Université catholique de Louvain
Laboratoire de Microélectronique

A Robust Nonlinear Projection Method

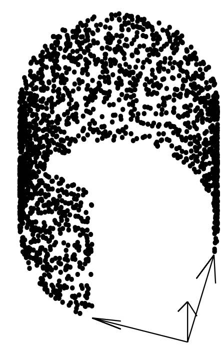
John Aldo Lee   Amaury Lendasse

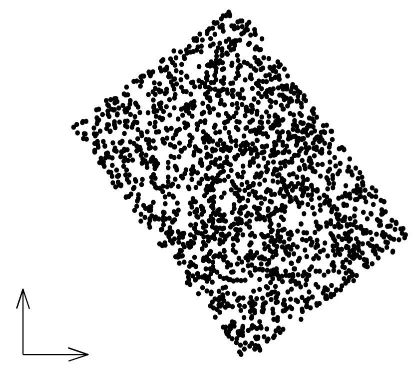
Nicolas Donckers   Michel Verleysen

Why & How to project databases?

Why?

- Data-mining;
- Classification;
- Time-series prediction;
- ...





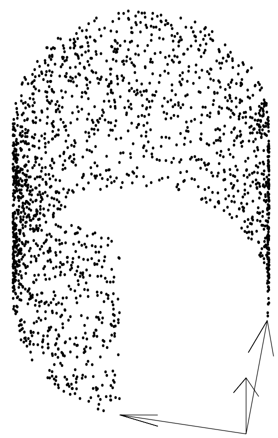
How?

Linear dependencies: PCA

Nonlinear dependencies: *not* PCA $\rightsquigarrow$???

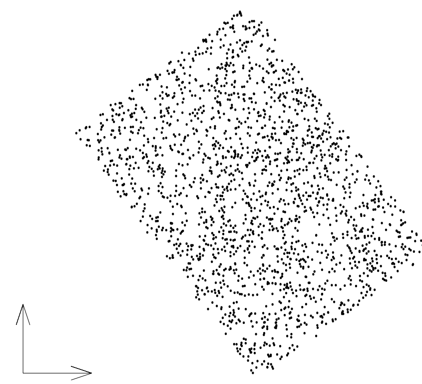
Problem Statement

Database | Mapping (unknown) | Projection



$\longleftarrow \mathcal{M} \longrightarrow$

$(n > p)$



$X^n = \{x_i^n \in \mathbb{R}^n\}$

$X^p = \{x_i^p \in \mathbb{R}^p\}$

Which criterion for $\mathcal{M}$?

Linear Solution: PCA

Criterion

Linear projection with maximal variance.

Advantages

- Well known method (strong theoretical background);
- Easily programmable & quickly computed;
- Only one parameter to set (loss of total variance).

Drawbacks

- Only linear projection (on a hyperplane).

1^{st} Nonlinear Solution: CCA

Criterion

$$\boxed{d^n_{i,j} = \|x^n_i - x^n_j\|} \longleftarrow \text{preservation} \longrightarrow \boxed{d^p_{i,j} = \|x^p_i - x^p_j\|}$$

Error function

$$\boxed{E_{CCA} = \frac{1}{2}\sum_{i \neq j}(d^n_{i,j} - d^p_{i,j})^2 F(d^p_{i,j})}$$

$\downarrow$

minimization

(specific gradient descent)

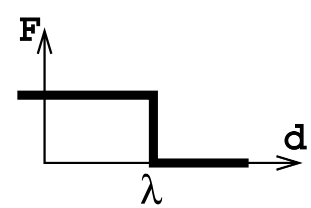
$\downarrow$

$$\boxed{X^p = \{x^p_i \in \mathbb{R}^p\}}$$

Parameters

$\alpha(t)$ = learning rate for gradient descent;

$\lambda(t)$ = neighborhood factor for weighting function F.



CCA: Advantages and Drawbacks

Advantages

- Nonlinear capabilities;
- More powerful than Sammon's mapping and nonlinear MDS;
- Easily programmable.

Drawbacks

- Computationally demanding (partially solved by VQ);
- No guarantee on result (local minimum, slow convergence);
- Difficult to parameterize (neighborhood factor $\lambda(t)$);
- Hard nonlinear database $\Rightarrow$ slow convergence.

2^{nd} Nonlinear Solution: CDA

Criterion

Curvilinear distance $\delta_{i,j}$ preservation.

Error function

$$\boxed{E_{CDA} = \frac{1}{2}\sum_{i\neq j}(\delta^n_{i,j} - d^p_{i,j})^2 F(d^p_{i,j})}$$

$$\downarrow$$

minimization

$$\downarrow$$

$$\boxed{X^p = \{x^p_i \in \mathbb{R}^p\}}$$

Parameters

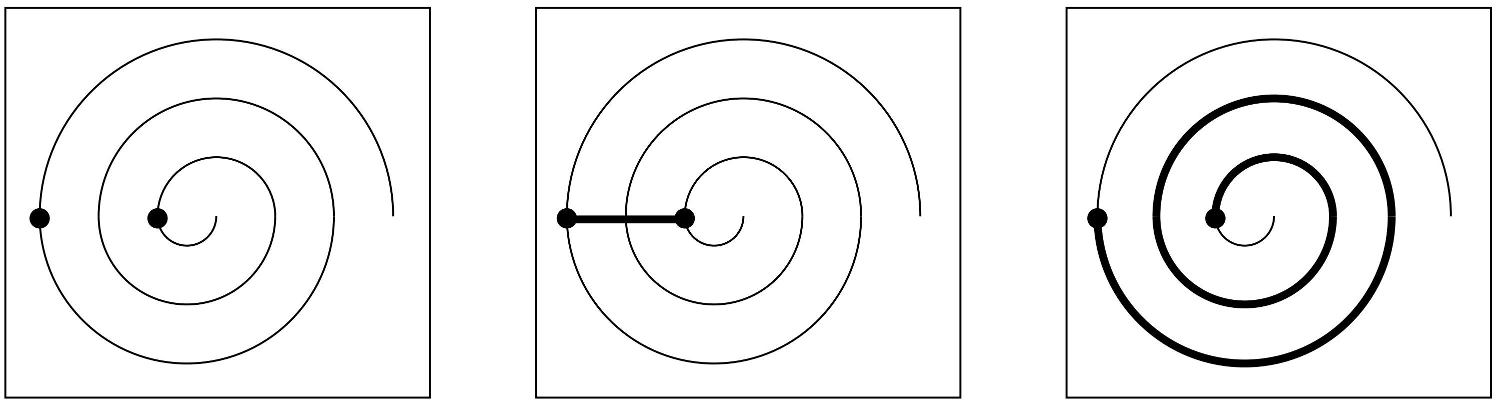
$\alpha(t)$ = learning rate for gradient descent;
$\lambda(t)$ = neighborhood factor for weighting function F;
$\omega(t)$ = Euclidian/curvilinear balance.

Curvilinear Distance

Theory

The distance between 2 vectors *along* an object.

Example



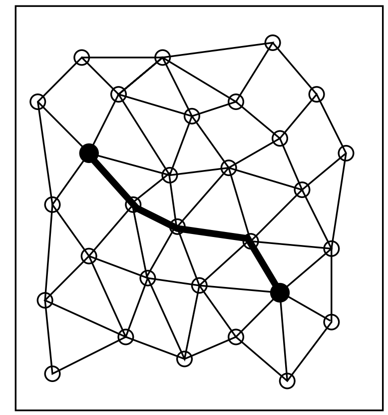
- a -   - b -   - c -

\- a - two points in a spiral,
\- b - Euclidian distance between them,
\- c - curvilinear distance.

Curvilinear Distance

In Practice

1. Vector Quantization on the database X^n;
2. Link the prototypes;
3. $\delta^n_{i,j} \approx$ shortest path between prototypes x^n_i and x^n_j.



Note

To keep full generality, replace $\delta^n_{i,j}$ by a generalized distance $D^n_{i,j}$:

$$D^n_{i,j} = (1-\omega(t))d^n_{i,j} + \omega(t)\delta^n_{i,j}$$

$$\downarrow$$

$$\boxed{E_{CDA} = \frac{1}{2}\sum_{i \neq j}(D^n_{i,j} - d^p_{i,j})^2 F(d^p_{i,j})}$$

CDA: Advantages and Drawbacks

Advantages

- Very good nonlinear capabilities;
- Easy to parameterize;
- Convergence less affected by hard nonlinear databases.

Drawbacks

- Computationally demanding (partially solved by VQ);
- No guarantee on result (local minimum);
- Complex implementation (VQ, links, shortest path, etc.).

CDA: Automatic Parameter Setting

Easy parameter setting for CDA

$\Downarrow$

Automation feasible

$\Downarrow$

Make CDA as easy-to-use as PCA.

How?

- Specific Vector Quantization methods;
- Analysis of VQ results (prototypes, links, LPCA, etc.);
- Analysis of distortions between $d_{i,j}^n$ and $\delta_{i,j}^n$.

CDA: Analysis of VQ results

Number of ...

- prototypes;
- links per prototypes;
- connected components;

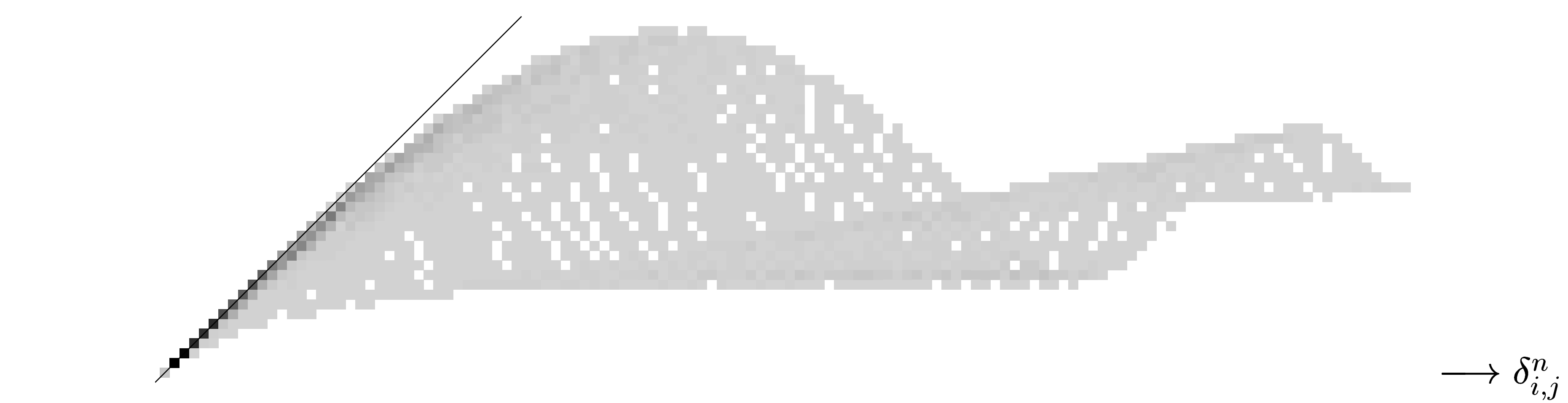
Local PCA

PCA on each Voronoi region

$$\Downarrow$$

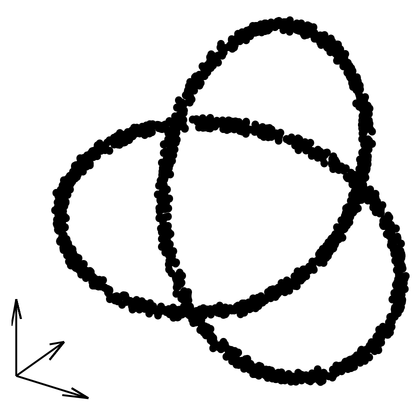
local dimension p of structure X^n

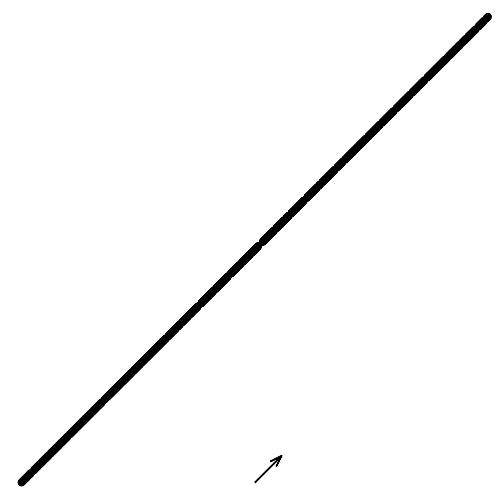
CDA: Distance Distortions



Distance distortions for a spiral.

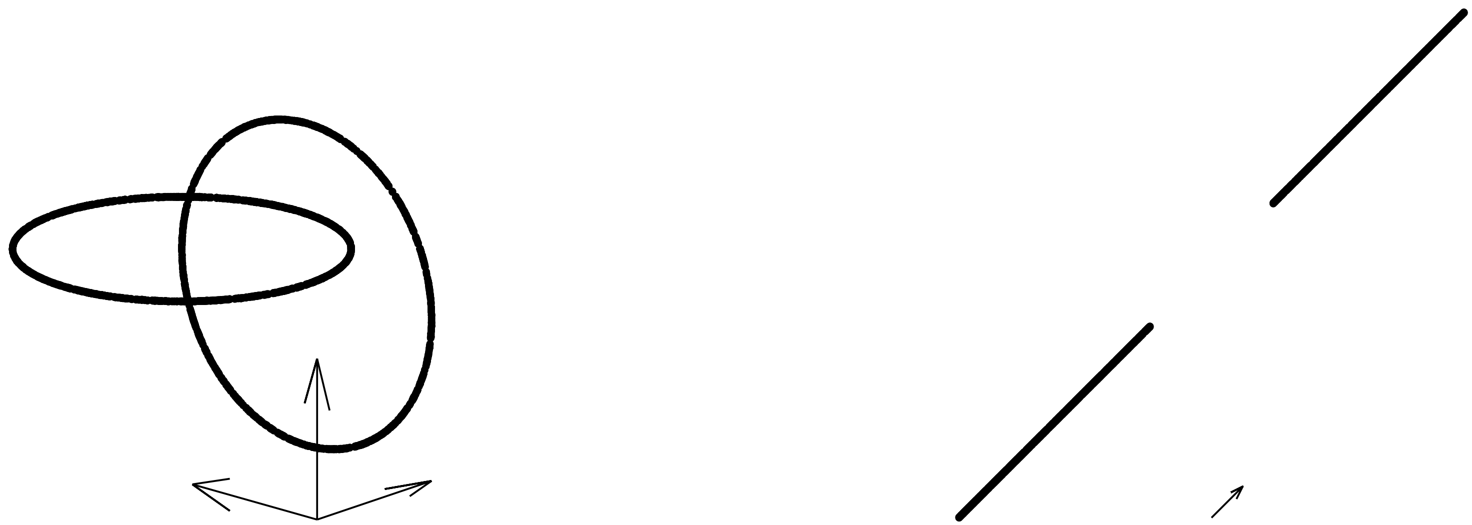
CDA: Example 1





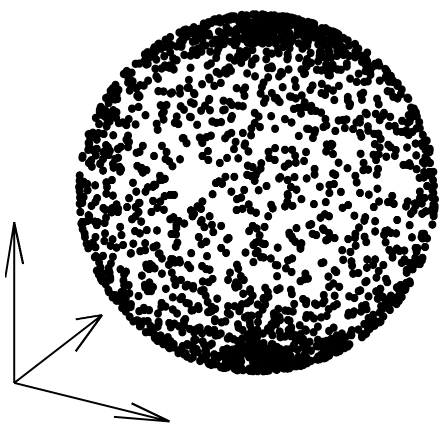
CDA projection of a noisy trefoil knot ($3D \rightarrow 1D$).

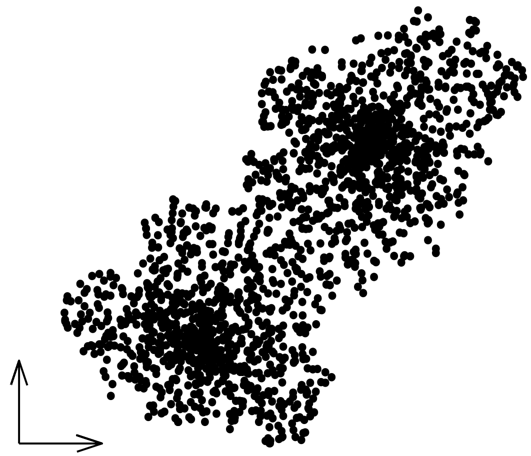
CDA: Example 2



CDA projection of the 'broken bicycle' ($3D \to 1D$).

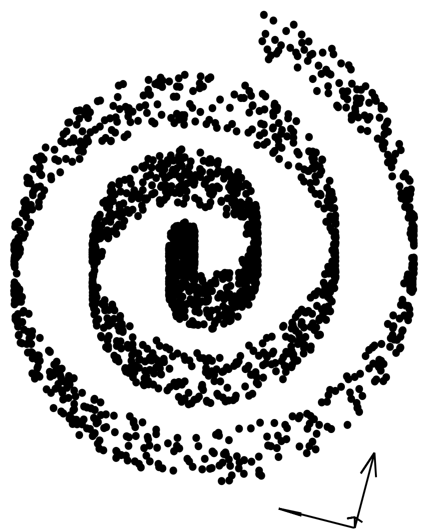
CDA: Example 3

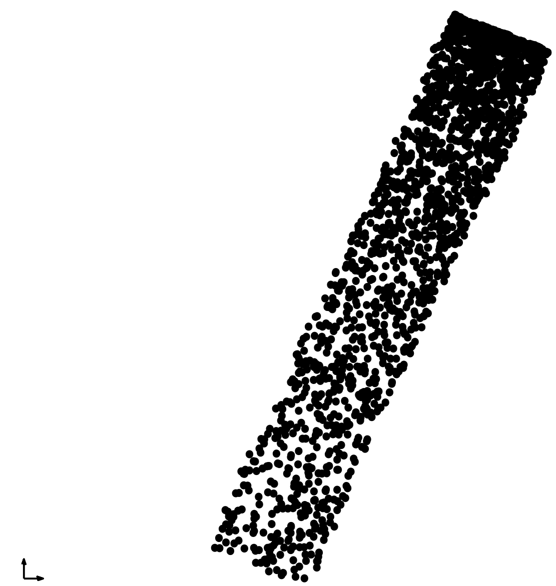




CDA projection of a sphere ($3D \to 2D$).

CDA: Example 4

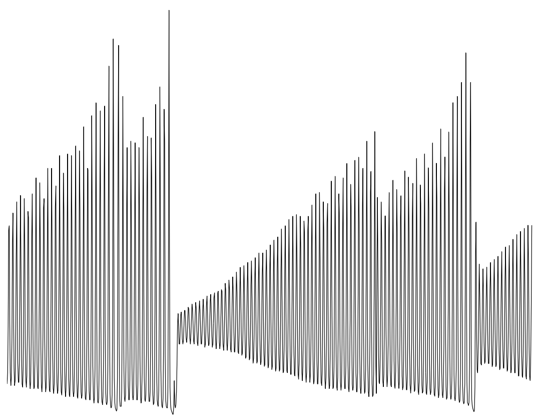




CDA projection of a ‘rolled red carpet’ ($3D \rightarrow 2D$).

CDA: Example 5

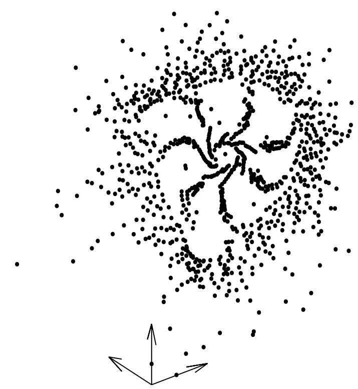
Time-series



Regressor length = 10

$\longleftarrow$ CDA $\longrightarrow$
bijective mapping
(rel.error $<$ 5%)

Regressor projection



CDA projection of Santa Fe A time-series ($10D \to 3D$).

Conclusion

Key idea

Curvilinear distances less distorted than Euclidian ones

$\Downarrow$

Simplified convergence of CCA/CDA error function

$\Downarrow$

Robustness

Consequences

- Easy parameter setting for CDA;
- Possible automation of the whole CDA process.