# On the number of modes of finite mixtures of elliptical distributions

**Grigory Alexandrovich[1], Hajo Holzmann[1] and Surajit Ray [2].**

[1] *Department of Mathematics and Computer Science, Marburg University, Germany,*
[2] *Department of Mathematics and Statistics, Boston University, USA*

**Abstract**

We extend the concept of the ridgeline from Ray and Lindsay (2005) to finite mixtures of general elliptical densities with possibly distinct density generators in each component. This can be used to obtain bounds for the number of modes of two-component mixtures of $t$ distributions in any dimension. In case of proportional dispersion matrices, these have at most three modes, while for equal degrees of freedom and equal dispersion matrices, the number of modes is at most two. We also give numerical illustrations and indicate applications to clustering and hypothesis testing.

*Key words:* finite mixtures, number of modes, elliptical distributions, $t$ distribution

## 1 Introduction

Finite mixtures are a popular tool for modeling heterogenous populations. In particular, multivariate finite mixtures are often used in cluster analysis, see e.g. McLachlan and Peel (2000). Here, analysis is mainly based on mixtures with multivariate normal components. However, mixtures of multivariate $t$-distributions offer an attractive, more flexible and more robust alternative, see McLachlan and Peel (2000).
An important feature of these mixtures are their analytic properties, in particular their modality structure. Modes are essential for a proper interpretability of the resulting density. For example, in cluster analysis, when there are less modes than components in a mixture, it is reasonable to merge several components into a single cluster based on their modality structure, see Hennig (2010). On the other hand, having more modes than components in a mixtures as can happen in dimensions $> 1$ is an undesirable feature.
The most important tools for assessing the number of modes of finite mixtures of multivariate normal distributions are the concepts of the ridgeline and the $\Pi$-function as introduced in Ray and Lindsay (2005). Recently, Ray and Ren (2011) showed that for two-component mixtures of normals in dimension $D$, the number of modes is at most $D+1$, and further constructed examples which achieved these bounds.
Here, we extend their concept of the ridgeline to finite mixtures of general elliptical densities with possibly distinct density generators in each component. This can be used to obtain bounds for the number of modes of two-component mixtures of $t$ distributions with possibly distinct

[1]Address for correspondence: Prof. Dr. Hajo Holzmann, Philipps-Universität Marburg, Fachbereich Mathematik und Informatik, Hans-Meerweinstr. D-35032 Marburg, Germany email: holzmann@mathematik.uni-marburg.de, Fon: + 49 6421 2825454

degrees of freedom in any dimension. In case of proportional dispersion matrices, we show that these have at most three modes, while for equal degrees of freedom and equal dispersion matrices, the number of modes is at most two.
The paper is structured as follows. In Section 2 we introduce the concept of the ridgeline and the Π-function for mixtures of general elliptical distributions, and state some basic properties. These are used in Section 3 to assess the model structure of two-component $t$-mixtures. In Section 4 we give numerical illustrations and indicate some statistical applications to clustering and hypothesis testing.

## 2 Ridgeline theory for general elliptical distributions

As indicated in Ray and Lindsay (2005), several of their results extend from finite mixtures of multivariate normal distributions to finite mixtures of general elliptical densities. In this section we formulate the relevant statements, for the proofs see Alexandrovich (2011).
First, we introduce some notation. A nonnegative measurable function $\varphi : [0,\infty) \to [0,\infty)$ for which $c_\varphi := \int_{\mathbb{R}^D} \varphi(x^T x)\, dx < \infty$ is finite is called a density generator of a $D$-dimensional spherical distribution. Evidently, $f(x) = c_\varphi^{-1}\varphi(x^T x)$ is then a $D$-dimensional density w.r.t. Lebesgue measure. If $\mu \in \mathbb{R}^D$ and $\Sigma > 0$ is a positive definite $D \times D$ matrix, then

$$f(x;\mu,\Sigma) = k\,\varphi\big((x-\mu)^\mathsf{T}\Sigma^{-1}(x-\mu)\big), \qquad k = \big(c_\varphi\ \det(\Sigma)^{1/2}\big)^{-1}$$

is a density from the associated family of elliptical distributions. For further details on elliptical distributions and their density generators see Fang et al. (1989). We consider general finite mixtures of elliptical densities with possibly distinct density generators in each component, i.e. densities of the form

$$g(x;\mu_i,\Sigma_i,\pi_i,\varphi_i, i=1,\dots,K) = \sum_{i=1}^{K} \pi_i\, k_i\, \varphi_i((x-\mu_i)^\mathsf{T}\Sigma_i^{-1}(x-\mu_i)), \tag{1}$$

where $\mu_i \in \mathbb{R}^D$, $\Sigma_i > 0$ are positive definite $D \times D$ matrices, $\varphi_i$ are density generators with $k_i = \big(c_{\varphi_i}\ \det(\Sigma_i)^{1/2}\big)^{-1}$ the appropriate normalizing constant, and $\pi_i \in [0,1]$ with $\sum_{i=1}^{K} \pi_i = 1$. Typically, the density generators $\varphi_i$ will all be equal as in case of normal mixtures, or at least belong to a parametric family of density generators such as t-distributions with distinct degrees of freedom. Set

$$\mathcal{S}_K := \Big\{\alpha = \big(\alpha_1,\dots,\alpha_K\big)^T \in \mathbb{R}^K : \alpha_i \in [0,1], \sum_{i=1}^{K} \alpha_i = 1\Big\}.$$

Ray and Lindsay (2005) introduced the map $x^* : \mathcal{S}_K \to \mathbb{R}^D$,

$$x^*(\alpha) = \big[\alpha_1\Sigma_1^{-1} + \dots + \alpha_K\Sigma_K^{-1}\big]^{-1}\, [\alpha_1\Sigma_1^{-1}\mu_1 + \dots + \alpha_K\Sigma_K^{-1}\mu_K],$$

the so-called *ridgeline function.* The next theorem summarizes the connection between the modes of the finite mixture $g$ in (1) and the ridgeline. For the proof in this general setting see Alexandrovich (2011).

**Theorem 2.1.** *Suppose that the density generators $\varphi_i$ in the finite mixture $g$ (see (1)) are continuously differentiable and strictly decreasing. Then*

*1. All critical points of $g$ as defined in (1) are contained in $x^*\big(\mathcal{S}_K\big)$, the image of $\mathcal{S}_K$ under the mapping $x^*$.*
*2. Set $h(\alpha) = g(x^*(\alpha))$, $\alpha \in \mathcal{S}_K$. Then $\alpha_{crit}$ is a critical point (resp. local maximum) of $h$ if and only if $x^*(\alpha_{crit})$ is a critical point (resp. local maximum) of $g$.*
*3. If $D > K-1$, then $g$ has no local minima, only local maxima and saddle points.*

Thus, looking for modes of $g$ it is sufficient to look for modes of $h$.

For a two component mixture, setting

$$\delta(x,i) = (x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i), \qquad i = 1,2, \tag{2}$$

we can write

$$g(x;\pi,\mu_1,\mu_2,\Sigma_1,\Sigma_2,\varphi_1,\varphi_2) = \pi\, k_1\varphi_1\big(\delta(x,1)\big) + (1-\pi)\, k_2\varphi_2\big(\delta(x,2)\big).$$

For the ridgeline, we write in slightly different notation than in the above section

$$x^*(\alpha) = S_\alpha^{-1}\left((1-\alpha)\Sigma_1^{-1}\mu_1 + \alpha\Sigma_2^{-1}\mu_2\right), \qquad S_\alpha = (1-\alpha)\Sigma_1^{-1} + \alpha\Sigma_2^{-1}. \tag{3}$$

As above, set $h(\alpha) = g\big(x^*(\alpha)\big)$. Then solving

$$\partial_\alpha h(\alpha) = \pi k_1\, \partial_\alpha\varphi_1\big(\delta\big(x^*(\alpha),1\big)\big) + (1-\pi)k_2\, \partial_\alpha\varphi_2\big(\delta\big(x^*(\alpha),2\big)\big) = 0$$

for $\pi$, where $\partial_\alpha$ is the derivative w.r.t. the real parameter $\alpha$, we get

$$\pi = \frac{k_2\, \partial_\alpha\varphi_2\big(\delta\big(x^*(\alpha),2\big)\big)}{k_2\, \partial_\alpha\varphi_2\big(\delta\big(x^*(\alpha),2\big)\big) - k_1\, \partial_\alpha\varphi_1\big(\delta\big(x^*(\alpha),1\big)\big)} =: \Pi(\alpha),$$

the so-called $\Pi$-*function*. Note that the $\Pi$-function depends on parameters $\mu_i, \Sigma_i, \varphi_i$, $i = 1,2$, but not on the weight $\pi$. For given $\pi$, it can be used to find the critical points of $g$. Further, it provides general bounds on the number of modes as follows.

**Theorem 2.2.** *a. $\Pi(0) = 1$, $\Pi(1) = 0$ and $\Pi(\alpha) \in [0,1]$.*
*Let $N$ be the number of zeros of the derivative $\partial_\alpha\Pi(\alpha)$ of $\Pi(\alpha)$ w.r.t $\alpha$ within the interval $[0,1]$. Then*
*b. $N$ is even, and for any $\pi \in [0,1]$ the equation $\Pi(\alpha) = \pi$ has at most $N+1$ solutions, the smallest of which, $\alpha_1$, gives a mode $x^*(\alpha_1)$ of $g$.*
*c. For any $\pi$, $g$ has at most $1 + N/2$ modes.*

We can compute general expressions for the $\Pi$-function and its derivative as follows. This will be refined for the $t$ distribution in the next section.

**Proposition 2.3.** *Let $\varphi_i'(t) = d\varphi_i/dt(t)$, $t \in \mathbb{R}$, $i = 1,2$ be the derivatives of the density generators. Then for $0 < \alpha < 1$*

$$\Pi(\alpha) = \frac{(1-\alpha)\, k_2\, \varphi_2'}{(1-\alpha)\, k_2\, \varphi_2' + \alpha\, k_1\, \varphi_1'}$$
$$\partial_\alpha\Pi(\alpha) = -k_1k_2 \frac{\varphi_1'\, \varphi_2' + 2\,\alpha(1-\alpha)\, p(\alpha)\big((1-\alpha)\varphi_1'\varphi_2'' + \alpha\varphi_2'\varphi_1''\big)}{\big((1-\alpha)\, k_2\, \varphi_2' + \alpha\, k_1\, \varphi_1'\big)^2} \tag{4}$$

*where $\varphi_2'$ and $\varphi_2''$ are evaluated at $\delta\big(x^*(\alpha),2\big)$ (see (2)), while $\varphi_1'$ and $\varphi_1''$ are evaluated at $\delta\big(x^*(\alpha),1\big)$, and*

$$p(\alpha) = (\mu_2-\mu_1)^\top \Sigma_1^{-1} S_\alpha^{-1} \Sigma_2^{-1} S_\alpha^{-1} \Sigma_2^{-1} S_\alpha^{-1} \Sigma_1^{-1} (\mu_2-\mu_1). \tag{5}$$

# 3 Modes of two components mixtures of $t$ distributions

In this section, based on the results of the previous section we give bounds on the number of modes of two-component $t$-mixtures. Observe that from Theorem 2.2 c., for given parameters $\mu_i, \Sigma_i$, $i = 1, 2$ (and degrees of freedom $n_i$ in case of the t distribution), the number of modes of the resulting mixture $g$ for any weight $\pi$ can be bounded by the number of zeros of $\partial_\alpha \Pi$ in $[0, 1]$. Thus, if we can bound this number of zeros in $[0, 1]$ for any parameter combination $\mu_i, \Sigma_i$ (and $n_i$), we obtain bounds in the number of modes of the mixture $g$.
For mixtures of $t$-distributions, the density generators are given by

$$\varphi(x; n_i) = k_i\Big(1 + \frac{x}{n_i}\Big)^{-(n_i+D)/2}, \qquad k_i = \frac{\Gamma\big(\frac{n_i+D}{2}\big)}{|\Sigma_i|^{\frac{1}{2}}\Gamma\big(n_i/2\big)(n_i\pi)^{D/2}}, \quad i = 1, 2,$$

where $n_i$ denotes the degrees of freedom in the $i^{\text{th}}$ component. The general two-component $t$-mixture is given by

$$g(x; \pi, \mu_1, \mu_2, \Sigma_1, \Sigma_2, n_1, n_2) = \pi k_1 \varphi\big(\delta(x, 1); n_1\big) + (1 - \pi)k_2\varphi\big(\delta(x, 2); n_2\big), \tag{6}$$

**Lemma 3.1.** *Consider a general t-mixture as in (6). Set*

$$\Sigma^* = \Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2}, \qquad \mu^* = \Sigma_2^{-1/2}(\mu_1 - \mu_2) \tag{7}$$

*and let $\Sigma^* = QD^*Q^T$, where $D = diag(\lambda_1^*, \ldots, \lambda_D^*)$ and $Q$ is an orthogonal matrix, denote the spectral decomposition of $\Sigma^*$. Then the number of modes of $g(x; \pi, \mu_1, \mu_2, \Sigma_1, \Sigma_2, n_1, n_2)$ is the same as that of $g(x, \pi, Q^T\mu^*, 0, D^*, I_D, n_1, n_2)$.*

This follows along similar lines as Theorem 4 in Ray and Ren (2011). Using this simplification, by bounding the number of zeros of $\partial_\alpha\Pi$-function one can obtain

**Theorem 3.2.** *1. Let $g(x; \pi, \mu_1, \mu_2, \Sigma_1, \Sigma_2, n_1, n_2) = \pi k_1\varphi\big(\delta(x, 1); n_1\big) + (1 - \pi)k_2\varphi\big(\delta(x, 2); n_2\big)$ be a two-component mixture of t distributions in dimension D, and let d be the number of distinct eigenvalues of the matrix $\Sigma_2^{-1/2}\Sigma_1\Sigma_2^{-1/2}$. Then the number of modes of g is at most $1 + 2d$.*
*2. Let $g(x; \pi, \mu_1, \mu_2, \Sigma, \Sigma/\lambda, n_1, n_2)$, $\lambda > 0$, be a two-component mixture of t distributions in dimension D with proportional covariance matrices. Then the number of modes of g in any dimension is at most three.*
*3. A two-component t-mixture with equal degrees of freedom and dispersion matrices, $g(x; \pi, \mu_1, \mu_2, \Sigma, n)$ has at most two modes in any dimension D.*

# 4 Illustrations and applications

### Numerical illustrations

We start by giving some numerical illustrations of some of the results in the paper.
1. First, we investigate the effect of varying the degrees of freedom in a mixture of two $t$-distributions while keeping the covariances of components fixed. We also consider a corresponding Gauss mixtures which can be considered as a limit case in which the degrees of freedom tend to $\infty$. Specifically, the parameters of the mixtures are

$$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mu_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0.05 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 0.05 & 0 \\ 0 & 1 \end{pmatrix}. \tag{8}$$

In the case of t-mixtures we scale the matrices $\Sigma_i$, $i = 1, 2$ with the factors $\frac{n_i - 2}{n_i}$ in order to retain equal covariances in each constellation of degrees of freedom.
Figure 1 contains plots of the Π-functions for various combinations of degrees of freedom, while Figure 2 has the corresponding for the weight $\pi = 0.65$. From Figure 1 we see that with decreasing degrees of freedoms, the range of mixture weights for which the mixture has three modes decreases as well. For the choice $\pi = 0.65$, the first (normal), second ($n_1 = n_2 = 10$) and forth ($n_1 = 10, n_2 = 3$) have three modes, otherwise there are only two.

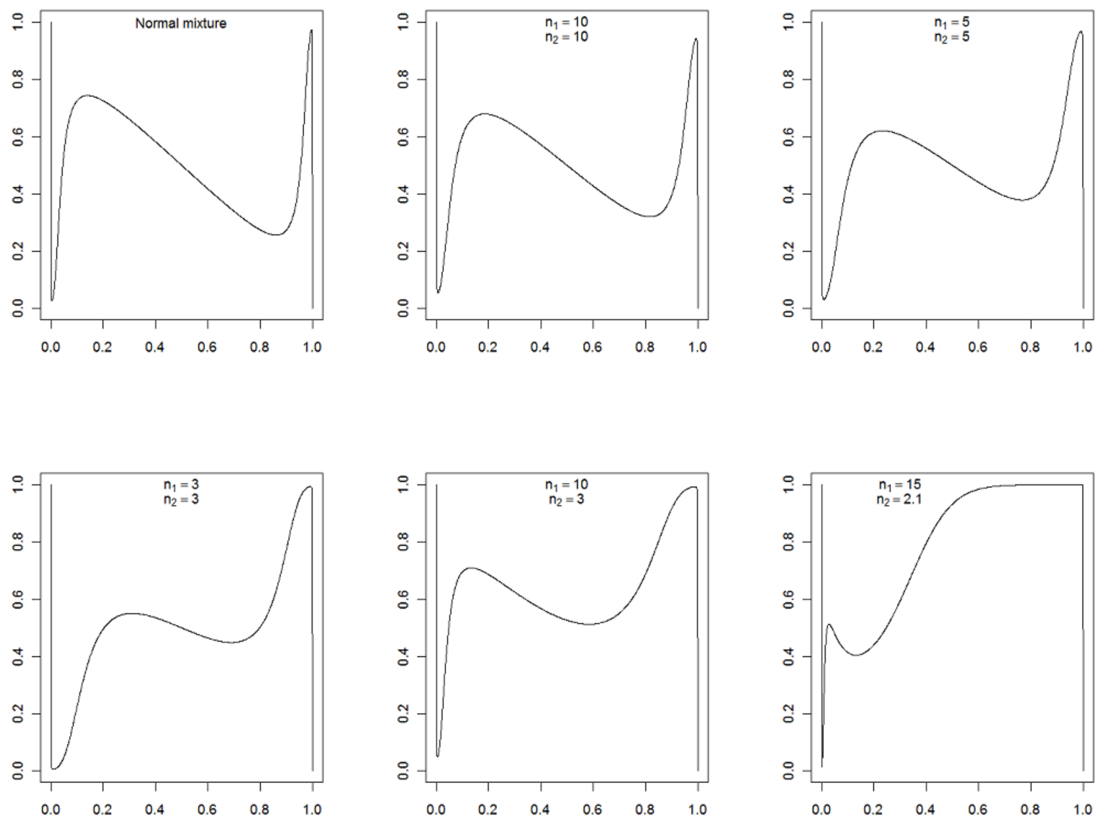


Figure 1: Π-functions for Gauss- and $t$-mixtures with various degrees of freedom

2. Second, we consider the transformation in Lemma 3.1 to diagonal dispersion matrices for a two-component $t$-mixture with 15 degrees of freedom and $\pi = 0.5$ for a special parameter combination. Specifically, consider

$$\mu_1 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}, \mu_2 = \begin{pmatrix} 1.5 \\ 1.5 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1 & 0.14 \\ 0.14 & 0.06 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 0.06 & 0.14 \\ 0.14 & 1 \end{pmatrix}.$$

Then the transformed parameters are given by

$$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mu_2 = \begin{pmatrix} -4.39 \\ -1.11 \end{pmatrix}, \Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma_2 = \begin{pmatrix} 24.19 & 0 \\ 0 & 0.041 \end{pmatrix}.$$

Figure 4 contains plots of the corresponding densities, which look quite distinct. Thus, it is not apparent that the transformation keeps the number of modes

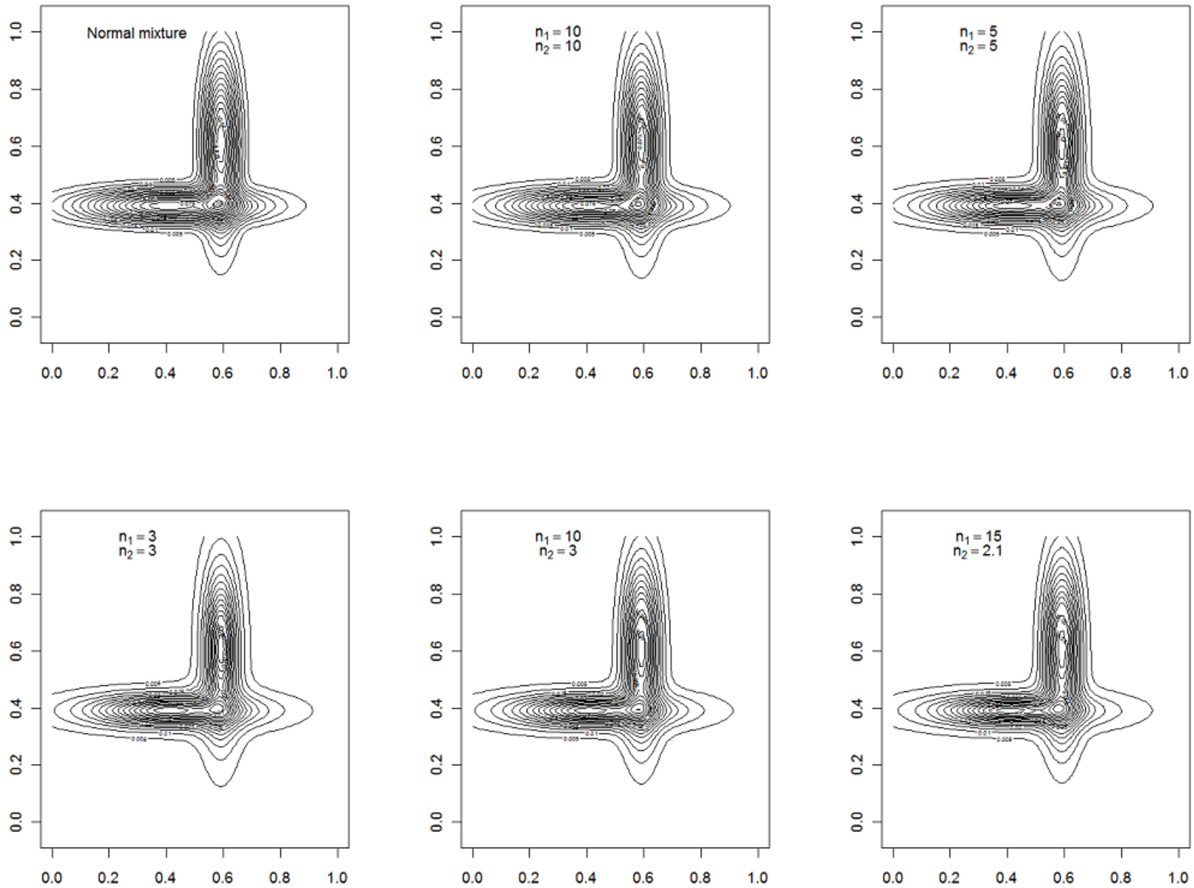


Figure 2: The corresponding contours for the mixtures for $\pi = 0.65$

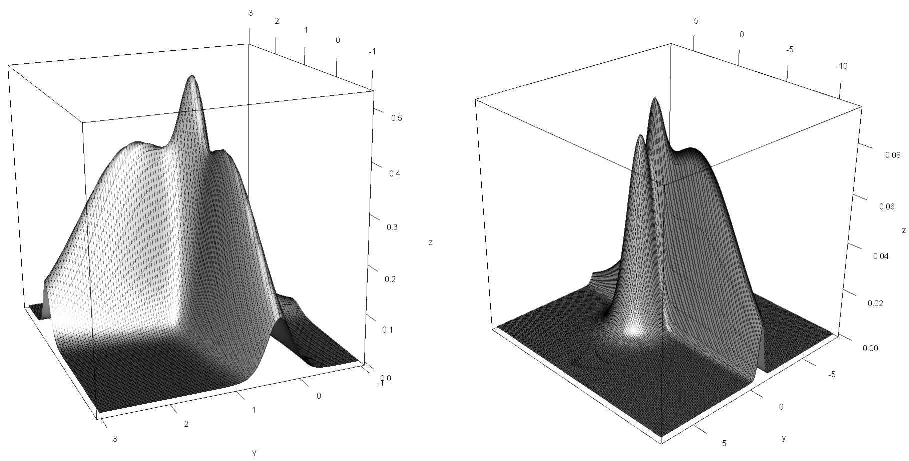

Figure 3: Two-component $t$-mixture before (left) and after (right) transformation to diagonal dispersion matrices

3. Third, we investigate the effect when rotating one component while keeping everything else fixed. We consider a two-component $t$-mixture with 15 degrees of freedom in each component, and parameters as in (8). We rotate the second component clockwise, with angles ranging from 45% up to 135% in equidistant setps. The corresponding densities are plotted in (4). In the process a third mode appears at an angle around 90% and vanishes again for higher angles.

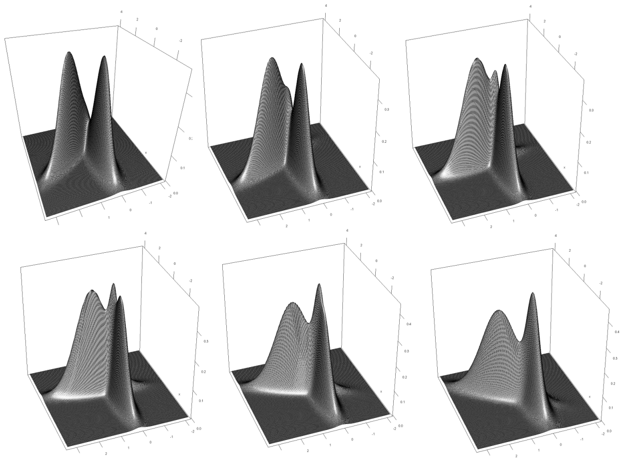

Figure 4: The clockwise rotation of one mixture component

**Statistical applications**

Finally, we indicate two potential statistical application of the above theory.

1. Merging components in mixtures of $t$-distributions.

McLachlan and Peel (2000) recommend the use of finite mixtures of $t$-distributions as a more robust alternative to normal mixtures. While $t$-mixtures allow for heavier tails of the components, asymmetry can still not be dealt with, and thus, the number of components may exceed the actual number of clusters in the data. Thus, modality-based merging algorithms like in Hennig (2010) for normal mixtures, based on the ridgeline as in Theorem 2.1, can be employed.

2. Testing for the number of modes.

If two-component mixtures under suitable parameter restrictions allow at most two modes, such as two-component normals with proportional covariances, or $t$-mixtures with equal degrees of freedom and covariances, one can use parametric methods to test for one against two modes in such a model by likelihood-ratio based methods, see Holzmann and Vollmer (2008) for univariate normal and von Mises mixtures. This requires explicit characterizations of the parameter constellations which yield unimodal or bimodal mixtures. For two-component normals with proportional covariances, these are given in Ray and Lindsay (2005), Corollary 4, while corresponding characterizations based on Theorem 3.2 2. and 3. still need to be derived.

# 5 References


Alexandrovich, G. (2011). *Analytische Eigenschaften von Mischungen elliptischer Verteilungen und deren Anwendung in der Clusteranalyse.* Diploma thesis, Philipps Universität Marburg

Fang, K.T., Kotz, S. and Ng, K.W. (1989). *Symmetric Multivariate and Related Distributions.*

Chapman & Hall.

Hennig, C. (2010). Methods for merging Gaussian mixture components. in: *Advances in Data Analysis and Classification.* doi: 10.1007/s11634-010-0058-3. URL http://dx.doi.org/10.1007/s11634-010-0058-3.

Holzmann, H., Vollmer, S. (2008). A likelihood ratio test for bimodality in two-component mixtures – with application to regional income distribution in the EU. *AStA - Advances in Statistical Analysis* **92**, 57-69.

McLachlan, G. J. and Peel, D. (2000) *Finite Mixture Models*, Wiley, New York.

Rahman, Q. I. and Schmeisser, G. (2002). *Analytic Theory of Polynomials*, Clarendon Press, Oxford.

Ray, S. and Lindsay, B.G. (2005). The topography of multivariate normal mixtures, *Ann. Statist.* **33**, 2042-2065

Ray, S. and Ren, D. (2011) How many modes can a two component mixture have? Preprint, Boston University.