Projection under pairwise distance controls

Alawieh Hiba, Nicolas Wicker, Christophe Biernacki

Noname manuscript No.
(will be inserted by the editor)

Projection under pairwise distance control

Alawieh Hiba · Nicolas Wicker · Christophe Biernacki

Received: date / Accepted: date

Abstract Visualization of high-dimensional and possibly complex 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 considered. 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 at the scale of 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.

1 Introduction

Several domains in science use data with large number of variables in their studies such as in biological [11,17], chemical [30], geographical [32], financial [21] studies and many others studies. These 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.

Different types of multivariate data analysis methods were developed to study these data as dimensionality reduction, variable selection, cluster analysis and others. Typically, dimension reduction is used to summarize the data, variable selection to choose the pertinent variables from the set of candidate

H. Alawieh, N. Wicker, C. Biernacki
Université Lille 1, - UFR de Mathématiques, cité scientifique, 59655 Villeneuve d'Ascq, France
E-mail: alawieh.hiba@gmail.com

variables and cluster analysis to group the objects or variables. In our study, we focus on dimension reduction. Dimension reduction techniques can be used in different ways like data dimensionality reduction that projects the data from a high-dimensional space to a low-dimensional space or data visualization that provides a simple interpretation of the data in $\mathbb{R}^2$ or $\mathbb{R}^3$.

Many data dimensionality reduction and data visualization methods have been proposed to drop the difficulties associated to the high dimensional data [26,16,12,24,9]. To quote a few, principal component analysis (PCA) [20], multidimensional scaling (MDS) [31], scatter plot matrix [13] and parallel coordinates [19] are some of the known used methods. Scatter plot matrix, parallel coordinates and Sammon's mapping methods are widely used to visualize multidimensional data sets. The first two methods have as inconvenience that when the number of dimensions grows, important dimensional relationships might not be visualized. Indeed, the quality of projection assessed by the percentage of variance that is conserved or by the stress factor is a global quality measure and 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 the 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 nonlinear projection. Moreover, PCA will fail to give a "good" representation in case of nonlinear configurations therefore, Kernel PCA has been developed to extract nonlinear principal components.

In this paper, we propose a new nonlinear 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 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 $i, j \in \{1, \ldots, n\}$. The task here is to project these points using distances into a reduced space $\mathbb{R}^m$ 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), Kernel PCA and multidimensional scaling is given to highlight the interest of our method.

2.1 Principal Component Analysis (PCA)

The PCA method is the most used one in the data visualization and dimensionality reduction. This method is a linear projection technique applied when the data is linearly separable. PCA can be stated as an optimization problem involving the squared Euclidean distances [26]. This optimization problem is the following:

$$\mathcal{P}_{\text{PCA}} : \begin{cases} \min\limits_{A \in \mathcal{M}_{p\times q}} \sum\limits_{1\leq i<j\leq n} |d_{ij}^2 - \|Ay_i - Ay_j\|^2| \\ s.t. \;\; \text{rank}(A) = m \\ \quad\;\; AA^T = I_p \end{cases}$$

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 $p \times q$ with q being the reduced space dimension. By construction, PCA cannot take into account nonlinear structures, since it describes the data in terms of a linear subspace. To deal with nonlinearity, it is possible to use kernel PCA, the reproducing kernel Hilbert space variant of PCA.

2.2 Kernel PCA (KPCA)

KPCA idea is to perform PCA in a feature space noted $\mathcal{F}$ obtained by a nonlinear mapping of data from its space into the feature space $\mathcal{F}$, where the low-dimensional latent structure is hopefully easier to discover. The mapping function noted Φ is considered as:

$$\begin{array}{rccc} \Phi : & \mathbb{R}^p & \rightarrow & \mathcal{F} \\ & X & \rightarrow & \Phi(X) \end{array}$$

The original data y_i is then represented in the feature space as a function $\Phi(y_i) = k(y_i, .)$, where $k(., .)$ is a positive kernel. Similarly to PCA, KPCA is based on finding the first m eigenvectors corresponding to the largest eigenvalues λ_i of the Gram matrix $K = (k_{ij})_{ij\in 1,\ldots,n}$. Let V_v for $v = 1, \ldots, m$ are the eigenvectors in the feature space and $P_{\Phi(y_i)}$ is the projection of $\Phi(y_i)$ onto the subspace $V_1, \ldots, V_m$. KPCA problem can be represented as a minimization problem of the following error:

$$\mathcal{E}_{\text{KPCA}} : \|\Phi(y) - P_{\Phi(y)}\|_2^2$$

with $P_{\Phi(y)} = \sum_{v=1}^{m} \langle \Phi(y), V_v \rangle V_v$

Furthermore, the only measures used to evaluate the projection quality of points are the squared cosine values as in PCA and theses values cannot be interpreted at the same time as the positions because the cosine values have not a specific unit. More precisely, the visualization of the projection in the reduced space using PCA and KPCA is not simple to be interpreted in term of original distances between the points. Indeed, in PCA, the cosine values do not give a quantitative assessment of the error made in considering the distances between the projected points, all the more in KPCA where the projected points are in the feature space so the term "distances" is not related to the distances between points in the original space.

2.3 Multidimensional Scaling (MDS)

Likewise PCA, Multidimensional scaling (MDS) consists in finding a new data configuration in a reduced space. The main difference between these two methods is that the input data in MDS are typically comprehensive measures of similarity or dissimilarity between objects, they are called "proximity". The key idea of MDS is to perform dimensionality reduction in a way to approximate high-dimensional distances noted δ_{ij} by the low-dimensional distances d_{ij} where d_{ij} is equal to the distance between x_i and x_j the coordinates of i and j in the reduced space. In the classical and simplest case of MDS, the least-squares loss function noted Stress is given as follows:

$$\text{Stress} = \sqrt{\sum_{1 \leq i < j \leq n} (d_{ij} - \|x_i - x_j\|)^2}.$$

By minimizing the stress function, we find the best configuration of $(x_1, \ldots, x_n) \in \mathbb{R}^p$ such that the distances fit to the initial distances. Now if we consider n variables $r_1, \ldots, r_n \in \mathbb{R}^+$, the sum of which bounds the stress function, the optimization problem $\mathcal{P}_{\text{MDS}}$ can be equivalently rewritten as:

$$\mathcal{P}_{\text{MDS}} : \begin{cases} \min\limits_{r_1,\ldots,r_n \in \mathbb{R}^+} \sum\limits_{i=1}^{n} r_i \\ s.t. \quad \sum\limits_{i=1}^{n} r_i \geq \dfrac{1}{n-1} \sqrt{\sum\limits_{1 \leq i < j \leq n} (|d_{ij} - \|x_i - x_j\|)^2} \end{cases}$$

A criterium to determine the local projection quality is proposed by Born and Groenen in [6] called Stress-per-point (SPP). The SPP of the point i is given by:

$$SPP_i = \frac{\frac{\sum_{j=1, j \neq i}^{n} (d_{ij} - \|x_i - x_j\|)^2}{\sum_{j=1, j \neq i}^{n} d_{ij}^2}}{Stress}$$

with $Stress = \frac{\sum_{1\leq i<j\leq n}^{n}(d_{ij} - \|x_i - x_j\|)^2}{\sum_{1\leq i<j\leq n}^{n} d_{ij}^2}$. Again, this is difficult to interpret directly on the projection as a distance error.

However, we can observe that the constraint on $\sum_{i=1}^{n} r_i$ can be modified to have a stronger control on each d_{ij} in the following way: $|d_{ij} - \|x_i - x_j\|| \leq r_i + r_j$ where x_i and x_j are projection coordinates of points i and j. It is important to note here that we could use instead the following inequality $|d_{ij} - \|x_i - x_j\||^2 \leq r_i^2 + r_j^2$ but in that case interpreting the results would be awkward compared to what will be done in section 2.4.1.

So, our objective is to propose a new nonlinear projection method that controls individually the projection of points and gives a graphical representation in the same metric as the original space one with an error associated to each point.

2.4 Method description

Let consider $x_1, \ldots, x_n$ be the coordinates of the projected points in $\mathbb{R}^m$ and $\|x_i - x_j\|$ is the distance between two projected points (i, j). Radii are introduced in this paper to assess how much $\|x_i - x_j\|$ is far from the given distance d_{ij}. Indeed for couple (i, j), what we aim at is that $\|x_i - x_j\|$ close to d_{ij} should imply small radii (r_i, r_j). Figure 1 depicts this idea: for each point $P_i \in \{1, \ldots, n\}$, the projection of P_i belongs to a sphere with center x_i and radius r_i such that for each couple $(i, j) \in \{1, \ldots, n\}$ we have $\|x_i - x_j\| - (r_i + r_j) \leq d_{ij} \leq \|x_i - x_j\| + r_i + r_j$.

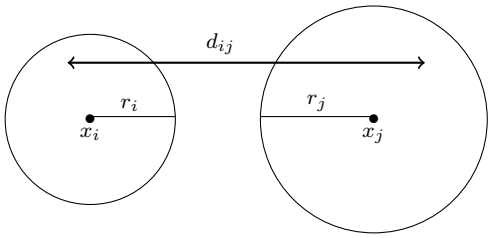


Fig. 1: Examples of radii for bounding the original distance d_{ij}

Radii for incertitude metric: The idea presented above can be expressed by finding the value of radii that satisfy these two constraints:

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

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

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

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

$$\mathcal{P}_{r} : \begin{cases} \min\limits_{r_1,\ldots,r_n \in \mathbb{R}^+} \sum\limits_{i=1}^{n} r_i \\ s.t \quad |d_{ij} - \|x_i - x_j\|| \leq r_i + r_j, \text{ for } 1 \leq i < j \leq n \end{cases}$$

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. So, for any method producing points in a reduced space as PCA for instance, we can compute the radii as a post-processing to assess the local quality of the projected points.

P_{rx} is a non-convex optimization problem: For any p, even $p = 1$, the optimization problem P_{rx} is not a convex problem. Indeed, if we take 4 aligned points with $x_i = 0$, $x_j = 2$, $y_i = 3$ and $y_j = 1$ and the distance between i and j equal to $d_{ij} = 2$. Let consider the function $g(w) = |\|w_i - w_j\| - d_{ij}|$ then we have $g(x) = 0$ and $g(y) = 0$ but $g(\frac{x+y}{2}) = |0 - 2| = 2 > \frac{g(x) + g(y)}{2} = 0$.

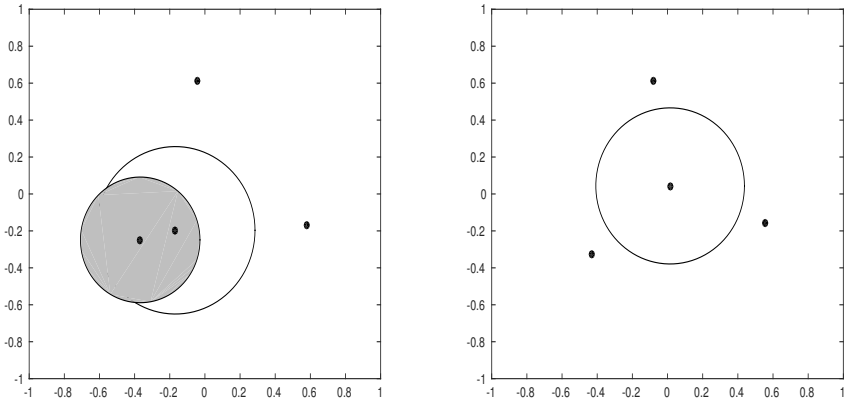
The problem P_{rx} is only a convex problem in dimension 1 if $x_1, \ldots, x_n$ are ordered. Indeed, let consider $x_j \leq x_i$ so we have $g(x) = |x_i - x_j - d_{ij}|$ and $g(y) = |y_i - y_j - d_{ij}|$ so that $\lambda, \mu \geq 0$ we have :

$$\begin{aligned} g(\frac{\lambda}{\lambda+\mu}x + \frac{\mu}{\lambda+\mu}y) &= \left| \frac{\lambda}{\lambda+\mu}(x_i - x_j - d_{ij}) + \frac{\mu}{\lambda+\mu}(y_i - y_j - d_{ij}) \right| \\ &\leq \frac{\lambda}{\lambda+\mu}g(x) + \frac{\mu}{\lambda+\mu}g(y) \end{aligned}$$

Therefore given an ordering we have each time a convex optimization that can be solved exactly so that the global optimum can be found by taking the minimum obtained for all permutations of $x_1, \ldots, x_n$. However, this is only working in dimension 1, in other dimensions, an approximate non-convex optimization is of course needed.

2.5 Visualization example

Let us apply our projection method to a simple example by taking a tetrahedron with all pairwise distances equal to 1. For problem $\mathcal{P}_r$, the coordinates x_i for $i = 1, \ldots, 4$ are obtained using multidimensional scaling. Using linear and nonlinear optimization packages in Matlab respectively for problems $\mathcal{P}_r$ and $\mathcal{P}_{r,x}$ gives 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 2a corresponds to the first solution and Figure 2b corresponds to the second one. In Figures 2*a* and 2*b*, 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 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*a*, half of the points are well projected whereas the other half have large radii indicating that they are not well projected. In Figure 2*b* just one circle appears indicating that the projection quality using problem $\mathcal{P}_{r,x}$ is better than $\mathcal{P}_r$. Moreover it is worth noting that as the three outer points have radii all equal to 0, this indicates that they are all perfectly placed with respect to one another.



(a) (b)

Fig. 2: Projected points after solving $\mathcal{P}_r$ and $\mathcal{P}_{r,x}$. (*a*) shows the projection obtained from the solution of $\mathcal{P}_r$ using MDS and (*b*) shows the one obtained from the solution of $\mathcal{P}_{r,x}$.

2.6 Link with other methods

Multidimensional fitting (MDF) [4] 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 a given set of 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 $X = \{x_1| \cdots |x_n\}$ 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 $\dfrac{1}{n-1} \displaystyle\sum_{1 \leq i < j \leq n} |d_{ij} - \|x_i - x_j\||$ where $x_1, \ldots, x_n$ is the optimum of the associated MDF problem.

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

$$\sum_{1 \leq i < j \leq n} |d_{ij} - \|x_i - x_j\|| \leq \sum_{1 \leq i < j \leq n} r_i + r_j = (n-1) \sum_{i=1}^{n} r_i$$

So, $\displaystyle\sum_{i=1}^{n} r_i \geq \frac{1}{n-1} \sum_{1 \leq < i < j \leq n} |d_{ij} - \|x_i - x_j\||$, which concludes the proof.

3 Optimization tools

3.1 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. We have essentially considered three possibilities. The first of them is the matrix obtained by PCA or another projection method. In what follows, we present the two other ones.

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

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

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

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

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 $\mathcal{P}_r$ This strategy is more involved. We need first two properties which provide a way to improve the optimization results of problem $\mathcal{P}_{r,x}$.

Property 2 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 satisfied. Then, the corresponding solution can be improved by moving x_i along the line $x_j - x_i$ in order 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, $\exists a \in [a_k^{'}, a_k^{''}]$ 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 perform a scale change by multiplying the coordinates x_i, for $i = 1, \ldots, n$, by a constant $a \in \mathbb{R}$. Thus, the new optimization problem is given by:

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

Property 3 Let $r_1, \ldots, r_n; x_1, \ldots, x_n$ 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 $\exists\ \tilde{r}_1, \ldots, \tilde{r}_n$ 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 $r_1, \ldots, r_n; x_1, \ldots, x_n$ a feasible solution of problem $\mathcal{P}_{r,x}$ and $a, \tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_n$ a solution of $\mathcal{P}_{r,a}$ where a is kept constant. 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 we can easily saturate it using property 2. So, we have:

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

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 $i = 1, \ldots, n$, 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 $\mathcal{P}_r$ by solving $\mathcal{P}_{r,a}$ and using property 3
- secondly, improve the solution of $\mathcal{P}_{r,a}$ using property 2.

3.2 Algorithm 1

Using the different initial values of matrix X presented above, we solve now problem $\mathcal{P}_{r,x}$. For this task, we use algorithm 1 which gives the best solution that can be obtained using the different initial values cited above. This algorithm consists in two steps an initialization step and an interior-point optimization step.

Algorithm 1

Input: D: distance matrix, N: number of iterations.
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 $\mathcal{P}_{r,a}$. Obtained solution: $(X^I_{\mathcal{P}_{r,a}}, r^I_{\mathcal{P}_{r,a}})$.
  $X_0 \leftarrow X^I_{\mathcal{P}_{r,a}}$.
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.

3.3 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 [22] allows us to build a Markov chain with the 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:

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

with E an application given by:

$$\begin{aligned} E: \quad & \mathbb{R}^n \longmapsto \mathbb{R} \\ & x = (x_1, \ldots, x_n) \longmapsto E(x) = \text{Solution of problem } \mathcal{P}_r \text{ with fixed } x. \end{aligned}$$

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 $q(X \to .)$ 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:

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

- 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 $x_i^* = x_i + |c_{ij^{(i)}}| L_{ij^{(i)}}$.
  - else $x_i^* = x_i - |c_{ij^{(i)}}| L_{ij^{(i)}}$.

  with $L_{ij^{(i)}} = \dfrac{x_i - x_j}{\|x_i - x_j\|}$.

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 car data sets.
- Categorical data: Soybean data set.
- Functional data: Coffee data set.

The Iris data set [1] is a famous data set and is presented to show that the projection works 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*). Car data set [28] 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 [29] 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 (D4).

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 [10]. 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 [8].

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 $i = 1, \ldots, n$, 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 [7] computed by the formula $p_{ij} = \sum_{t=1}^{Q} w_t p_{ij}^t$ where $p_{ij}^t = \begin{cases} 1 & \text{if } i^t = j^t \\ \dfrac{n_k^2}{n_k^2 + 2} & \text{else} \end{cases}$, 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 dissimilarities into similarities: $s_{ij} = 1 - p_{ij}$, the distances can be obtained by the standard transformation formula converting similarities to distances: $d_{ij} = \sqrt{s_{ii} - 2s_{ij} + s_{jj}}$.

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 [18]. In that 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 as 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 problems $\mathcal{P}_r$ and $\mathcal{P}_{r,a}$, we apply firstly PCA

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

	$\sum r_i^{\text{Algo 1}}$	$\sum r_i^{\text{MH}}$
Iris	16.19	17.2
Cars	3.27	3.35
Soybean	3.98	3.93
Coffee	21.68	21.97

for quantitative data and MDS for categorical and functional data, 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 nonlinear optimization problems, therefore we use a nonlinear programming package to solve it selecting the active-set algorithm to obtain the best values of $(x_1, \ldots, x_n)$ and $(r_1, \ldots, r_n)$. 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 [33].

Our proposed Metropolis-Hastings algorithm can provide a good solution if parameters λ, σ and T are chosen adequately. For instance, λ should be such that the points belonging to unsaturated constraints are chosen with small probabilities. Therefore, we take it equal to 100. For the other parameters σ and T, we take their values respectively in a range from 0.01 to 100.

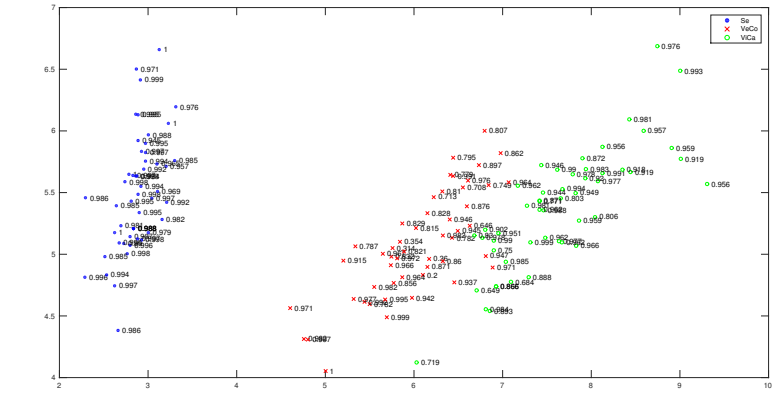
As we have mentioned in the section of visualization, the visualization of the projection of each point i in $\mathbb{R}^2$ is presented as a circle having x_i as center and r_i as radius so as the projected point belongs to this circle and this is the specificity of our method. For each data set, we 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 and KPCA, we use the squared cosine values as projection quality and for MDS the Stress-per-point (SPP).

4.3 Results

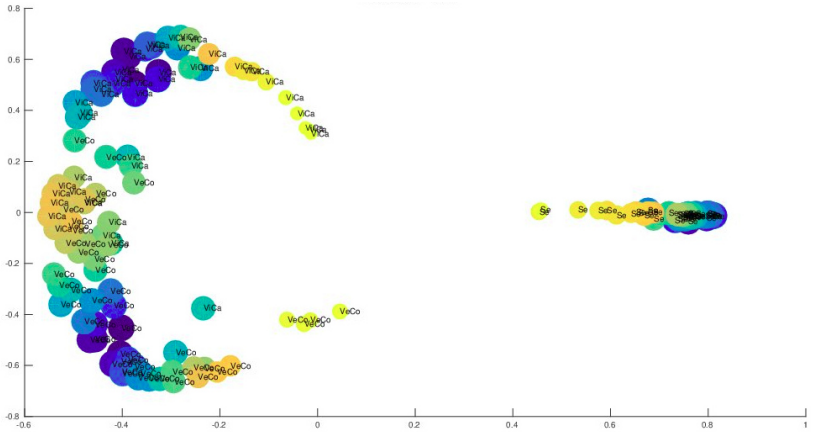
4.3.1 Visualization data in $\mathbb{R}^2$

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.

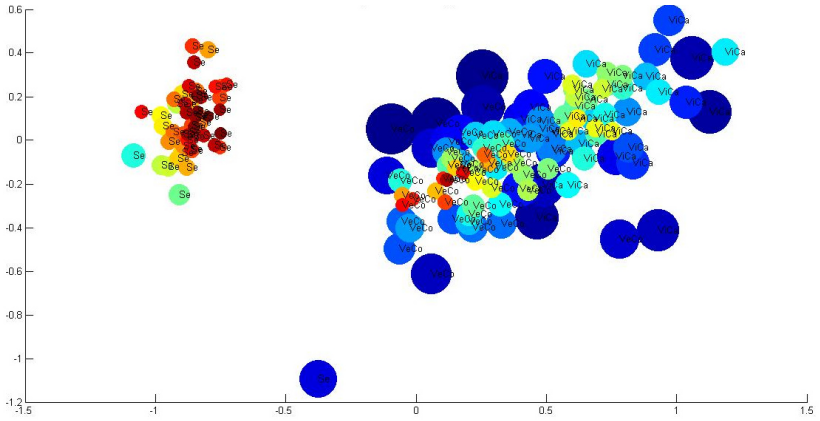
Figures 4 and 6 depict the results of projection under pairwise distance control for quantitative data. This projection is compared with the projection given by PCA, KPCA and MDS. For PCA and KPCA, we have plotted the projection of the points indexed by their squared cosine values. For MDS, we have used the smacof package in R to compute the stress-per-point and to plot the bubble plot represented the stress-per-point.



(a)



(b)



(c)

Fig. 3: Projection of Iris data set using PCA (a), KPCA (b) and MDS (c) respectively. The color convention stands as follows: the more red a disk is, the better the projection. Inversely, the more blue a disk is the worse the projection.

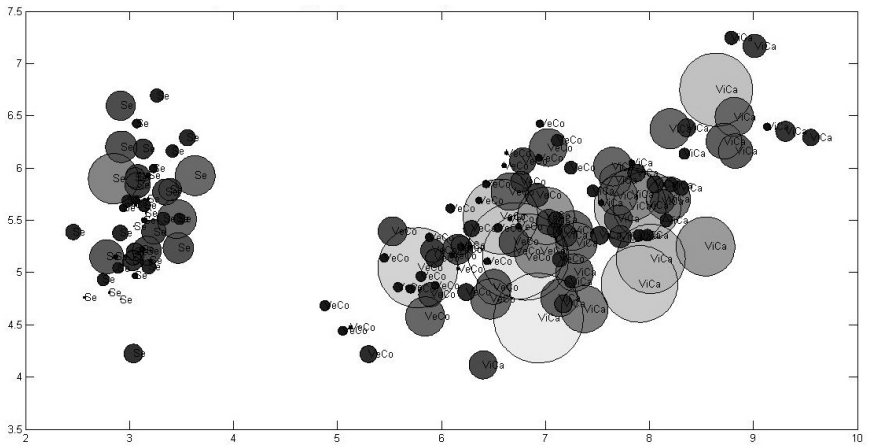
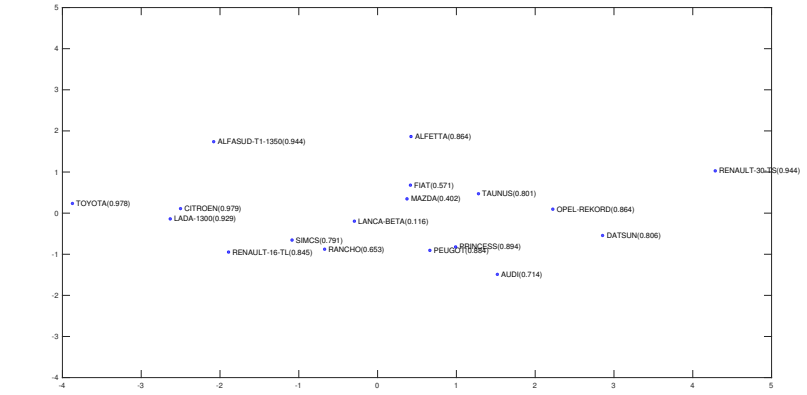


Fig. 4: Projection of Iris data set using projection under pairwise distance control method. Two well separated groups can be observed.

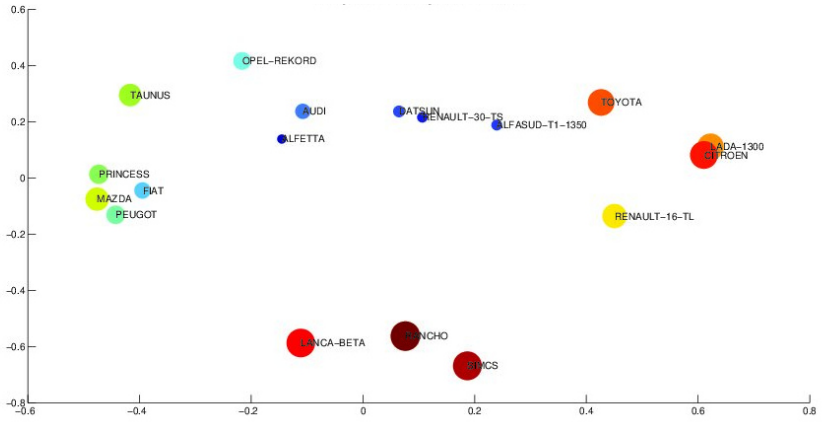
In the projection of Iris data set showed in Figure 4, it is interesting to remark that appealingly two areas are well separated. This corresponds to the well-known fact that Iris versicolor and virginica are close whereas the species Iris setosa are more distant. Refering to original data, the Iris data set contains three classes corresponding to the three types of iris plants and one class is linearly separable from the other two classes. This result clearly appears in our projection. Concerning the car data set, the projection of points using projection under pairwise distance control is given in Figure 6. The expensive cars as "Audi 100", "Alfetta-1.66", "Datsun-200L", "Renault 30" are well-separated from the low-standard cars as "Lada-1300", "Toyota Corolla", "Citroen GS Club", "Simca 1300".

For these two data sets, we want to compare the projection quality for each method presented in Figures 3 and 5 with the projection quality obtained using projection under pairwise distance control in Figures 4 and 6. Compared to PCA, we can say that our method projected the points without giving any importance to any group. Indeed, Figure 3a 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, we can assert that the projected points obtained using projection under pairwise distance control method are well separated as there is no circle intersection.

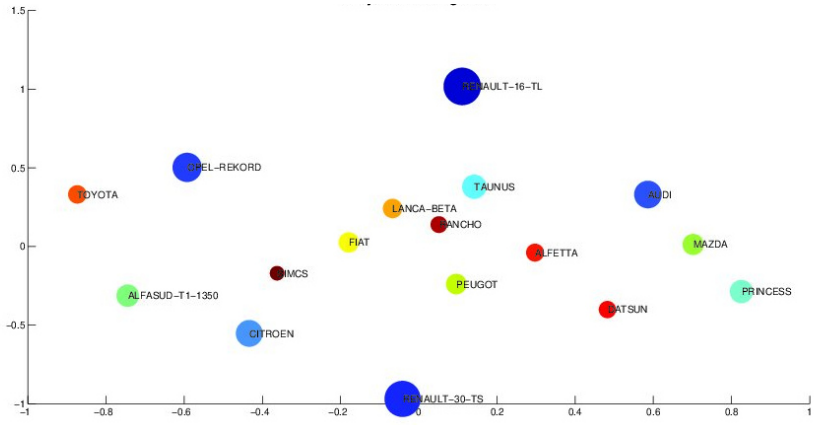
For KPCA, we have plotted the squared cosine as a circle to make the representation clearer especially for Iris data set as the Iris setosa species are projected next to each other. From Figure 3b we can conclude that in each category, the points which have close quality values are located side by side. The same conclusion can be drawn for cars data set in Figure 5b, we can see that the points with navy circle are located almost around the same Y



(a)



(b)



(c)

Fig. 5: Projection of car data set using PCA (a), KPCA (b) and MDS (c).

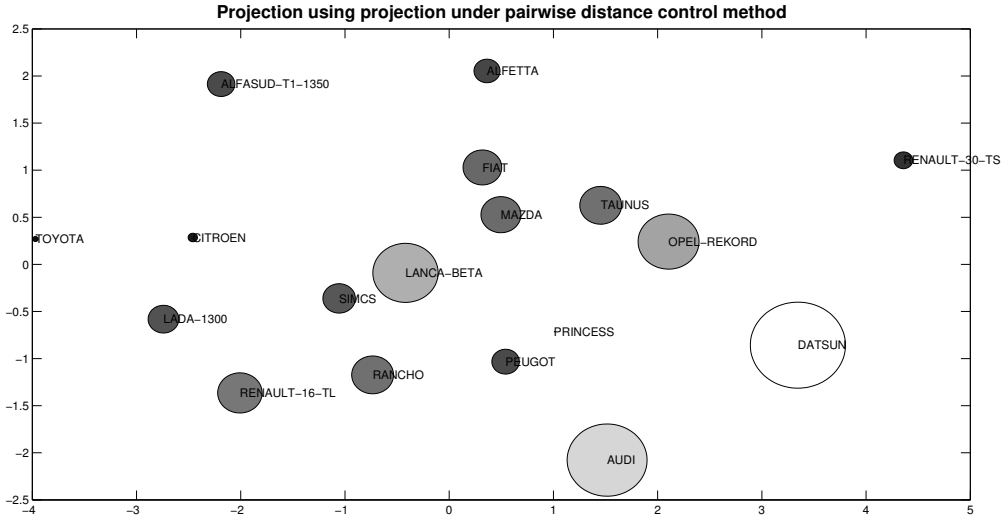


Fig. 6: Projection of car data set using projection under pairwise distance control.

axis coordinates and the sames goes for the red circles. So the local quality in KPCA is dependent on the points position.

Furthermore, by comparing our method projection for the Iris data set with the one obtained by MDS, we can draw the conclusions as when using PCA, the points are projected by giving more importance to Iris setosa group. Indeed, almost all the red circles (indicating a very good projection) are given to the Iris setosa species. Besides, the comparison of the position of the points in the reduced space in term of distance between points cannot be viewed in the classical method as the points in the reduced space is not in the same metric of the initial distances whereas in our method we have conserved the metric of the initial distances.

As for the car data set, in Figure 5b we notice that the cars Princess, Mazda, Fiat and Peugeot are in the same zone with small circles. So from this, the only conclusion that we can make is related to the size of circles and then to the quality of the projected points but we cannot say anything about the closeness of these 4 points as the distances here are in the feature space and are not related to the original space. Whereas, in Figure 6 we can conclude that the two cars Mazda and Fiat are well projected in the reduced space and they have similar characteristic as these two cars are close. The same conclusion can be made for Peugeot and Princess cars.

So, the pairwise distances are meaningful in our method and give an interpretation about the distances between points whereas the distances between the projected points using PCA, KPCA and MDS are not interpretable as the

cosine values and the Stress-per-point cannot be interpreted as distances in PCA and KPCA and MDS. This is the particular strength of our method. Hence, projection under pairwise distance control suggests an absolute interpretation whereas the other methods give a relative one. From this, we can conclude from Figure 6 that there is a big difference between the two cars "Toyota" and "Renault 3" as the distance between this two cars is very important. Conversely, the distance between "Lada1300" and "Citroen" cars is 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 of the distance between them.

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 squared cosine as quality measure 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 disease 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 other classes as there is no intersection between the circles of class 2 and the circles of other classes. Moreover, class 1 can be considered as well separated class from classes 3 and 4 if we do not take into account the largest D_3 circle.

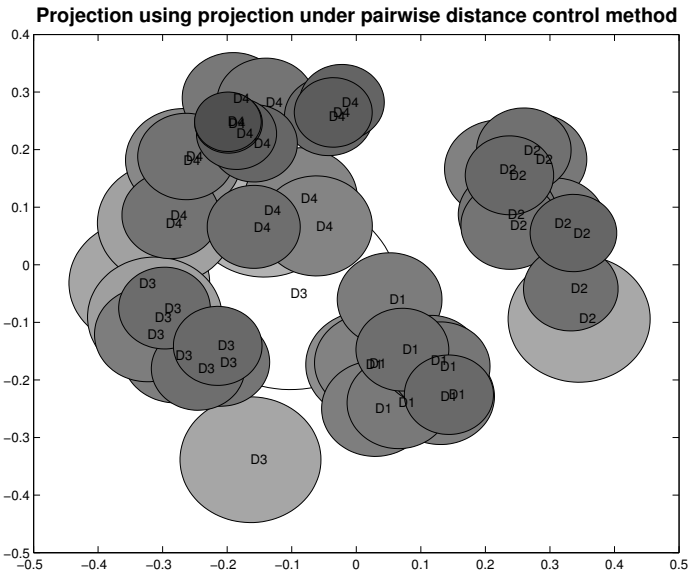


Fig. 7: Projection under pairwise distance control for soybean dat set. Four groups are presented, indexed by D1, D2, D3 and D4.

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 [29] which labels as "normal" the first two classes and "irrelevant" the latter two classes. The comparison of projection under pairwise distance control result with PCA and KPCA is not possible for this data set because the matrix B is not semi-definite positive.

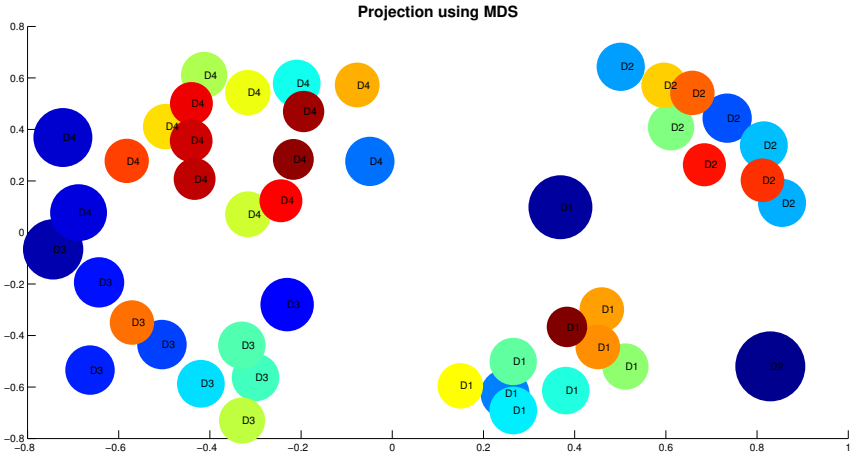
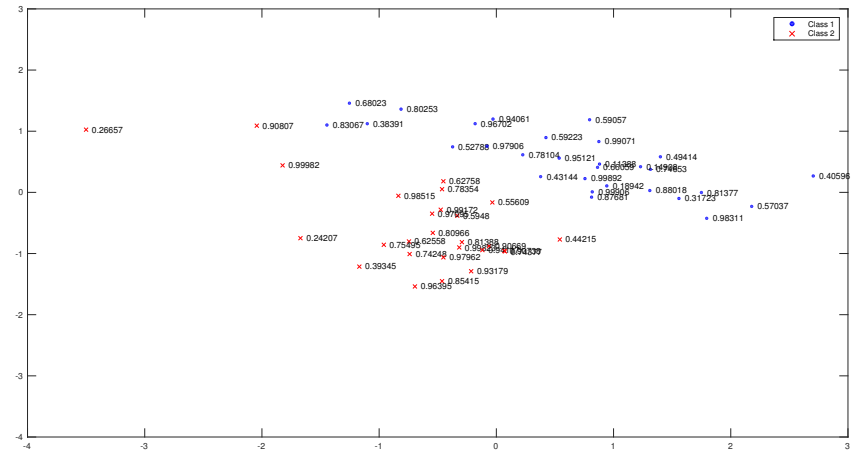


Fig. 8: MDS for soybean dat set. Four groups are presented, indexed by D1, D2, D3 and D4.

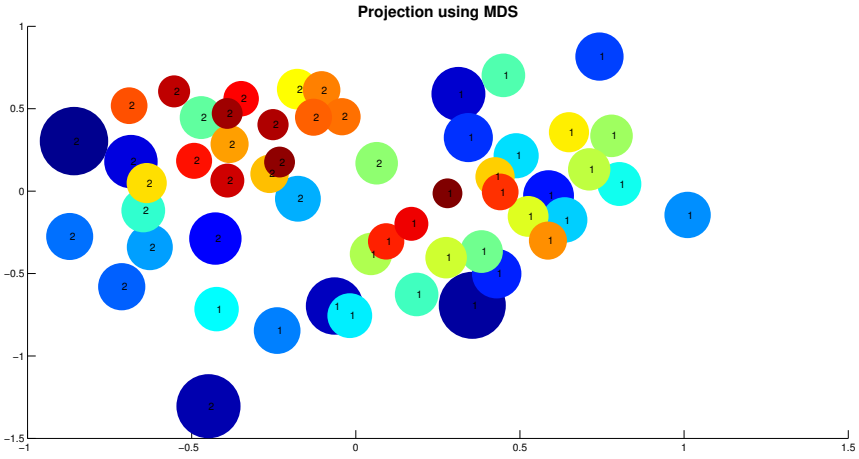
The coffee data set has been studied in several articles ([8,3]) and different classification methods have shown the different groups contained in this data set. We can see clearly in Figures 9 and 10 the grouping structure that is obtained.

In Figure 10, 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 simple points gives more information about the real class of points and shows the points who are at the risk of being misplaced in a class.

Figure 9a and 9b show the projection quality using PCA and MDS respectively. As all the eigenvalues of matrix B are positive, 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 9a and 10, we can observe that the quality of projection of the set of points is pretty steady.



(a)



(b)

Fig. 9: Projection of coffee data set using PCA and MDS.

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 11 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. Thus, the obtained radii are guessed to be close to the optimum.

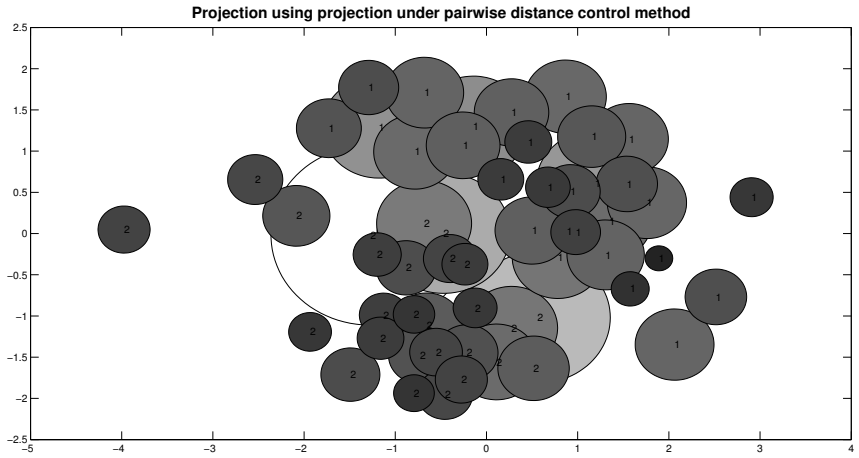


Fig. 10: Projection of coffee data set using projection under pairwise distance control. Two clusters indexed 1 and 2 indicate respectively Arabica and Robusta classes.

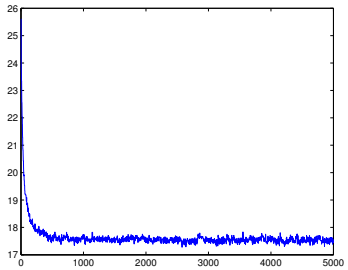
4.3.2 Dimensionality reduction results

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

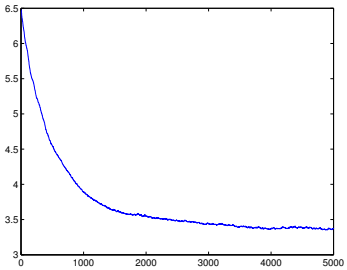
Here, we have solved the problem $\mathcal{P}_{r,x}$ using the different possible dimension values. We have plotted in Figure 12 the values of $\sum_{i=1}^{n} r_i$ as a guide for choosing the reduced number of variables. This figure shows the values of $\sum_{i=1}^{n} r_i$ for the different data sets using different dimensions. It is clear to see that the value of $\sum_{i=1}^{n} r_i$ decreases when the dimension 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 kept. Many methods have been discussed in the literature [23,5] to determine the dimension of the 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.

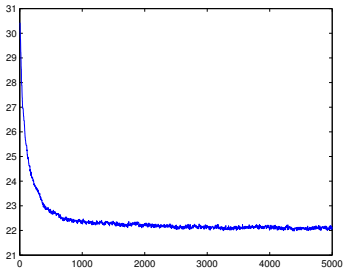
Concerning the quantitative data sets (Iris and car), if the main objective of the user is to obtain a very good projection quality then a choice of three components against 4 for iris and 6 for cars can be a good choice as the value of $\sum_{i=1}^{n} r_i$ is small and there is not a big difference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



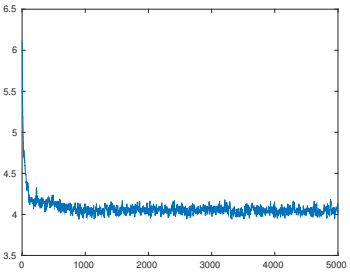
(a) Iris data set



(b) car data set



(c) Coffee data set



(d) Soybean data set

Fig. 11: Trace plots of Metropolis Hastings for different data sets. The x-axis corresponds to the iteration number and y-axis to the value of $\sum_{i=1}^{n} r_i$.

from 56 sample time series down to 6 simple extracted features is considered as a good choice. As for soybean data set, a reduced space dimension equal to 4 dimensions can be considered as an appropriate reduced space.

A comparison of our results with the existent results shows a coherence between them. For the Iris data set, articles [14] and [25] 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. For the car data set, Saporta in his book [28] (Table 7.4.1 page 178) notices that the conservation of two dimensions leads to the explanation of 88% of inertia where the inertia term reflects the importance of a component. So, these results seem very similar to our results, the important decrease is located between dimensions 1 and 2. The other decrements are negligible for these two data sets. Selection variables is studied on time series coffee data set in [2]. 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 [15] 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

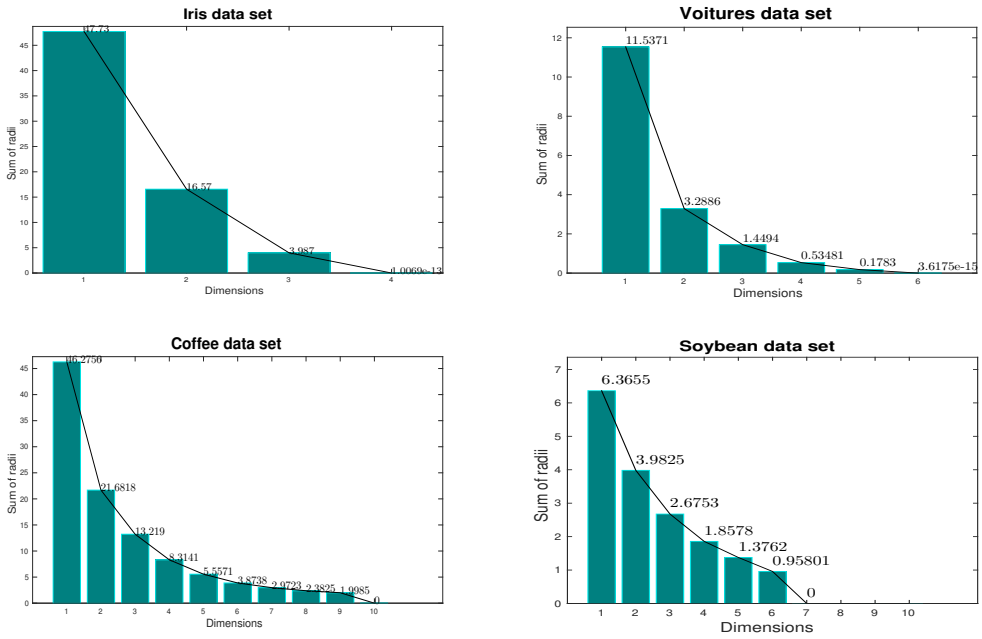


Fig. 12: The scree plot of $\sum_{i=1}^{n} r_i$ for different dimensions for the four data sets.

presented results confirm that we can reduce the dimension nonlinearly and at the same time assess a reasonable number of dimensions.

4.4 Advantages of projection under pairwise distance control method

As we have seen, our presented method has several advantages. To summarize: first, it is a nonlinear projection method which takes into account the projection quality of each point individually. Next, 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. Then, the quality projection of each point can even be used outside our method, that is as a post-processing of PCA or MDS. Finally, it also looks efficient for selecting the number of dimensions in dimension reduction.

5 Conclusion

The purpose of this article was to outline a new nonlinear 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 a stochastic optimization method. As perspectives, a lower bound for the optimisation problem is needed and this *radii* approach could also be applied to other methods as kernel PCA for example.

References

1. Anderson E, The Irises of the Gaspé Peninsula, *Bull. Am.* Iris Soc., 59, 2–5 (1935)
2. Andrews, J. L. and McNicholas, P.D., Variable Selection for Clustering and Classification, *Journal of Classification*, 31, 136-153 (2014)
3. Bagnall, A., Davis, L., Hills, J., and Lines, J., Transformation Based Ensembles for Time Series Classification, *Proceedings of the 12th SIAM International Conference on Data Mining*, 307–319 (2012)
4. Berge, C., Froloff, N., Kalathur, RK., Maumy, M., Poch, O., Raffelsberger, W. and Wicker, N., Multidimensional fitting for multivariate data analysis. *Journal of Computational Biology*, 17, 723–732 (2010)
5. Besse, P., PCA stability and choice of dimensionality, *Statistics & Probability Letters*, 13, 405-410, (1992)
6. Borg, I., Groenen, P. Modern Multidimensional Scaling: Theory and Applications (2nd ed.). *New York: Springer-Verlag* (2005).
7. Boriah, S., Chandola, V., and Kumar, V., Similarity Measures for Categorical Data: A Comparative Evaluation, *Proceedings of the SIAM International Conference on Data Mining* (2008)
8. Briandet, R., Kemsley, E. K., and Wilson, R. H., Discrimination of arabica and robusta in instant coffee by fourier transform infrared spectroscopy and chemometrics, *J. Agric. Food Chem.*, 44 (1), 170–174 (1996)
9. Chan, W. W-Y., A survey on multivariate data visualization in Science and technology, *Department of Computer Science and Engineering Hong Kong, University of Science and Technology,* 8(6), 1–29 (2006)
10. Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A. and Batista, G., *The UCR Time Series Classification Archive*, `www.cs.ucr.edu/~eamonn/time_series_data/` (2015).
11. Cheung, L.W., Classification approaches for microarray gene expression data analysis, *Methods Mol Biol.*, 802, 73–85 (2012)
12. Chinchilli, V.M., Sen, P.K., Multivariate Data Analysis: Its Methods, *Chemometrics and Intelligent Laboratory Systems*, 2, 29–36 (1987)
13. Cleveland, W.S., McGill, M.E., Dynamic Graphics for Statistics, *Wadsworth and Brooks/Cole*, Pacific Grove, CA, (1988)
14. Chiu, S. L., Method and Software for Extracting Fuzzy Classification Rules by Subtractive Clustering, *In Proceedings of North American Fuzzy Information Processing Society Conference* (1996)
15. Dela Cruz, G. B., Comparative Study of Data Mining Classification Techniques over Soybean Disease by Implementing PCA-GA, *International Journal of Engineering Research and General Science*, 3(5), 6–11 (2015)
16. Dempster, A.P., An overview of multivariate data analysis, *Journal of Multivariate Analysis*, 1(3), 316–346, (1971)

17. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D. and Lander, E. S., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, *Science*, 286, 531–537 (1999)
18. Ieva, F., Paganoni, A.M., Pigoli, D., and Vitelli., V., Multivariate functional clustering for the analysis of ECG curves morphology, Journal of the Royal Statistical Society, *Applied Statistics,* series C., 62(3), 401–418 (2012)
19. Inselberg, A., The Plane with Parallel Coordinates, *Special Issue on Computational Geometry, The Visual Computer*, 1, 69–91 (1985)
20. Jackson, J., *A Users Guide to Principal Components*, John Wiley & Sons, New York (1991)
21. Jagannathan, R. and Ma, T., Risk reduction in large portfolios: why imposing the wrong constraints helps, *J. Finan.*, 58, 1651–1683 (2003)
22. Johansen, A. M. and Evers, L. *Monte Carlo Methods*, Department of Mathematics, University of Bristol (2007)
23. Jollife, I.T., *Principal Component Analysis*, Springer, New York (1986)
24. Keim, D. A. and Kriegel, H. P., Visualization Techniques for Mining Large Databases: A Comparison, *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 923-938 (1996)
25. Liu, H., Setiono, R., Chi2: feature selection and discretization of numeric attributes, *In Proceedings., Seventh International Conference on Tools with Artificial Intelligence* (TAT'95) (1995)
26. Mardia, K.V., Kent, J.T. and Bibby, J.M., *Multivariate analysis*, Academic Press, London (1979)
27. Sammon, J., A nonlinear mapping for data structure analysis, *IEEE Transactions on Computers*, 18(5), 401–409 (1969)
28. Saporta, G., Probabilités, analyse des données et statistique, *Technip* (2006)
29. Stepp, R., Conjunctive conceptual clustering, Doctoral dissertation*, department of computer science, university of Illinois*, Urbana-Champaign, IL (1984)
30. Svante W., C. Albano, W. J. DunnIII, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg , M. Sjostrom., Multivariate Data Analysis in Chemistry, *Chemometrics,* 138, 17–95 (1984)
31. Togerson, W.S., *Theory and methods of scaling*, New York: Wiley (1958)
32. Van der Hilst, R., de Hoop, M., Wang, P., Shim, S.-H., Ma, P. and Tenorio, L., Seismostratigraphy and thermal structure of earth's core-mantle boundary region, *Science*, 315, 1813–1817 (2007)
33. Wong, E., *Active-Set Methods for Quadratic Programming*, Ph.D. thesis, university of California, San Diego (2011)