Using a Multi-objective Genetic Algorithm for Curve Approximation

by Tim Sabsch, Christian Braune, Alexander Dockhorn, and Rudolf Kruse

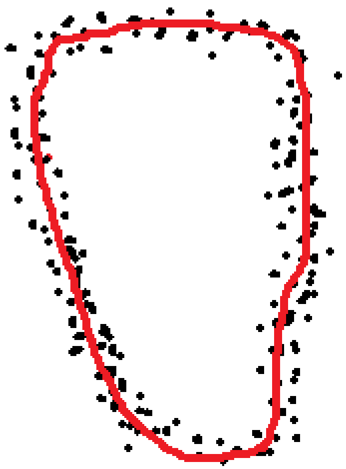
Institute for Intelligent Cooperating Systems
Department for Computer Science, Otto von Guericke University Magdeburg
Universitaetsplatz 2, 39106 Magdeburg, Germany

Email: {tim.sabsch, christian.braune, alexander.dockhorn, rudolf.kruse}@ovgu.de

Contents

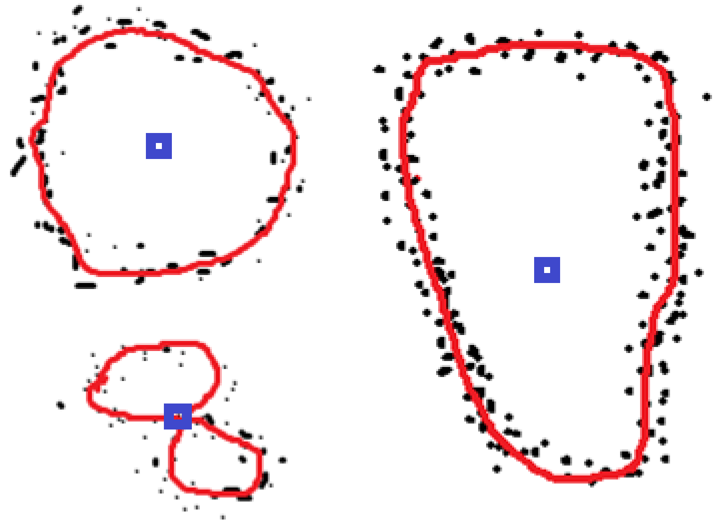
Curve Fitting

- Approximate an (un-)ordered set of points / voxels with a curve
- Many algorithms require vector data such as lines, curves or surfaces
- B-splines are of special interest for applications in computer-aided geometric design
  - Their piecewise polynomial structure makes them flexible and quick to compute
  - Their structure can be easily extended to surface and hyperplanes



What do we need it for?

- This work was done to find cluster descriptors for density-based clusters.
- Those can be efficiently mined, but most prototype methods provide unsuitable approximations.
- E.g. consider the mean vector as it is commonly used for clusters found by k-means



Mean vector

True approximation of the cluster form

Previous strategies – Related Work I

- Regression methods
  - Requires the order of points to be known
- Genetic algorithms have been successfully applied for the determination of optimal number and location of knots
  - Requires the order of points to be known
- Iterated Thinning / Skeleton construction / Graph construction
  - Restricted to low dimensional data
- … or cannot fit data of certain geometric properties

Previous strategies – Related Work II

- Local Tangential Flow
  - Deals with self intersections, sharp corners and high dimensions
  - But fails in point clouds of varying density

- PCA used for ordering the points on a polynomial chain
  - Fast and deals will high dimensions and self-intersections
  - Struggles with point clouds of varying density

- Nevertheless, density based clusters need efficient descriptors as well

B-Spline Curve Fitting

- Task: find a B-spline curve that approximates the shape best

$$C(t) = \sum_{i=0}^{n} N_{i,k}(t) P_i$$

$P = (P_0, P_1, \dots, P_n)$ is the set of control points

$N_{i,k}(t) = (N_{0,k}(t), N_{1,k}(t), \dots, N_{n,k}(t))$ is the set of B-spline basis functions defined over a degree *k* - 1 and a non-decreasing sequence of real numbers

$T = (u_0, u_1, \dots, u_{n+k})$ is the knot vector.

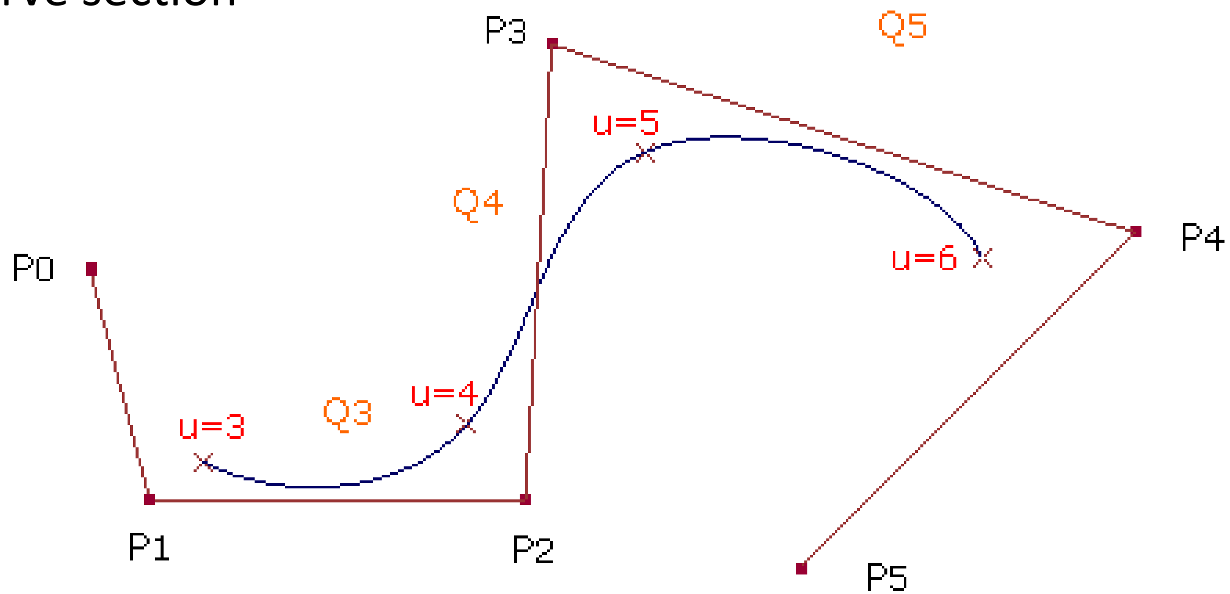
B-Spline Curve Fitting

- A curve is said to be clamped uniform, if the first and last knot in the knot vector each has multiplicity k, and the remaining knots are evenly spaced.
  - Such a curve starts at the first control point and ends at the last control point.

- Due to the definition of the basis functions, a control point P_i only influences the curve in the interval $[u_i; u_{i+k})$ (local control property).

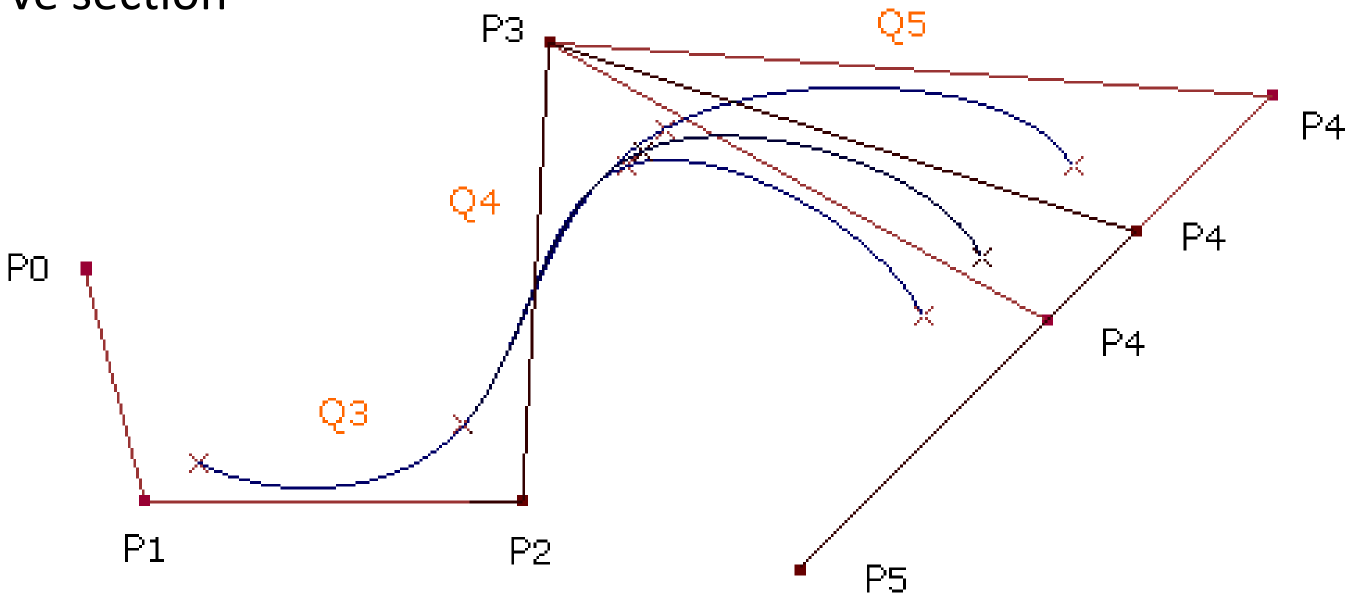
B-splines by example I

- P = control point
- u = knot vector
- Q = curve section



B-splines by example II

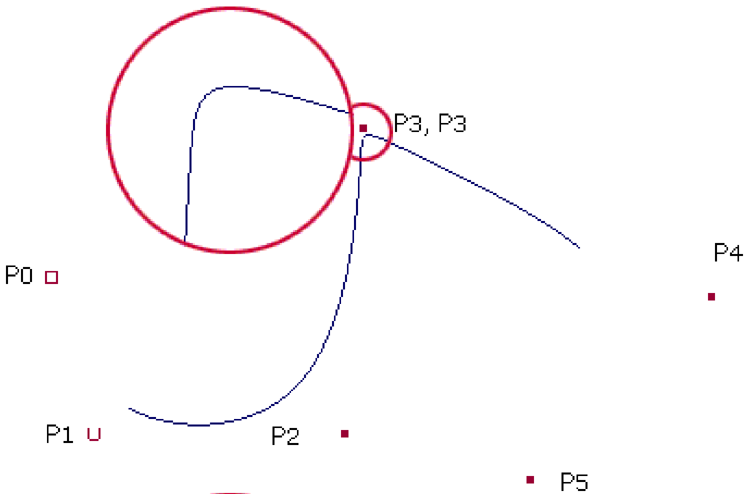
- P = control point
- u = knot vector
- Q = curve section

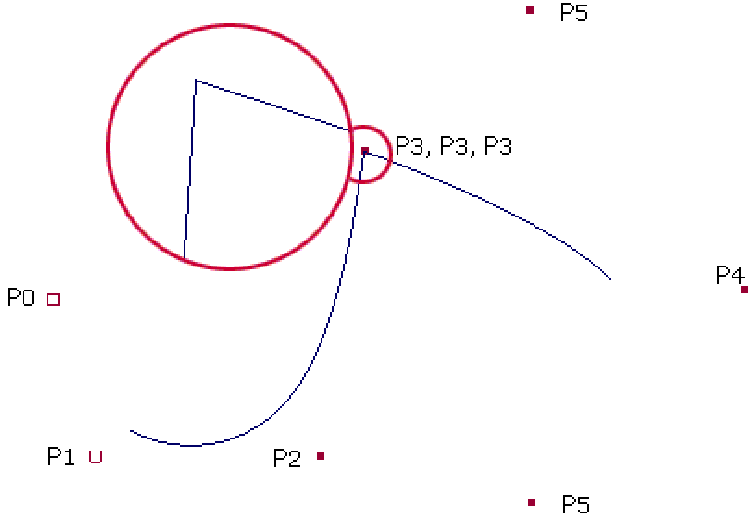


B-splines by example III

- P = control point
- u = knot vector
- Q = curve section

- Playing multiple control points at the same position, sharpens the curve at this section





Multi-objective Optimization with NSGA-II

- During the multi-objective optimization we want to find
  - the best fitting set of control points per b-spline
  - The best set of b-splines to approximate the dataset
- For smoothness: the curves developed in the algorithm are fixed to be cubic, i.e. of degree 3
- For control: changing a control point will only influence the curve in an interval of four knots
- The knot vector

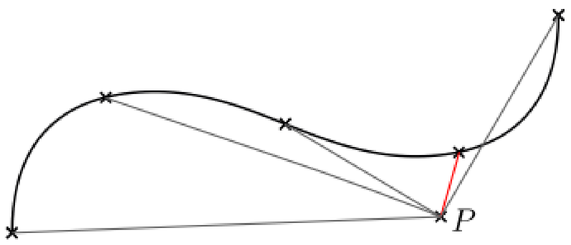
Parameter	Value
Number of generations	1000
Population size	100
Crossover probability	0.10
Probability of mutating control point number	0.01
Probability of mutating control point location	0.25

Objectives

- Objectives used in this study:
  - Number of control points
  - Distance minimization
  - (Length of the curve)

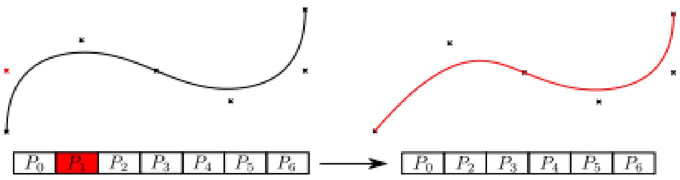
- Other possible objectives:
  - Near-parallel segments
  - Segments outside the points cloud

- Distance calculation:
  - Equidistant sampling of a number of points on the curve
  - Shortest distance to the target point is considered to be the distance of the curve and the point

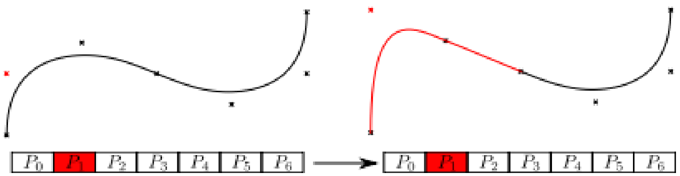


Mutation and Crossover for B-splines

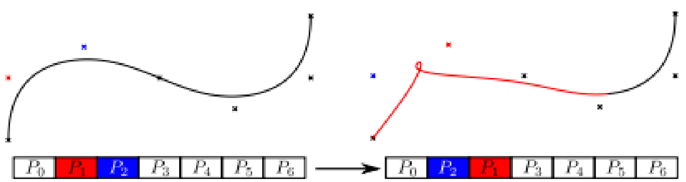
- Mutations:
  - Removing/adding a control point



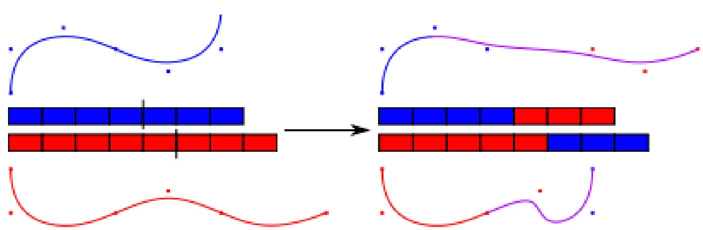
  - Changing a control point via random gaussian position change



  - Swapping the position of two control points

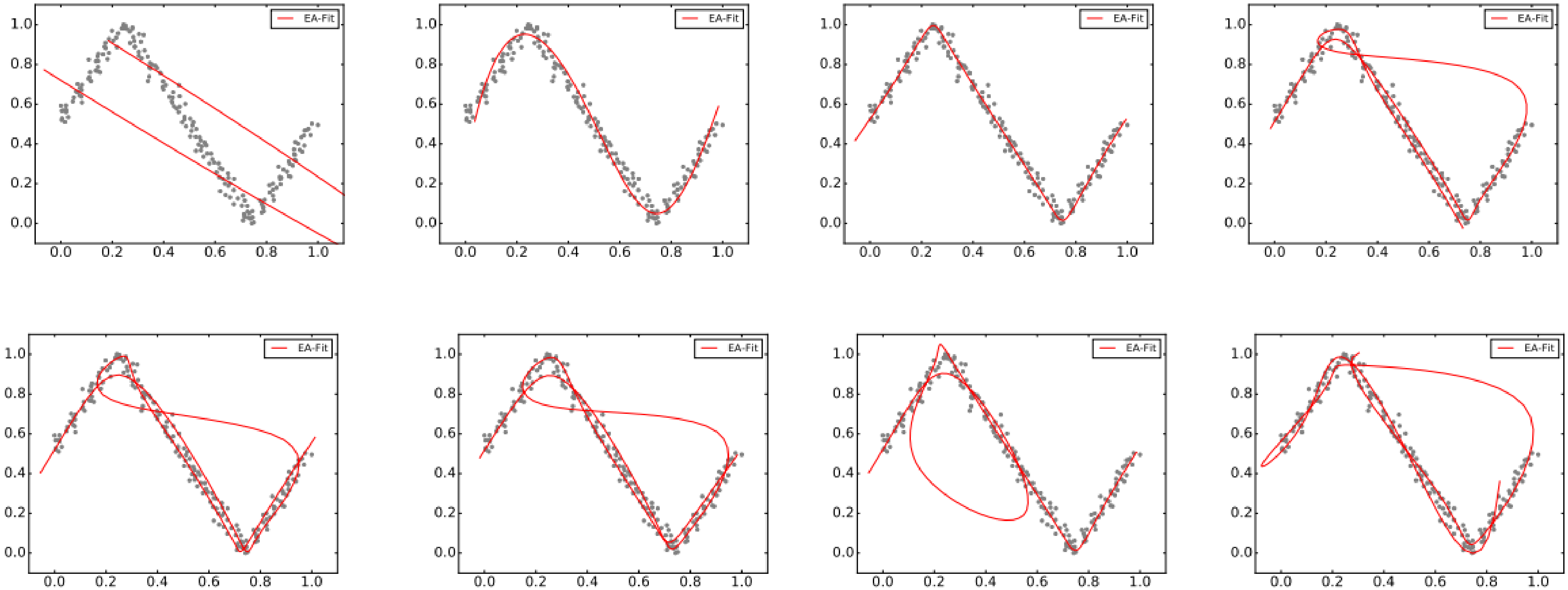


- Crossover:
  - One-point-Crossover



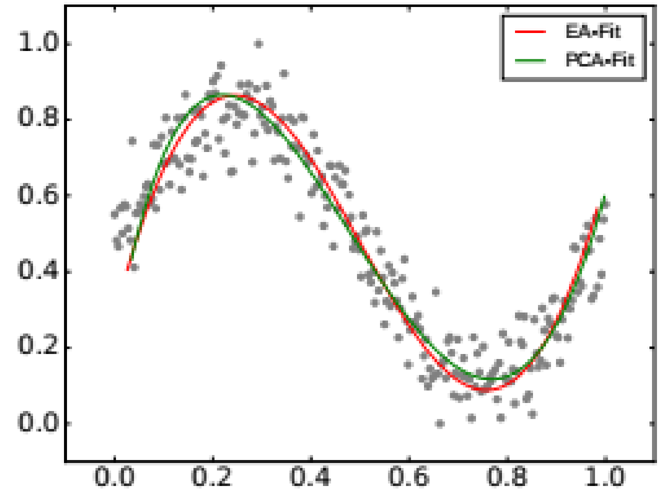
Results of one Pareto Front

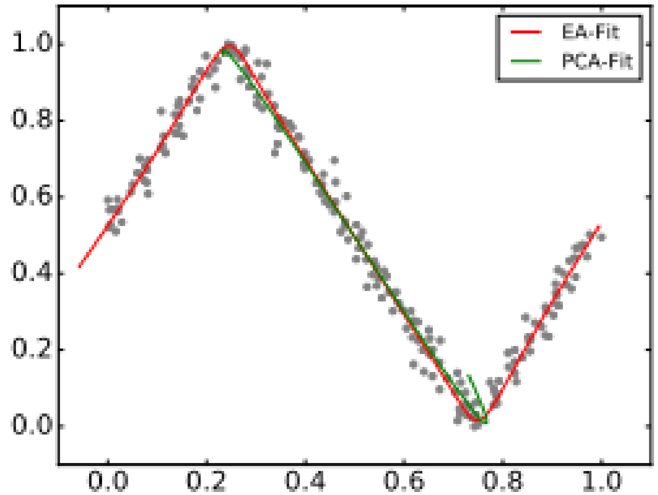
- Models are differing in complexity and error-rate

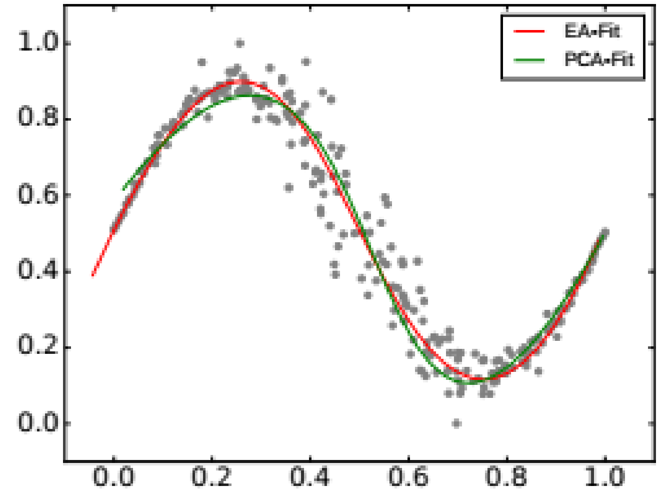


Comparing the Results - I

- Stable results for non-intersecting point-clouds
- Even with varying density and sharp edges, no problems arise





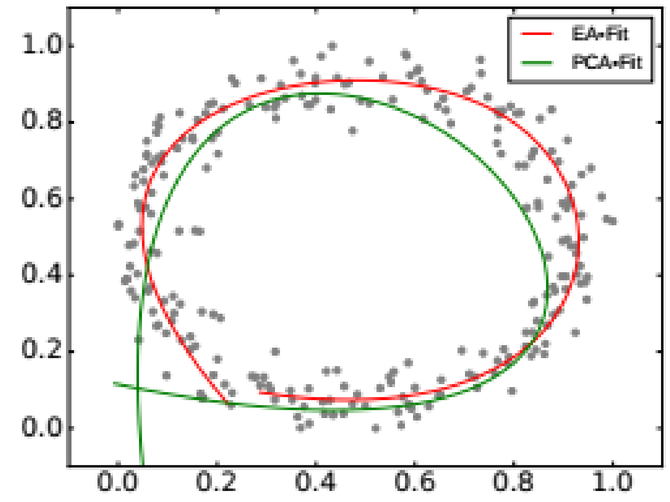


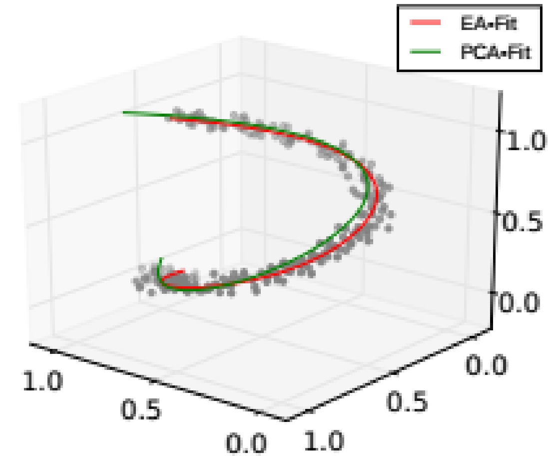
Comparing the Results - II

- Closed forms can be approximated very well

- Few dimensions can be handled without problems

- Multiple dimensions can affect the objectives due to the curse of dimensionality

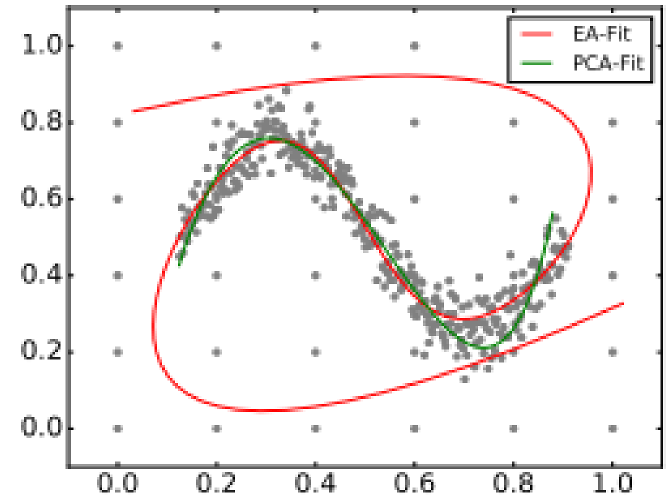


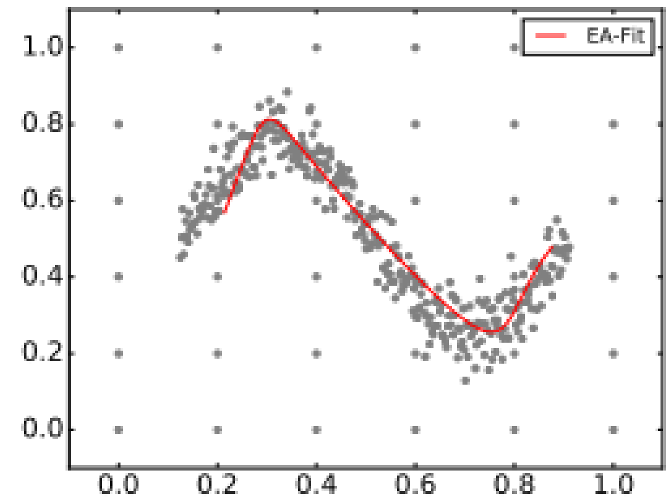


OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG INF FAKULTÄT FÜR INFORMATIK

Comparing the Results - III

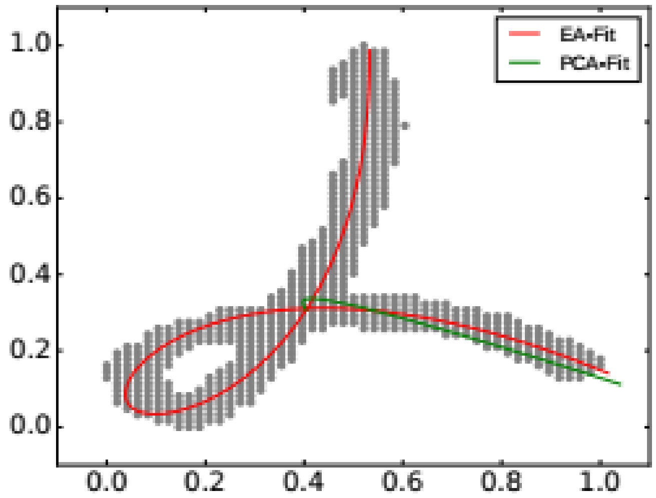
- Background noise can increase the complexity of the resulting model
- Dense areas will be fitted very well, whereas sparse areas are only fitted approximately





Comparing the Results - III

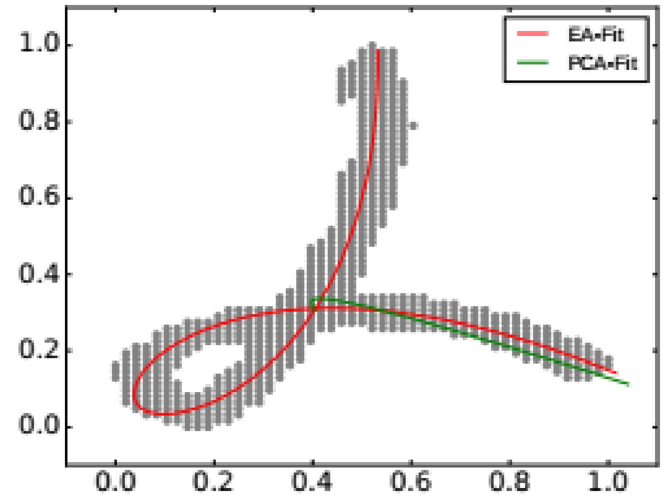
- Intersection can be succesfully modeled
- Nevertheless it becomes exceedingly hard to choose the best fitting model in this case.



Comparing the Results - III

- Intersection can be succesfully modeled

- Nevertheless it becomes exceedingly hard to choose the best fitting model in this case.



Comparing the Results - IV

- Sum of squared errors per algorithm on the presented datasets

Data Set	EA-Fit	PCA-Fit
Open Point Cloud	**0.4314**	0.4806
Closed Point Cloud	**0.4744**	1.3197
Varying Density	**0.2500**	0.3617
Sharp Corners	**0.0577**	6.0421
Self-Intersections	**0.6195**	53.5088
Three-Dimensional Data	**0.6819**	1.1549
Background Noise	**0.8018**	2.5212

Conclusions

- A survey of existing research revealed that many publications deal with the problem of approximation, but often struggle with point clouds of high dimensions or specific geometric characteristics.

- Two properties of fitting curves were defined as objective functions for the evolutionary
algorithm: the distance to the point cloud as well as the number of control points to preserve simplicity

- based on the multi-objective genetic algorithm NSGA-II

- Critical characteristics, such as non-uniform noise, sharp corners and self-intersections as well as higher dimensional data was fitted properly

Limitations and Future Work

- optimization process is constrained to the control point vector
  - Include the knot vector as well
- The best individual is currently chosen by hand or by the average rating
  - study the influence of each criterion and find better weights thereby
  - Find a method to receive the most suitable individual in the pareto-front
- For a higher number of objectives, one should consider using a Many-Objective Genetic Algorithm, such as NSGA-III
- b-splines cannot represent simple curves like circles and ellipses
  - Adapt the system to non-uniform rational b-splines (NURBS)

Thank you for your attention!

by Tim Sabsch, Christian Braune, Alexander Dockhorn, and Rudolf Kruse

Institute for Intelligent Cooperating Systems
Department for Computer Science, Otto von Guericke University Magdeburg
Universitaetsplatz 2, 39106 Magdeburg, Germany

Email: {tim.sabsch, christian.braune, alexander.dockhorn, rudolf.kruse}@ovgu.de