Visualization of multivariate density estimates with shape trees

Jussi Klemelä

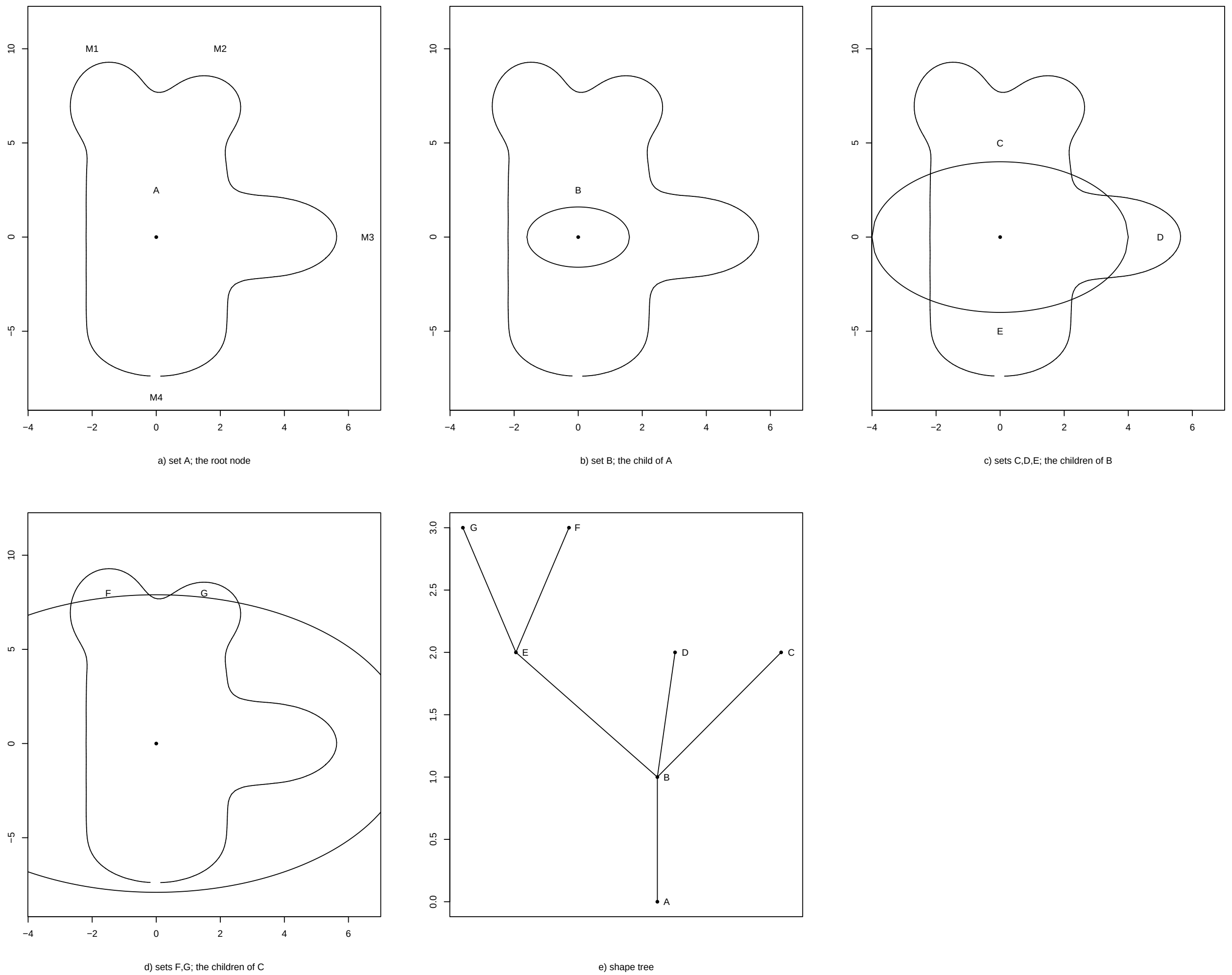
University of Mannheim

- We define a "shape transform" $S : \{A \subset \mathbf{R}^d\} \to \{f : \mathbf{R} \to \mathbf{R}\}$.
- The transform preserves certain shape characteristics of sets and thus may be used to visualize multivariate sets.
- Multivariate functions may be visualized by visualizing their level sets.
- Exploratory data analysis is possible through visualization of multivariate density estimates.

Definition of a shape transform

1. define a shape tree of a connected set, associated with a grid of radiuses and a reference point.
2. associate a piecewise constant function to the tree
3. let the grid of radiuses become finer.

Step 1: Shape tree



a) set A; the root node

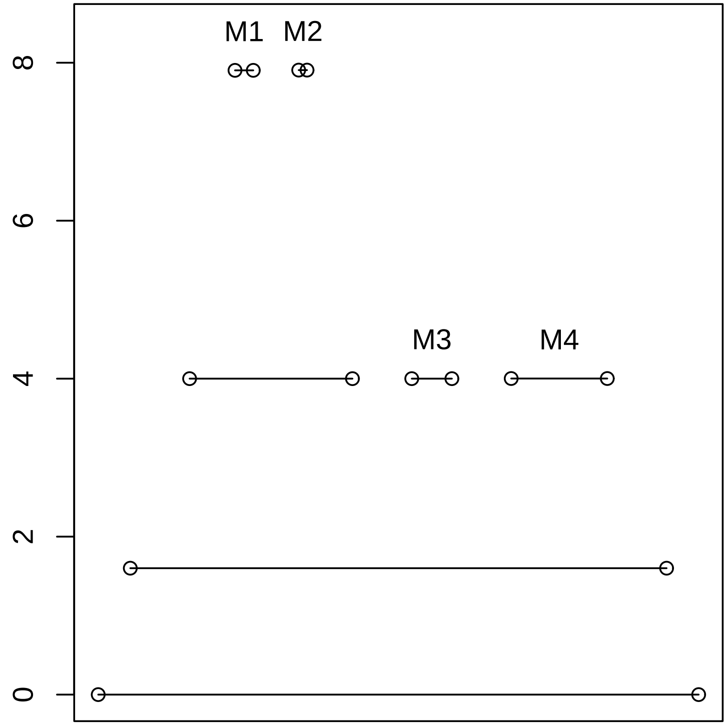
b) set B; the child of A

c) sets C,D,E; the children of B

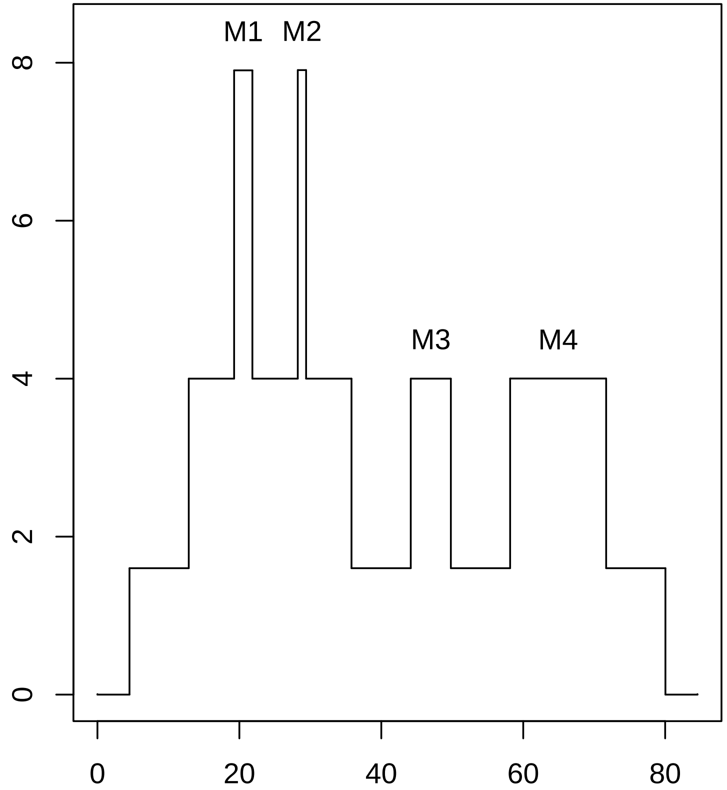
d) sets F,G; the children of C

e) shape tree

Step 2: Associate a function to a tree

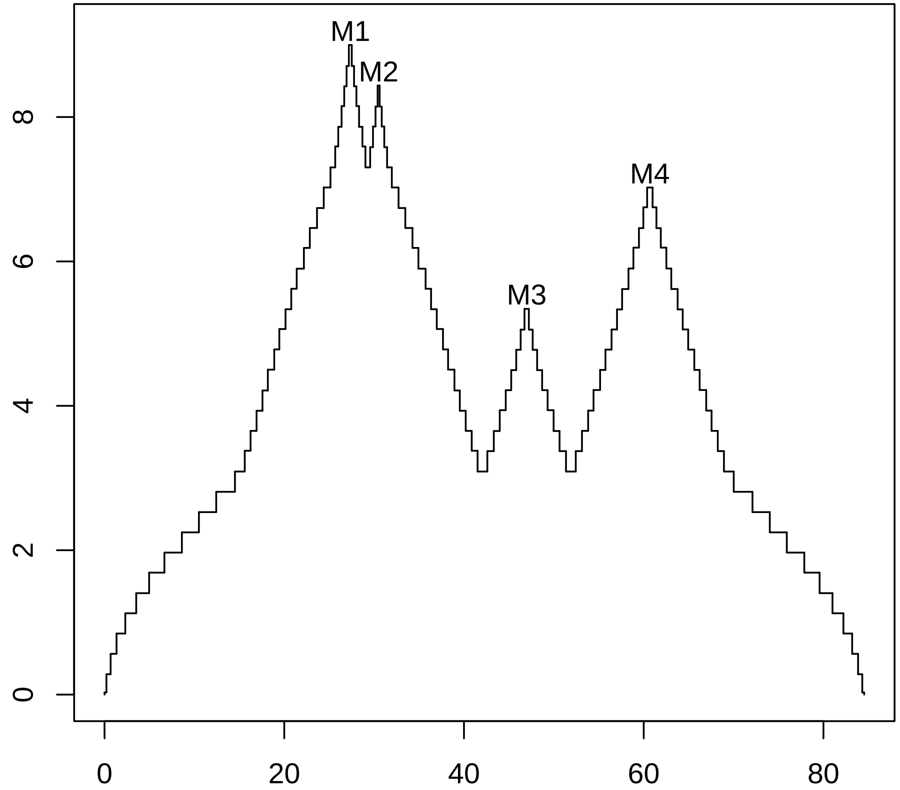


a) intervals associated with the nodes



b) radius plot

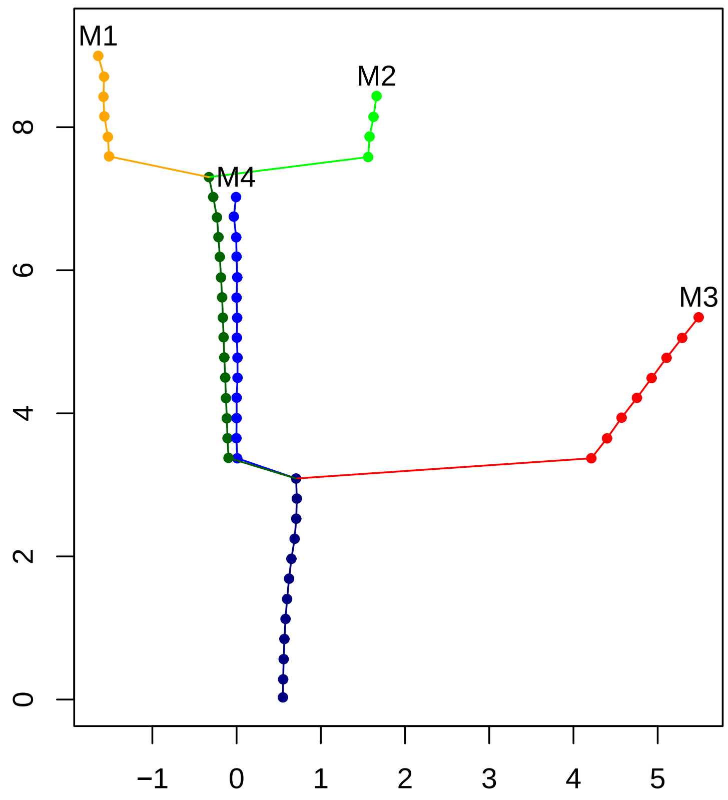
Step 3: let the grid get finer



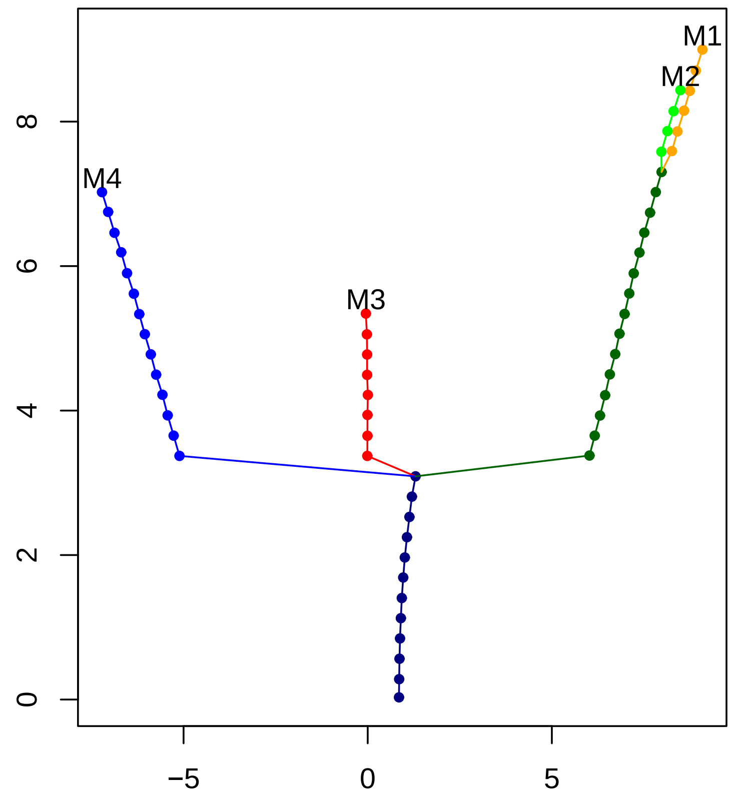
Free parameters

1. Choice of the reference point
2. Choice of the height and length of the lines: radius transform, probability content transform, volume transform.
   - In radius transform the length of the line is equal to the volume of the corresponding set and the height is equal to the distance of the corresponding set from the reference point.

Spatial location through a location plot

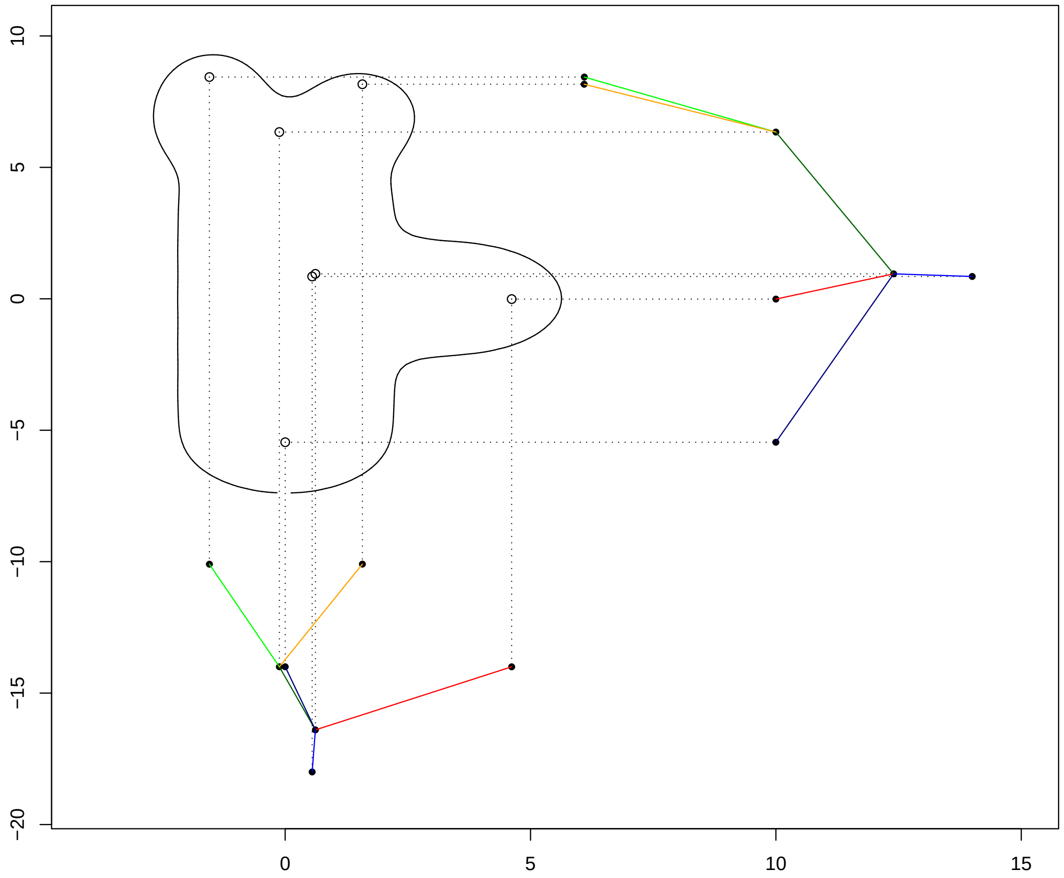


a) location plot, 1st coordinate



b) location plot, 2nd coordinate

Location plot, explanation



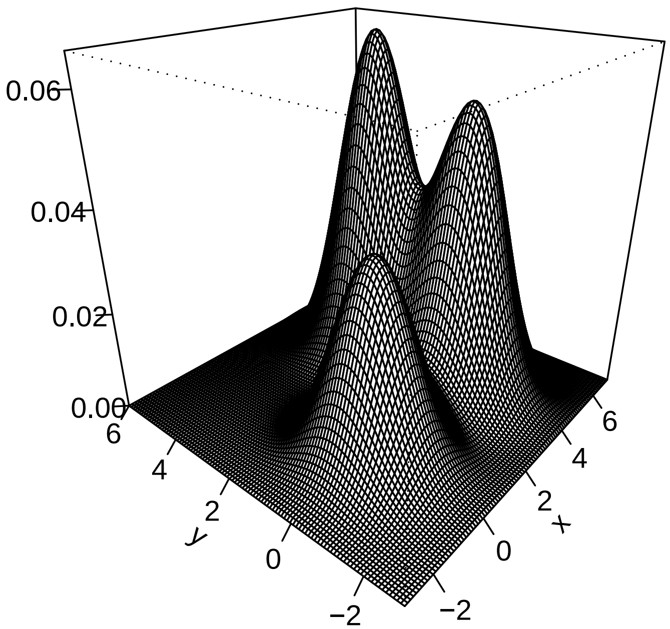
Visualizing functions

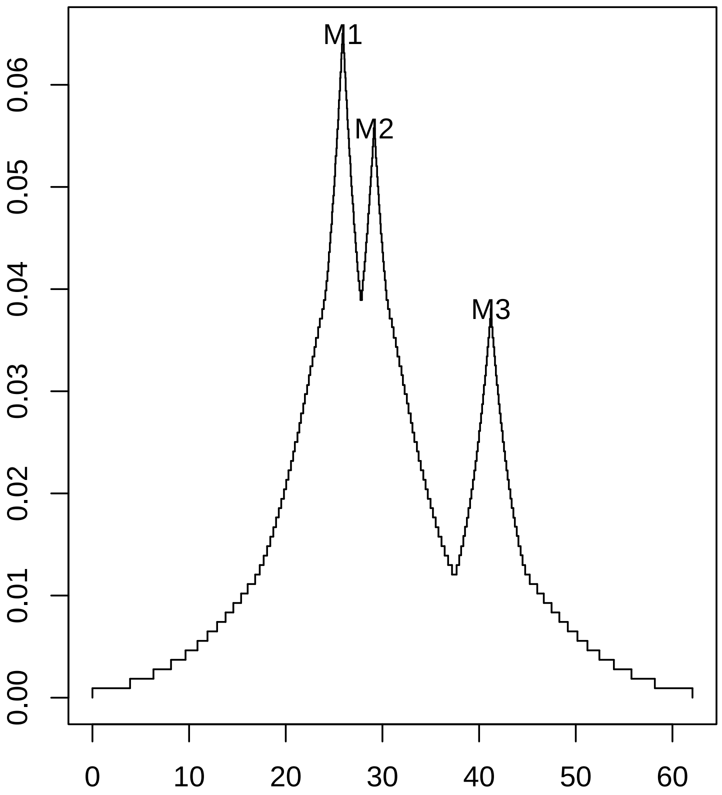
Look at the level sets of $f : \mathbf{R}^d \to \mathbf{R}$:

$$\{x \in \mathbf{R}^d : f(x) \geq \lambda\}.$$

1. Start with the level set trees; look at the multimodality and at the spacing of level sets (kurtosis and skewness).
2. Finally look at the shape of the level sets with shape transforms.

Level set transforms: volume transform of a function

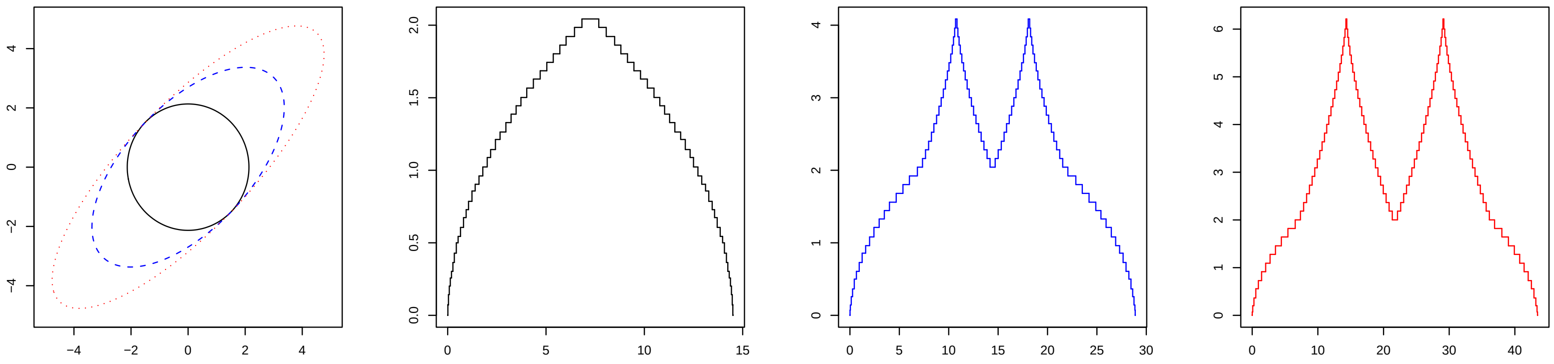




Exploratory data analysis

1. Start with data $X_1, \ldots, X_n \in \mathbf{R}^d$.
2. Smooth the data: calculate a density estimate $\hat{f} : \mathbf{R}^d \to \mathbf{R}$.
   (a) Look at the level set transforms of $\hat{f}$.
   (b) Look at the shape transforms of level sets of $\hat{f}$.

Examples of radius transform: ellipsoids



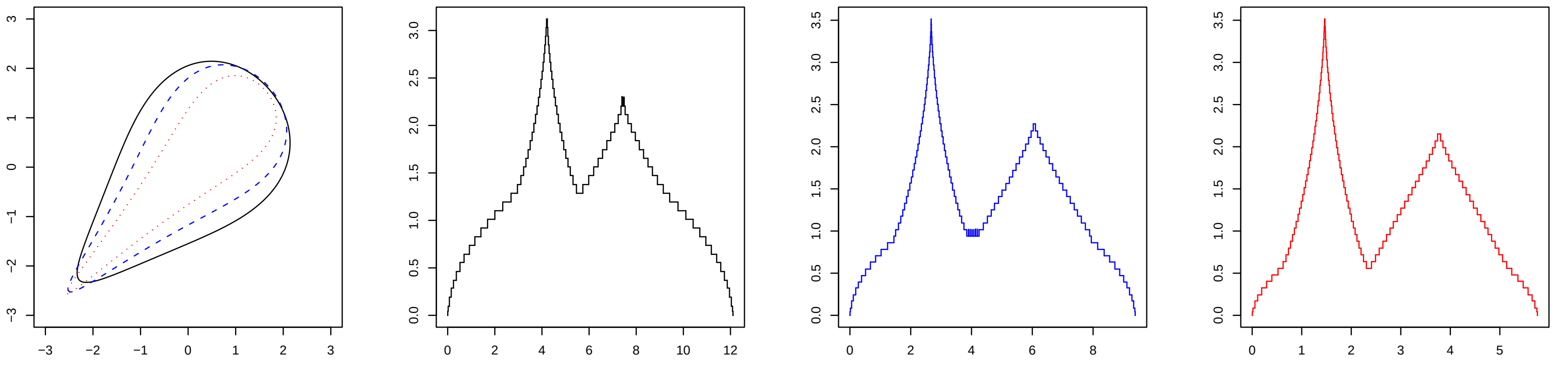
a) 3 Gaussian level sets

b) a radius plot of the ball

c) a radius plot of the shorter ellips

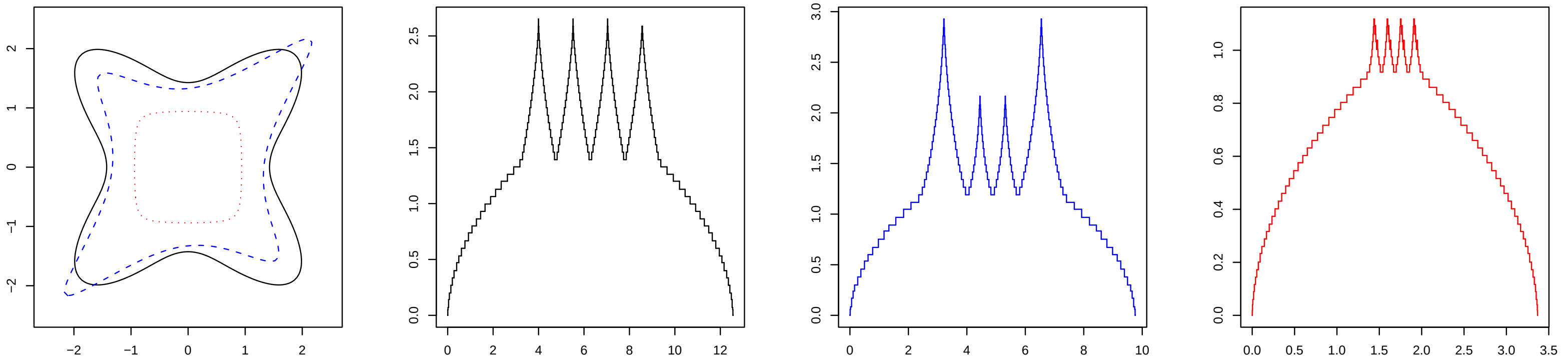
d) a radius plot of the longer ellips

Examples of radius transform: tail dependence



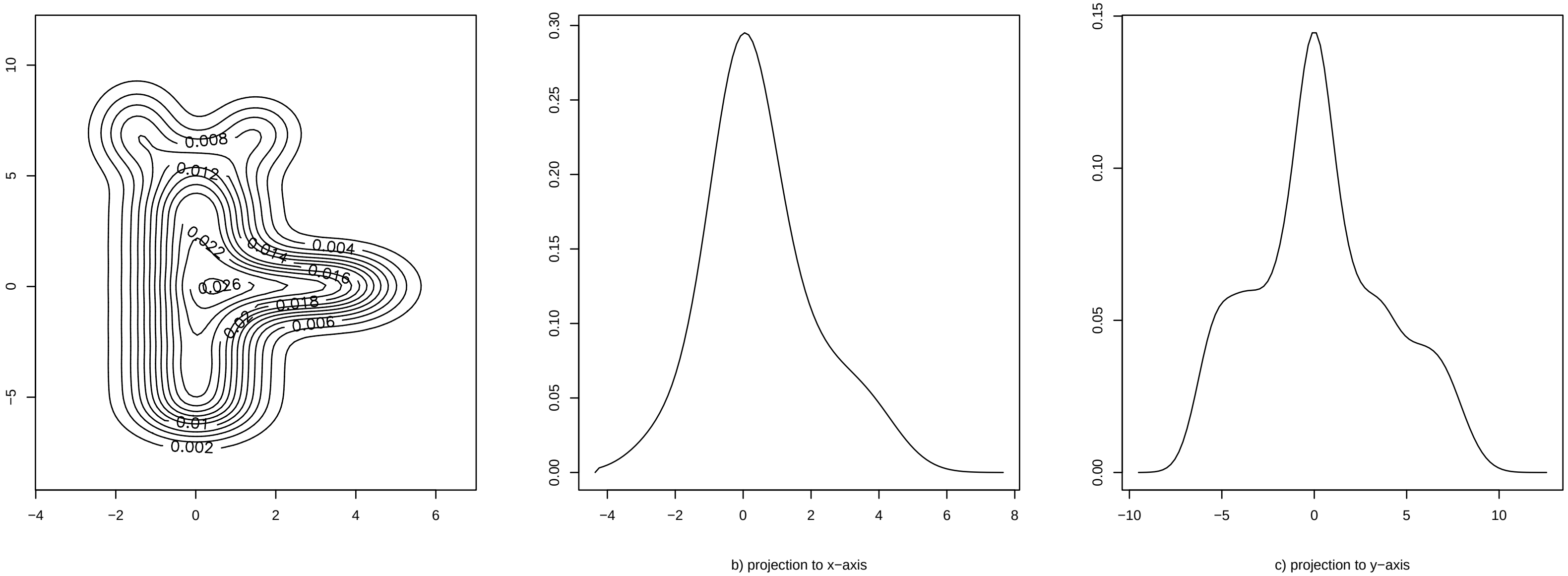
a) 3 level sets   b) set with solid boundary   c) set with dashed boundary   d) set with dotted boundary

Examples of radius transform: multimodality of sets

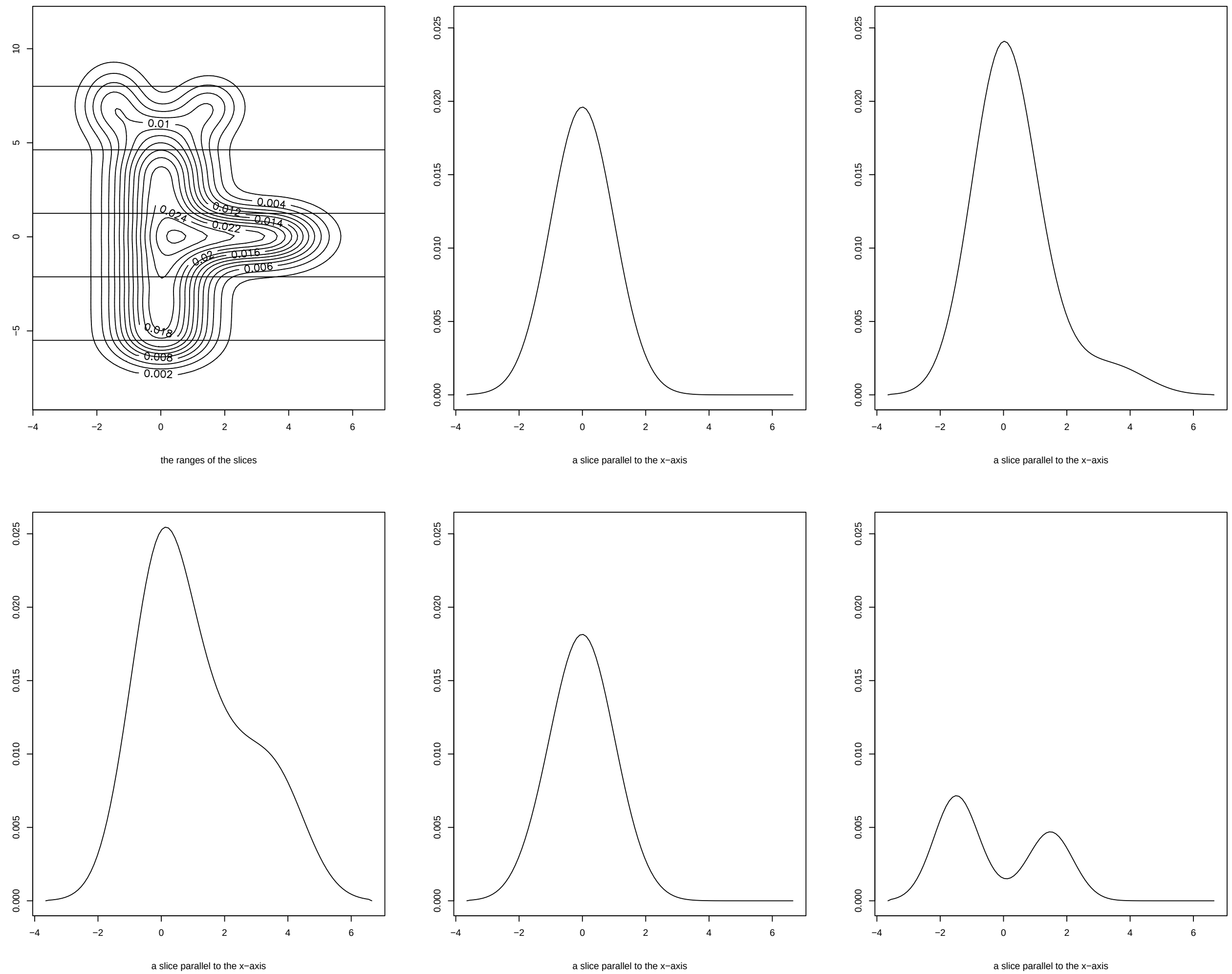


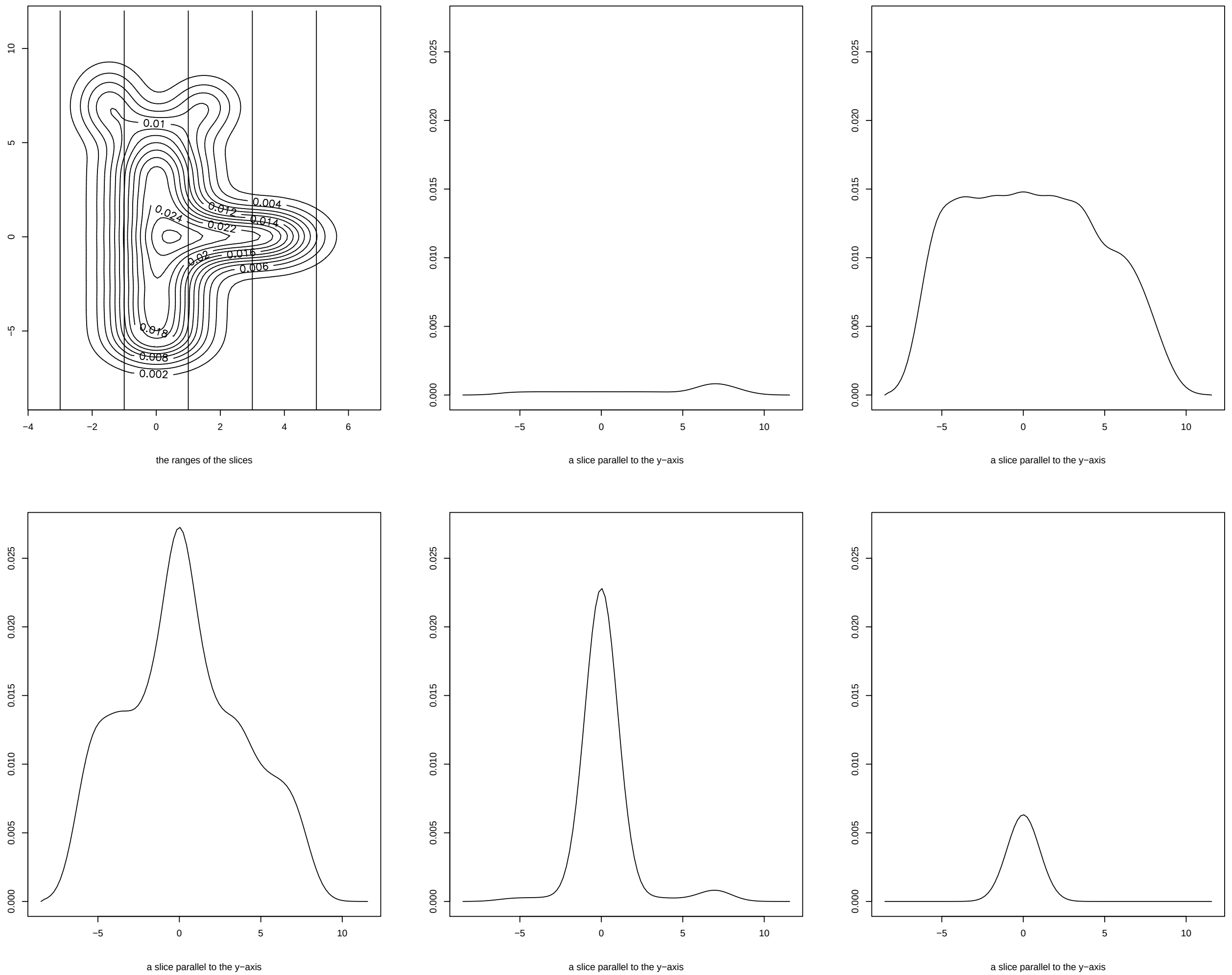
a) 3 level sets   b) mixture of 2 Gaussians   c) Student copula   d) Bartlett−Epanechnikov

Why not marginal densities?

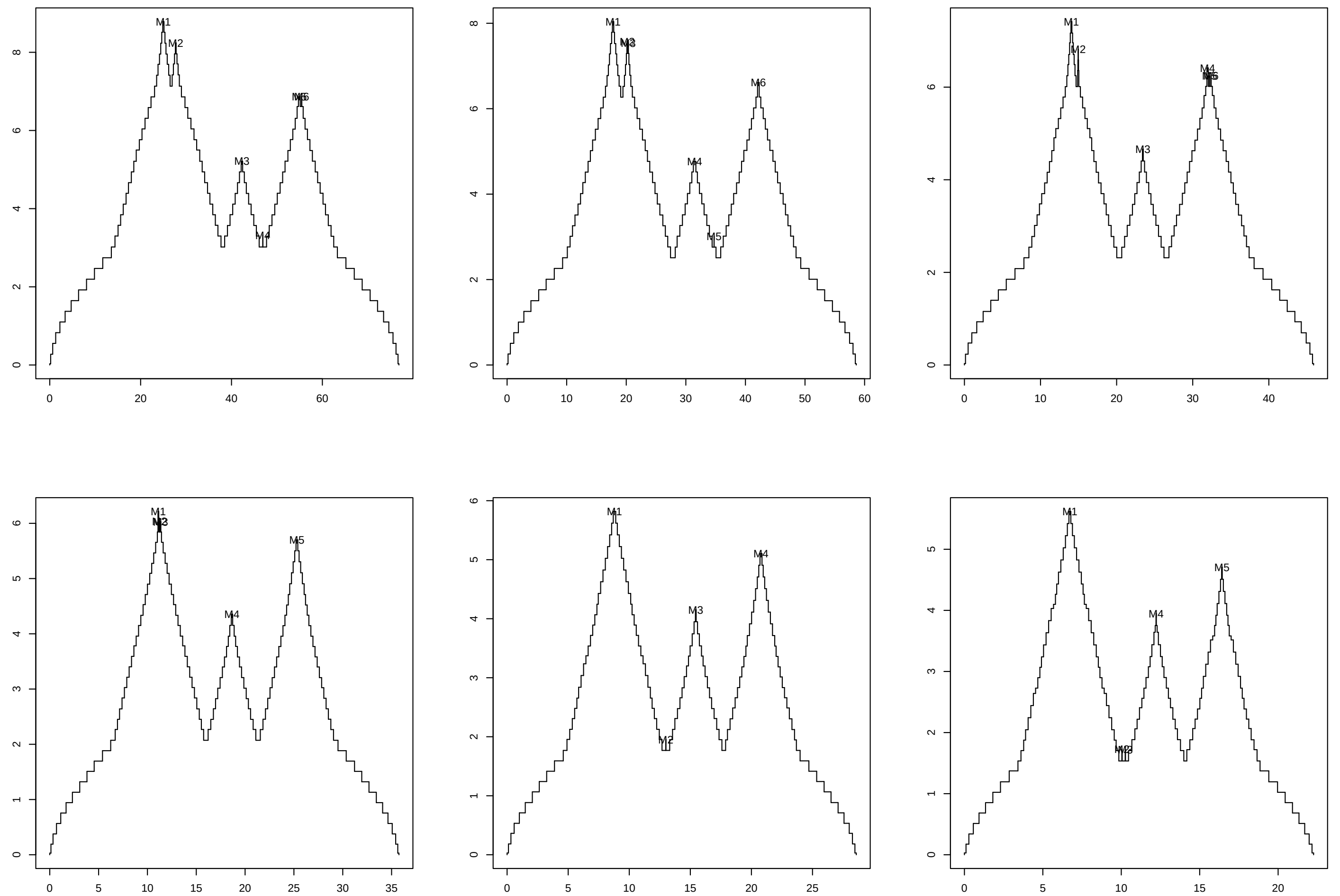


Why not slices?




the ranges of the slices
a slice parallel to the y−axis
a slice parallel to the y−axis
a slice parallel to the y−axis
a slice parallel to the y−axis
a slice parallel to the y−axis

Radius transforms of 10%-60% level sets



Conclusion

- We have defined a transform of multivariate sets to 1D functions which preserves "mode structure" of sets
- When visualizing functions one has to take into account multimodality of functions and the spacing of level sets.