Detecting multivariate outliers using projection pursuit with particle swarm optimization

Anne Ruiz-Gazen   Alain Berro   Souad Larabi Marie-Sainte

University of Toulouse 1 - Capitole (TSE - IRIT - IMT)

COMPSTAT, Paris, September 2010

Introduction

What is Exploratory Projection Pursuit?

search for “interesting” linear low dimensional projections of high dimensional multivariate data

Introduction

What is Exploratory Projection Pursuit?

search for "interesting" linear low dimensional projections of high dimensional multivariate data

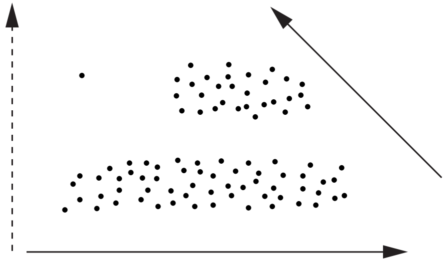
- Interesting structures:
  - outliers,
  - clusters, ...

Introduction

What is Exploratory Projection Pursuit?

search for “interesting” linear low dimensional projections of high dimensional multivariate data

- Interesting structures:
  - outliers,
  - clusters, ...
- Two ingredients:
  - projection interestingness: projection index I
  - optimization of the index: algorithm



Introduction

- EPP usually known by statisticians but not used!
- Well-known statistical softwares do NOT propose PP procedures (some routines in Fortran, Splus, Matlab and GGobi).

Introduction

- EPP usually known by statisticians but not used!
- Well-known statistical softwares do NOT propose PP procedures (some routines in Fortran, Splus, Matlab and GGobi).
- Recent applications in the domain of anomalies detection in hyperspectral imagery (Achard et al., 2004, Malpika et al., 2008, Smetek and Bauer, 2008).

Introduction

Mathematically

- denote X data matrix $n \times p$, X_i observation $p \times 1$, continuous variables,
- data are centered and scaled (divided by standard deviation or made spherical),
- consider one-dimensional projections from R^p to R : $z = X\alpha$,
  - where α is a p-dimensional projection vector $\alpha'\alpha = 1$,
  - z is a n-dimensional vector: coordinates of the projected observations,
- define a projection index function $I : \alpha \rightarrow I(\alpha)$,
- find projection vectors α : $\max_{\{\alpha \in R^p | \alpha'\alpha = 1\}} I(\alpha)$

Contents

Contents

Contents

Contents

Contents

Contents

First proposals

- Definition of an "interesting" projection discussed in the founding papers on PP (Friedman and Tukey, 1974, Huber, 1985, Jones and Sibson, 1987, and Friedman, 1987).
- Several arguments: "gaussianity is uninteresting".
- Any measure of departure from normality = a PP index.
- Objective more general than looking for projections that reveal outlying observations. However, several indices very sensitive to departure from normality in the tails of the distribution and reveal outliers in priority.

First proposals

- Friedman-Tukey (1974):

$$I_{FT}(\alpha) = \frac{1}{n^2 h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\alpha'(X_i - X_j)}{h}\right)$$

with $K(u) = \frac{35}{32}(1-u^2)^3 I_{\{|u|\leq 1\}}$ and $h = 3.12N^{-\frac{1}{6}}$ (Klinke, 1997).

First proposals

- Friedman-Tukey (1974):

$$I_{FT}(\alpha) = \frac{1}{n^2 h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\alpha'(X_i - X_j)}{h}\right)$$

with $K(u) = \frac{35}{32}(1-u^2)^3 I_{\{|u|\leq 1\}}$ and $h = 3.12N^{-\frac{1}{6}}$ (Klinke, 1997).

- Friedman (1987): index based on the L^2 distance between the projected data distribution and the Gaussian distribution (using expansions based on Legendre polynomials).

First proposals

- Friedman-Tukey (1974):

$$I_{FT}(\alpha) = \frac{1}{n^2 h^2} \sum_{i=1}^{n} \sum_{j=1}^{n} K\left(\frac{\alpha'(X_i - X_j)}{h}\right)$$

with $K(u) = \frac{35}{32}(1-u^2)^3 I_{\{|u|\leq 1\}}$ and $h = 3.12N^{-\frac{1}{6}}$ (Klinke, 1997).

- Friedman (1987): index based on the L^2 distance between the projected data distribution and the Gaussian distribution (using expansions based on Legendre polynomials).
- Kurtosis:

$$I_{kurt}(\alpha) = \sum_{i=1}^{n} (\alpha' X_i)^4$$

(Huber, 1985, Peña and Prieto, 2001)

Other indices

- Measure of outlyingness (Stahel-Donoho): for each observation $i = 1, \dots, n$,

$$I_i(\alpha) = \frac{|\alpha' X_i - \mathrm{med}_j(\alpha' Xj)|}{\mathrm{mad}_j(\alpha' X_j)}$$

where "med" $=$ median, "mad" $=$ median absolute deviation of the projected data from the median.

Other indices

- Measure of outlyingness (Stahel-Donoho): for each observation $i = 1, \ldots, n$,

$$I_i(\alpha) = \frac{|\alpha' X_i - \text{med}_j(\alpha' Xj)|}{\text{mad}_j(\alpha' X_j)}$$

  where "med" = median, "mad" = median absolute deviation of the projected data from the median.
- Dispersion-based indices: robust dispersion estimator (Li and Chen, 1985, Croux and Ruiz-Gazen, 2005), defines a robust principal component analysis.

Other indices

- Measure of outlyingness (Stahel-Donoho): for each observation $i = 1, \ldots, n$,

$$I_i(\alpha) = \frac{|\alpha' X_i - \text{med}_j(\alpha' Xj)|}{\text{mad}_j(\alpha' X_j)}$$

  where "med" = median, "mad" = median absolute deviation of the projected data from the median.
- Dispersion-based indices: robust dispersion estimator (Li and Chen, 1985, Croux and Ruiz-Gazen, 2005), defines a robust principal component analysis.
- Juan and Prieto (2001) for concentrated contamination patterns
- Hall and Kay (2005): non parametric atypicality index
- Indices adapted to time series, ...

Other indices

- Measure of outlyingness (Stahel-Donoho): for each observation $i = 1, \ldots, n$,

$$I_i(\alpha) = \frac{|\alpha' X_i - \text{med}_j(\alpha' Xj)|}{\text{mad}_j(\alpha' X_j)}$$

where "med" = median, "mad" = median absolute deviation of the projected data from the median.
- Dispersion-based indices: robust dispersion estimator (Li and Chen, 1985, Croux and Ruiz-Gazen, 2005), defines a robust principal component analysis.
- Juan and Prieto (2001) for concentrated contamination patterns
- Hall and Kay (2005): non parametric atypicality index
- Indices adapted to time series, ...

Many complementary definitions of indices but ...

the main problem with PP: pursuit computationally intensive.

Contents

Strategy for the first proposals

Usually complex structure needs several one-dimensional projections to be revealed $\Rightarrow$ several interesting optima of the projection index.

Strategy for the first proposals

Usually complex structure needs several one-dimensional projections to be revealed ⇒ several interesting optima of the projection index.

The usual strategy

- Global optimization algorithm ⇒ one optimum projection.
- Remove the structure found from the data set.
- Iterate the procedure.

Example: For the kurtosis index, Peña and Prieto (2001) uses a modified version of Newton's optimization method and remove the structure by projecting the data on the space orthogonal to the projection found. The procedure is iterated p times (the number of dimensions).

Strategy for the first proposals

Limitations:

- Global optimization based on repeated local optimization usually quite costly.
- Local optimization algorithms rely on regularity conditions on the projection index.
- Structure removal may miss some interesting projections (Huber, 1985, Friedman, 1987).

Other strategies

Finite number of projection vectors

replace the maximization problem: $\max_{\{\alpha \in R^p | \alpha'\alpha = 1\}} I(\alpha)$ by:

$$\max_{\{\alpha \in \mathcal{A} | \alpha'\alpha = 1\}} I(\alpha)$$

where $\mathcal{A}$ contains a finite number of directions and

calculate $I(\alpha)$ for all $\alpha \in \mathcal{A}$.

Limitations: This strategy may miss interesting projections by not exploring enough the space of solutions.

Contents

A new strategy

Find directly several local optima

- Run several times local optimization algorithms.
- Use heuristics algorithms for local optimization (no need of regularity condition and better exploration of the space of solutions)
- No need for global optimization and structure removal.

A new strategy

Find directly several local optima

- Run several times local optimization algorithms.
- Use heuristics algorithms for local optimization (no need of regularity condition and better exploration of the space of solutions)
- No need for global optimization and structure removal.

Difficulty: not easy to know the extent to which a new view reflects a similar or a different structure compared with the previous views (Friedman in Jones and Sibson, 1987, discussion) $\Rightarrow$ exploratory tools to analyze the different projections.

Heuristics optimization algorithms

Two families of heuristic optimization methods (Gilli and Winker, 2008):

- the trajectory methods (e.g. simulated annealing or Tabu search) which consider one single solution at a time,
- the population based methods (e.g. genetic algorithms or Particle Swarm Optimization) which update a whole set of solutions simultaneously. Focus on this second family of methods (exploration of the whole search space sometimes more efficient).

Heuristics optimization algorithms

Two families of heuristic optimization methods (Gilli and Winker, 2008):

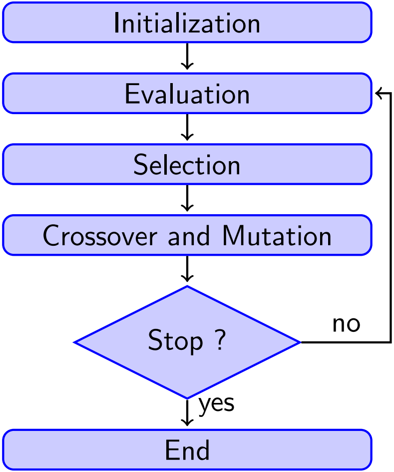
- the trajectory methods (e.g. simulated annealing or Tabu search) which consider one single solution at a time,
- the population based methods (e.g. genetic algorithms or Particle Swarm Optimization) which update a whole set of solutions simultaneously. Focus on this second family of methods (exploration of the whole search space sometimes more efficient).

Main characteristics:

- Heuristics optimization methods can tackle optimization problems that are not tractable with classical optimization tools.
- They usually mimic some behavior found in nature.

Implementation of GA, PSO and Tribes (adaptive PSO algorithm).

GA

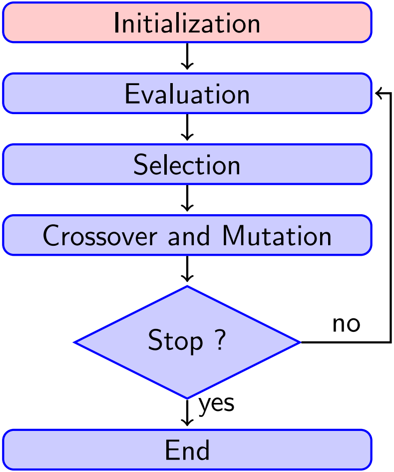


Model

- An individual represents a projection vector
- The fitness function is the projection index

- Random initialization
- Tournament selection with 3 participants
- probability of mutation $= 0.05$
- probability of crossover $= 0.65$
- Number of iterations

GA

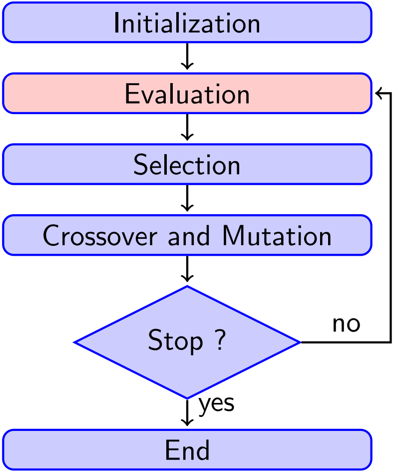


Model

- An individual represents a projection vector
- The fitness function is the projection index

- Random initialization
- Tournament selection with 3 participants
- probability of mutation $= 0.05$
- probability of crossover $= 0.65$
- Number of iterations

GA

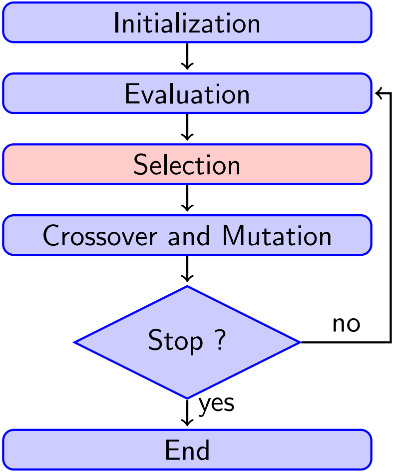


Model

- An individual represents a projection vector
- The fitness function is the projection index

- Random initialization
- Tournament selection with 3 participants
- probability of mutation $= 0.05$
- probability of crossover $= 0.65$
- Number of iterations

GA

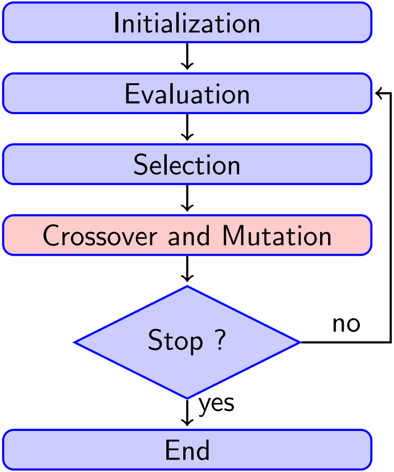


Model

- An individual represents a projection vector
- The fitness function is the projection index

- Random initialization
- Tournament selection with 3 participants
- probability of mutation = 0.05
- probability of crossover = 0.65
- Number of iterations

GA

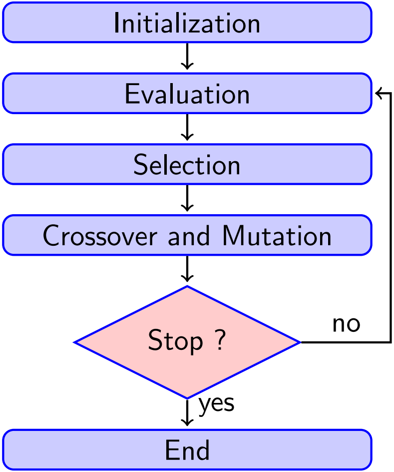


Model

- An individual represents a projection vector
- The fitness function is the projection index

- Random initialization
- Tournament selection with 3 participants
- probability of mutation $= 0.05$
- probability of crossover $= 0.65$
- Number of iterations

GA



Model

- An individual represents a projection vector
- The fitness function is the projection index

- Random initialization
- Tournament selection with 3 participants
- probability of mutation $= 0.05$
- probability of crossover $= 0.65$
- Number of iterations

PSO

Particle Swarm Optimization: Kennedy and Eberhart (1995)

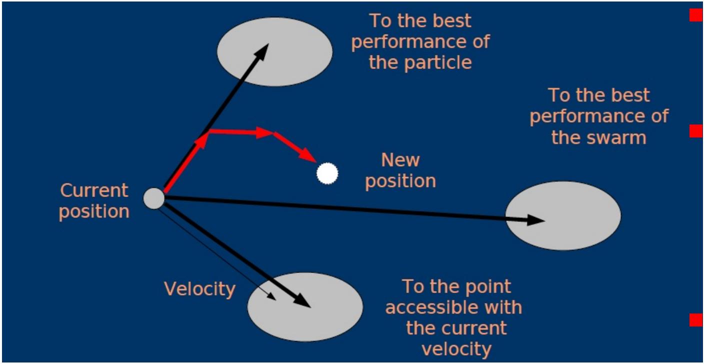
- Stochastic method
- Biological inspiration (fish schooling and bird flocking)
  - Each bird seems to move randomly
  - The communication between birds is limited

PSO

Particle Swarm Optimization: Kennedy and Eberhart (1995)

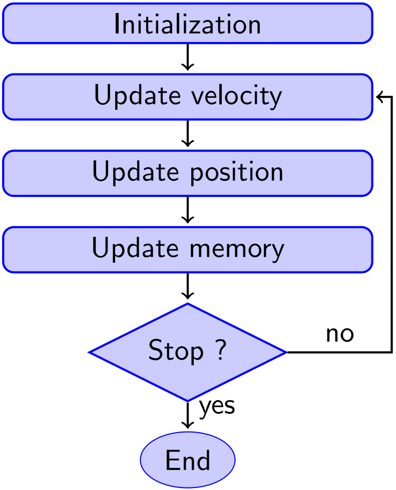
- Stochastic method
- Biological inspiration (fish schooling and bird flocking)
  - Each bird seems to move randomly
  - The communication between birds is limited
  - However, a swarm is able to find food

PSO: strategy of displacement of a particle



- Generation of a swarm of particles
- A fitness is associated to each particle
- Particles move according to their own experience and that of the swarm
- Convergence made possible by the cooperation between particles

PSO Algorithm



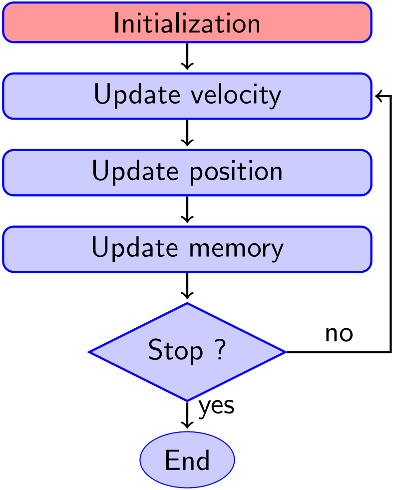
Model

- A particle represents a projection vector
- The fitness function is the projection index

Parameters of the particle i

- $\overrightarrow{x_i}$: position
- $\overrightarrow{v_i}$: velocity
- $\overrightarrow{pbest_i}$ best solution
- $\overrightarrow{gbest}$ best solution of the swarm

PSO Algorithm

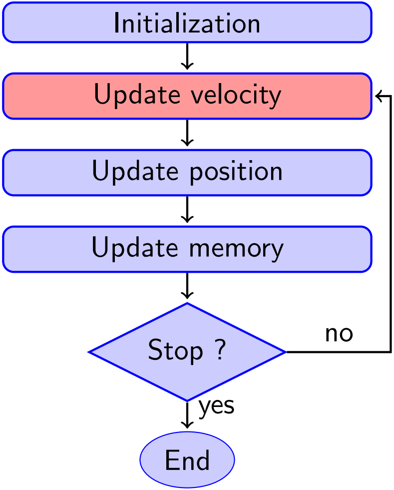


- Random initialization
- Velocity equation :

$$\begin{aligned}\overrightarrow{v_i} \quad \leftarrow \quad & \omega \cdot \overrightarrow{v_i} \\ & + c_1 \cdot \overrightarrow{r_1} \otimes (\overrightarrow{pbest_i} - \overrightarrow{x_i}) \\ & + c_2 \cdot \overrightarrow{r_2} \otimes (\overrightarrow{gbest} - \overrightarrow{x_i})\end{aligned}$$

where : ω, c_1, c_2 : parameters and $\overrightarrow{r_1}$, $\overrightarrow{r_2}$: random values

PSO Algorithm

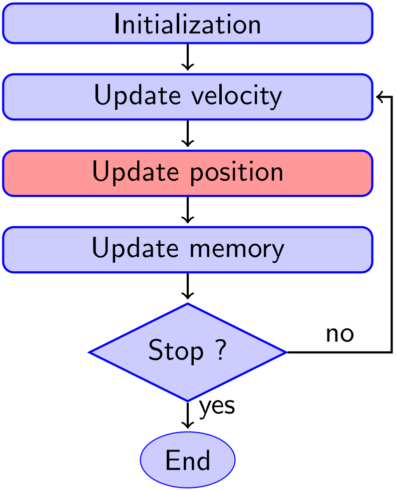


- Random initialization
- Velocity equation :

$$\begin{aligned}\overrightarrow{v_i} \quad \leftarrow \quad & \omega \cdot \overrightarrow{v_i} \\ & + c_1 \cdot \overrightarrow{r_1} \otimes (\overrightarrow{pbest_i} - \overrightarrow{x_i}) \\ & + c_2 \cdot \overrightarrow{r_2} \otimes (\overrightarrow{gbest} - \overrightarrow{x_i})\end{aligned}$$

where : ω, c_1, c_2 : parameters and $\overrightarrow{r_1}$, $\overrightarrow{r_2}$: random values

PSO Algorithm

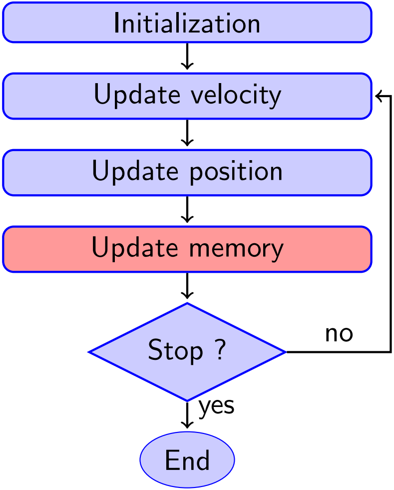


- position equation :

$$\overrightarrow{x_i} \leftarrow \overrightarrow{x_i} + \overrightarrow{v_i}$$

- I is the projection index
- update $pbest_i$ and $gbest$
- Number of iterations

PSO Algorithm

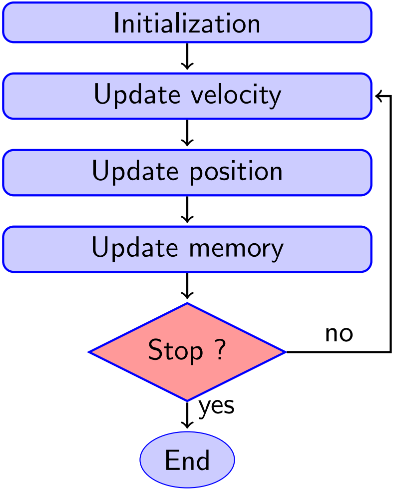


- position equation :

$$\overrightarrow{x_i} \leftarrow \overrightarrow{x_i} + \overrightarrow{v_i}$$

- I is the projection index
- update $pbest_i$ and $gbest$
- Number of iterations

PSO Algorithm



- position equation :

$$\overrightarrow{x_i} \leftarrow \overrightarrow{x_i} + \overrightarrow{v_i}$$

- I is the projection index
- update $pbest_i$ and $gbest$
- Number of iterations

Tribes method

GA and PSO compared in Berro et al. (2010). Results quite similar but several parameters to tune.

Tribes method

GA and PSO compared in Berro et al. (2010). Results quite similar but several parameters to tune.

- Tribes (Clerc, 2006) is the first parameter-free particle swarm optimization algorithm. It is an adaptive algorithm.
- Principle:
  - Swarm divided in "tribes"
  - At the beginning, the swarm is composed of only one particle
  - According to tribes' behaviors, particles are added or removed in the tribes
  - According to the performances of the particles, their strategies of displacement are adapted

The Tribes method

Tribes has been compared with usual PSO in Larabi et al. (2010). Tribes is more interesting in the EPP context for two reasons:

- No parameter to settle except the stopping criterion,
- It converges very quickly to local optima.

Contents

The interface EPP-Lab

Implemented in Java (A. Berro, S. Larabi, E. Chabbert, I. Griffith).

Contents

- Indices for outliers detection: Friedman-Tukey, Friedman and kurtosis.
- Optimization algorithms: GA, PSO and Tribes.

The interface EPP-Lab

Implemented in Java (A. Berro, S. Larabi, E. Chabbert, I. Griffith).

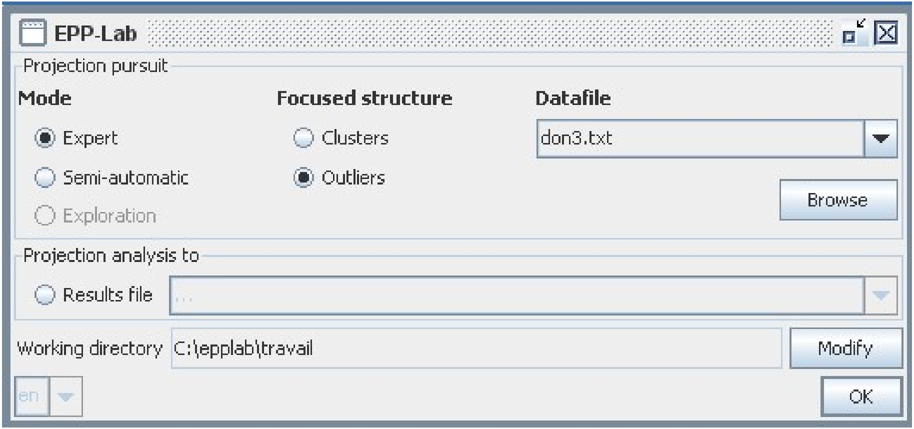
Contents

- Indices for outliers detection: Friedman-Tukey, Friedman and kurtosis.
- Optimization algorithms: GA, PSO and Tribes.

EPP-Lab in action

- First step in the analysis: the pursuit (may be time consuming but no need for the statistician to be in front of the computer $\neq$ GGOBI),
- Second step: from the first step (projections saved), analysis of the results (imediate)

The interface EPP-Lab in action



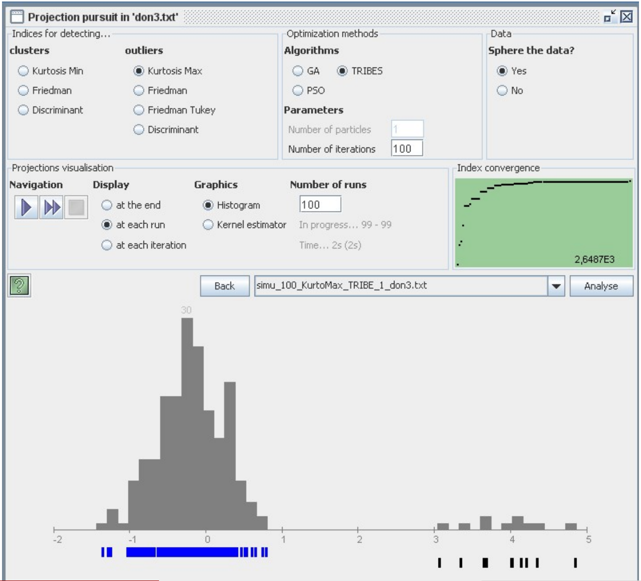
The interface EPP-Lab in action

The file don3.txt is 200×8 with:

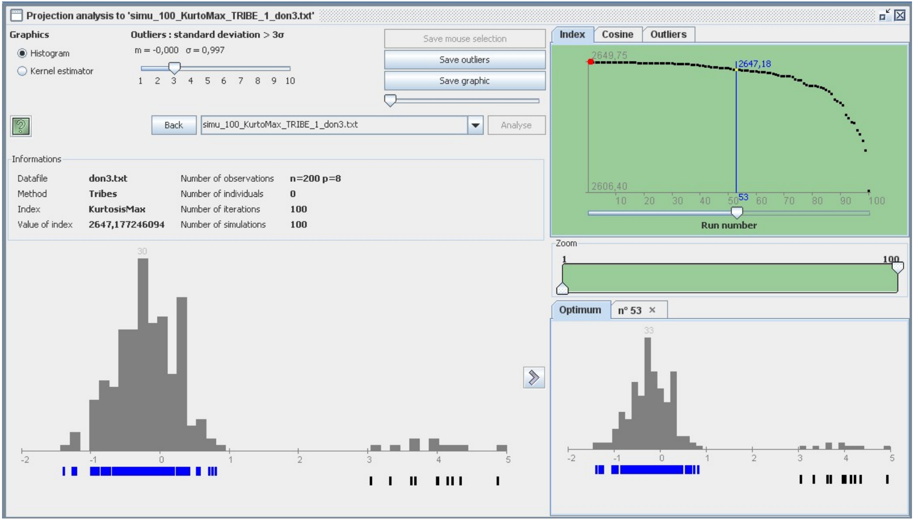
- the first 190 observations follow a $\mathcal{N}(0, I_8)$ distribution
- and the last 10 follow a $\mathcal{N}((10, 0, \ldots, 0)', I_8)$ distribution.

Observations 191 to 200 are outlying.

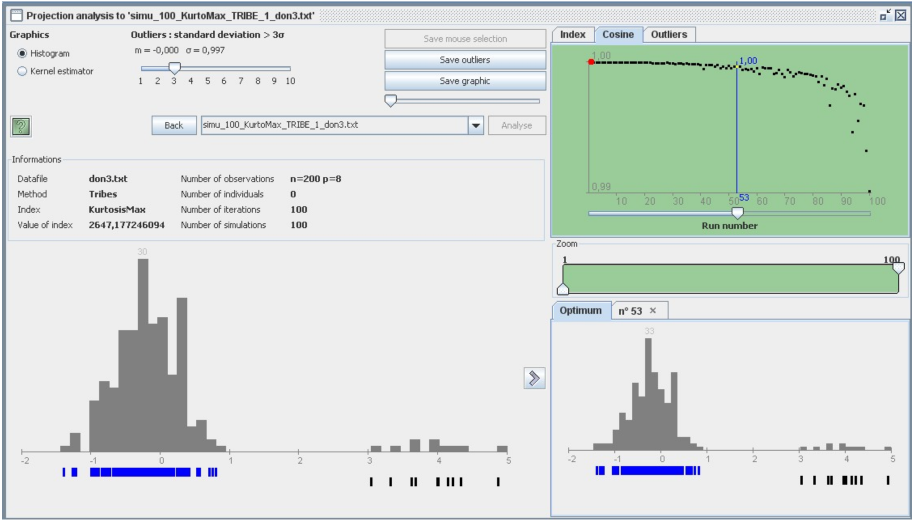
The interface EPP-Lab in action



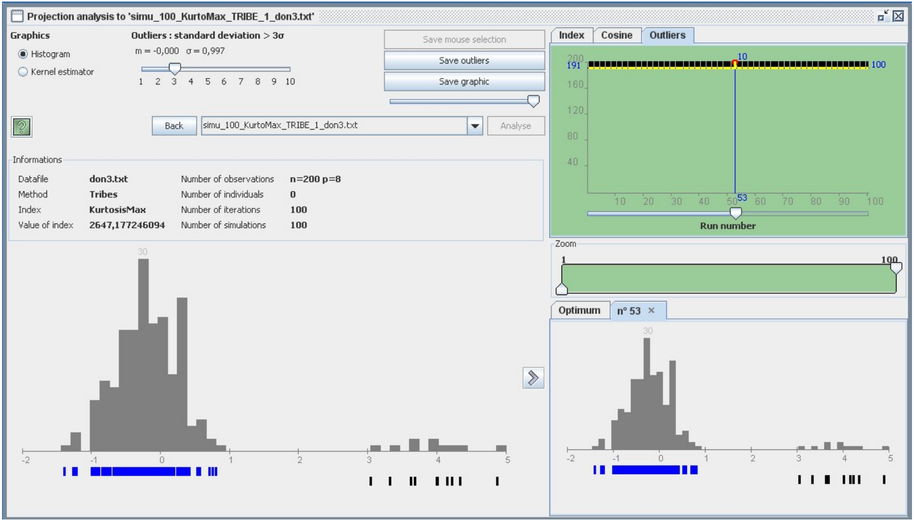
The interface EPP-Lab in action



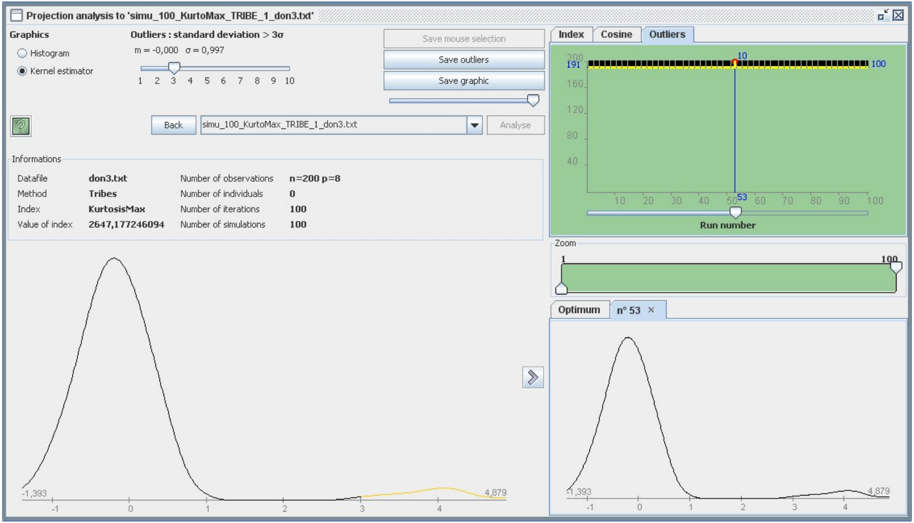
The interface EPP-Lab in action



The interface EPP-Lab in action



The interface EPP-Lab in action



Contents

Conclusion and perspectives

- **Conclusion**
  - Heuristics algorithms interesting in the context of EPP with a new strategy based on local optimization.
  - Possibility of using free-parameter algorithms such as Tribes.

Conclusion and perspectives

- **Conclusion**
  - Heuristics algorithms interesting in the context of EPP with a new strategy based on local optimization.
  - Possibility of using free-parameter algorithms such as Tribes.
- **Perspectives**
  - Interface EPP-Lab to improve:
    - study and implement several stopping criteria,
    - implement Stahel-Donoho index,
    - implement robust selection,
    - Develop and implement statistical tools to summarize the different projections (clustering of variables, principal components analysis, sum of projectors, . . .)

Conclusion and perspectives

- **Conclusion**
  - Heuristics algorithms interesting in the context of EPP with a new strategy based on local optimization.
  - Possibility of using free-parameter algorithms such as Tribes.
- **Perspectives**
  - Interface EPP-Lab to improve:
    - study and implement several stopping criteria,
    - implement Stahel-Donoho index,
    - implement robust selection,
    - Develop and implement statistical tools to summarize the different projections (clustering of variables, principal components analysis, sum of projectors, . . .)
  - Introduce multiobjective optimization in order to deal with spatial data sets.
  - . . .

Biblography I

Berro, A., Larabi Marie-Sainte, S. and Ruiz-Gazen, A.
Genetic Algorithms and Particle Swarm Optimization for Exploratory Projection Pursuit.
Annals of Mathematics and Artificial Intelligence, to appear, 2010

Larabi Marie-Sainte S., Berro, A. and Ruiz-Gazen, A.
An Efficient Optimization Method for Revealing Local Optima of Projection Pursuit Indices.
Ants 2010, Seventh International Conference on Swarm Intelligence, 2010

Caussinus, H. and Ruiz-Gazen, A.
Exploratory Projection Pursuit.
In Data Analysis (Digital Signal and Image Processing series), Ed. G. Govaert, ISTE, 2009

Thank you for your attention!