Projection under pairwise distance control

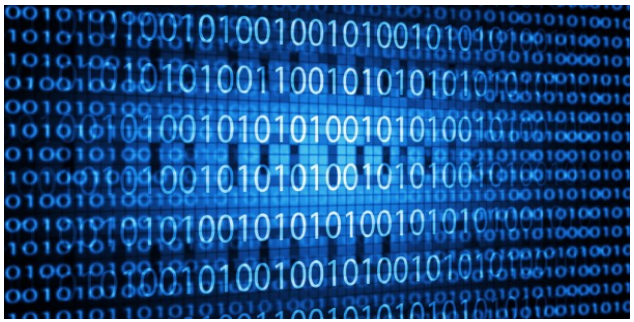
Hiba Alawieh, Nicolas Wicker, Christophe Biernacki

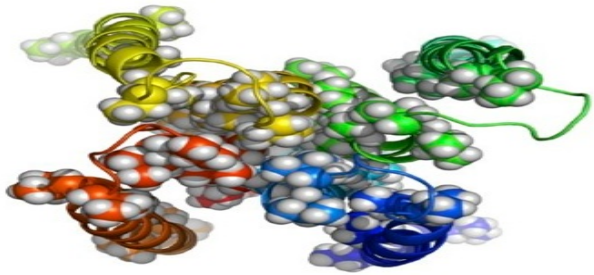
Laboratoire Paul Painlevé, Université Lille 1, 59655, Villeneuve d'Ascq, France.

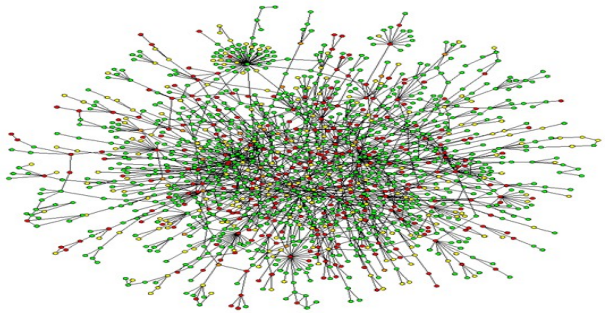
CMStatistics 2016
9-11 December 2016

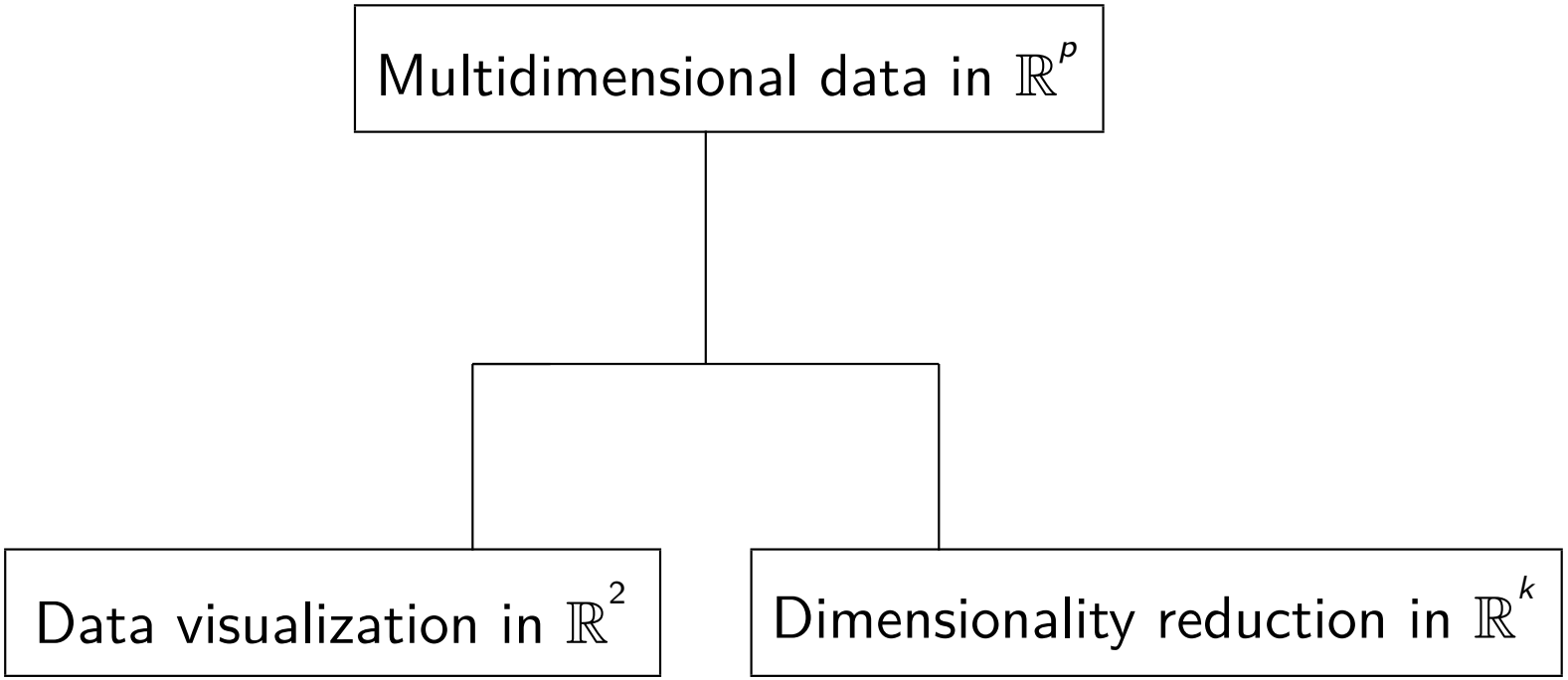
Table of contents

Multidimensional Data








Multidimensional data in $\mathbb{R}^p$
Data visualization in $\mathbb{R}^2$
Dimensionality reduction in $\mathbb{R}^k$

Short reminder of the PCA method

- **Aim:** Project the original data set in a reduced space by preserving as much of the variance from the original data set as possible.

Short reminder of the PCA method

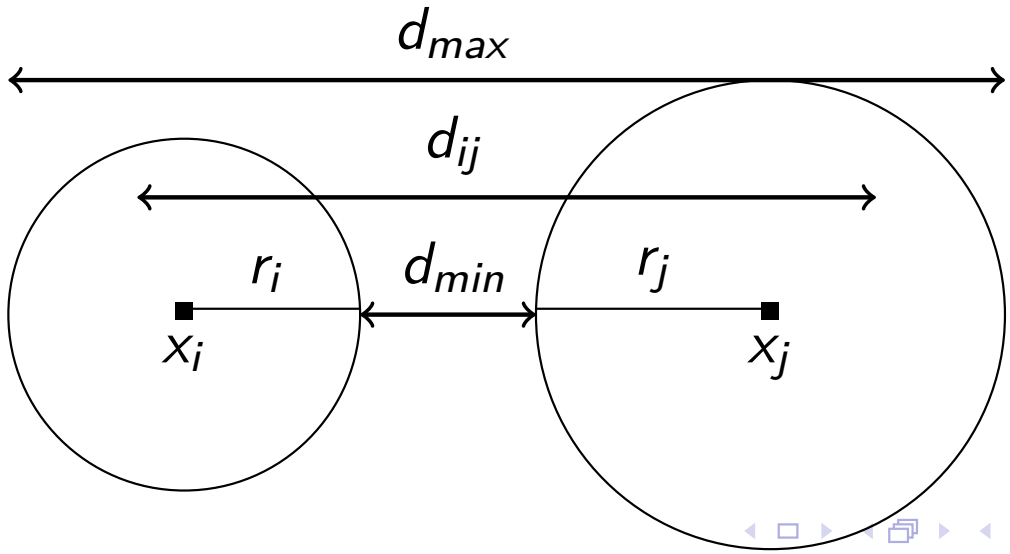
- **Aim:** Project the original data set in a reduced space by preserving as much of the variance from the original data set as possible.
- **Local quality:** Squared cosine of angle between the principal space and the vector of the point gives the local measure.
  - usable in linear projection.
  - unusable in non-linear projection.
  - Not interpretable as distances.

Presentation of the method

Our objective

Propose a new **non-linear** projection method taking into account the **local projection quality** that is **interpretable** as distances.

- The idea is to bound the distance d_{ij} by a minimal and maximal distances calculated on the projected points and the radii.



Basic of the method:

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n1} & \cdots & d_{nn} \end{bmatrix} \xrightarrow[(r_1, \cdots, r_n)]{} X = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}$$

- The variables $(r_1, \ldots, r_n)$ are called radii given for each point i such that $r_i \in \mathbb{R}$ for all $i = 1, \ldots, n$.

Characteristics of Radii

- Radii are important elements introduced to assess how much the distance $\|x_i - x_j\|$ is far from d_{ij}.
- (r_i, r_j) are small $\Rightarrow$ $\|x_i - x_j\|$ is close to d_{ij}.
- Radii indicate to which extent the projection of each point is accurate.
- The local quality is then given by the values of the radii.
- Both units of d_{ij}'s and r_i's are identical.

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

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

Problem $\mathcal{P}_{r,x}$ is a hard non-linear optimization problem.

Simplification of problem $\mathcal{P}_{r,x}$:

- $(x_1, \ldots, x_n)$ are fixed using PCA or any other projection method.
- Problem $\mathcal{P}_r$ is a new linear optimization problem obtained after fixing $(x_1, \ldots, x_n)$.
- It can easily be solved in $(r_1, \ldots, r_n)$ using linear programming.

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

- Solution of $\mathcal{P}_r$ is not in general the optimum of problem $\mathcal{P}_{r,x}$.

Different ways to find a solution of problem $\mathcal{P}_{r,x}$:

- Lower bound.
- Optimization.
- Simulation.

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

- Let $x_1, \cdots, x_n; r_1, \cdots, r_n$ a feasible solution of $\mathcal{P}_{r,x}$, and $M \in \mathbb{R}$ such that:

$$M = \max_{(i,j)} \{ \| x_i - x_j \| \}.$$

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

- Let $x_1, \cdots, x_n; r_1, \cdots, r_n$ a feasible solution of $\mathcal{P}_{r,x}$, and $M \in \mathbb{R}$ such that:

$$M = \max_{(i,j)} \{ \| x_i - x_j \| \}.$$

- The objective is to find three functions noted f, g, h depending on M such that:

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

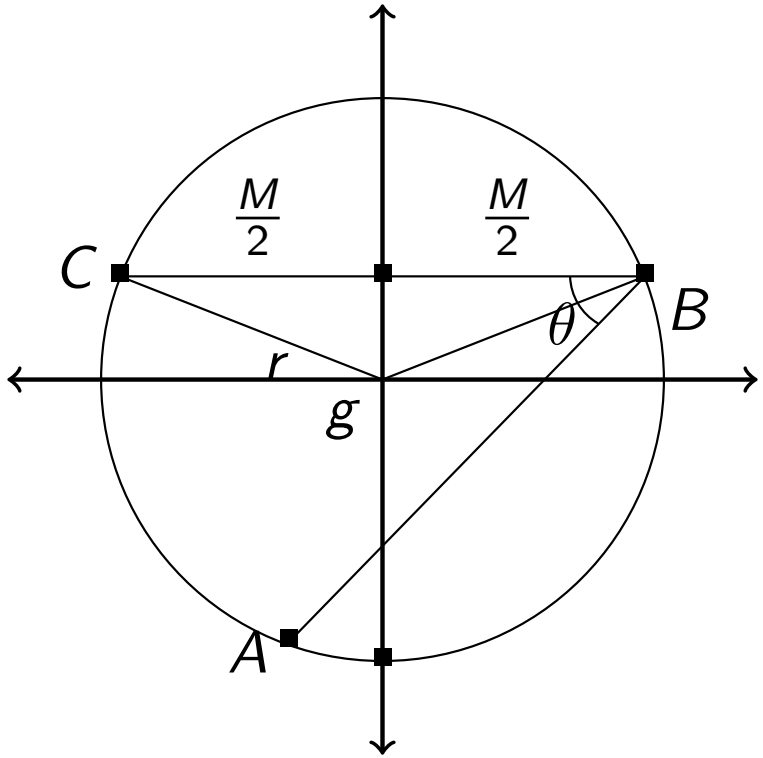
Computation of the functions

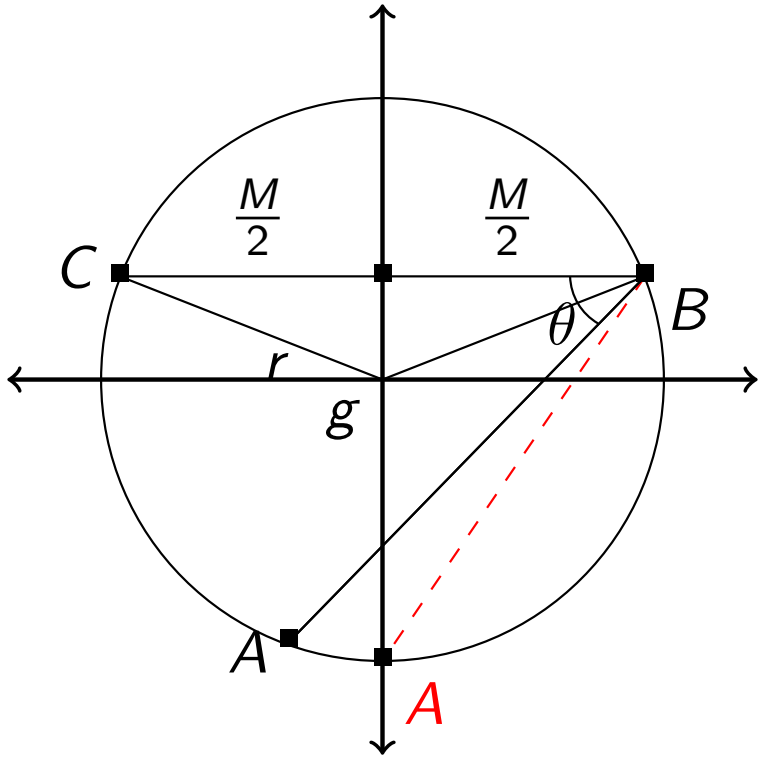
Function $f(M)$: This function is obtained by:

- summing all the squared constraints $d_{ij}^2 \leq (\|x_i - x_j\| + r_i + r_j)^2$.
- bounding $\sum_{1 \leq i < j \leq n} \|x_i - x_j\|^2$ by $\frac{n(n-1)}{3} M^2$ after maximizing the inertia of the projected points x_i under constraints: $\|x_i - x_j\| \leq M$ for all (i, j).

Indeed, we consider:

- (C) be the smallest circle containing the n points
- A, B and C belong to (C) and $\|x_B - x_C\| = M$.


$\frac{M}{2}$
$\frac{M}{2}$
C
B
θ
r
g
A


$\frac{M}{2}$
$\frac{M}{2}$
C
B
θ
r
g
A
A

- We have proved that the maximum of the inertia $\sum_{i=1}^{n} \| y_i - g \|^2$, under the constraint that $y_1, \ldots, y_n$ are inside (C), is equal to $nr^2 = n\frac{M^2}{3}$.
- Thus,

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

Function $g(M)$: This function is obtained using the maximal distance of d_{ij}.

$$g(M) = |M - d_{max}|$$

Function $h(M)$: This function is obtained by taking four distinct points i, j, k and l such that:

- For couple (i, j), $\|x_i - x_j\| = M$ and x_i or x_j is equal to zero.
- We consider the following linear combinations:

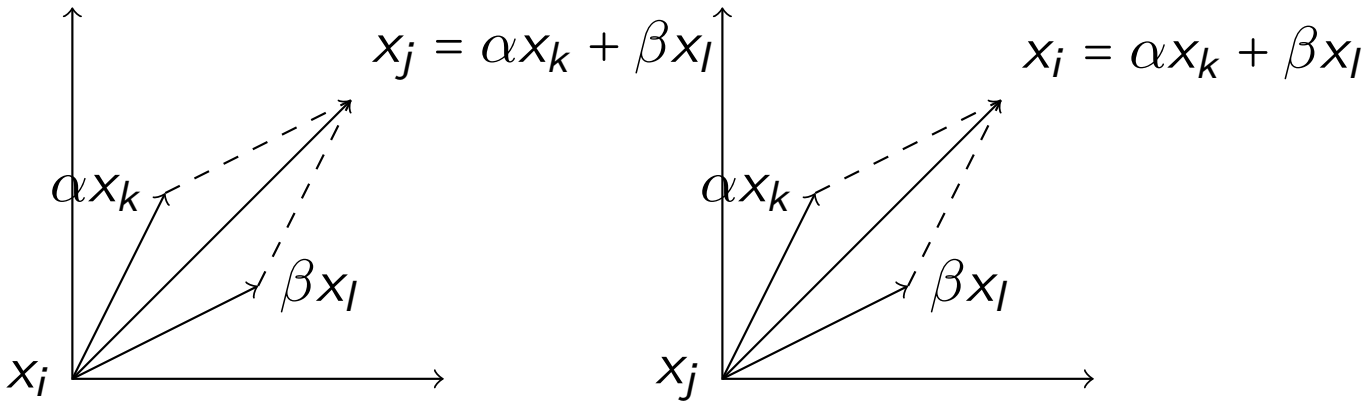
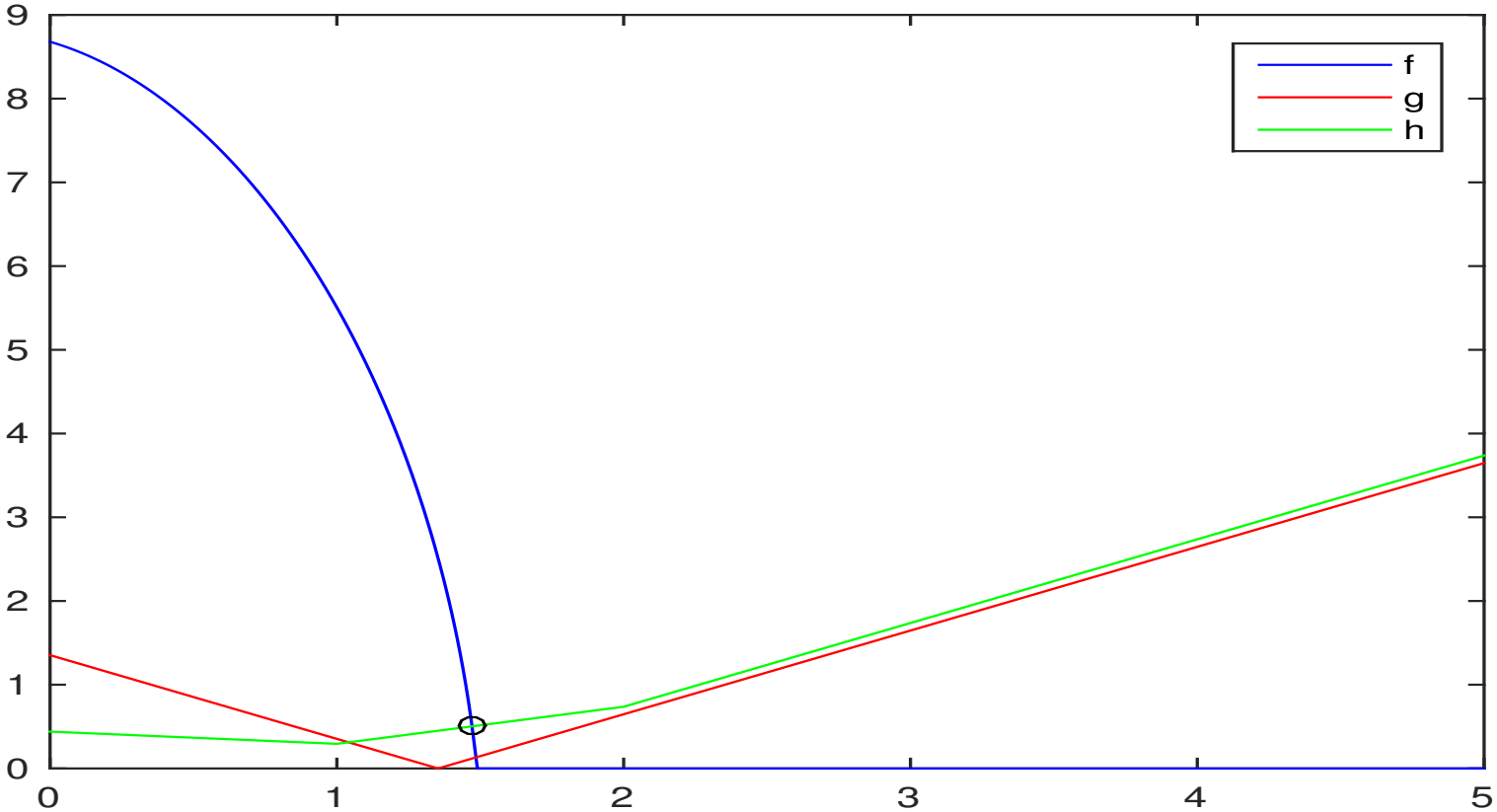


Illustration of the three function f, g and h.



Simulation algorithm

Stochastic optimization method ⇒ Metropolis-Hastings algorithm .

Simulation algorithm

Stochastic optimization method $\Rightarrow$ Metropolis-Hastings algorithm .

Target distribution

- The target distribution is related with the objective function of problem $\mathcal{P}_{r,x}$.
- An application E is given by:

$$\begin{array}{rccc} E: & \mathcal{M}_{n\times p} & \longmapsto & \mathbb{R} \\ & X=(x_1,\ldots,x_n) & \longmapsto & E(X)=\text{Solution of } \mathcal{P}_r \text{ with } x \text{ fix.} \end{array}$$

- The target distribution is: $\pi(s) \propto \exp\left(-E(x)/T\right)$.

The proposal distribution $q(X \to .)$ has been constructed by giving priority to the selection of points involved in saturated constraints.

The proposal distribution $q(X \to .)$ has been constructed by giving priority to the selection of points involved in saturated constraints.

Proposal distribution

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

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

- 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 $d_{ij^{(i)}} - \left\|x_i - x_{j^{(i)}}\right\| > 0$ then $x_i^* = x_i + \left|c_{ij^{(i)}}\right| L_i$.
  - else $x_i^* = x_i - \left|c_{ij^{(i)}}\right| L_i$.

  with $L_i = \dfrac{x_i - x_{j^{(i)}}}{\left\|x_i - x_{j^{(i)}}\right\|}$.

Algorithm 1 Metropolis-Hastings Algorithm:

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

Generate X^* from the *proposal distribution* $q(X^{t-1} \to X^*)$.

Solve linear optimization problem $\mathcal{P}_r$.

Calculate $$\alpha = \frac{g(s^*)q(X^* \to X^{t-1})}{g(s^{t-1})q(X^{t-1} \to X^*)} = \frac{g(s^*)\prod_{i=1}^{n} P_j^{i^*}}{g(s^{t-1})\prod_{i=1}^{n} P_j^{i^{t-1}}}.$$

if $\alpha = 1$ **then**

Take $X^t = X^*$.

else

$u = \mathcal{U}(0;1)$.

if $u \leq \alpha$ **then**

Take $X^t = X^*$

else

$X^t = X^{t-1}$.

end if

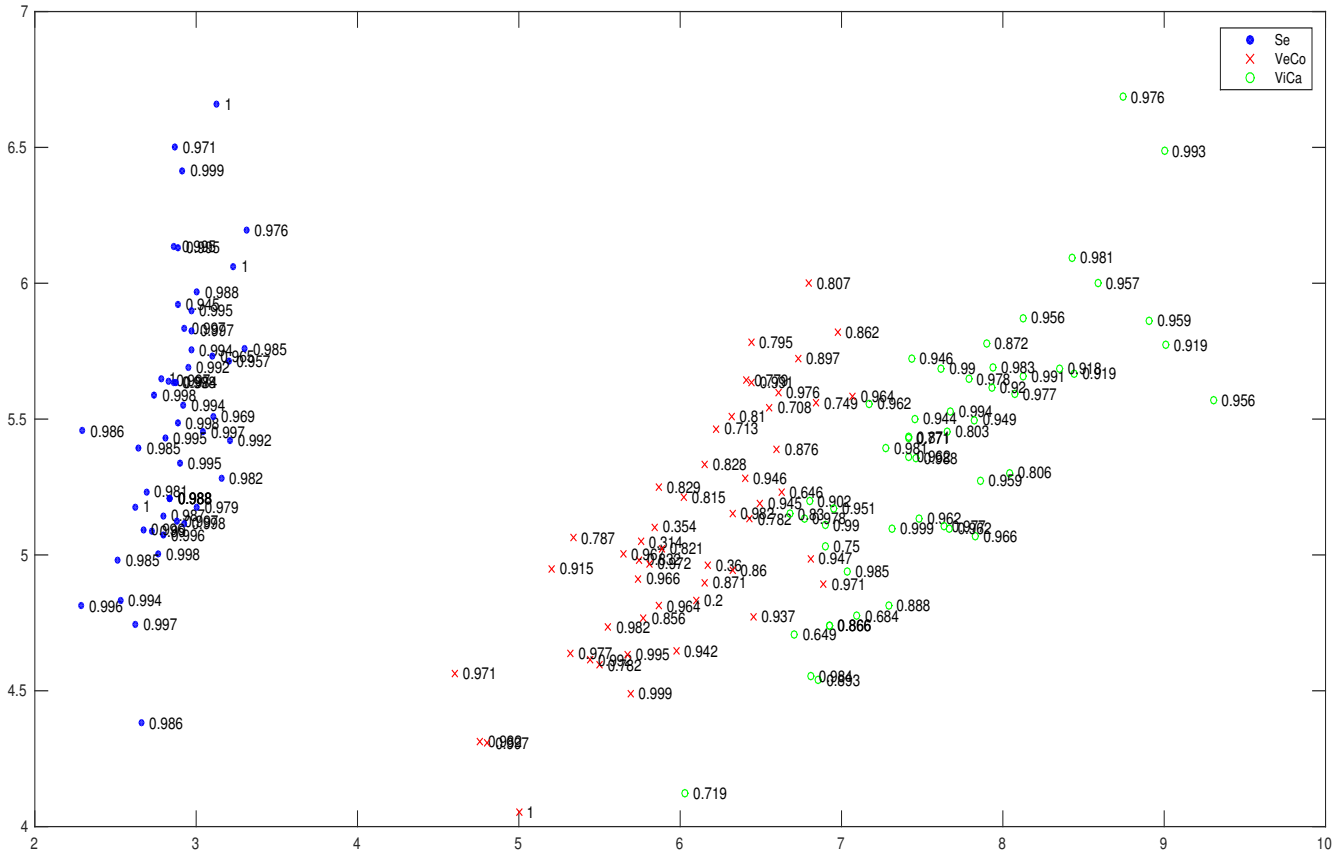
end if

end for

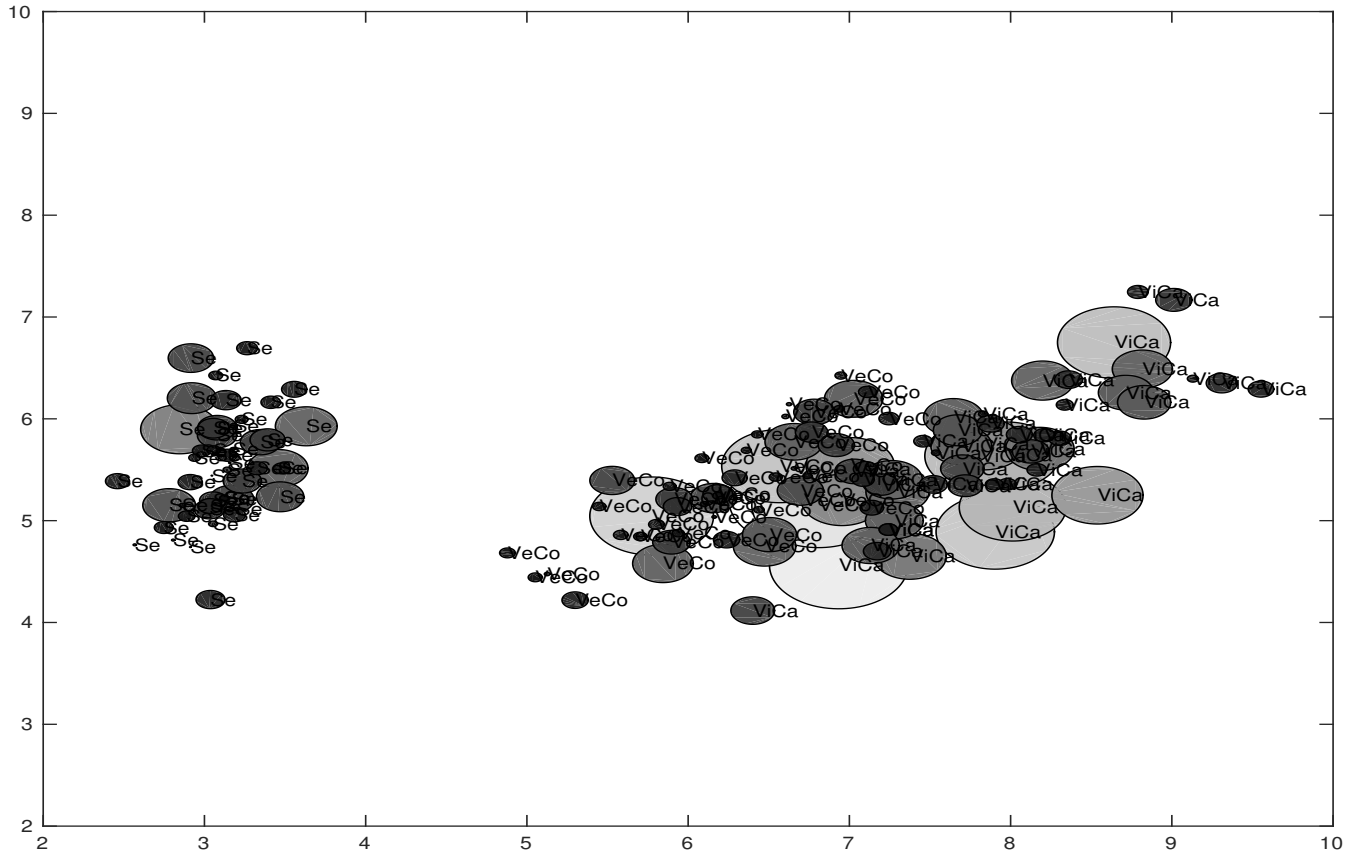
Application

- Different types of real data sets are used.
  - For example: Quantitative data(Iris data set).
- Parameters of Metropolis-Hastings:
  - Parameter $\lambda = 100$.
  - The standard deviation $\sigma = 0.01$.
  - Temperature $T = 100$

PCA



Projection under pairwise distance



Conclusions & Perspectives

Conclusions

- new non-linear projection method based on a new local measure of projection quality.
- The quality of projection is given here by additional variables called radii.
- Radii enable to give a bound on the original distances.
- The idea can be written as an optimization problem in order to minimize the sum of the radii under some constraints.
- Different algorithms and a lower bound for the objective function are developed.

Conclusions & Perspectives

Conclusions

- new non-linear projection method based on a new local measure of projection quality.
- The quality of projection is given here by additional variables called radii.
- Radii enable to give a bound on the original distances.
- The idea can be written as an optimization problem in order to minimize the sum of the radii under some constraints.
- Different algorithms and a lower bound for the objective function are developed.

Perspectives

- Improvement of the lower bound in order to assess how close the algorithms are from the minimum.

Thank you!