Visualization of multivariate density estimates with shape trees

Jussi Klemelä*
Department of Statistics, Economics Faculty
University of Mannheim
L 7 3-5 Verfügungsgebäude
68131 Mannheim, Germany
Email: klemela@rumms.uni-mannheim.de
Fax +49 621 1811931

Abstract

We introduce graphical tools for visualizing multivariate functions, specializing to the case of visualizing multivariate density estimates. We visualize a density estimate by visualizing a series of its level sets. From each connected part of a level set we form a *shape tree*. A shape tree is a tree whose nodes are associated with regions of the level set. With the help of a shape tree we define a transformation of a multivariate set to a univariate function. We visualize shape trees with the *shape plots* and the *location plot*. By studying these plots one may identify the regions of the Euclidean space where the probability mass is concentrated. An application of shape trees to visualize the distribution of stock index returns is presented.

Mathematics Subject Classifications (AMS 2000): 62G07, 62-09
Key Words: Exploratory data analysis, Level set trees, Nonparametric density estimation, Visualization of functions
Short title: Visualization of multivariate density estimates

---

*Writing of this article was financed by Deutsche Forschungsgemeinschaft under project MA1026/8-1.

1 Introduction

One may visualize 2 dimensional functions with perspective plots and contour plots, and 3 dimensional functions may be visualized with perspective plots of 3D level sets. Higher than 3 dimensional functions may be visualized by visualizing 2 or 3 dimensional slices of the original function. Density functions may be visualized by visualizing 2 or 3 dimensional marginal densities and in general, multivariate functions may be visualized by visualizing 2 or 3 dimensional Radon transforms of the function. Finding a series of informative slices or marginal densities (projections) becomes increasingly difficult as the dimension grows.

We present visualization tools which are less dimension-sensitive than the use of slices and projections. We may visualize functions with arbitrary high dimension – up to the availability of efficient estimators and the computational complexity. We visualize "in one step" some important characteristics of multivariate functions, bypassing the problem of looking for informative slices and projections. We specialize to the case of visualizing probability density functions and in particular to the case of visualizing density estimates, based on a sequence of random vectors $X_1, \ldots, X_n \in \mathbf{R}^d$.

A dimension-insensitive method for the visualization of the mode structure of a density was introduced in Klemelä (2004*b*), where densities are visualized by visualizing the *level set tree* of the density. Klemelä (2004*b*) introduces the *volume plot* for the visualization of the relative largeness of the modes and the *barycenter plot* for the visualization of the barycenters of the level sets, in particular the locations of the modes. However, beyond the mode structure, these plots give only a limited insight into the shape of the density. We need additional visualization tools since also unimodal densities may have a large variety of shapes: the probability mass is distributed only in some regions of the huge multivariate Euclidean space.

The visualization tools which we present are based on a *shape tree* of a set. We visualize a density by visualizing a series of level sets of the density. We form a series of shape trees from a series of level sets and apply our graphical tools to visualize the shape trees. A shape tree may be formed from a connected set, where by a connected set we mean a set which cannot be expressed as a union of two sets which do not touch each other. Level sets of a unimodal density are connected. When a density is multimodal some of its level sets, possibly at higher levels, are not connected, and in this case we should visualize separately each connected component, or restrict ourselves to the visualization of lower level sets.

A shape tree of a connected set is a tree which is formed by looking at the deviations of the set from the balls centered at a given reference point, where the reference point is typically the location of the mode of the density, or the barycenter of the set. The root of the tree is identified with the set itself. The non-root nodes correspond to the separated components of the intersection of the set with the complement of a ball centered at the reference point.

We present two types of plots: the *shape plots* and the *location plot*. The shape plots visualize only the shape of the set and the location plot visualizes spatial information showing how the shape is located in the multivariate Euclidean space.

The shape plots include the *radius plot*, the *probability content plot*, and the *volume plot*. These plots are plots of 1D functions which are obtained from the shape tree. Thus we define transforms of a multivariate set to 1D functions. For example, a radius transform reveals basically the following features of the set.

1. **Qualitative information.** The modes of the radius transform correspond to the "extensions" or "tails" of the set, where with extensions of the set we mean those parts of the set which do not fit inside a given ball. We call a connected set "multimodal" when the radius transform is multimodal.

2. **Quantitative information.**

   (a) The lengths of the level sets of the radius transform are equal to the volumes of the corresponding regions of the original set.

   (b) The levels of the level sets of the radius transform are equal to the distance of the corresponding regions of the original set from the reference point.

The location plot visualizes the locations of the extensions of the set. In the location plot we plot the barycenters of the sets associated with the nodes of the shape tree.

Shape trees may be used in exploratory data analysis, in the spirit of Tukey (1977). We propose a two step approach to be used in exploratory data analysis: (1) first we make inference on the multimodality of the density with the *level set tree technique*, as introduced in Klemelä (2004*b*), and (2) second we make inference on the shape of the level sets (unimodal case) or on the shape of the connected parts of the level sets (multimodal case), with the *shape tree technique*. One may interpret this approach as a search for deviations from Gaussianity: the multimodality is the most severe deviation from Gaussianity and in the unimodal case the density deviates from a Gaussian density when the level sets are not ellipsoids, in particular when the level sets have more than two modes. Multimodal densities may often be modeled as a mixture of Gaussians with varying mode locations. “Multimodality” of level sets may sometimes be modeled with a mixture of Gaussians with the same mode but varying covariance matrices.

In Section 2 we define the basic concepts: Section 2.1 gives the definition of a shape tree, Section 2.2 contains the definitions of the shape plots, and Section 2.3 contains the definition of the location plot. In Section 3 we discuss the choice of the parameters of the plots, visualize a collection of examples with radius plots, and discuss the differences between using shape trees in visualization as compared to using marginal densities and slices. In Section 4 we discuss the implementation of the tools: Section 4.1 describes an algorithm for the calculation of shape trees and Section 4.2 gives an impression of radius plots and location plots for some potentially useful multivariate density estimators. In Section 5 we analyze a financial data set. Section 6 contains a discussion.

We use the following notations and terminology. The ball with center $\mu \in \mathbf{R}^d$ and radius $r \geq 0$ is $B_r(\mu) = \{x \in \mathbf{R}^d : \|x - \mu\| \leq r\}$ where $\|\cdot\|$ is the Euclidean distance. For two sets $A, B \subset \mathbf{R}^d$, we denote the set difference with $A \setminus B = A \cap B^c$. Lebesgue measure of set $A \subset \mathbf{R}^d$ is denoted with $\mathrm{vol}(A) = \mathrm{vol}_d(A) = \int_A dx$. We denote with I_A, $A \subset \mathbf{R}^d$, the function for which $I_A(x) = 1$, when $x \in A$ and 0 otherwise. We denote with $\#\mathcal{P}$ the cardinality of a finite set $\mathcal{P}$. A level set will be defined in (5) and the barycenter of a set will be defined in (7). Term “grid of radii” denotes a set $\{r_0 < \cdots < r_L\}$ of real numbers.

Computations and graphics in this article have been made with an R-package ”denpro”, which may be downloaded from http://denstruct.net. Klemelä (2004*a*) contains further proofs, definitions, discussion, pseudo code, illustrations, and examples.

2 Visualization concepts and tools

We define the concept of a shape tree in Section 2.1. With the help of this concept we define the graphical tools. Section 2.2 contains the definitions of the shape plots and Section 2.3 contains the definition of the location plot.

2.1 Shape tree of a set

A shape tree is a tree structure associated with a connected set, with a reference point inside this set, and with a grid of radii. The nodes of the tree are associated with subsets of the set and with the radii in the grid. The tree has a single root node and we associate the root with the set itself. To find the children of a given node, we increase the radius by one step, make the intersection of the set associated with the node with the complement of the ball with this new radius, and associate the children with the separated components of this intersection.

Separated sets and a connected set. We say that sets $B, C \subset \mathbf{R}^d$ are *separated* if $\inf\{\|x-y\| : x \in B,\ y \in C\} > 0$, where $\|\cdot\|$ is the Euclidean distance. Thus two sets are separated if there is some space between these sets. Then we say that set $A \subset \mathbf{R}^d$ is *connected* if for every nonempty B, C such that $A = B \cup C$, B and C are not separated. Thus a set is connected if it cannot be written as a union of two separated sets. Note that if $A = A_1 \cup \cdots \cup A_M$ where each A_i is connected and sets A_i are pairwise separated, then this decomposition of A to pairwise separated sets is maximal: since each set A_i is connected, we cannot find a decomposition of larger cardinality to separated sets.

Definition 1 A shape tree *of a connected set $A \subset \mathbf{R}^d$, associated with reference point $\mu \in A$, and set of radii $\mathcal{R} = \{0 = r_0 < r_1 < \cdots < r_L\}$, is a tree whose nodes are associated with subsets of A and radii in $\mathcal{R}$ in the following way.*

1. *The tree has a single root node, and this root node is associated with set A, and the radius of this node is $r_0 = 0$.*

2. *Let node m be associated with set $B \subset A$ and radius $r_l \in \mathcal{R}$, $0 \le l < L$.*

   (a) *If $B \setminus B_{r_{l+1}}(\mu) = \emptyset$, then node m is a leaf node.*

   (b) *Otherwise, write*

$$B \setminus B_{r_{l+1}}(\mu) = C_1 \cup \cdots \cup C_M \tag{1}$$

   where sets C_i are pairwise separated, and each is connected. Then node m has M children, which are associated with sets C_i, $i = 1, \ldots, M$, respectively, and each child is associated with the same radius r_{l+1}.

Figure 1 illustrates Definition 1. Frame a) shows set $A \subset \mathbf{R}^2$, and the reference point at the origin. Set A is associated to the root node of the shape tree drawn in frame e). Frame b) shows set $B = A \setminus B_{r_1}(\mu)$, which is associated to the single child node of the root node. Frame c) shows that set $B \setminus B_{r_2}(\mu) = C \cup D \cup E$ has three separated components, denoted with

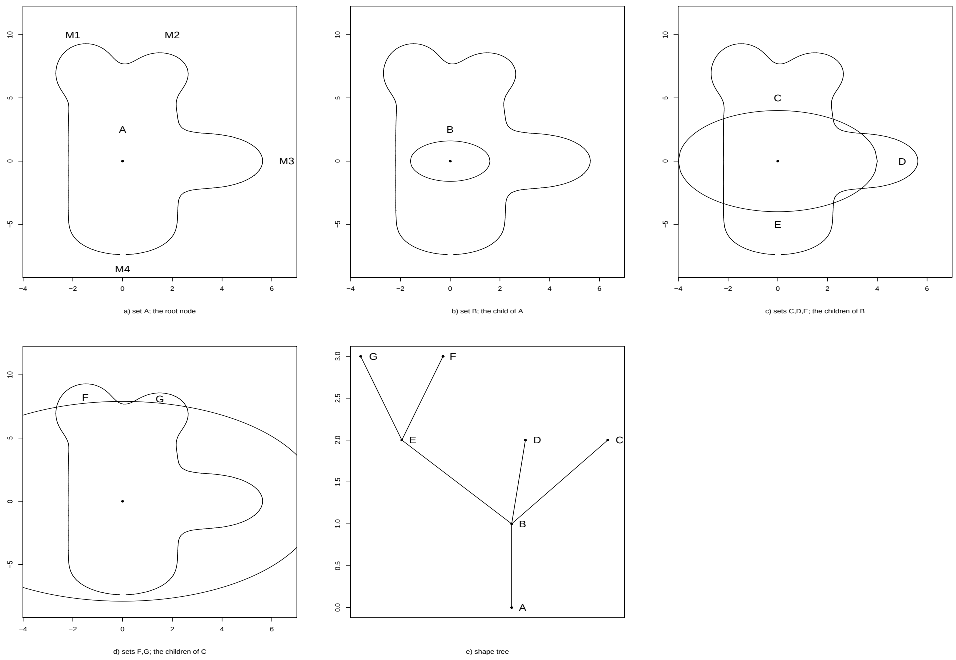


Figure 1: Illustration of a shape tree.

C, D, and E, which are associated to the three child nodes of the node corresponding to set B. Frame d) shows that set $C \setminus B_{r_3}(\mu) = F \cup G$ has two separated components, denoted with F and G, which correspond to two leaf nodes of the tree. Also the nodes associated with sets D ane E are leaf nodes.

Regularity conditions. There exists connected sets for which it is not possible to make the finite decomposition to pairwise separated and connected components as in (1), for some radius and some reference point. These sets do not have an associated shape tree. Under some regularity conditions a shape tree always exists. For example, a sufficient condition is that the set is star shaped with a smooth boundary function. We do not try to characterize the collection of sets for which shape trees exist. In practice we consider always discrete sets, such as unions of rectangles. For these sets a shape tree may always be constructed.

Modes of a connected set. The basic qualitative features of a set which are represented with a shape tree are *modes of the set*.

Definition 2 Modes of a connected set*, associated with a reference point and a grid of radii, are those sets which are associated with the leafs of the corresponding shape tree of the set.*

As the grid of radii is becoming finer, the modes of a connected set are typically shrinking to sets containing only one point of the Euclidean space. Using the term "modes of a set" is motivated by the fact that there is a unique correspondence between the modes of a shape transform and with the modes of the original set. This fact is illustrated in Section 2.2.2 and in Section 2.2.3.

2.2 Shape plots and transforms

In Section 2.2.1 we explain how to construct functions from trees. In Section 2.2.2 we define a radius plot and transform of a connected set and in Section 2.2.3 we define a probability content plot and transform of a connected level set of a density. Klemelä (2004*a*), Section B.2, contains the definition of a volume plot and transform of a star shaped set. We use the term *shape transform* to mean a radius transform, a probability content transform, or a volume transform, and a similar use is made of the term *shape plot*. Section 2.2.4 contains the definition of a limit shape transform, as the steps of the grid of radii of shape trees are becoming smaller.

2.2.1 Function generating trees

To define the plots we apply a general technique of associating 1D functions to a tree. This technique was applied in Klemelä (2004*b*) to define a volume plot of a level set tree.

Definition 3 A function generating tree *is a tree satisfying the following properties.*

1. *The root nodes and the children of each node have been ordered.*

2. *Each node is associated with a real valued height and the height of a child is larger than the height of the parent.*

3. *Each node is associated with a positive length and the sum of the lengths of children is smaller than or equal to the length of the parent.*

We may associate intervals to the nodes of a function generating tree. We define these intervals for the trees with a single root node. The lengths of the intervals are determined by the lengths associated to the nodes. The root node is associated with the interval whose left end point is at the origin and we create a nested family of intervals so that the intervals associated to the children are inside the interval associated to the parent. The child intervals are disjoint and equally spaced.

Definition 4 Intervals associated to the nodes of a function generating tree, *when this tree has a single root node, with equal spacing, are defined with the following rules.*

1. *The root node is associated with interval*

$$[0, L], \tag{2}$$

   where L is the length associated to the root.

2. *Let a node be associated with interval $[a, b]$ and assume that this node has $M \geq 1$ children, which are associated with lengths L_i, $i = 1, \dots, M$. Define the distance between the child intervals as*

$$\delta = \frac{(b-a) - \sum_{i=1}^{M} L_i}{M+1}.$$

   The i-th child is annotated with interval $[a_i, b_i]$, where $a_1 = a + \delta$, $b_1 = a + L_1$, and for $i = 2, \dots, M$,

$$a_i = b_{i-1} + \delta, \qquad b_i = a_i + L_i. \tag{3}$$

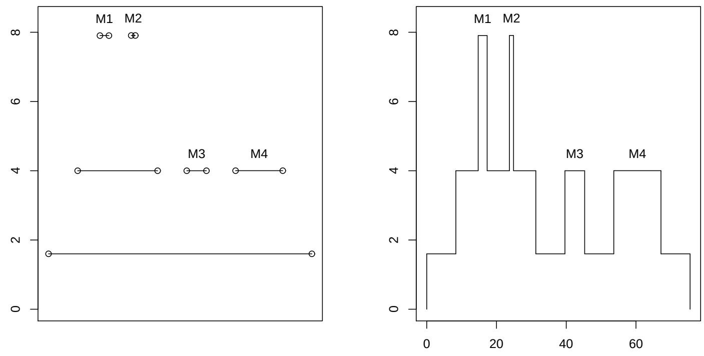
We give a formal definition of the 1D function associated with a function generating tree. The basic idea is that a level set of the 1D function is equal to the union of those intervals which are associated to the nodes whose height is larger than or equal to the level of the level set.

Definition 5 *The function $g : \mathbf{R} \to \mathbf{R}$, generated by the function generating tree T, is such that for level $\lambda \in \mathbf{R}$, the level set $\{x \in \mathbf{R} : g(x) \geq \lambda\}$ of g is equal to*

$$\bigcup \{\mathcal{I}_m : m \text{ is such node of } T \text{ that } H_m \geq \lambda\}, \tag{4}$$

where H_m is the height associated to node m and $\mathcal{I}_m$ is the interval associated to node m, as defined in Definition 4.

We illustrate this definition in Figure 2. Figure 2a shows the intervals which are associated to the nodes of the tree shown in Figure 1e. We have drawn these intervals at the heights associated with the nodes. Figure 2b shows the corresponding function. The graph of the function generated by



a) intervals associated with the nodes   b) radius plot

Figure 2: Frame a) shows the intervals associated with the nodes and Frame b) shows the function generated by the tree. Frame b) shows a radius plot of the set in Figure 1a.

the tree may be thought to be constructed so that we draw vertical lines at the end points of the intervals, joining a child to the parent, but delete the intervals themselves.

We define the radius plot, the probability content plot, and the volume plot by associating the nodes of the shape tree in different ways with lengths and heights.

A shape tree was not defined as an ordered tree, but we may define an ordering for the siblings based on the barycenters of the sets associated with the nodes. We discuss ordering rules and define the ordering rule used in this article in Klemelä (2004*a*), Section A.

2.2.2 Radius plot and transform of a connected set

We define a *radius plot* of a connected set as a plot of the function generated by a shape tree, when we choose the height of a node of the shape tree to be equal to the radius associated with the node, and the length of the node to be equal to the volume of the set associated with the node.

Definition 6 *A* radius plot *is a plot of a radius transform. A* radius transform *of a connected set A is the 1D function generated by a shape tree of A when we choose the height and the length in Definition 3 in the following way.*

1. *The height associated to a node of the shape tree is equal to the radius associated with the node.*

2. *The length associated to node m of the shape tree is equal to* $\mathrm{vol}(A_m)$ *where A_m is the set associated to node m.*

Figure 2b shows a radius plot of the set shown in Figure 1a. The reference point and the grid of radii is the same as in Figure 1.

The main qualitative features of a set which are visualized with a radius plot are the modes of the set, as defined in Definition 2. There is a unique correspondence between the modes of the set and the modes of the radius transform. In Figure 1a we have labeled the modes of the set with M1-M4 so that they correspond to the labeling of the modes of the radius transform in Figure 2b.

We visualize two quantitative features with the a radius plot: (1) volumes of the sets associated with the nodes, since the lengths of the level sets of the radius transform are equal to these volumes, and (2) distances of the sets associated with the nodes from the reference point, since the levels of the level sets of the radius transform are equal to these distances.

2.2.3 Probability content plot and transform of a connected level set

A probability content plot is defined for the level sets of a density function. By the level set of function $f : \mathbf{R}^d \to \mathbf{R}$ at level $\lambda \in \mathbf{R}$ we mean the set

$$\{x \in \mathbf{R}^d : f(x) \geq \lambda\}. \tag{5}$$

In the definition of a probability content plot we modify the definition of a radius plot by modifying the definition of the height associated with the nodes. The heights of the nodes are determined so that the probability content plot visualizes the probability content inside the level set. The length of a node is taken to be the volume of the associated set, like in a radius plot.

Definition 7 *A* probability content plot *is a plot of a probability content transform. A* probability content transform *of connected level set A of a density $f : \mathbf{R}^d \to \mathbf{R}$ is the 1D function generated by a shape tree of A when we choose the height and the length in Definition 3 in the following way.*

1. *Let node m of the shape tree be associated with set B, let the parent of node m be associated with height H, and let the children of node m be associated with sets $C_1, \ldots, C_M$. The height of node m is*

$$H + \frac{P_f(B) - \sum_{i=1}^{M} P_f(C_i)}{\mathrm{vol}(B)}, \tag{6}$$

where we take $H = 0$ when m is the root node. We denoted above with P_f the probability measure corresponding to density f.

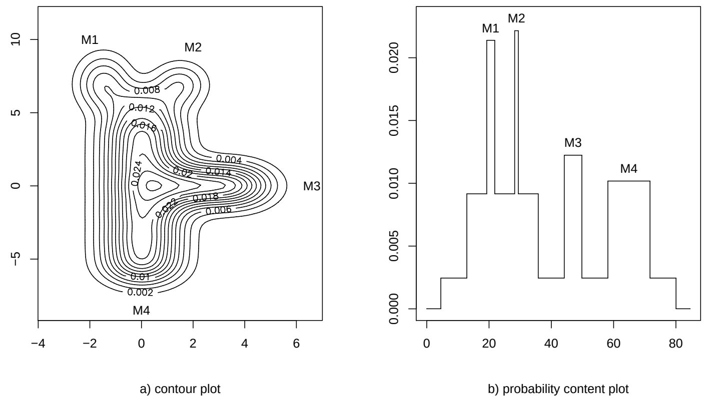


Figure 3: Frame a) shows a contour plot of a density. Frame b) shows the probability content plot of the lowest level set of the contour plot.

2. *The length associated to node m of the shape tree is equal to* $\mathrm{vol}(A_m)$ *where A_m is the set associated to node m.*

Figure 3 illustrates Definition 7. Figure 3a shows a contour plot of a density. Figure 3b shows a probability content plot of the lowest level set in the contour plot. This level set is shown in Figure 1a and the level of this level set is 0.002. The reference point and the grid of radii is the same as in Figure 1.

Similarly as for a radius plot, the basic qualitative features which are visualized with a probability content plot are the modes of the set.

We visualize two quantitative features with a probability content plot: (1) similarly as for a radius plot, the volumes of the sets associated with the nodes, and (2) the probability content of the "tail" parts of the level sets, by the choice of the levels of the level sets of the probability content transform. The excess masses of a probability content transform are equal to the probability content of the corresponding regions of the level set. We will formulate more precisely the visualization of the probability content in Theorem 1.

Theorem 1 says that the tail probabilities of density f on level set A are equal to the excess masses of the probability content transform of A. To formulate the theorem we use the following notations. When m is a node of a shape tree, then we denote with C_m the set associated with node m. When the shape tree is augmented to be a function generating tree then we denote with H_m the height and with $\mathcal{I}_m$ the interval (as defined in Definition 4) associated with node m. We denote with $\mathrm{parent}(m)$ the parent of node m.

Theorem 1 *Let m_0 be a node of a shape tree of level set $A \subset \mathbf{R}^d$ of density $f : \mathbf{R}^d \to \mathbf{R}$. Let $g : \mathbf{R} \to \mathbf{R}$ be the probability content transform of A, as defined in Definition 7. Then the probability $P_f(C_{m_0})$ of the tail region $C_{m_0} \subset A$ is equal to the excess mass of g at interval $\mathcal{I}_{m_0}$ and with level $H_{\mathrm{parent}(m_0)}$:*

$$\int_{C_{m_0}} f = \int_{\mathcal{I}_{m_0}} \left(g - H_{\mathrm{parent}(m_0)} \right)$$

where for the case that m_0 is the root node we denote $H_{\mathrm{parent}(m_0)} = 0$.

A proof and an illustration of Theorem 1 is given in Klemelä (2004*a*), Section B.1.

As a corollary of Theorem 1 we have that the probability of level set A is equal to the integral of the probability content transform: when g is the probability content transform of A, then

$$P_f(A) = \int_{-\infty}^{\infty} g.$$

2.2.4 The limit shape transform

A shape transform is a radius transform, a probability content transform, or a volume transform. We have associated a shape transform always to a shape tree of A, that is, to a finite grid of radii. We may define the limit shape transform by considering a sequence of shape trees whose grid of radii is becoming finer, and defining the shape transform to be the limit of the shape transforms associated with these shape trees.

Definition 8 The shape transform $v(A) : \mathbf{R} \to \mathbf{R}$ *of a bounded set $A \subset \mathbf{R}^d$, associated with reference point $\mu \in A$, is defined by*

$$v(A)(x) = \lim_{k \to \infty} v(A; T_k)(x), \qquad x \in \mathbf{R},$$

*where $v(A; T_k)$ is a shape transform of A associated with shape tree T_k, as defined in Definition 6, Definition 7, or in Klemelä (2004*a*), Definition 12, T_k, $k = 1, 2, \ldots$, is a sequence of shape trees of A, where each tree is associated with reference point $\mu \in A$, and with grid of radii $\mathcal{R}_k = \{r_{k,1} < \cdots < r_{k,k}\}$, and the sequence of grid of radii has the property*

$$\lim_{k \to \infty} \max\{r_{k,i} - r_{k,i-1} : i = 1, \ldots, k+1\} = 0$$

where we denote $r_{k,0} = 0$ and $r_{k,k+1} = \sup_{x \in A} \|x - \mu\|$ for $k = 1, 2, \ldots$.

2.3 Location plot

With a location plot we visualize the locations of the sets associated with the nodes of a shape tree. We draw the barycenters of these sets, when with the barycenter of a set $A \subset \mathbf{R}^d$ we mean the d-dimensional vector

$$\text{barycenter}(A) = \frac{1}{\text{vol}(A)} \int_A x \, dx \in \mathbf{R}^d. \tag{7}$$

The barycenter is the "center of mass" of the set: it is the expectation of the random vector which is uniformly distributed on the set.

Since barycenters are d-dimensional vectors we need d windows to draw the barycenters. Each window shows one coordinate of barycenters. We associate each location plot to a shape plot, and choose the vertical positions of the nodes in the location plot to be equal to the vertical positions in the associated shape plot.

Definition 9 The location plot *of a shape tree, associated with a shape plot, consists of* d *windows. The nodes of the shape tree are drawn as bullets.*

1. *The horizontal position of a node in the* i*-th window,* $i = 1, \ldots, d$*, is equal to the* i*-th coordinate of the barycenter of the set associated with the node.*

2. *The vertical positions of the nodes are the same as in the associated shape plot.*

3. *The parent-child relations are expressed by the line joining a child with the parent.*

Figure 4 shows a location plot of the set in Figure 1a, corresponding to the radius plot of Figure 2b.

It is enough to label the leaf nodes to identify the nodes between different windows of a location plot (and between the associated shape plot and the location plot). However, to ease the identification of nodes across different windows, we also color the nodes. We choose first distinct colors for the leaf nodes and then travel towards the root node, changing the color always when two branches are merging. We color also the lines joining a child and a parent. The color of a line will be the same as the color of the child node which is at the child end (upper end) of the line.

Figure 5 illustrates the location plot of Figure 4. We have drawn as circles the 7 barycenters of the sets associated with the 7 nodes of the shape tree. These barycenters are joined with dotted lines to the corresponding nodes of the trees related to the 1st and 2nd window of the location plot.

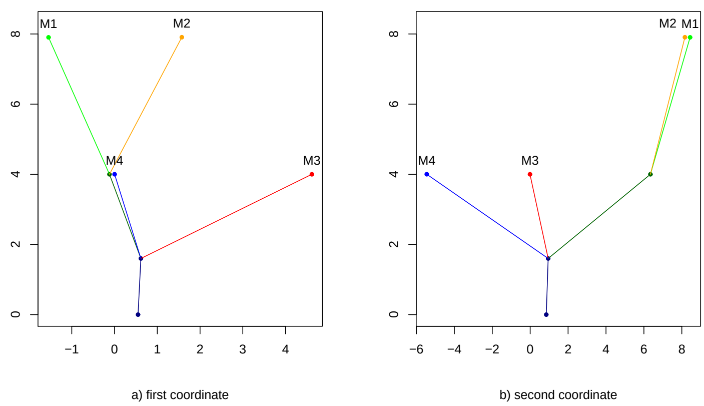


Figure 4: Location plot of the level set in Figure 1a, corresponding to the radius plot of Figure 2b.

The leaf nodes of a shape tree correspond to the certain most extreme boundary points of the set, and the leaf nodes in a location plot show the locations of these extreme points. Thus a location plot visualizes a delineator of the set.

3 Interpretation and application of the tools

In Section 3.1 we discuss the effect of the choice of the parameters of shape trees. In Section 3.2 we discuss how the shape of a density is reflected by the level sets and we illustrate how shapes of sets are reflected by radius plots. In Section 3.3 we discuss why shape trees are sometimes easier to use in visualization than marginal densities and slices.

3.1 Choice of the parameters

We discuss the choice of the grid of radii, the choice of the reference point, and differences between a radius plot and a probability content plot. Klemelä (2004*a*), Section C.1, discusses the choice of the metric in the definition of a ball.

3.1.1 Grid of radii

Figure 6 shows a radius plot and the corresponding location plot for the set of Figure 1a, when a grid of 30 radii is used, and the set is approximated

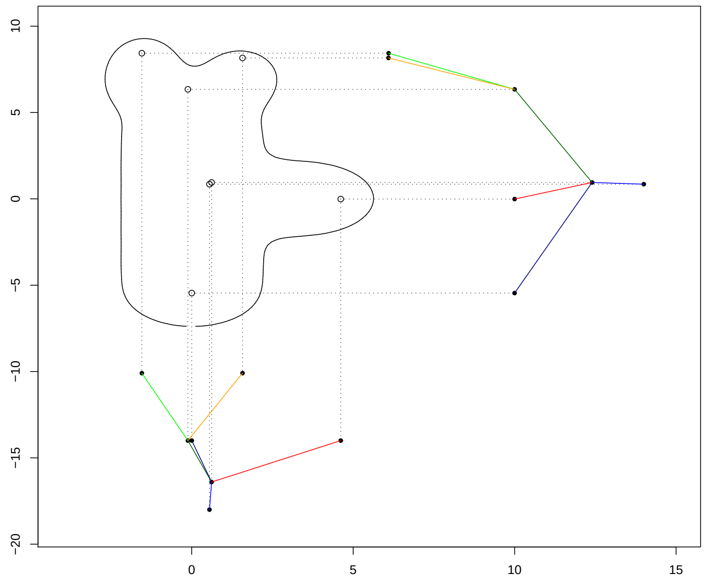


Figure 5: An illustration of the location plot in Figure 4.

with a grid of size 100^2. Note that in Figure 1, Figure 2, and Figure 3 we applied a grid of 4 radii and these radii were not equispaced.

Section 2.2.4 gives the definition of the limit shape transform, as a limit of a sequence of shape transforms, when the steps of the grid of radii are becoming smaller. In practise, using a grid of radii that is too fine may result to a messiness of plots. In fact, when we approximate a smooth boundary with some discrete approximation, and use a fine grid of radii, then the resulting shape transforms may have some spurious modes which do not correspond to any real modes in the original smooth set. In this article we approximate sets with unions of rectangles. The "spurious" modes do not typically have disturbing effects on shape plots. However, location plots look messy with a large number of spurious modes. One should avoid a too fine grid of radii to avoid messiness of location plots. On the other hand, with a too sparse grid of radii some real modes may not show up in the shape tree. Discretization effects are further illustrated in Klemelä (2004*a*), Section D.1, Section D.2.

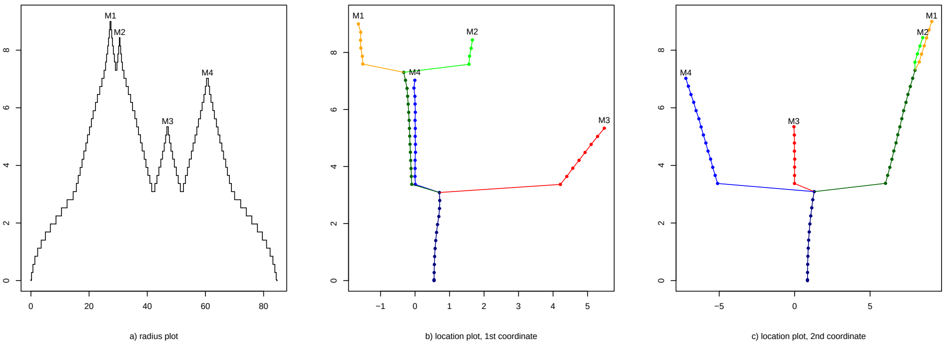


Figure 6: A radius plot and the corresponding location plot for the set of Figure 1a.

3.1.2 Reference point

The choice of the reference point has a substantial influence on the plots. Natural choices for the reference point are either the mode of the density, or the barycenter of the level set. The barycenter of a level set is not equal to the mode when the density is skewed. Before applying shape trees, it is useful to investigate the skewness of the density with a barycenter plot of the level set tree, as introduced in Klemelä (2004*b*). The expectation could also be used as a reference point if it exists, and belongs to the level set. The mode is the only point which belongs to every level set of a unimodal density, and it may be helpful to apply the same reference point for every level set when we are studying a series of level sets.

One may get additional information by trying several reference points. We illustrate this with a skewed density, shown in Figure 7a. We define the density in Klemelä (2004*a*), Section C.4, following Azzalini and Dalla Valle (1996). Frame a) shows a contour plot where the mode is shown with a blue square and the barycenter of the level set with level 0.02 is shown with a red bullet. Frame b) shows a radius plot of the 0.02 level set when the reference point is the mode. Frames c) and d) show the corresponding location plot. The radius plot is unimodal although the level set is not a ball, because the reference point is far from the barycenter of the level set. The skewness in the location plot reveals the fact that the chosen reference point is far from the barycenter.

In Figure 8 the reference point is the barycenter. Figures 8a-c show a radius plot and the corresponding location plot of the 0.02 level set of the skewed density of Figure 7a. Now the radius plot is bi-modal. The north-east extension is closer to the barycenter, with more excess volume, than the the

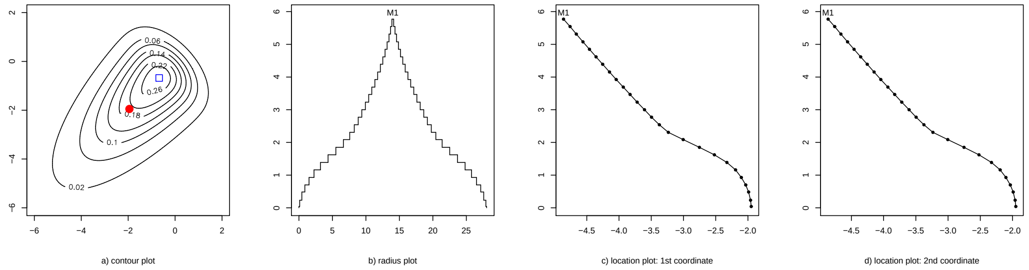


Figure 7: Skewed density; the mode as the reference point for the 0.02 level set; a radius plot and the corresponding location plot.

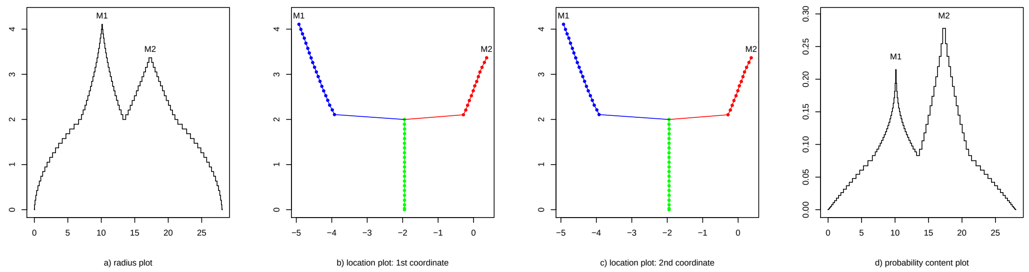


Figure 8: Skewed density; the barycenter as the reference point for the 0.02 level set. Frame a) shows a radius plot, frames b) and c) show the location plot corresponding to the radius plot, and frame d) shows a probability content plot.

south-west extension, and thus the corresponding mode in the radius plot is shorter and wider than the mode corresponding to the south-west extension.

3.1.3 Radius plot vs. probability content plot

A radius plot is easier to understand than a probability content plot. However, a probability content plot visualizes alternative information, showing the probability mass inside the level set. Thus a probability content plot might be relevant for making inference whether the modes of a level set of a density estimate correspond to the true modes of the level set of the underlying density function.

A probability content plot may be useful in visualizing skewed densities. Figure 8d shows a probability content plot for the skewed density of Figure 7a, again for the 0.02 level set, when the reference point is the barycenter. We may compare the probability content plot with the radius plot in Figure 8a.

In the probability content plot of Figure 8d the mode corresponding to the north-east extension of the level set is clearly bigger than the mode corresponding to the south-west extension. In the radius plot of Figure 8a these extensions correspond to modes of roughly the same size. Thus the probability content plot has visualized the fact that the north-east direction contains more probability mass. The location plot corresponding to the probability content plot of Figure 8d is shown in Klemelä (2004*a*), Section C.4.

3.2 Interpretation of the graphics

3.2.1 Level sets and the shape of the density

The idea of visualizing densities with level sets appears in Scott (1992). This idea has been used to gain one more dimension: 3D functions may be visualized with perspective plots of 3D level sets. Also the shape tree based visualizations utilize the fact that the shape of the density is reflected by the shape of the level sets. However, besides the shape of the level sets one has to take into account the spacing of the level sets. The spacing of the level sets affects the "kurtosis" and "skewness" of the density.

1. We use the term "kurtosis" in an informal sense to refer to the spacing of the level sets in the vertical direction. For example, all spherically symmetric densities have ball-shaped level sets. Spherically symmetric densities may be written as $f(x) = g(\|x\|)$, where $g : [0, \infty) \to [0, \infty)$ and $\|\cdot\|$ is the Euclidean distance. The volumes of the level sets are however changing differently as the function of the level, depending on function g.

2. Level sets may be unequally spaced in the horizontal direction. This happens for skewed densities.

Level set trees may be used to diagnose kurtosis and skewness. A volume plot of a level set tree visualizes the change of the volumes of level sets as the function of the level, and thus it visualizes the kurtosis. See Klemelä (2004*a*), Section C.2. A barycenter plot of the level set tree may be used to detect skewness of the density.

3.2.2 Uni- and bi-modality

Ellipsoidal level sets. A shape transform of a ball is a unimodal function. Shape transforms of ellipses are bi-modal functions, when the reference point is not close to the boundary of the ellipse. Figure 9 shows three sets and

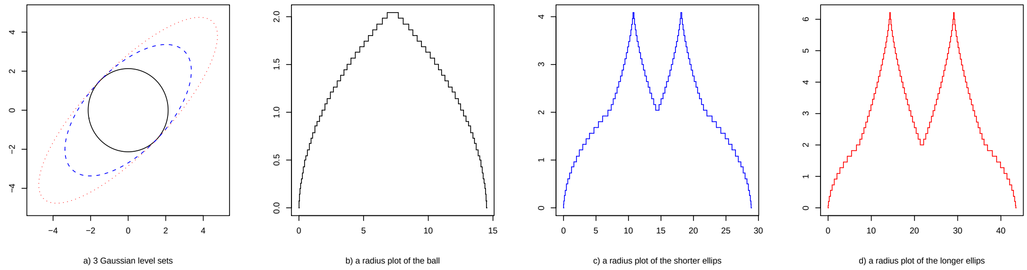


Figure 9: Frame a) shows a ball and 2 ellipses, frame b) shows a radius plot of the ball, frame c) shows a radius plot of the shorter ellipse, frame d) shows a radius plot of the longer ellipse.

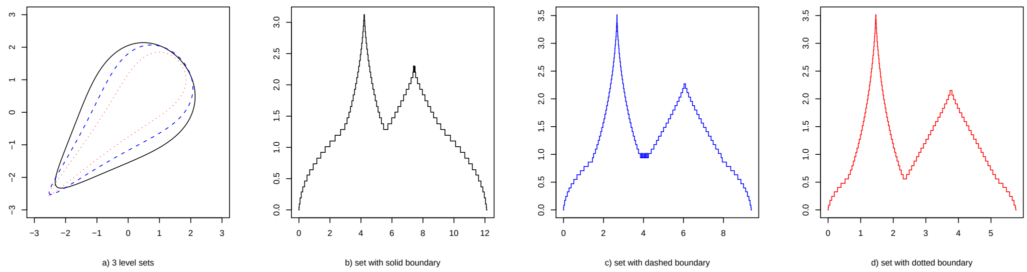


Figure 10: Frame a) shows 3 Claytonian level sets and frames b)-d) show the corresponding radius plots. Frame b) corresponds to the set with the black solid boundary, frame c) to the set with the blue dashed boundary, and frame d) to the set with the red dotted boundary.

their radius transforms, when the reference point is at the origin. Frame a) shows a ball and 2 ellipses, frame b) shows a radius plot of the ball, frame c) shows a radius plot of the shorter ellipse, and frame d) shows a radius plot of the longer ellipse. The modes of the radius transforms are becoming more distinguished when the ellipses are becoming longer.

Non-symmetric tail dependence. Figure 10a shows 3 Claytonian level sets. The densities in the Clayton family have non-symmetric tail dependence; the dependence is larger in the negative orthant. We chose the marginal densities to be standard Gaussian. The Clayton family was discussed by Clayton (1978) and we define this family in Klemelä (2004*a*), Section C.5.

Figures 10b-d show radius plots corresponding to the level sets. The radius plots visualize the asymmetry of dependence: the mode corresponding

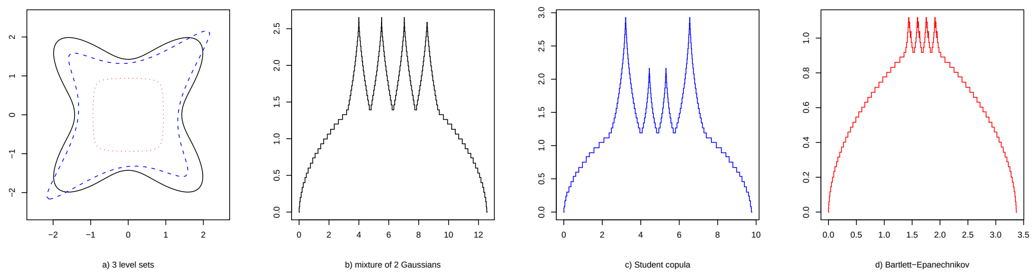


Figure 11: Frame a) shows 3 level sets and frames b)-d) show the corresponding radius plots. Frame b) corresponds to the set with the black solid boundary, frame c) to the set with the blue dashed boundary, and frame c) to the set with the red dotted boundary.

to the area in the negative orthant is higher but the mode corresponding to the area in the positive orthant is wider.

3.2.3 Multimodality of level sets

Figure 11a shows three level sets which have 4-modal shape transforms. The black solid line delineates a level set of a mixture of two Gaussian densities, the blue dashed line delineates a level set of a density with Student copula and with standard Gaussian margins, and the red dotted line delineates the 10% level set of the Bartlett-Epanechnikov product kernel. The densities with Student copula are defined in Klemelä (2004a), Section C.6. The Bartlett-Epanechnikov density is $(x_1, \ldots, x_d) \mapsto (3/4)^d \Pi_{i=1}^d \max\{0, 1 - x_i^2\}$.

Figures 11b-d show the corresponding radius plots. The 4 tails of the mixture of Gaussians show up in the radius transform as 4 modes of equal size. The density with Student copula has also 4 tails but of unequal size, and they show up in the radius transform as modes of unequal size. The 10% level set of Bartlett-Epanechnikov density has a slightly rectangular form and thus the radius transform has 4 modes but these modes are of small size.

3.3 Marginal densities and slices

Multivariate density estimation has been applied in exploratory data analysis and data mining by looking at 2 or 3 dimensional slices or marginal densities of the original multivariate density estimate. These slices or marginal densities may be visualized with perspective plots (2 dimensional case) or with perspective plots of density contours (3 dimensional case). For example, 4 dimensional densities may be visualized by visualizing 3D density contours as the fourth variable is changed over its range, see Scott (1992) and Härdle and Scott (1992).

There are two difficulties with this approach. First, there exists a huge number of projections from which to construct marginal densities, and even larger number of possible slices. Grand tour and projection pursuit may be applied to find some informative marginal densities, see for example Asimov (1985), Cook, Buja and Cabrera (1993), Cook, Buja, Cabrera and Hurley (1995). (Note that we may first project data to 2 or 3 dimension and only after that smooth the data, see Wegman and Luo (2002).) The second difficulty is to deduce from marginal densities and slices the shape of the original function: even when we had a sufficient collection of marginal densities and slices, it is a non trivial effort to make conclusions about the shape of the original density based on these low dimensional views. Note that looking only at marginal densities may hide some features and often we need a combination of slices and marginal densities, see Furnas and Buja (1994).

We argue that it is practically impossible to use marginal densities to visualize multimodality of level sets and that it is very difficult to visualize the multimodality of level sets with slices, due to the large number of possible slices.

Marginal densities. Multimodality of densities may sometimes be detected by visualizing marginal densities, problems arising when the modes are close to each other or when there are a large number of modes. On the other hand, detecting multimodality of level sets with marginal densities is essentially more difficult. Figure 12 shows the marginal densities along the coordinate axes of the density shown in Figure 3a. The precise shape would be difficult to reconstruct even with a large number of marginal densities.

Slices. With skillfully chosen slices it is possible to visualize multimodality of level sets. However, the main problem is to find the right slices and to keep records of the information provided by the slices. Note that in the d-dimensional Euclidean space there are $d(d-1)/2$ ways to choose two coordinate directions, but we need also to consider other than coordinate directions, and for each direction we need a grid of slices. Klemelä (2004a), Section C.7, illustrates the difficulty of using slices to visualize the density shown in Figure 3a.

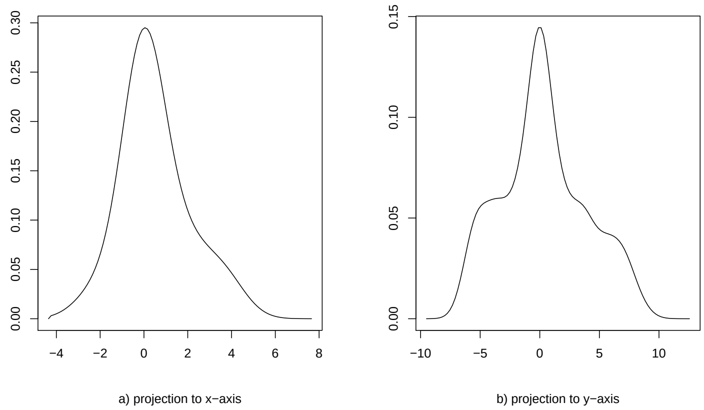


Figure 12: The marginal densities on the coordinate axes of the density of Figure 3a.

4 Implementation of the graphical tools

We present an algorithm for the calculation of a shape tree in Section 4.1. In Section 4.2 we give examples of applying shape trees in the case of some useful density estimators.

4.1 Algorithms

We assume that the density function (density estimate) to be visualized is a rectangularwise constant function:

$$f(x) = \sum_{R \in \mathcal{P}} f_R I_R(x), \qquad x \in \mathbf{R}^d, \tag{8}$$

where $f_R \in \mathbf{R}$, and $\mathcal{P}$ is a finite collection of almost everywhere disjoint rectangles.

We present algorithm **LeafsFirst** for the calculation of a shape tree. This algorithm starts building the tree from the leaf nodes. First we find the rectangle which is furthest away from the reference point. (We define the distance of a rectangle to a point to be the distance of the point to the boundary of the rectangle.) This rectangle is associated to the first leaf node. We go through rectangles in such a way that we always encounter first rectangles which are further away from the reference point, and check whether the new rectangle touches some of the previous collections of rectangles. If it does not touch any previous collections of rectangles, then we create a new leaf node and associate this leaf node with the rectangle. Otherwise, if it touches some previously encountered rectangle, then we associate it to the corresponding collection of rectangles, and if it touches several rectangles corresponding to different branches of the tree, then we join those branches.

We give below a pseudo code for the algorithm **LeafsFirst**. This algorithm creates a shape tree for a level set of a piecewise constant function as in (8). Since the function is already assumed to be discrete, we let the function determine the grid of radii of the shape tree. We define a grid of radii of cardinality $\#\mathcal{P} + 1$, corresponding to reference point $\mu \in \mathbf{R}^d$, by taking $r_0 = 0$, and assuming we have defined r_l, $0 \le l < \#\mathcal{P}$, let

$$r_{l+1} = \inf\{D(\mu, R) : R \in \mathcal{P}, D(\mu, R) > r_l\} \tag{9}$$

where $D(\mu, R) = \inf\{\|x - \mu\| : x \in R\}$.

1. **Input** of the algorithm is a piecewise constant function f as in (8), a level $\lambda \in \mathbf{R}$, and a reference point $\mu \in A = \{x \in \mathbf{R}^d : f(x) \ge \lambda\}$.

2. **Output** of the algorithm is a shape tree of A, as defined in Definition 1, with the grid of radii defined with the rule (9).

ALGORITHM **LeafsFirst**

1. Find the collection of rectangles $\mathcal{P}_\lambda$ forming the level set with level λ: $\mathcal{P}_\lambda = \{R \in \mathcal{P} : f_R \ge \lambda\}$.

2. Order rectangles $R \in \mathcal{P}_\lambda$ according to the distance to the reference point μ:
$$R_1 \preceq R_2 \Leftrightarrow D(\mu, R_1) \ge D(\mu, R_2).$$

3. Find the first rectangle $R \in \mathcal{P}_\lambda$ in the ordering $\preceq$, create the first leaf node of the shape tree, the set associated with this node is R, and the radius is $D(\mu, R)$.

4. We go through rectangles $R \in \mathcal{P}_\lambda$ in the ordering $\preceq$. Assume we have encountered rectangle $R \in \mathcal{P}_\lambda$. Find which sets, associated with the current root nodes (those nodes which do not yet have a parent), touch rectangle R.

    (a) If rectangle R does not touch any sets associated with the current root nodes, then create a new leaf node to the tree. The set associated with this node is R, and the radius is $D(\mu, R)$.

(b) If rectangle R touches sets $C_1, \ldots, C_M$ associated with the current root nodes, then create a parent to these nodes which are touched. The set of this parent node is $R \cup C_1 \cup \cdots \cup C_M$, and the radius is $D(\mu, R)$.

We illustrate the algorithm in Klemelä (2004*a*), Section D.1.

In step 1 of the algorithm we need to go through all rectangles in $\mathcal{P}$, which takes $O(\#\mathcal{P})$ steps. In step 2 we need to calculate the distances between reference point μ and the rectangles in $\mathcal{P}_\lambda$, and order these rectangles. Calculation of the distance between μ and a rectangle takes d flops. Thus step 2 requires $O(d \cdot \#\mathcal{P}_\lambda)$ flops. In the worst cases the step 4 of the algorithm requires the pairwise comparison of all rectangles in $\mathcal{P}_\lambda$, to find which rectangles touch each other, which takes $O(d \cdot (\#\mathcal{P}_\lambda)^2)$ steps. Thus the worst case complexity of the algorithm is $O\left(\#\mathcal{P} + d \cdot (\#\mathcal{P}_\lambda)^2\right)$. Typically $\#\mathcal{P}_\lambda$ is very large and the naive version of the algorithm is often not feasible. However, we may enhance the algorithm with the *bounding box technique*.

In the bounding box technique we associate the nodes with the bounding box of the set associated with a node. The bounding box of set $A \subset \mathbf{R}^d$ is the smallest rectangle containing A, such that the sides of the rectangle are parallel to the coordinate axes. In step 4 of algorithm **LeafsFirst** we find which rectangles, associated with the current root nodes, are touched by rectangle R. If rectangle R does not touch the bounding box of those rectangles, then it does not touch any rectangles inside the bounding box. Only if it does touch the bounding box, then we have to travel further towards the leaf nodes and find whether R touches some of the smaller bounding boxes. With the bounding box enhancement the worst case complexity of step 4 is still $O(d \cdot (\#\mathcal{P}_\lambda)^2)$, but with this technique we achieve considerable improvements in typical cases.

Klemelä (2004*a*), Section D.3, contains a pseudo code for the bounding box enhancement of step 4 of algorithm **LeafsFirst**. Klemelä (2004*a*), Section D.4, contains comments helping the practical application of the algorithms.

4.2 Density estimators

We give examples of shape trees for the kernel estimator, CART histogram and bootstrap aggregated estimator.

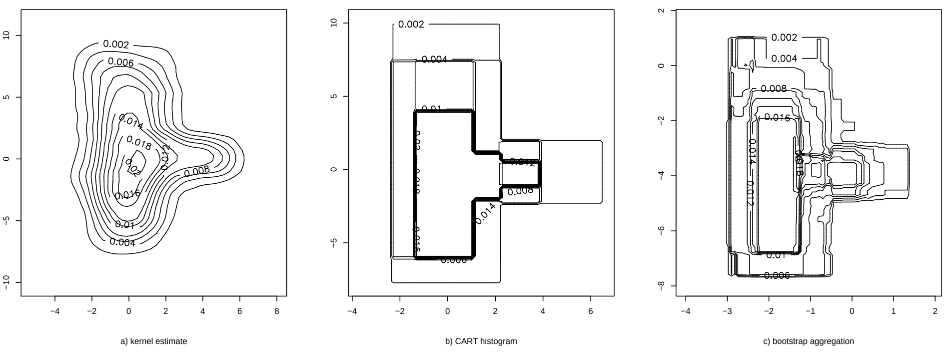


Figure 13: Contour plots of three estimates: a) kernel estimate, b) CART histogram, and c) bootstrap aggregated estimate. We estimate the density of Figure 3a.

4.2.1 Kernel estimator

The kernel estimator, based on sample $X_1, \ldots, X_n \in \mathbf{R}^d$, is defined as

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K((x - X_i)/h), \qquad x \in \mathbf{R}^d, \tag{10}$$

where $h > 0$ is the smoothing parameter and $K : \mathbf{R}^d \to \mathbf{R}$ is the kernel function. We evaluate the kernel estimate on a grid which lies on a rectangle which contains the support of the estimate. Let G be the set of grid points where the estimate is positive. We consider the discretized kernel estimator

$$\bar{f}(x) = \sum_{g \in G} \hat{f}(g) I_{R(g)}(x), \qquad x \in \mathbf{R}^d,$$

where $R(g)$ is the rectangle whose center is g and whose side lengths are equal to the steps of the grid, and $\hat{f}$ is defined in (10).

Figure 13a shows a contour plot of a kernel estimate. We generated a sample of size 500 from the density whose contour plot is shown in Figure 3a. We applied the Bartlett-Epanechnikov product kernel $K(x) = (3/4)^d \Pi_{i=1}^d (1 - x_i^2)_+$, where $(a)_+ = \max\{0, a\}$. The smoothing parameter was $h = 1.8$, and the kernel estimate was discretized with a grid of 32^2 grid points.

Figure 14 shows a radius plot and the corresponding location plot for the kernel estimate of Figure 13a, for the level set of the kernel estimate with level 0.002. The reference point is the origin and we used a grid of 26 radii.

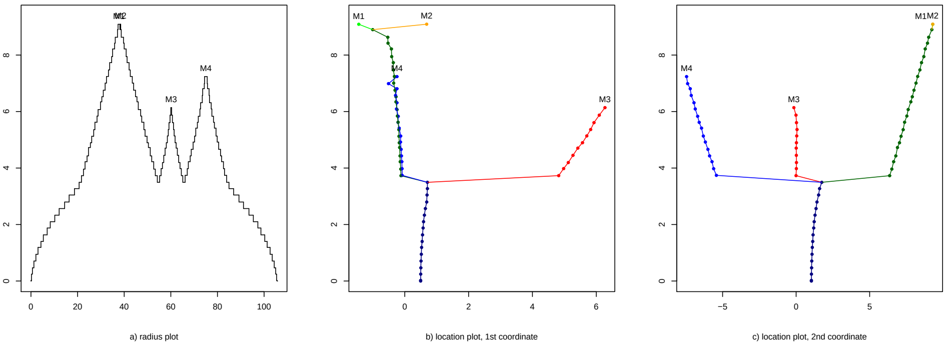


Figure 14: Kernel estimate: a radius plot and location plot for the estimate of Figure 13a.

4.2.2 CART histogram

A histogram is an estimate of the form

$$\hat{f}(x, \mathcal{P}) = \sum_{R \in \mathcal{P}} \frac{n_R}{n \operatorname{vol}(R)} I_R(x), \qquad x \in \mathbf{R}^d, \tag{11}$$

where $\mathcal{P}$ is a finite partition to rectangles of a rectangle estimated to contain the support of the true density, and n_R is the number of observations $X_1, \ldots, X_n$ in R. An CART histogram is a histogram whose partition is chosen in a data-dependent way. Typically the partition is chosen by maximizing the likelihood of the estimate in a greedy fashion, followed by the complexity penalized pruning. For details and references, see Holmström, Hoti and Klemelä (2005).

Figure 13b shows a contour plot of a CART histogram for a sample of size 1000 from the density whose contour plot is shown in Figure 3a. The fine partition was grown so that the cells with less than 5 observations were not splitted, and then the fine partition was pruned to contain 15 rectangles.

Figure 15 shows a radius plot and the corresponding location plot for the CART histogram of Figure 13b, for the level set with level 0.002. The reference point is the origin and we let the estimate to determine the grid of radii.

4.2.3 Bootstrap aggregation

In bootstrap aggregation we generate several bootstrap samples from the original sample, construct an CART histogram based on each sample, and the final estimate is the average of the estimates based on the bootstrap samples. For details, see Holmström et al. (2005).

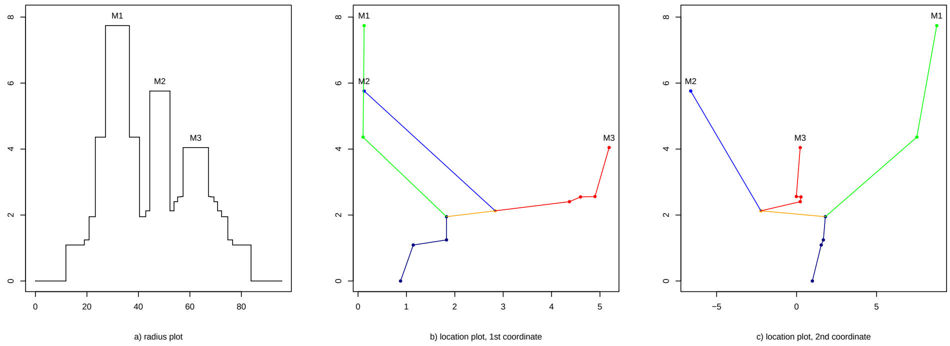


Figure 15: CART histogram: a radius plot and location plot for the estimate of Figure 13b.

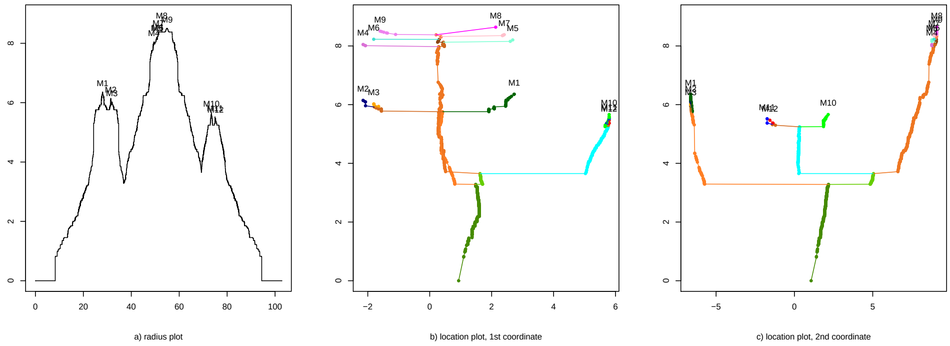


Figure 16: Bagged estimate: a radius plot and location plot for the estimate of Figure 13c.

Figure 13c shows a bagged estimate for a sample of size 1000 from the density whose contour plot is shown in Figure 3a. We took the average of 10 CART histograms, each bootstrap sample was constructed with $n/2$-out-of-n without-replacement bootstrap. CART histograms were grown so that the cells with less than 5 observations were not splitted and the fine partition was pruned to contain 10 rectangles.

Figure 16 shows a radius plot and the corresponding location plot for the bagged estimate of Figure 13c, for the level set with level 0.002. The reference point is the origin and we used a grid of 600 radii.

5 Example: Stock index returns

We study the distribution of the daily returns of stock indeces SP500, DAX30, FTSE100, and Nikkei225. We collected data from the one year time period 4.1.2003-4.1.2004. The data consists of the daily returns, $X_i = (X_{i1}, \ldots, X_{id})$, $X_{ij} = (S_{i,j} - S_{i-1,j})/S_{i-1,j}$, where $S_{i,j}$ is the closing price at the ith day of the jth index, $i = 1, \ldots, n$, $j = 1, \ldots, d$. The number of observations is $n = 231$ and the dimension of the data is $d = 4$.

We assume that the distribution of the returns is changing continuously as a function of time, so that the returns are not identically distributed. We apply the time-localized kernel estimator

$$\hat{f}(x) = h^{-d} \sum_{i=1}^{n} p_{g,i} K((x - X^{(i)})/h), \qquad x \in \mathbf{R}^d, \tag{12}$$

where $X^{(1)}, \ldots, X^{(n)}$ are the observations ordered according to time so that $X^{(i)}$ corresponds to the previous in time observation than $X^{(j)}$ when $i < j$, and the weights are the Gaussian weights

$$p_{g,i} = \frac{q_{g,i}}{\sum_{i=1}^{n} q_{g,i}}, \qquad q_{g,i} = \begin{cases} \exp\{-(n-i)^2/(2g^2)\}, & \text{when } (n-i)/g \le 4 \\ 0, & \text{otherwise} , \end{cases}$$

$i = 1, \ldots, n$. We give more weight to the recent observations and gradually decrease the weights for the more distant observations. Parameter $g > 0$ is the time localization parameter so that when g is small the estimate is effectively based only on recent observations. We choose g reflecting the rate of change of the density as a function of time. The kernel estimator in (10) is a special case of the estimator in (12) when $p_{g,i} = n^{-1}$ for $i = 1, \ldots, n$. For the properties of a time-localized kernel estimator, see Klemelä (2005).

We applied the kernel estimator defined in (12) with smoothing parameter $h = 1.1$, with the Bartlett-Epanechnikov product kernel, and the estimate was discretized with a grid of 16^4 points. We chose the time localization parameter $g = 40$ so that the weights are positive for 161 most recent observations. We normalized the marginal variances to unity. The mode of the estimate is $(0.09, 0.14, -0.09, 0.67)$. A barycenter plot of a level set tree of the estimate indicates that the estimate is skewed in the 4th coordinate (NIKKEI225), see Klemelä (2004a), Section E. Thus we chose the barycenters to be the reference points of the shape trees.

Figure 17 shows 3 radius plots with the barycenters as the reference points: Frame a) shows the level set with level $0.01 \cdot \|\hat{f}\|_\infty \approx 0.0004$, with barycenter $(0.10, 0.04, 0.08, -0.11)$ as the reference point. Frame b) shows

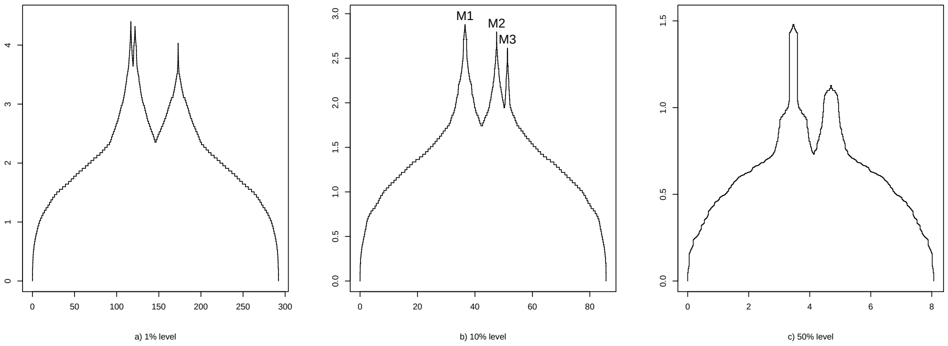


Figure 17: Radius plots of 3 level sets of a kernel estimate from the stock index return data.

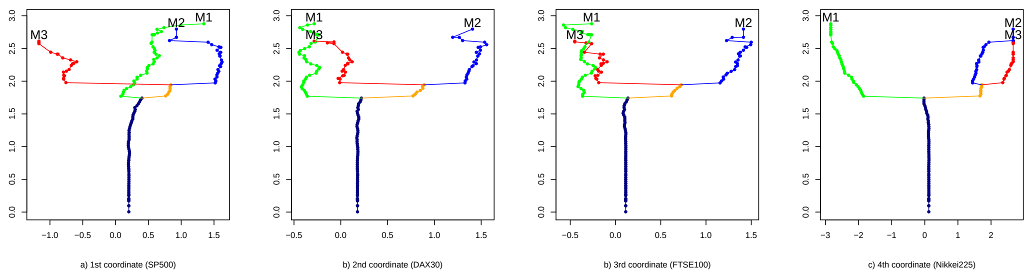


Figure 18: A location plot for the 10% level set, corresponding to the radius plot in Figure 17b.

level $0.1 \cdot \|\hat{f}\|_\infty \approx 0.004$, with barycenter $(0.20, 0.18, 0.11, 0.12)$ as the reference point. Frame c) shows level $0.5 \cdot \|\hat{f}\|_\infty \approx 0.021$, with barycenter $(0.14, 0.11, 0.06, 0.34)$ as the reference point.

Figure 18 shows the location plot for the 10% level set, corresponding to the radius plot in Figure 17b.

One expects that the index returns are correlated and that the level sets of the density have ellipsoidal shape, extending from the negative orthant to the positive orthant. The ellipsoidal shape of the level sets would imply two modes for the radius plot. The bi-modal shape of the level sets was conformed by the radius plots in Figures 17a and c. Note however that these figures reveal a certain egg-shapedness of the level sets.

An additional deviation from the ellipsoidal shape is seen from Figure 17b. Figure 17b shows that the radius plot for the 10% level set has three apparently non-negligible modes. The mode labeled as M3 corresponds to negative returns for the SP500 but to positive returns for Nikkei225, as can be seen from Figure 18a, and Figure 18d, respectively. Klemelä (2004*a*), Section E, shows a 2D slice for variables SP500 and Nikkei225.

The location plot in Figure 18 shows that the mode M1 corresponds to negative returns and the mode M2 corresponds to positive returns. Mode M1 is located at $(1.35, -0.28, -0.26, -2.84)$ and mode M2 is located at $(0.93, 1.41, 1.41, 2.67)$, as can be seen from Figure 18.

In summary, the radius plots indicate that the distribution of the returns is not multivariate Gaussian, for any covariance matrix, due to the egg-shapedness and the 3-modality of the level sets.

6 Discussion

Shape tree technique. With the shape tree technique we may visualize at one step certain important features of the density. With a shape tree we visualize a single level set of the density. It is often enough to visualize only few level sets to get an accurate impression from the shape of the density.

Shape trees visualize connected sets: the density is assumed to be unimodal, or else we restrict ourselves to some connected part of a level set, or to lower levels. We may apply the following steps in exploratory data analysis.

1. Construct non-parametric density estimates from the data. One should calculate several estimates corresponding to a scale of smoothing parameters.

2. Apply the level set tree based techniques (the volume plot and barycenter plot), as introduced in Klemelä (2004*b*), to visualize the density estimates.

   (a) Make inference on the multimodality of the density. At this step we may also choose the values of the smoothing parameters of the estimator.

   (b) Make inference on the skewness and the kurtosis of the density. If the density is concluded to be multimodal we study the skewness and the kurtosis of the density only on a region where the level sets are connected.

3. Apply the shape tree based techniques to visualize level sets of a density estimate.

   (a) If we have concluded that the density is unimodal, then we may apply shape trees to visualize the complete collection of level sets of the estimate.

(b) If the density is concluded to be multimodal, we may restrict ourselves to a lower level set, or to a single mode (single connected component of a level set), and apply shape trees to make inference on the shape of the density restricted to this set.

We have presented the shape plots and the location plot for the visualization of a shape tree. The shape plots include the radius plot and the probability content plot. The shape plots show "modes" of the level set. Modes of the set are the tail regions of the set which do not fit inside a ball centered at a reference point. In addition, the shape plots visualize various quantitative information concerning the tail regions of the level set. The location plot visualizes the locations of the tail regions of the level set, showing the barycenters of the tail regions. The location plot visualizes also a "delineator of the set".

Density estimators. We have considered three types of estimates: kernel estimates, CART histograms, and bootstrap aggregated estimates. The kernel estimator suffers from the computational complexity and the curse of dimensionality. In some moderate dimensional cases (d=3,4) we may succeed in calculating shape trees from kernel estimates. CART histograms are computationally attractive and somewhat resistant to the curse of dimensionality. However, level sets of CART histograms are often quite inaccurate estimates of the level sets of the true density. We may increase statistical accuracy of CART histograms with bootstrap aggregation, with some cost of computational complexity.

Acknowledgments

I wish to thank Lasse Holmström, Enno Mammen, and Günther Sawitzki for helpful comments.

References

Asimov, D. (1985), 'The grand tour: a tool for viewing multidimensional data', *SIAM J. of Scient. and Statist. Comp.* **6**, 128–143.

Azzalini, A. and Dalla Valle, A. (1996), 'The multivariate skew normal distribution', *Biometrika* **86**, 715–726.

Clayton, D. G. (1978), 'A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence', *Biometrika* **65**, 141–151.

Cook, D., Buja, A. and Cabrera, J. (1993), 'Projection pursuit indexes based on orthonormal function expansions', *J. Comput. Graph. Statist.* **2**(3), 225–250.

Cook, D., Buja, A., Cabrera, J. and Hurley, C. (1995), 'Grand tour and projection pursuit', *J. Comput. Graph. Statist.* **4**(3), 155–172.

Furnas, G. W. and Buja, A. (1994), 'Prosection views: Dimensional inference through sections and projections (with discussion)', *J. Comput. Graph. Statist.* **3**(4), 323–385.

Härdle, W. and Scott, D. (1992), 'Smoothing by weighted averaging of rounded points', *Comput. Statist.* **7**, 97–128.

Holmström, L., Hoti, F. and Klemelä, J. (2005), 'Flexible multivariate histograms and level set tree plots'. Manuscript.

Klemelä, J. (2004*a*), 'Appendix of the shape tree article'. http://www.denstruct.net.

Klemelä, J. (2004*b*), 'Visualization of multivariate density estimates with level set trees', *J. Comput. Graph. Statist.* **13**(3), 599–620.

Klemelä, J. (2005), 'Density estimation with locally identically distributed data and with locally stationary data'. Submitted, http://www.rni.helsinki.fi/~jsk/ps/locsta.ps.

Scott, D. W. (1992), *Multivariate Density Estimation*, Wiley.

Tukey, J. (1977), *Exploratory Data Analysis*, Addison-Wesley.

Wegman, E. J. and Luo, Q. (2002), 'On methods of computer graphics for visualizing densities', *J. Comput. Graph. Statist.* **11**(1), 137–162.

A Ordered shape trees

To define shape transforms we need to define a shape tree as an ordered tree: the children of a given parent has to be ordered. We do not include a definition of the ordering in the proper definition of a shape tree given in Definition 1 for two reasons: (1) there are several almost equally natural ways to define an ordering for the siblings, and (2) with such an ordering we may express less useful information than otherwise with the shape trees.

We give an ordering rule based on the barycenters of the separated components associated with the nodes of a shape tree. The barycenter of a set is defined in (7).

Definition 10 Origin based barycenter ordering *of a set of siblings of the shape tree is given by the following rule.*

1. *The first node among the siblings is the one with the largest Euclidean distance of the barycenter from the origin.*
2. *After that, the sibling nodes are ordered according to the distance of their barycenters from the barycenter of the first node; the second node is the one with the closest Euclidean distance of the barycenter from the barycenter of the first node. We continue ordering by finding the distance to the first node.*

In practise shape trees are such that each node has either 0, 1, or 2 nodes. Only when we use few levels to build a shape tree, it may happen that some node has 3 or more children. Thus, to define an ordering we need to concentrate on finding the first sibling: the rule for ordering the remaining siblings is not important. Thus item 2 of Definition 10 is less important than item 1.

We may express useful information with the ordering if the ordering of the leaf nodes reflects somehow the spatial location of the sets associated with the leaf nodes. However, a tree may be ordered only by starting from the root nodes. To partially solve this problem we associate every node m with a vector $\mathrm{da}(m)$, which is chosen in a certain way by choosing a barycenter from all the barycenters of the descendants of the node. When we order the siblings with the help of vectors $\mathrm{da}(m)$, then the ordering takes better into account the location of the sets associated with the leaf nodes.

Definition 11 Origin based descendant-barycenter ordering *of a set of siblings of the shape tree is given by the following rule. Annotate node m with* *that barycenter* $da(m)$ *of its descendants which has the largest Euclidean distance from the origin:*

$$\mathrm{da}(m) = \mu \in \mathbf{R}^d$$

when μ *is such that*

$$\|\mu\| = \max\left\{\|\mathrm{barycenter}(A)\| : \textit{ set } A \textit{ is associated with a descendant of } m\right\}.$$

1. *Node* m *is the first node among the siblings if* $\|\mathrm{da}(m)\|$ *is larger than* $\|\mathrm{da}(m')\|$ *for other siblings* m'*.*

2. *Node* m' *is the second node if* $\|\mathrm{da}(m')-\mathrm{da}(m)\|$ *is smaller than* $\|\mathrm{da}(m'')-\mathrm{da}(m)\|$ *for other siblings* m''*, where* m *is the first node. We continue ordering by finding the distance to the first node.*

Again, in Definition 11 item 1 is the important part. Note that Definition 11 would be equivalent to Definition 10 if we would define $\mathrm{da}(m) = \mathrm{barycenter}(A)$, where A is the set associated with node m. The origin based descendant-barycenter ordering is not essentially more complex computationally than the origin based barycenter ordering, when one calculates the vectors $\mathrm{da}(m)$ for each node m at the same time when the tree is grown.

We have used the ordering rule of Definition 11 in this article.

B Shape plots

B.1 Proof and illustration of Theorem 1

Proof of Theorem 1. We say that a node is a descendant of a node m if it is a children of m or a children of other descendant of m. We have that

$$\begin{aligned}
\int_{C_{m_0}} f &= \sum\Big\{P_f(C_m) - \sum\{P_f(C_{m'}) : m' \text{ is a child of } m\} : \\
&\qquad\qquad m \text{ is a descendant of } m_0 \text{ or } m = m_0\Big\} && (13)\\
&= \sum\Big\{\mathrm{vol}(C_m)\Big(H_m - H_{\mathrm{parent}(m)}\Big) : \\
&\qquad\qquad m \text{ is a descendant of } m_0 \text{ or } m = m_0\Big\} && (14)\\
&= \int_{\mathcal{I}_{m_0}} \Big(g - H_{\mathrm{parent}(m_0)}\Big). && (15)
\end{aligned}$$

□

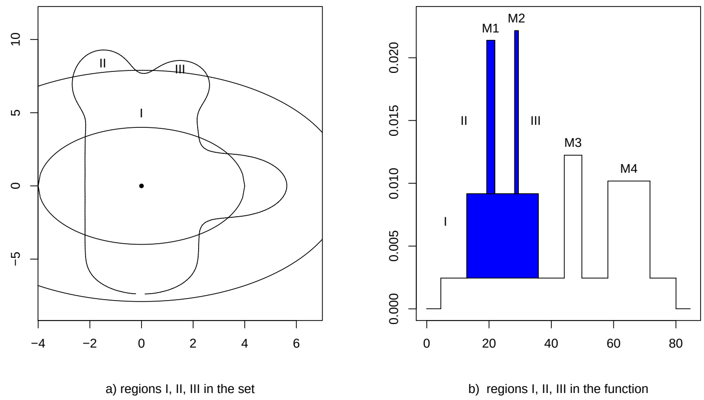


Figure 19: Steps of the proof of Theorem 1.

Illustration of Theorem 1. Figure 19 illustrates Theorem 1. Node m_0 is the node associated with set C in Figure 1c. Step (13) in the proof of Theorem 1 follows from Figure 19a: $P_f(C_{m_0})$ is equal to the sum of the probabilities of regions I, II, and III. Step (14) follows from the definition of the height of the nodes given in (6). Step (15) follows from Figure 19b: excess mass of function g in interval $\mathcal{I}_{m_0}$ is equal to the the volume of the colored region, and this volume is equal to the sum of the volumes of the three rectangles whose union is forming the colored region.

B.2 Volume plot of a star shaped set.

To define the *volume plot* we modify the definition of the radius plot by modifying the definition of the length associated with the nodes. We take the length to be equal to the volume of the pre-image of the (outer) part of the boundary of the set associated with the node, in terms of the polar coordinate representation of the boundary. The heigth of a node is taken to be its radius, like in the radius plot.

The volume plot is defined only for the star shaped sets. Set $A \subset \mathbf{R}^d$, $d \geq 2$, is called star shaped when there exists reference point $\mu \in A$ and boundary function $g : \Theta \to [0, \infty)$ so that

$$A = \{\mu + r\Psi(\theta) : r \in [0, g(\theta)],\ \ \theta \in \Theta\},$$

where $\Theta = [0, 2\pi] \times [0, \pi]^{d-2}$,

$$\Psi(\theta) = (\psi_1(\theta), \ldots, \psi_d(\theta)), \qquad \psi_i(\theta) = \cos\theta_{i-1} \sin\theta_i \cdots \sin\theta_{d-1},$$

$\cos\theta_0 = 1$, $\theta = (\theta_1, \ldots, \theta_{d-1})$. For example, when $d = 2$, then $\Psi(\theta_1) = (\sin\theta_1, \cos\theta_1)$, $\theta_1 \in [0, 2\pi]$.

Definition 12 *A* volume plot *is a plot of a volume transform. A* volume transform *of a star shaped set A is the 1D function generated by a shape tree of A when we choose the height and the length in Definition 3 in the following way.*

1. *The height associated to a node of the shape tree is equal to the radius associated with the node.*
2. *The length associated to node m of the shape tree is equal to*

$$\mathrm{vol}_{d-1}\left(\{\theta \in \Theta : \mu + r\Psi(\theta) \in A_m \textit{ for some } r \in [0, g(\theta)]\}\right)$$

where A_m is the set associated to node m.

Note that the visualization of a star shaped set is equivalent to the visualization of the boundary function of the set.

1. In the two dimensional case, when $d = 2$, the volume plot is closely related to the plot of the boundary function. The boundary function g is equal to a scaled limit of volume transforms, when we let the step size of the grid of radii vanish, and define the ordering of nodes of the shape tree of A suitably, possibly in a different way than we did in Section A.

   For example, the boundary function of a 2D ball, centered at the origin, when the reference point is at the origin, is $RI_{[0,2\pi]}$ where R is the radius of the ball. On the other hand, the limit of volume transforms is $RI_{[0,2\pi R]}$. Note that a radius transform of a 2D ball is drawn in Figure 9b. This radius transform is a unimodal function but it is not an indicator of any interval. The analytical expression for the limit of radius transforms of a 2D ball is given in (18).

2. In the multivariate case, when $d \geq 2$, the volume plot of A is equal to the volume plot of a piecewise constant approximation of the boundary function of A, where the volume plot of a piecewise constant function was defined in Klemelä (2004*b*).

The volume plot is defined only for the star shaped sets. The other shape plots are defined for (some) connected sets. However, in order shape plots to give easily interpretable information, we might have to assume that the sets are close to being star shaped.

Even when we would not apply the volume plot in practise, Definition 12 of the volume plot shows the connection to the boundary function in the 2 dimensional case, and it is shows the connection to the volume plot as defined in Klemelä (2004*b*).

C Interpretation of the graphics

C.1 Shape of the generating sets.

We may replace the sequence of balls $(B_r(\mu), r \in \mathcal{R})$ in Definition 1 of a shape tree with some other sequence of sets. We may consider sequences $(S_r(\mu), r \in \mathcal{R})$, where $S_r(\mu) = \{x \in \mathbf{R}^d : \|x - \mu\|_{alt} \le r\}$ is a ball with some other norm than the Euclidean norm. For example, we may take $\|x\|_{alt} = \max_{i=1,\ldots,d} |x_i|$, or $\|x\|_{alt}^2 = x^T Ax$, for some matrix A, so that $S_r(\mu)$ is a rectangle or an ellips. We call the sets in sequence $(S_r(\mu), r \in \mathcal{R})$, the *generating sets* of the shape tree.

The shape transforms of $A \subset \mathbf{R}^d$ are unimodal when $A = S_r(\mu)$ for some $r \in \mathcal{R}$, and $(S_r(\mu), r \in \mathcal{R})$ are the generating sets of the shape tree. Thus we may test the hypothesis of a certain shape of the level set by constructing a shape tree with a sequence of generating sets containing this shape. If the shape transforms are multimodal we reject the hypothesis.

There are at least two reasons to prefer the Euclidean distance: (1) the balls are the only symmetric sets, with shape trees we visualize in a sense deviations from the basic shape determined by the norm, and it is more easy to understand deviations from a symmetric shape than from an asymmetric shape, (2) the hypothesis of standard Gaussianity is justified in many applications.

C.2 Level sets and the shape of the density

Figure 20 shows volume plots of level set trees for different dimensions of the Bartlett density. Figure 21 shows volume plots of level set trees of Gaussian densities for different dimensions. We define the multivariate Bartlett density as

$$f(x) = C\kappa^d \left(1 - \|\kappa x\|^2\right)_+$$

where $(x)_+ = \max\{0, x\}$, $\kappa > 0$, C is the normalization constant,

$$C = \frac{d(d+2)}{2\mu(S_d)} \tag{16}$$

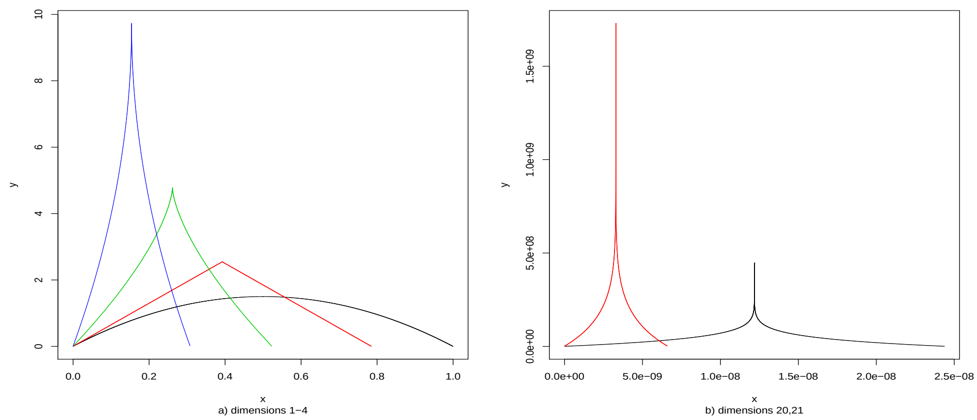


Figure 20: Volume plots of level set trees of Bartlett densities with $\kappa = 2$ for dimensions $1 - 4$, 20, and 21.

and $\mu(S_d)$ is the volume of the unit sphere $S_d = \{x \in \mathbf{R}^d : \|x\| = 1\}$:

$$\mu(S_d) = 2\pi^{d/2}/\Gamma(d/2). \tag{17}$$

C.3 Unimodality

We may find an analytical expression for the radius transform of a ball. The radius transform of a ball with radius R is

$$g(x) = \begin{cases} \left(R^2 - (2/\pi)(x - (\pi/2)R^2)\right)^{1/2}, & 0 \le x \le (\pi/2)R^2 \\ \left(R^2 + (2/\pi)(x - (\pi/2)R^2)\right)^{1/2}, & (\pi/2)R^2 \le x \le \pi R^2. \end{cases} \tag{18}$$

In Figure 9b we have drawn the radius transform by approximating the ball with a union of rectangles.

C.4 Multivariate skew normal density

C.4.1 The definition

The multivariate skew normal density is

$$2f_{\mu,\Sigma}(x)\Phi(((x_1 - \mu_1)/\sigma_1, \ldots, (x_d - \mu_d)/\sigma_d) \cdot \alpha),$$

where $f_{\mu,\Sigma}$ is the Gaussian density with the expectation $\mu = (\mu_1, \ldots \mu_d)$ and covariance matrix Σ, whose diagonal is $(\sigma_1^2, \ldots, \sigma_d^2)$, Φ is the distribution function of the standard Gaussian density, and $\alpha \in \mathbf{R}^d$ is the skewness parameter, see Azzalini and Dalla Valle (1996).

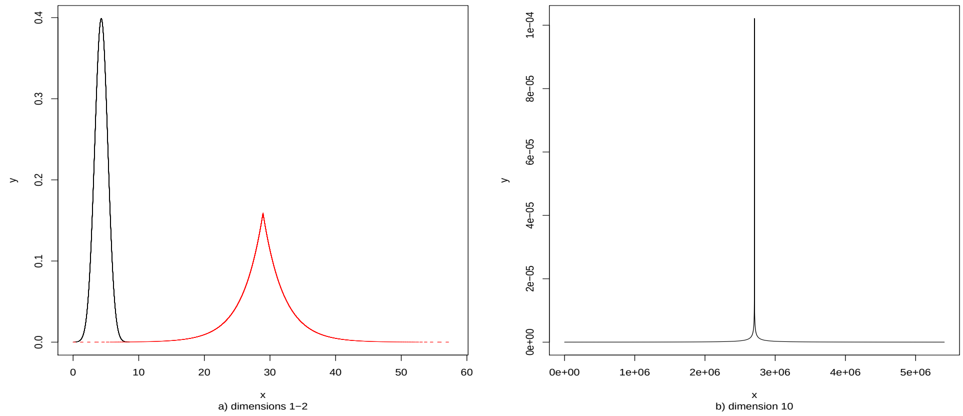


Figure 21: Volume plots of level set trees of standard Gaussian densities for dimensions $1-2$ and 10.

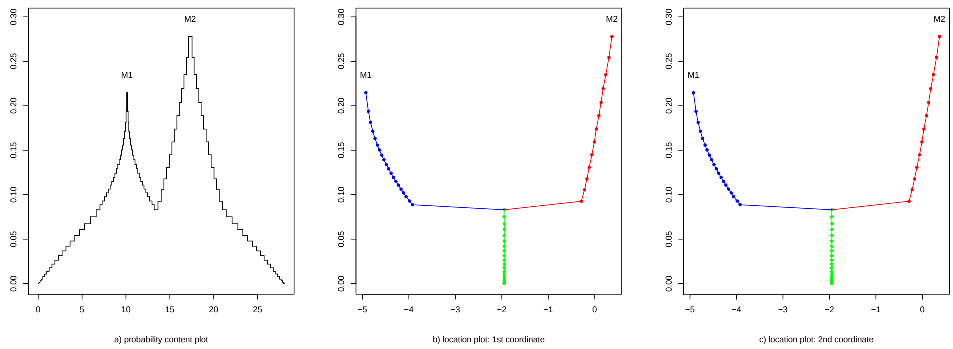


Figure 22: Skewed Gaussian density; a probability content plot and a location plot for the 0.02 level set, when the barycenter is the reference point.

C.4.2 The parameters of Figure 8

In Figure 7 we chose $\mu = (0,0)$, $(\sigma_1, \sigma_2) = (3,1)$, $\alpha = (6,0)$, and then this density was rotated 135 degrees in the clockwork direction.

C.4.3 The location plot corresponding to the probability content plot in Figure 8d

The location plot corresponding to the probability content plot of Figure 8d is shown in Figure 22, frames b and d. The probability content plot itself is shown again in Figure 22a.

C.5 Clayton family

C.5.1 The definition

Clayton density $f_{c,\theta} : \mathbf{R}^2 \to \mathbf{R}$, with standard Gaussian marginals, with parameter $\theta > 0$, is defined as

$$\begin{aligned} & f_{c,\theta}(x_1, x_2) \\ &= (1+\theta)\left[\Phi(x_1)^{-\theta} + \Phi(x_2)^{-\theta} - 1\right]^{-2-1/\theta} (\Phi(x_1)\Phi(x_2))^{-\theta-1}\phi(x_1)\phi(x_2), \end{aligned}$$

where $(x_1, x_2) \in \mathbf{R}^2$, Φ is the standard Gaussian distribution function and ϕ is the standard Gaussian density function.

C.5.2 The parameters of Figure 10

Figure 10a shows the contour plots of the 10% level sets, with $\theta = 1$, $\theta = 2$, and $\theta = 4$, and frames b, c, and d show the corresponding radius plots. When the parameter θ increases, then the dependence between coordinate variables increases.

C.6 Densities with Student copula

C.6.1 Definition

The density of the student copula, with parameters ν and $\rho \in (-1, 1)$, with standard Gaussian margins, is defined by

$$\begin{aligned} f_{s,\nu,\rho}(x_1, x_2) &= \frac{\Gamma((\nu+2)/2)}{\Gamma(\nu/2)\nu\pi\sqrt{1-\rho^2}} \frac{\phi(x_1)\phi(x_2)}{t_\nu(z(x_1))t_\nu(z(x_2))} \\ &\times \left(1 + \frac{z(x_1)^2 + z(x_2)^2 + 2\rho z(x_1)z(x_2)}{\nu(1-\rho^2)}\right)^{-(\nu+2)/2}, \end{aligned}$$

where $(x_1, x_2) \in \mathbf{R}^2$, $z(u) = T_\nu^{-1}(\Phi(u))$, Φ is the standard Gaussian distribution function, ϕ is the standard Gaussian density function, T_ν^{-1} is the inverse of the distribution function of the Student distribution with degrees of freedom ν, and t_ν is the density of the Student distribution with degrees of freedom ν.

C.6.2 The parameters of Figure 11

In Figure 11 we have $\nu = 3$ and $\rho = 0.3$.

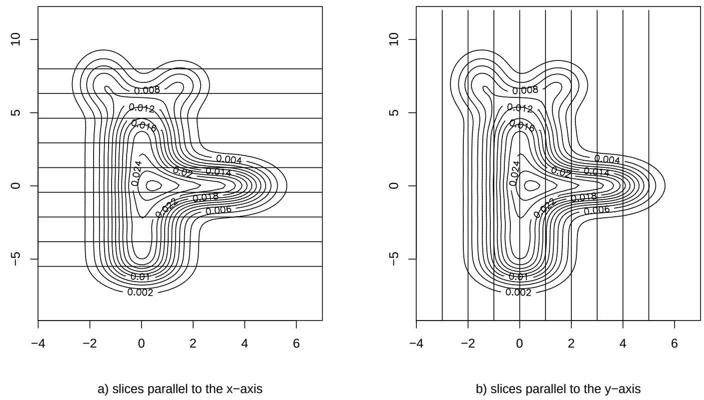


Figure 23: The positions of the slices.

C.7 Marginal densities and slices

We draw slices of the 2D density shown in Figure 3a. Note that it is the simplest case to use 1D marginal densities and slices to make inference on the shape a two dimensional density and it is essentially more difficult to make inference from, say a 5-dimensional density using 2-dimensional marginal densities.

Slices along the coordinate axes. Figure 23 shows a contour plot of the density and the positions of the slices. Figure 24 shows the slices parallel to the x-axis and Figure 25 shows the slices parallel to the y-axis.

Rotated slices. To rotate the slices is equivalent to rotating the density. We will rotate the density 45 degrees and look at the slices parallel to coordinate axes. Figure 26 shows the positions of the slices. Figure 27 shows the slices parallel to the x-axis and Figure 28 shows the slices parallel to the y-axis.

D Algorithms

D.1 Illustration of the LeafsFirst algorithm

Figure 29 illustrates the algorithm. We have approximated the ball with a union of 30^2 rectangles. Figure 29a shows colored the rectangle which is furthest away from the origin, and will be the rectangle associating the

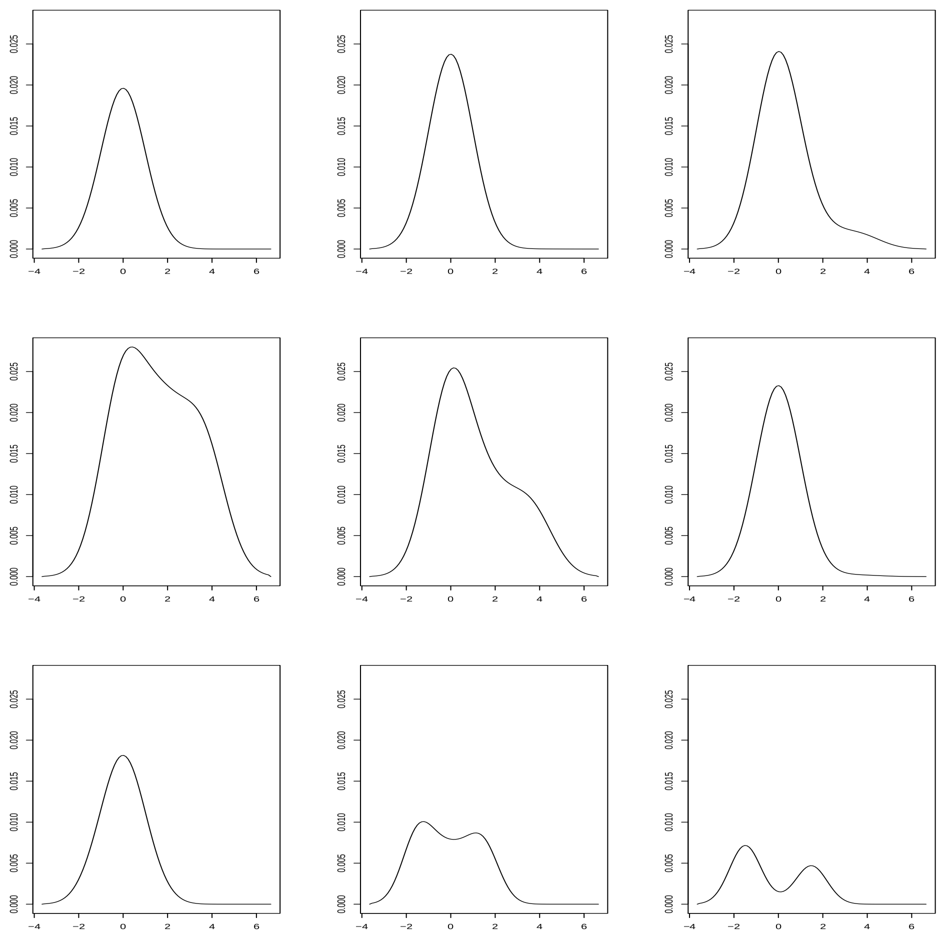


Figure 24: Slices parallel to the x-axis.

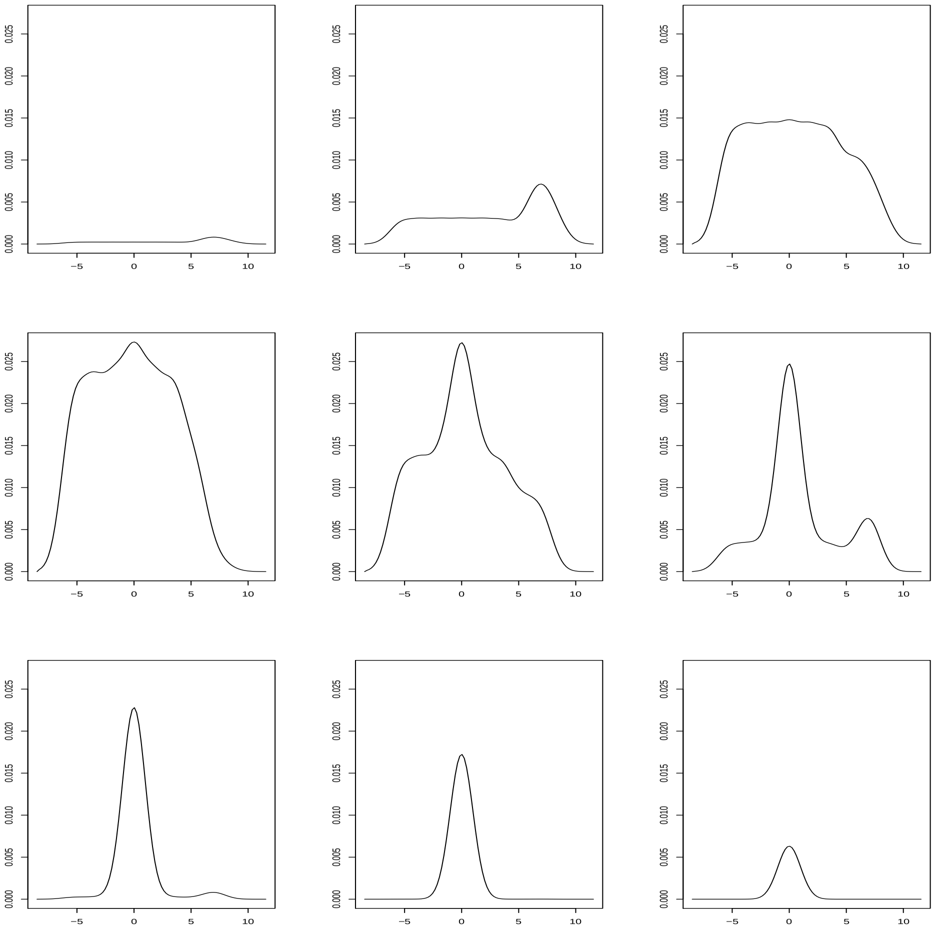


Figure 25: Slices parallel to the y-axis.

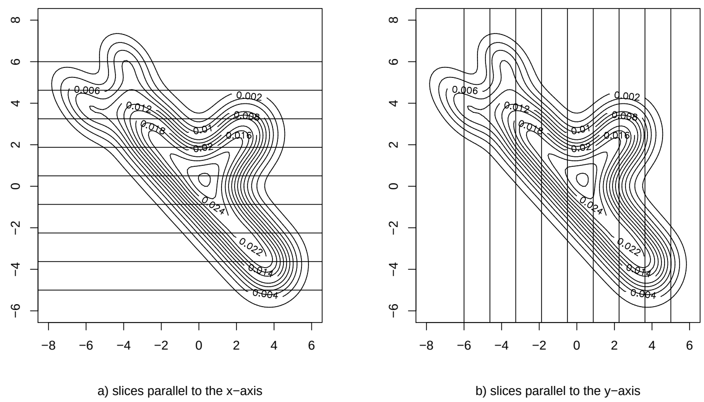


Figure 26: The positions of the slices.

first leaf node. Figure 29b shows a discretization effect: the "true shape tree", which is the limit of the approximations when we let the number of rectangles grow and the step in the grid of radii become smaller, would have 1 leaf. However, for the discretized set we have some spurious leafs. Figure 29c shows the phase just before we have joined all the branches of the tree. Figure 29d shows the phase where we have almost reached the root node.

Figure 30 shows the radius plot and the corresponding location plot of the shape tree which is grown with the steps shown in Figure 29.

D.2 Grid of radii

We have shown in Section D.1 that shape trees may contain spurious modes. The method and accuracy of the approximation of the set affects the number of spurious modes. The second factor affecting the number of spurious modes is the fineness of the grid of radii.

Figure 31 shows the effect of the discretization in the case of a ball. We have approximated the ball with 120^2 rectangles. Frame a) shows the radius plot with a grid of 10 radii. The radius transform is unimodal. Frame b) shows the radius plot with a grid of 29 radii, and the radius transform is still unimodal. Frame c) shows the radius plot with a grid of 30 radii and the radius transform has now 8 modes.

One notes that the "spurious" modes do not have disturbing effects on the radius plot. However, the location plot looks messy with the large number of spurious modes. Figure 31d shows the location plot for the 1st coordinate corresponding to the radius plot of Figure 31c.

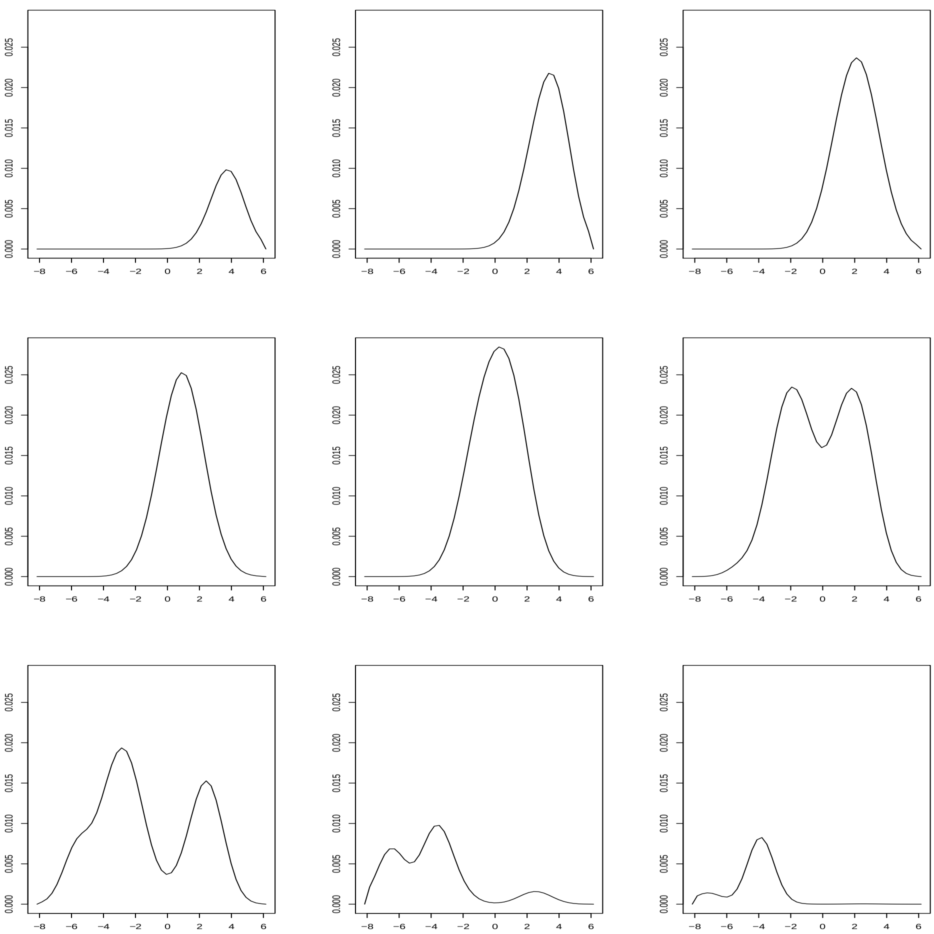


Figure 27: Slices parallel to the x-axis.

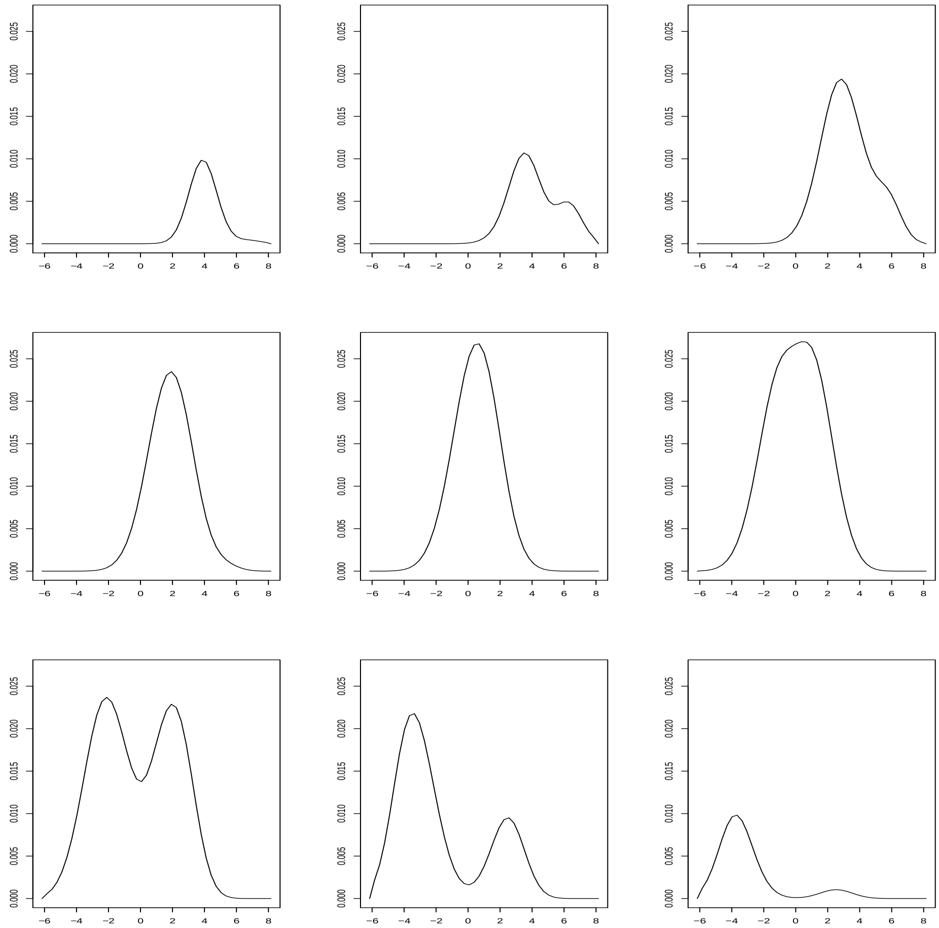


Figure 28: Slices parallel to the y-axis.

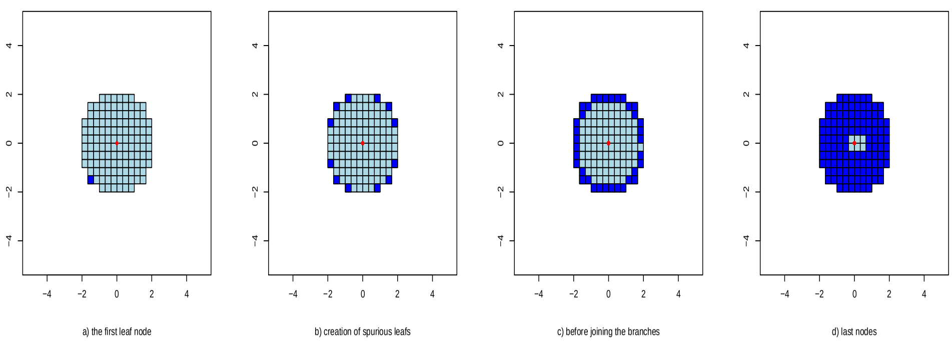


Figure 29: Illustration of the algorithm **LeafsFirst**.

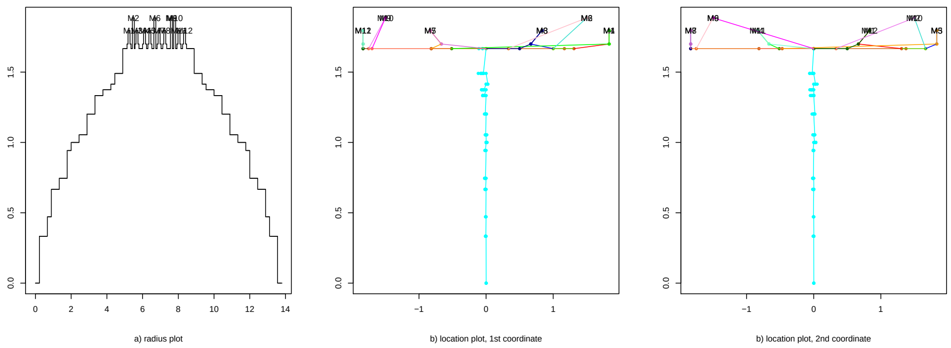


Figure 30: The radius plot and the location plot corresponding to the shape tree grown in Figure 29.

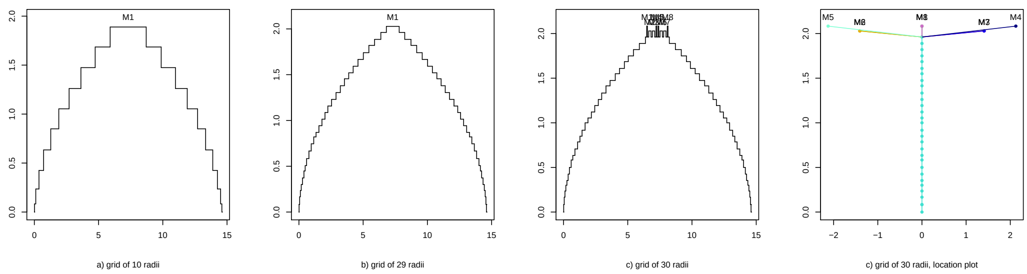


Figure 31: Radius plots with different grid of radii. The radius plots visualize the sphere.

D.3 The bounding box enhancement of step 4 of algorithm LeafsFirst

We present a pseudo code for enhancing step 4 of the algorithm **LeafsFirst**.

1. **Input** is a rectangle R and the collection of the current root nodes of the shape tree (we are building the shape tree starting from the leaf nodes and thus the current root nodes of the unfinished shape tree are the nodes without parent).

2. **Output** is the list of those current root nodes which are touched by rectangle R (list of those current root nodes whose associated set is touched by rectangle R).

3. An additional **internal data structure** of the algorithm are the annotations of each node of the (unfinished) shape tree with the bounding box of the set associated with the node.

ALGORITHM **Bounding box enhancement**
(of step 4 of algorithm LeafsFirst)

1. **answer=emptyList** (at the beginning we assume that there are no touches);

2. **loop 1**: **go** through the current root nodes, assume we have encountered current root node m_r;

3. **loop 2**: **go** through the nodes of the tree whose root is m_r, starting with node m_r, in such a way that the parent is always encountered before children;

4. consider node m associated with rectangle R_m and bounding box $\mathcal{B}_m$;

   (a) **if** rectangle R does not touch bounding box $\mathcal{B}_m$, **then** (conclude that R does not touch the set associated with node m, and thus does not touch the set associated with m_r), **goto loop 1**;

   (b) **else** (R touches bounding box $\mathcal{B}_m$) **if** R touches the rectangle R_m, **then** (conclude that R touches the set associated with node m, and thus touches the set associated with node m_r) **concatenate** m_r to the **answer**, **goto loop 1**;

   (c) **else** (R touches the bounding box, but does not touch rectangle R_m) **continue loop 2**;

5. **return answer**.

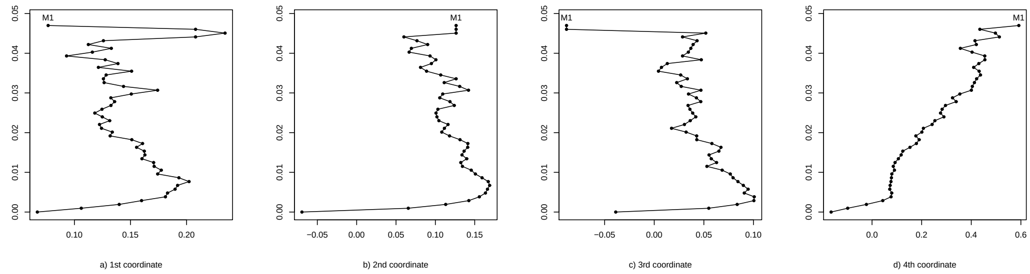


Figure 32: Barycenter plot of a level set tree of the estimate of the stock index return data. Frame a) shows SP500, frame b) DAX30, frame c) FTSE100, and frame d) Nikkei225.

D.4 Practical application of the algorithms

We make two remarks concerning the practical application of the algorithms.

1. Sometimes density estimates have minor modes, for example in the tail regions of the density, and we may want to apply the shape trees to visualize these estimates without the prior clustering of the level set to connected components. When we apply algorithm **LeafsFirst** to a set which is not connected, then the output of the algorithm will be a tree with several root nodes. The additional root nodes do not typically have any major disturbing effects on the visualizations.

2. A reasonable approach for calculating a shape tree is to first use **LeafsFirst** algorithm to calculate a shape tree with the finest grid of radii, given by the rule (9). After that one prunes the shape tree, creating a smaller shape tree with a sparser grid of radii. Pruning the tree may give two advantages: (1) we get rid of some of the spurious modes, and (2) the size of the data structures representing the shape tree may be reduced considerably without a major decrease in the quality of the plots. One should try several grid of radii and be careful not to delete some true modes by using too sparse grid of radii.

E Example:Stock index data, 2D slice

E.1 A barycenter plot of a level set tree of the estimate

Figure 32 shows a barycenter plot of a level set tree of the estimate. One may note that the estimate is skewed in the 4th coordinate (Nikkei225).

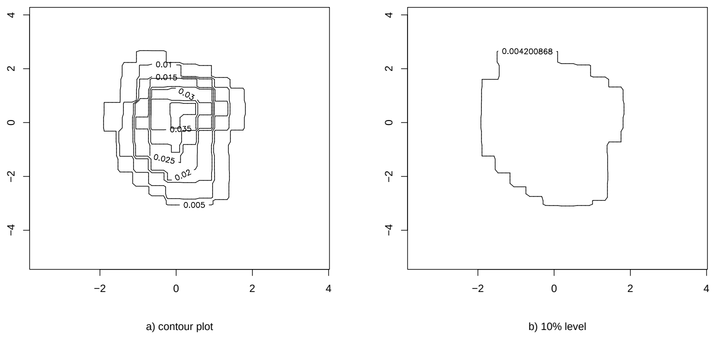


Figure 33: Frame a shows a contour plot of a 2D slice (SP500 and Nikkei225), and frame b shows the 10% level set of this 2D slice.

E.2 A slice of the estimate

Figure 33a shows the 2D slice $g(x_1, x_4) = \hat{f}(x_1, 0.13, -0.09, x_4)$ corresponding to SP500 and Nikkei225. Figure 33b shows the 10% level set of the slice. We see that the probability mass is spread widely over the positive region of Nikkei225.

F Further discussion

Raw data vs. smoothing. A large part of the statistical visualization literature discusses the visualization of the raw data (scatter plots). We discussed the visualization of density estimates: we first smooth the data and then visualize the density estimate. Multivariate functions are much more complex objects than data matrices. However, we claim that with the help of level set trees and shape trees we have efficient visualization tools to visualize multivariate density estimates and thus to make inference on the shape of multivariate densities.

Decoupling the shape information and the spatial information. Two dimensional star shaped sets may be visualized with 1D boundary functions. It is a natural question how to generalize this visualization to higher dimensional cases. We have solved this problem by decoupling the shape information and the spatial information. The shape plots show the pure shape information and the location plot nails down the shape at the certain locations.

Visualization of sets. We have visualized multivariate functions by visualizing their level sets. However, there are also many situations where we want to visualize sets which are not level sets of any function. For example, we may want to visualize the classification sets of a supervised classification procedure (discriminant analysis). Connected sets may be visualized with the radius plot and with the corresponding location plot.

Hypothesis testing. Statistical inference may proceed by formulating a null hypothesis, and designing a formal testing procedure; finding a test statistic whose distribution is known under the null hypothesis. For example, one may take as the null hypothesis the assumption of unimodality, or the assumption of Gaussianity. The graphical tools based on shape trees may accompany such statistical tests. For example, when we reject the hypothesis of Gaussianity, the result of the test does not give information how the hypothesis is violated. Visualization of density estimates gives us clues how the true density is differing from a Gaussian density.

Parsimonious tools. Given a multivariate density we might be interested whether the the variation of the function is equally large in every direction; whether the tails of the function are equally fat in every direction. In the d-dimensional Euclidean space there are 2^d directions where the tails of the density may be extending, and it is not feasible to study every direction separately. However, when the level sets of the density are given as unions of rectangles, then algorithms for calculating shape trees give a parsimonious way to find and visualize the tail behavior of the density.