Bernd Hamann
Stratovan

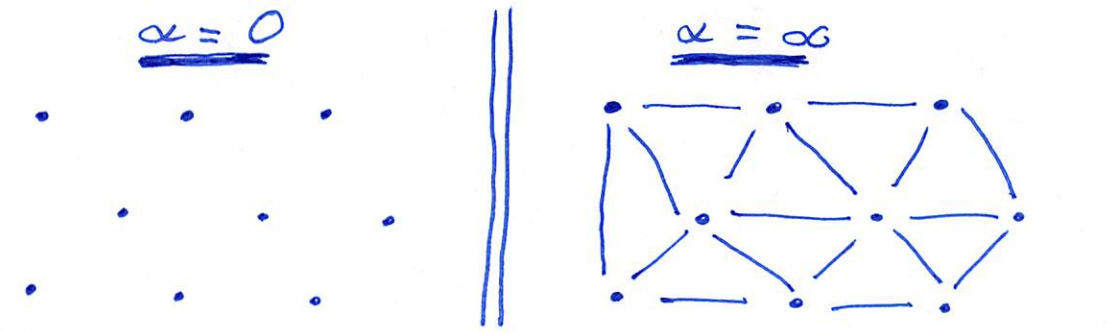
7 NOV 2016

CLUSTERING OF POINT DATA FOR CLASSIFICATION

- Distance Thresholds / Graphs / Complexes / Clusters

(Points = 'Feature Points' Defining a Material / Object)

- **Problem**: Given a set of (unconnected) feature points, determine a 'meaningful' set of point clusters...
- **Idea**: Generate various α-COMPLEXES (= subsets of the point set's Delaunay-complex) to define 'a few meaningful' implied cluster representations.
  [Understand α as threshold for the maximal length of an edge in an α-complex.]
- **Example**:

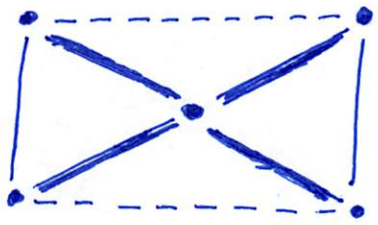


→ Extremal case: All points unconnected.

→ Extremal case: Connectivity of Delaunay complex

⇒ SIMPLY DETERMINE DEL-COMPLEX & 'RECORD' EDGE LENGTHS:

⇒ EDGE LENGTH DETERMINE THE IMPLIED COMPLEX:



— Length $< t_1$

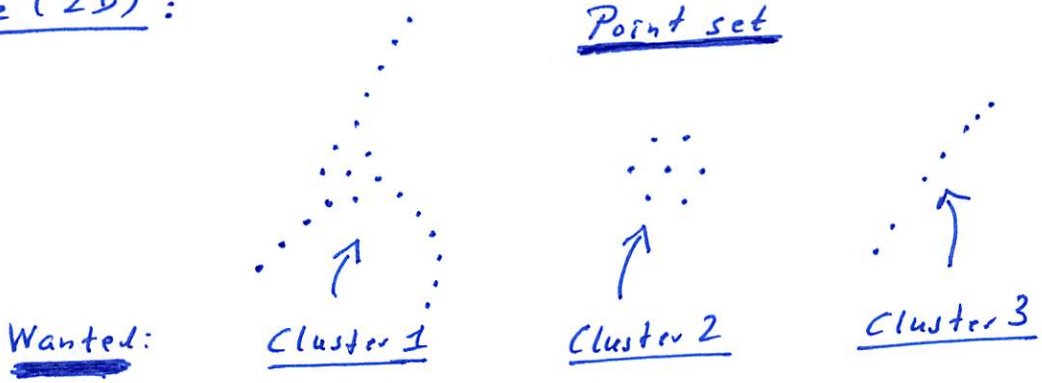
— " $< t_2$

--- " $< t_3$

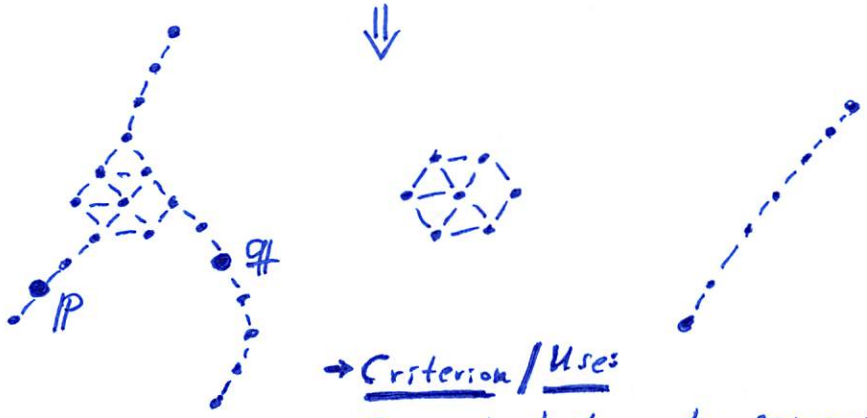
⇒ By varying value of edge length threshold, one can "smoothly" transition from one extremal case to the other (unconnected points ↔ DEL-complex).

Using Sub-complexes of the Delaunay-complex for Point Set Clustering & Classification

• Example (2D):



These 3 clusters are defined by computing the α-complex using an edge length limit of —



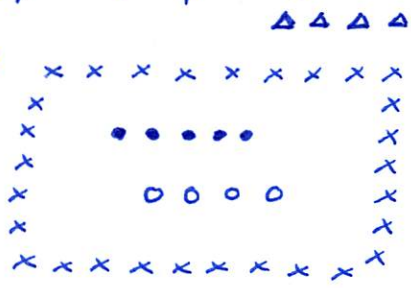
→ Criterion / Use:
2 points belong to same cluster
⇔ One can reach q from p via a connected poly-line (of edges).

→ USE SPECTRAL SUB-SPACE CLUSTERING?

??? • BUT: Can one avoid computing, storing and relying on this graph structure for determining clusters?

⇒ In principle, complex and intricate point clusters can be obtained:

• Ex:



→ Here: Edge length threshold = —

⇒ {•}, {o}, {x}, {Δ} = clusters 1, 2, 3 and 4

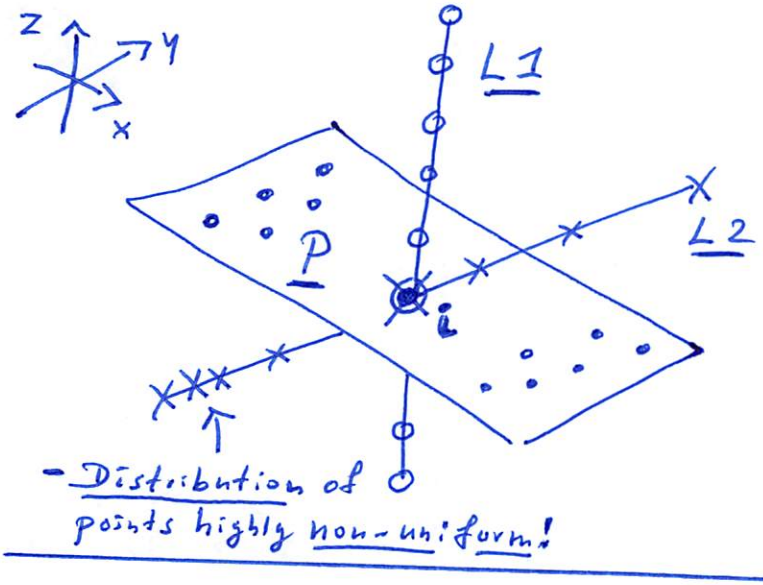
[Edges in complex are not drawn.]

■ Needed / Desirable: Robust and Efficient SUB-SPACE CLUSTERING for High-dim. Data

→ Goal: Find and define clusters for high-dim. feature vector data (defining material types) by computing lower-dim. spaces in which data sub-sets lie – via SUB-SPACE clustering (finding linear or non-linear sub-spaces containing data sub-sets)

* → Software: SUB-SPACE clustering via R ? Matlab ?

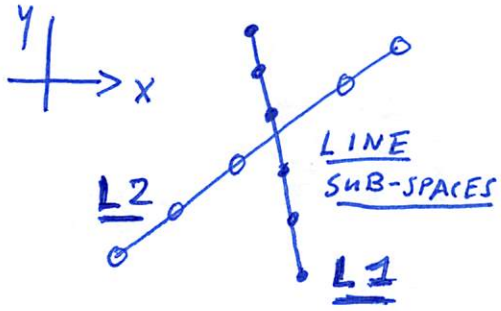
→ Examples of prototypical cases to be handled:

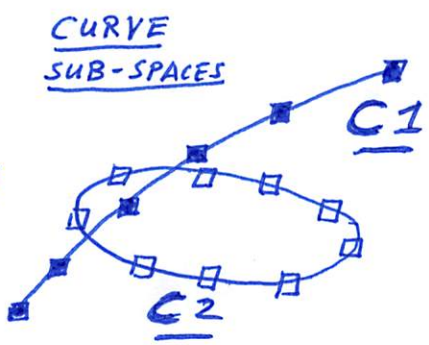


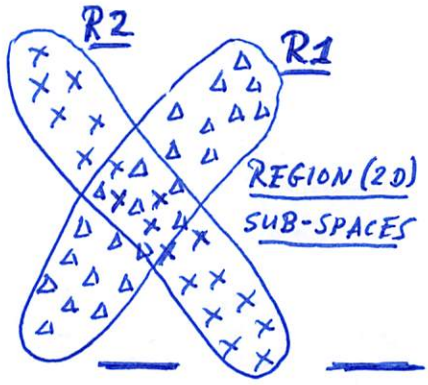
- Distribution of points highly non-uniform!

{•} 2D sub-space P
{o} 1D sub-space L1
{x} 1D " L2

⇒ Must determine 3 (linear) sub-spaces in 3D space, one being a plane, two being lines
⇒ Here: intersection point i lies in all sub-spaces.







- L1, L2: linear 1D sub-spaces
- C1, C2: non-linear/curved 1D "
- R1, R2: regional / 2D "

⇒ DETERMINE LINEAR / NON-LINEAR MATHEMATICAL MODELS FOR ALL SUB-SPACES & USE FOR CLASSIFICATION