, \dots, x_n$  be the points maximizing  $\sum_{i=1}^n \|x_i - g\|^2$ . So, Using lemma A.1 and A.2, we can conclude that  $\|x_i - g\| = r$  for all  $i = 1, \dots, n$ .

Let now consider three points noted  $x, y$  and  $z$  among the  $n$  points. This three points belong to the sphere ( $S$ ). We suppose that the distance between  $y$  and  $z$  is equal to  $M$  i.e.  $\|y - z\| = M$  and by hypothesis, we have  $\|x - y\| \leq M$ .

We note  $\theta$  the angle between  $(zy)$  and  $(xy)$ ,  $\eta_1$  the angle between  $(zy)$  and  $(yg)$  and  $\eta_2$  the angle between  $(yg)$  and  $(yx)$ . Figure 12 illustrates the situation.

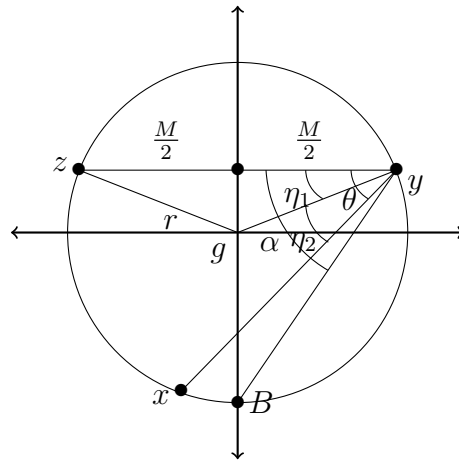


Figure 12: Representation of the points on the circle.

Furthermore, we have:

- $\theta = \eta_1 + \eta_2$

- $\cos \eta_1 = \frac{M}{2r}$
- $\cos \eta_2 = \frac{\|x - y\|}{2r}$

Consequently:

$$\begin{aligned}
\|x - y\| &= 2r \cos \eta_2 = 2r \cos(\eta_1 - \theta) \\
&= 2r (\cos \eta_1 \cos \theta + \sin \eta_1 \sin \theta) \\
&= 2r \left( \frac{M}{2r} \cos \theta + \sin \theta \sqrt{1 - \frac{M^2}{4r^2}} \right) \\
&= M \cos \theta + \sin \theta \sqrt{4r^2 - M^2}
\end{aligned}$$

We have  $\|x - y\| \leq M$ , then:

$$\begin{aligned}
M &\geq M \cos \theta + \sin \theta \sqrt{4r^2 - M^2} \\
M(1 - \cos \theta) &\geq \sin \theta \sqrt{4r^2 - M^2} \\
M^2(1 - \cos \theta)^2 &\geq \sin^2 \theta (4r^2 - M^2) \\
M^2(1 - \cos \theta) &\geq 2r^2(1 - \cos^2 \theta) \\
M^2 &\geq 2r^2(1 + \cos \theta) \\
r^2 &\leq \frac{M^2}{2(1 + \cos \theta)} \text{ as } \theta \neq \pi
\end{aligned}$$

Let  $B$  a point having coordinates  $(0, -r)$  and  $\alpha$  the angle between  $(zy)$  and  $(yB)$ . Without loss of generality, we suppose that  $x$  is in the third quadrant. By hypothesis, we have  $\|x - y\| \leq M$ ,  $\|x - z\| \leq M$  and  $\|y - B\| \leq M$  and additionally,  $\|y - B\| = \frac{M}{2 \cos \alpha}$  thus  $\alpha \leq \frac{\pi}{3}$  that gives  $\theta \leq \frac{\pi}{3}$  as  $x$  is in the third quadrant. So, we obtain:

$$r^2 \leq \frac{M^2}{2(1 + \cos \theta)} \leq \frac{M^2}{3}.$$

Hence,

$$\|x_i - g\|^2 \leq \frac{M^2}{3},$$

and then,

$$\sum_{i=1}^n \|x_i - g\|^2 \leq \frac{nM^2}{3}$$

which concludes the proof.  $\square$



## Appendix B: Three functions used to compute the lower bound

Recall optimization problem  $\mathcal{P}_{r,x}$ :

$$\left\{ \begin{array}{l} \min_{r_1, \dots, r_n, x_1, \dots, x_n} \sum_{i=1}^n r_i \\ \text{s.t. } d_{ij} - \|x_i - x_j\| \leq r_i + r_j \\ \|x_i - x_j\| - d_{ij} \leq r_i + r_j \end{array} \right.$$

### Appendix B.1: Function $f(M)$

Using the first constraint, we have:

$$\begin{aligned} d_{ij} &\leq \|x_i - x_j\| + r_i + r_j \\ d_{ij}^2 &\leq (\|x_i - x_j\| + r_i + r_j)^2 \\ \sum_{i < j} d_{ij}^2 &\leq \sum_{i < j} \|x_i - x_j\|^2 + \sum_{i < j} (r_i + r_j)^2 + 2 \sum_{i < j} (\|x_i - x_j\|)(r_i + r_j) \\ \sum_{i < j} d_{ij}^2 &\leq \sum_{i < j} \|x_i - x_j\|^2 + \sum_{i < j} (r_i + r_j)^2 + 2M \sum_{i < j} (r_i + r_j) \text{ as } \|x_i - x_j\| \leq M \end{aligned} \quad (5)$$

Let  $g$  be the center of gravity of the projected points, so:

$$\begin{aligned} \|x_i - x_j\| &= \|x_i - g - x_j + g\| \\ \|x_i - x_j\|^2 &= \|x_i - g\|^2 + \|x_j - g\|^2 + 2(x_i - g)'(x_j - g) \\ \sum_{i < j} \|x_i - x_j\|^2 &= \sum_{i < j} (\|x_i - g\|^2 + \|x_j - g\|^2) + 2 \sum_{i < j} (x_i - g)'(x_j - g) \end{aligned}$$

As  $\sum_{i < j} (x_i - g)'(x_j - g) = 0$  then:

$$\sum_{i < j} \|x_i - x_j\|^2 = (n-1) \sum_{i=1}^n \|x_i - g\|^2 \quad (6)$$

Replacing equation (6) in (5) gives :

$$(n-1) \sum_{i=1}^n \|x_i - g\|^2 + \sum_{i < j} (r_i + r_j)^2 + 2M \sum_{i < j} (r_i + r_j) - \sum_{i < j} d_{ij}^2 \geq 0 \quad (7)$$

The quantity  $\sum_{i=1}^n \|x_i - g\|^2$  is the inertia of projected points  $x_i$ . As long as we want to conserve the initial information, the inertia must be maximal under the constraint  $\|x_i - x_j\| \leq M$ .

Recalling equation (7) and by using Lemma A.3, we obtain:

$$\begin{aligned} \frac{n(n-1)}{3}M^2 + \sum_{i<j} (r_i + r_j)^2 + 2M \sum_{i<j} (r_i + r_j) - \sum_{i<j} d_{ij}^2 &\geq 0 \\ \frac{n(n-1)}{3}M^2 + (n-1) \left( \sum_{i=1}^n r_i \right)^2 + 2(n-1)M \sum_{i=1}^n r_i - \sum_{i<j} d_{ij}^2 &\geq 0 \\ \left( \sum_{i=1}^n r_i \right)^2 + 2M \left( \sum_{i=1}^n r_i \right) + \frac{n}{3}M^2 - \frac{1}{n-1} \sum_{i<j} d_{ij}^2 &\geq 0. \end{aligned} \quad (8)$$

The discriminant of equation (8) is given by:  $\Delta = 4 \left(1 - \frac{n}{3}\right) M^2 + \frac{4}{n-1} \sum_{i<j} d_{ij}^2$  and as  $r_i \geq 0, \forall i = 1, \dots, n$  we get:

$$\sum_{i=1}^n r_i \geq \sqrt{\left(1 - \frac{n}{3}\right) M^2 + \frac{1}{n-1} \sum_{i<j} d_{ij}^2} - M.$$

We note  $f(M) = \sqrt{\left(1 - \frac{n}{3}\right) M^2 + \frac{1}{n-1} \sum_{i<j} d_{ij}^2} - M$ .

## Appendix B.2: Function $g(M)$

Two situations are possible:

1.  $\exists(i', j')$  such that  $\|x_{i'} - x_{j'}\| = M$ , that gives:

$$r_{i'} + r_{j'} \geq \|x_{i'} - x_{j'}\| - d_{i'j'} \geq M - d_{i'j'} \geq M - d_{max}.$$

As  $\sum_{i=1}^n r_i \geq r_{i'} + r_{j'}$ , we obtain:

$$\sum_{i=1}^n r_i \geq M - d_{max}$$

2.  $\exists(i^*, j^*)$  such that  $d_{i^*j^*} = d_{max}$ , that gives:

$$r_{i^*} + r_{j^*} \geq d_{i^*j^*} - \|x_{i^*} - x_{j^*}\| \geq d_{max} - M.$$

Then, we obtain:

$$\sum_{i=1}^n r_i \geq d_{max} - M.$$

Hence:

$$\sum_{i=1}^n r_i \geq |D - d_{max}|. \quad (9)$$

We note  $g(M) = |D - d_{max}|$ .

### Appendix B.3: Function $h(M)$

Let us consider four distinct points  $i, j, k$  and  $l$ . We suppose that there is a couple  $(i, j)$  such that  $\|x_i - x_j\| = M$  and one of their coordinates is equal to zero ( $x_i = 0$  or  $x_j = 0$ ).

We distinguish two cases:

1.  $x_i = 0$ .
2.  $x_j = 0$ .

**Case 1:** For  $x_i = 0$ , we take  $x_j = \alpha x_k + \beta x_l$  with  $\alpha, \beta \in [0, 1]$ . The constraints related to these four points are the following:

$$\left\{ \begin{array}{l} \|x_j\| - d_{ij} \leq r_i + r_j \quad (C1) \\ \|x_k\| - d_{ik} \leq r_i + r_k \quad (C2) \\ \|x_l\| - d_{il} \leq r_i + r_l \quad (C3) \\ \|x_j - x_k\| - d_{jk} \leq r_j + r_k \quad (C4) \\ \|x_j - x_l\| - d_{jl} \leq r_j + r_l \quad (C5) \\ \|x_k - x_l\| - d_{kl} \leq r_k + r_l \quad (C6) \end{array} \right.$$

Firstly, using constraints (C4) and (C6) we obtain:

$$2 \sum_{t=1}^n r_t \geq d_{jk} - d_{kl} + \|x_k - x_l\| - \|x_j - x_k\|.$$

Additionally, as  $x_j = \alpha x_k + \beta x_l$  with  $\alpha, \beta \in [0, 1]$ , then

$$\|x_k - x_j\| = \|x_k - \alpha x_k - \beta x_l\| = \|(1 - \alpha)x_k - \beta x_l\| \leq \|x_k - x_l\|,$$

which gives

$$\sum_{t=1}^n r_t \geq \frac{d_{jk} - d_{kl}}{2}. \quad (10)$$

Secondly, using constraints (C5) and (C6) we obtain:

$$2 \sum_{t=1}^n r_t \geq d_{jl} - d_{kl} + \|x_k - x_l\| - \|x_j - x_l\|.$$

As  $\|x_l - x_j\| \leq \|x_k - x_l\|$  we obtain:

$$\sum_{t=1}^n r_t \geq \frac{d_{jl} - d_{kl}}{2} \quad (11)$$

Thirdly, constraints (C1) and initial hypothesis  $\|x_i - x_j\| = M$  lead to:

$$\sum_{t=1}^n r_t \geq |d_{ij} - M|. \quad (12)$$

Consequently, equations (10), (11) and (12) involve:

$$\sum_{t=1}^n r_t \geq \max \left\{ \frac{d_{jk} - d_{kl}}{2}; \frac{d_{jl} - d_{kl}}{2}; |d_{ij} - M| \right\} \text{ denoted } L_{ijkl}^1$$

**Case 2:** For  $x_j = 0$ , we take  $x_i = \alpha x_k + \beta x_l$  with  $\alpha, \beta \in [0, 1]$ . By analogy with case 1, we obtain:

$$\sum_{t=1}^n r_t \geq \max \left\{ \frac{d_{ik} - d_{kl}}{2}; \frac{d_{il} - d_{kl}}{2}; |d_{ij} - M| \right\} \text{ denoted } L_{ijkl}^2.$$

Due to the choice of one case among cases 1 and 2, we take the minimum of  $L_{ijkl}^1$  and  $L_{ijkl}^2$ .

Thus:

$$\sum_{t=1}^n r_t \geq \min \{L_{ijkl}^1; L_{ijkl}^2\}.$$

Moreover, for a given  $i, j$ , this inequality is verified. So that:

$$\sum_{t=1}^n r_t \geq \min_{i < j} \max_{k < l; k, l \neq i, j} \min \{L_{ijkl}^1; L_{ijkl}^2\}.$$

We note  $h(M) = \min_{i < j} \max_{k < l; k, l \neq i, j} \min \{L_{ijkl}^1; L_{ijkl}^2\}$ .