# School of Informatics, University of Edinburgh

## Institute for Adaptive and Neural Computation

# An isotropic Gaussian mixture can have more modes than components

by

Miguel Carreira-Perpinan, Chris Williams

**Informatics Research Report EDI-INF-RR-0185**

**School of Informatics** **December 2003**
http://www.informatics.ed.ac.uk/

# An isotropic Gaussian mixture can have more modes than components

Miguel Carreira-Perpinan, Chris Williams

Informatics Research Report EDI-INF-RR-0185

SCHOOL *of* INFORMATICS
Institute for Adaptive and Neural Computation

December 2003

**Abstract :**

Carreira-Perpinan and Williams (2003) conjectured that a homoscedastic Gaussian mixture of M components in d $>$ 1 dimensions has at most M modes. Prof. J. J. Duistermaat (personal communication, 2003) provided the counterexample of a 3–component mixture in d = 2 where the Gaussians are located at the vertices of an equilateral triangle; for a certain range of variances modes are present near to the vertices and also at the centre of the triangle. In this paper we illustrate the nature of the counterexample and compute the range of variances for which there are more than 3 maxima. We also extend the construction to the regular simplex with M vertices and show that for M$>$2 there is always a range of variances for which M+1 modes are present.

**Keywords** : Gaussian mixture model, scale-space

# An isotropic Gaussian mixture can have more modes than components

Miguel Á. Carreira-Perpiñán
Dept. of Computer Science
University of Toronto
miguel@cs.toronto.edu

Christopher K. I. Williams
School of Informatics
University of Edinburgh
c.k.i.williams@ed.ac.uk

December 22, 2003

**Abstract**

Carreira-Perpiñán and Williams (2003) conjectured that a homoscedastic Gaussian mixture of $M$ components in $d > 1$ dimensions has at most $M$ modes. Prof. J. J. Duistermaat (personal communication, 2003) provided the counterexample of a 3–component mixture in $d = 2$ where the Gaussians are located at the vertices of an equilateral triangle; for a certain range of variances modes are present near to the vertices and also at the centre of the triangle. In this paper we illustrate the nature of the counterexample and compute the range of variances for which there are more than 3 maxima. We also extend the construction to the regular simplex with $M$ vertices and show that for $M > 2$ there is always a range of variances for which $M + 1$ modes are present.

## 1 Introduction

The Gaussian kernel is used in a number of fields. For example, Gaussian mixtures (GMs) are a flexible density model in statistics and machine learning, allowing the representation of complex, multimodal density functions in several dimensions. In scale space theory (Lindeberg, 1994), convolution of a signal with a Gaussian kernel at different scales is a convenient way to obtain blurred versions of an signal, whose maxima can be used to locate meaningful objects. It is known that blurring a 1D function with the Gaussian kernel cannot create new modes as the scale increases, and that the Gaussian is the only kernel with this property, see e.g. Koenderink (1984), Babaud et al. (1986), Yuille and Poggio (1986). However, in 2D or higher the Gaussian kernel can also create modes. Lifshitz and Pizer (1990) gave an example of this where two "mountains", one slightly higher than the other, are connected by a narrow ridge. For zero blurring there is only one mode, but as the amount of blurring is increased the lower mountain also becomes a mode.

A related question, originally considered by Carreira-Perpiñán (1999), is this: can a GM with $M$ components have more than $M$ modes? Noting that convolving a delta mixture with a Gaussian kernel gives a GM and using the earlier scale-space theory result, Carreira-Perpiñán and Williams (2003) proved that in 1D any GM (possibly with different variances) has at most $M$ modes. Since the creation of modes in 2D is a rare event, they conjectured that a GM with $M$ isotropic components in $d > 1$ dimensions could not have more than $M$ modes. J. J. Duistermaat (personal communication, 2003) provided a counterexample for this conjecture which we describe in section 2. It is particularly interesting as (i) it shows that mode creation can occur when the original function (before convolution) is a finite number of delta functions and (ii) as opposed to Lifshitz and Pizer's construction it is an analytic example that can be fully analysed with the usual tools of calculus.

Below we consider a homoscedastic GM with $M$ isotropic components in $d > 1$ dimensions, of the form:

$$p(\mathbf{t}) = \sum_{m=1}^{M} \frac{1}{M(2\pi\sigma^2)^{d/2}} e^{-\frac{1}{2}\left\| \frac{\mathbf{t}-\boldsymbol{\mu}_m}{\sigma} \right\|^2} \tag{1}$$

for scale $\sigma > 0$ and centroids $\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_M \in \mathbb{R}^d$. For $\sigma \to 0$, $p(\mathbf{t})$ has $M$ modes (near the centroids) while for $\sigma \to \infty$, $p(\mathbf{t})$ has 1 mode near the centre of mass $\frac{1}{M} \sum_{m=1}^{M} \boldsymbol{\mu}_m$. We seek values of $\sigma$ for which more than $M$ modes exist.

The structure of the remainder of the paper is as follows. In section 2 we illustrate Duistermaat's counterexample, and in section 3 we provide a analysis of general case of the regular simplex with $M$ vertices, with a Gaussian at each vertex.

## 2 Duistermaat's counterexample

Duistermaat's counterexample consists of a 2D Gaussian mixture with $M = 3$ isotropic components in the vertices of an equilateral triangle[1]. We take the centroids at $(1, 0)$, $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$ and $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})$. For a narrow interval of scales $\sigma_1 < \sigma < \sigma_3$ (where $\sigma_1 = \frac{\sqrt{2}}{2} \simeq 0.7071$ and $\sigma_3 \simeq 0.7391$) there are 4 modes, one at the triangle centre and 3 near the centroids (see Figure 1). The central mode is difficult to discern because the density is very flat around it. As $\sigma$ increases through $\sigma_1$ from below, the three saddle points shown in the top plot of the right column of Figure 1 merge with the local minimum at the centre and turn into a local maximum. As $\sigma$ increases further to $\sigma_3$, the three saddle points illustrated in the middle row of Figure 1 merge with the central local maximum and the three original maxima disappear.

Rather than analysing the specific counterexample, we analyse the more general case of a $M$-vertex simplex Gaussian mixture; Duistermaat's counterexample corresponds to $M = 3$.

## 3 A $M$-vertex simplex Gaussian mixture

We first describe the construction of a regular simplex and then analyse the corresponding Gaussian mixture.

Consider the $M$-dimensional space $\mathbb{R}^M$, and set $D = M - 1$. For $m = 1, \dots, M$ let the $m$th centroid be $\boldsymbol{\mu}_m = \sqrt{(D+1)/D}\,\mathbf{e}_m$, where $\mathbf{e}_m$ is the unit vector along the $m$th direction. Then the centre of the simplex is at $\mathbf{c} = \mathbf{1}/\sqrt{D(D+1)}$ (where $\mathbf{1}$ is the vector of ones) and $\|\boldsymbol{\mu}_m - \mathbf{c}\| = 1$, so the simplex has unit circumradius. The scale value $\sigma$ is then relative to the circumradius. Notice that all of the centroids and $\mathbf{c}$ obey the equation $\mathbf{1}^T\mathbf{t} = \sqrt{(D+1)/D}$. For example for $M = 3$ we obtain the equilateral triangle lying in the plane $\mathbf{1}^T\mathbf{t} = \sqrt{3/2}$.

As all the vertices and the centre of the simplex lie in the hyperplane $\mathbf{1}^T\mathbf{t} = \sqrt{(D+1)/D}$, the $M$-vertex simplex can be embedded in $\mathbb{R}^{M-1}$. Note that the density falls off as a Gaussian along the $\mathbf{1}$ direction. One consequence of this is that when the centre of the $M$-vertex simplex is a minimum in $M-1$ dimensions it will be a saddle point in $M$ dimensions. For example, for the triangle ($M = 3$), in 2D we get a minimum at the centre for $\sigma < \sigma_1$, as shown in Figure 1. However, considering the triangle embedded in 3D with 3D Gaussians we obtain a saddle point at the centre. If we use the above construction in $d > M$ dimensions then the density will also have a Gaussian decay on the additional dimensions.

Given the simplex defined above in $M = D + 1$ dimensions, we locate a Gaussian at each vertex to obtain

$$p(\mathbf{t}) = K \sum_{m=1}^{M} e^{-\frac{1}{2}\left\|\frac{\mathbf{t}-\boldsymbol{\mu}_m}{\sigma}\right\|^2} \qquad K = \frac{1}{M(2\pi\sigma^2)^{M/2}}. \tag{2}$$

Then one can compute the gradient and Hessian at the simplex centre (see appendix A):

$$\mathbf{g}(\mathbf{c}) = \mathbf{0} \tag{3}$$

$$\mathbf{H}(\mathbf{c}) = \beta\left((D+1)(1 - D\sigma^2)\mathbf{I} - \mathbf{1}\mathbf{1}^T\right) \tag{4}$$

for $D \geq 1$, where $\beta = Ke^{-\frac{1}{2\sigma^2}}/D\sigma^4$. The Hessian has eigenvalues $-\beta(D+1)D\sigma^2$ associated with the eigenvector $\mathbf{1}$ (multiplicity 1) and $\beta(D+1)(1 - D\sigma^2)$ associated with the degenerate eigenvector space $\mathbf{1}^\perp$ (multiplicity $D$). Thus the centre is always a critical point, in particular a saddle, catastrophe and maximum if $\sigma^2 < \frac{1}{D}$, $\sigma^2 = \frac{1}{D}$ and $\sigma^2 > \frac{1}{D}$, respectively.

Now let us consider the points inside the simplex. All the critical points of $p$ are here, since the simplex is the convex hull of the centroids, which contains all the GM critical points (Carreira-Perpiñán and Williams, 2003). By symmetry we need only consider points along the axis through $\boldsymbol{\mu}_1$ and $\mathbf{c}$. Further, we are only interested in points between $\boldsymbol{\mu}_1$ and $\mathbf{c}$ (which is where the mode coming from the $\boldsymbol{\mu}_1$ centroid lies). Thus, we reduce the problem to a one-dimensional one by transforming $\mathbf{t} = (1-x)\mathbf{c} + x\boldsymbol{\mu}_1$ for $x \in [0, 1]$. Now $x$ indicates the distance relative to the distance between the simplex centre and vertex. Simplifying eq. (2) we obtain

$$\begin{aligned}
p(x) &= Ke^{-\frac{1}{2}\left(\frac{x-1}{\sigma}\right)^2}\left(1 + De^{-\left(\frac{D+1}{D}\right)\frac{x}{\sigma^2}}\right) \\
p'(x) &= -\frac{K}{\sigma^2}e^{-\frac{1}{2}\left(\frac{x-1}{\sigma}\right)^2}\left(x - 1 + (1 + Dx)e^{-\left(\frac{D+1}{D}\right)\frac{x}{\sigma^2}}\right) \\
p''(x) &= -\frac{K}{\sigma^4}e^{-\frac{1}{2}\left(\frac{x-1}{\sigma}\right)^2}\left(\sigma^2 - (x-1)^2 + e^{-\left(\frac{D+1}{D}\right)\frac{x}{\sigma^2}}\left(D\sigma^2 - (1 + Dx)\left(x + \frac{1}{D}\right)\right)\right).
\end{aligned}$$

[1]A related construction (though not for GMs) appeared in Duistermaat (1984, pp. 96–97).

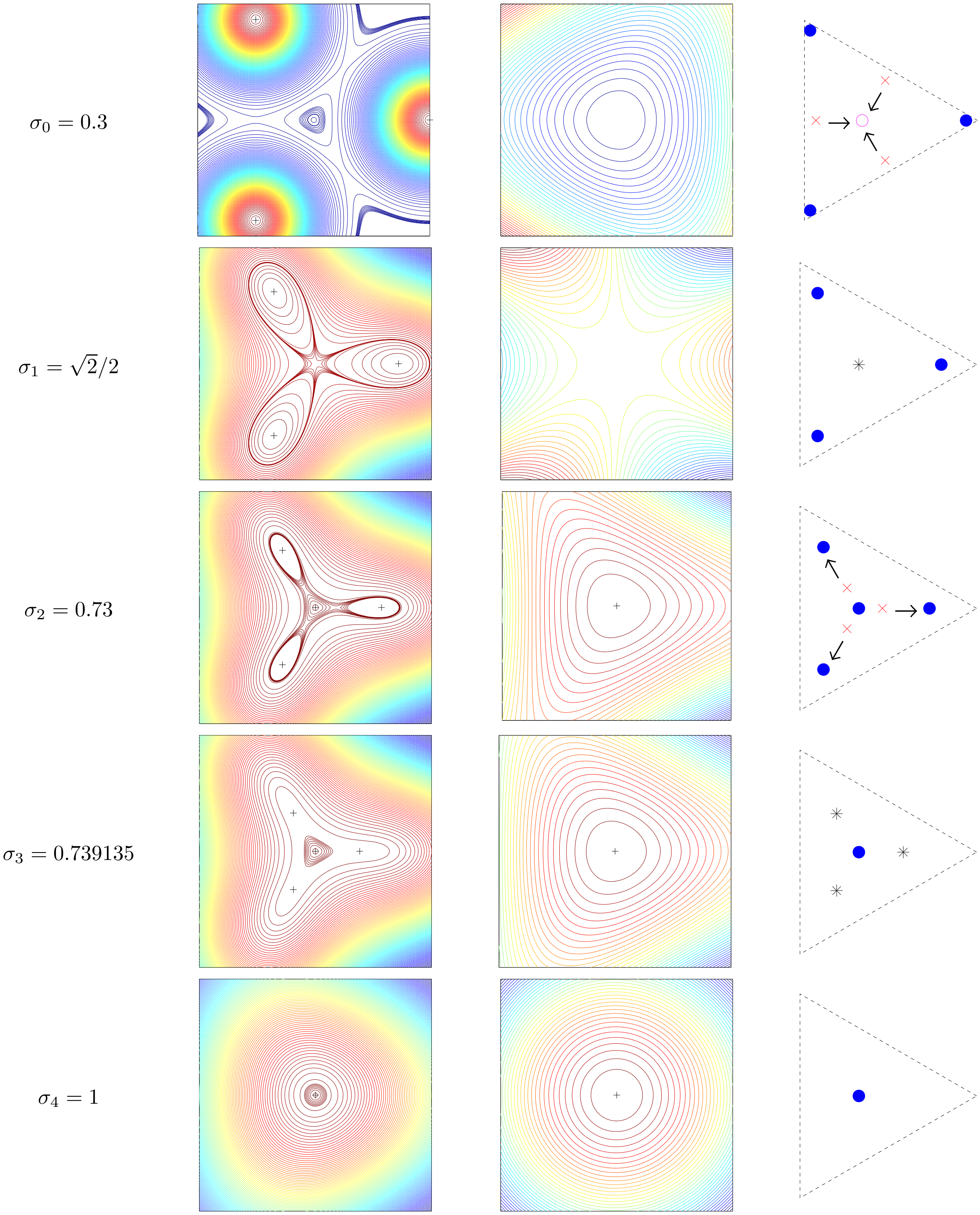


Figure 1: Duistermaat's counterexample: a GM in $D = 2$ dimensions with $M = 3$ homoscedastic components centred at the vertices of an equilateral triangle, for interesting values of the scale $\sigma$. We take the GM centroids at $(1, 0)$, $(-\frac{1}{2}, \frac{\sqrt{3}}{2})$ and $(-\frac{1}{2}, -\frac{\sqrt{3}}{2})$. *Left*: contour plot of the square $x_1,\ x_2 \in [-1, 1]$. The GM has 4 modes (marked $+$) for scales $\sigma \in (\sigma_1, \sigma_3)$. Note that the spacing between the coutours is reduced near to the origin since the density is almost flat there. *Middle*: contour plot of the square $x_1,\ x_2 \in [-0.1, 0.1]$ to zoom into the origin. The origin is a catastrophe for $\sigma = \sigma_1 = \frac{\sqrt{2}}{2}$; it passes from being a minimum for $\sigma < \sigma_1$ to being a maximum for $\sigma > \sigma_1$. *Right*: schematic evolution of the critical points (maxima $\bullet$, minima $\circ$, saddles $\times$, catastrophes $*$). The arrows indicate the direction of movement of the saddles with increasing $\sigma$ (the maxima from the centroids always move towards the origin).

We are interested in the critical points of $p(x)$ (and so of $p(\mathbf{t})$) as a function of the scale parameter $\sigma > 0$ and the dimension $D$. The critical points occur for $p'(x) = 0$ and so must be solutions of the following implicit equation:

$$F(x, \sigma, D) = x - \frac{1 - e^{-\left(\frac{D+1}{D}\right)\frac{x}{\sigma^2}}}{1 + De^{-\left(\frac{D+1}{D}\right)\frac{x}{\sigma^2}}} = 0. \tag{5}$$

As a function of $x \geq 0$, the second term in $F(x, \sigma, D)$ is convex in $(0, \frac{D\sigma^2}{D+1} \log D)$ and concave otherwise.

Consider the point $x = 0$ (the simplex centre). This verifies $p'(0) = 0$ and $p''(0) \propto \left(\frac{1}{D} - \sigma^2\right)$. Thus, it is always a critical point, in particular a saddle, catastrophe and maximum if $\sigma^2 < \frac{1}{D}$, $\sigma^2 = \frac{1}{D}$ and $\sigma^2 > \frac{1}{D}$, respectively (paralleling the earlier result for the Hessian). Call $\sigma_1 = \frac{1}{\sqrt{D}}$ the scale value for the first catastrophe.

Since $F$ is continuously differentiable with respect to $x$ and $\sigma$ and $\partial F/\partial \sigma \neq 0$ for $\sigma > 0$ and $x > 0$, eq. (5) locally defines a curve $\sigma = \sigma(x)$ by the implicit function theorem. Likewise, if $\partial F/\partial x \neq 0$ we have locally a curve $x = x(\sigma)$; for each critical point we have one such curve, which gives its scale-space trajectory. When $\partial F/\partial x = 0$ at a catastrophe point, the curves change topology (e.g. the critical points merge).

For $D = 1$ ($M = 2$ components) equation (5) becomes $x = \tanh(x/\sigma^2)$ which has at most one solution in $(0, 1)$; see Figure 2 (left). This corresponds to a catastrophe at $\sigma = 1$ where two maxima coming from the centroids annihilate with the minimum at the centre to create a maximum at the centre. Thus, in this case the number of modes is always at most 2, in agreement with earlier results.

For $D > 1$ the equation (5) can have at most two solutions in $(0, 1)$, one a maximum coming from the $\boldsymbol{\mu}_1$ centroid and the other one a minimum[2], while the centre is a maximum. This corresponds to the appearance of an extra, $(M + 1)$th mode at the centre. The sequence of events with $0 < \sigma_0 < \sigma_1 < \sigma_2 < \sigma_3 < \sigma_4$ is as follows:

| $\sigma$ | critical points in $(0, 1)$ |
|---|---|
| $\sigma_0$ | minimum at $x = 0$, maximum near $x = 1$ |
| $\sigma_1 = \frac{1}{\sqrt{D}}$ | catastrophe at $x = 0$ (minimum at $x = 0$ merges with saddle coming from $x < 0$), maximum at $x < 1$ |
| $\sigma_2$ | maximum at $x = 0$, saddle, maximum at $x < 1$ |
| $\sigma_3$ | maximum at $x = 0$, catastrophe at $x = x^*$ (saddle annihilates with maximum at $x < 1$) |
| $\sigma_4$ | maximum at $x = 0$. |

Here, $\sigma_1$ and $\sigma_3$ are determined by $D$, while $\sigma_0$, $\sigma_2$ and $\sigma_4$ are any value in $(0, \sigma_1)$, $(\sigma_1, \sigma_3)$ and $(\sigma_3, \infty)$, respectively.

Figure 2 (right) shows the $F(x, \sigma, D)$ curves for $\sigma_0, \ldots, \sigma_4$, for $D = 4$. Figure 3 shows the corresponding density functions $p(x)$ (along the centre-vertex axis); the inset reveals for $\sigma_2$ the presence of a maximum at $x = 0$ and another at $x < 1$. Figure 4 shows the motion of the critical points in scale space for $D = 4$. Figure 1(right) shows the evolution of all critical points (not just along the centre-vertex axis) for $D = 2$.

Let us now compute the value $\sigma_3$ for the second catastrophe, for fixed $D$. For $\sigma > \sigma_3$ there is only one critical point (at the origin), i.e., the curve $F(x, \sigma, D)$ does not intersect the $x$-axis for $x > 0$. Hence, $\sigma_3$ corresponds to the largest $\sigma$ for which $F(x, \sigma, D) = 0$ for $x > 0$; let the intersection of $F$ with the $x$-axis be at $x^*$. Solving for $\sigma^2$ in eq. (5), differentiating with respect to $x$ and equating to zero we obtain an equation for $\sigma_3^2$ and another for $x^*$ ((8)–(9) below). An alternative way is to require $F(x^*, \sigma_3, D) = 0$ (eq. (5)) and $\partial F/\partial x|_{x=x^*, \sigma=\sigma_3} = 0$ (when the condition for the implicit function theorem fails, at the catastrophe) to hold simultaneously. At $x^*$ we have (from (5))

$$x^* = \frac{1 - e^{-\gamma x^*}}{1 + De^{-\gamma x^*}} \tag{6}$$

where $\gamma = (D + 1)/D\sigma_3^2$. This can be rearranged to give $e^{-\gamma x^*} = (1 - x^*)/(1 + Dx^*)$. Differentiating $F$ wrt $x$ and using (6) we obtain

$$1 - \frac{\gamma e^{-\gamma x^*}}{1 + De^{-\gamma x^*}} - \frac{D\gamma x^* e^{-\gamma x^*}}{1 + De^{-\gamma x^*}} = 0. \tag{7}$$

Substituting for $e^{-\gamma x^*}$ from above we obtain after some manipulation

$$\sigma_3^2 = (1 - x^*)(1 + Dx^*)/D. \tag{8}$$

By taking logs of $e^{-\gamma x^*} = (1 - x^*)/(1 + Dx^*)$ and using (8) we finally obtain the self-consistent equation

$$\log \frac{1 + Dx^*}{1 - x^*} = \frac{(D + 1)x^*}{(1 - x^*)(1 + Dx^*)}. \tag{9}$$

[2]This critical point is a minimum along the $x$-axis but is a saddle point in $D$ dimensions.

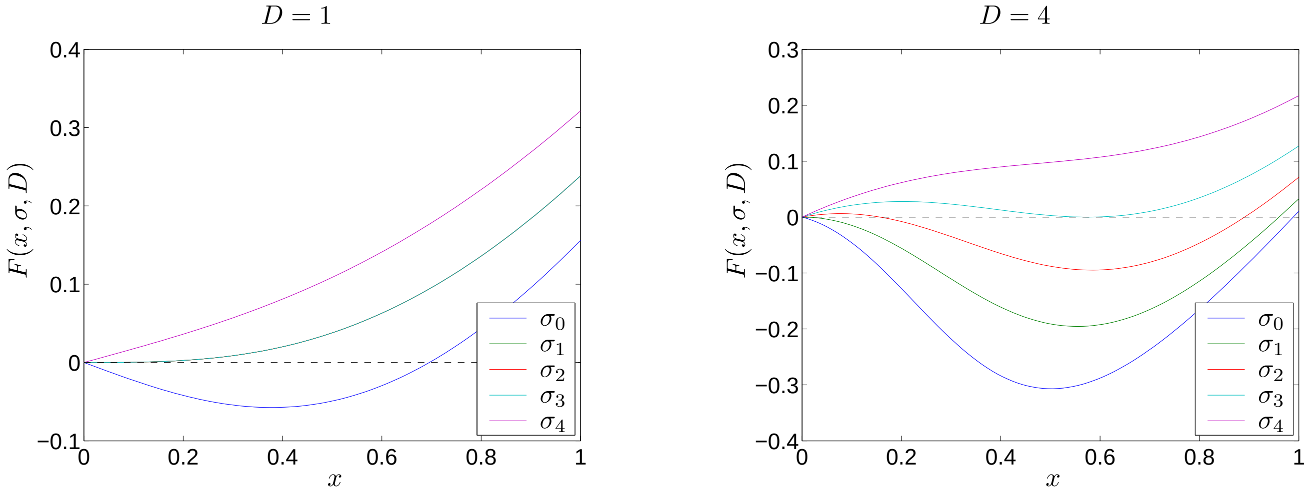


Figure 2: Evolution of the curve $F(x, \sigma, D)$ from eq. (5) for fixed dimension $D$ as a function of $\sigma$, for 5 representative values of $\sigma$, $\sigma_0 < \sigma_1 < \sigma_2 < \sigma_3 < \sigma_4$, where $\sigma_1$ and $\sigma_3$ are the first and second catastrophes, respectively. *Left*: for $D = 1$ ($M = 2$ components), the curve $F(x, \sigma, D)$ intersects the $x$ axis at most once in $(0, 1)$, so $\sigma_1 = \sigma_2 = \sigma_3$ and there are only 3 curves. *Right*: for $D > 1$, $F(x, \sigma, D)$ intersects the $x$ axis twice in $(0, 1)$ for $\sigma \in (\sigma_1, \sigma_3)$.

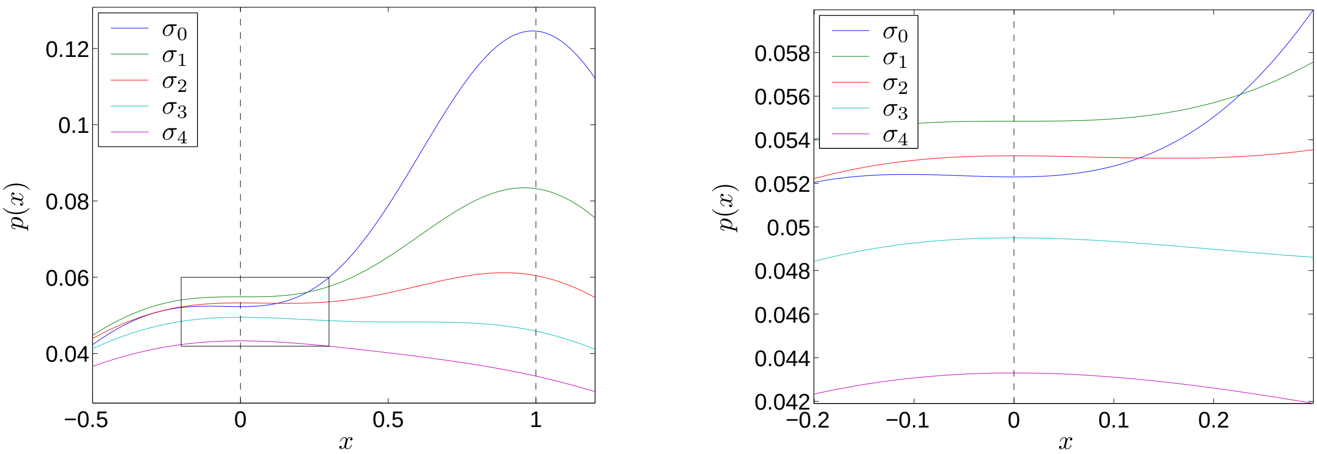


Figure 3: Plot of the Gaussian mixture density along the line passing through the simplex centre and a vertex, $p(x)$, for $D = 4$ (i.e., $M = 5$ components). The right graph is a blowup of the rectangular area in the left one. The curves correspond to the same $\sigma$ values as in fig. 2 (right).

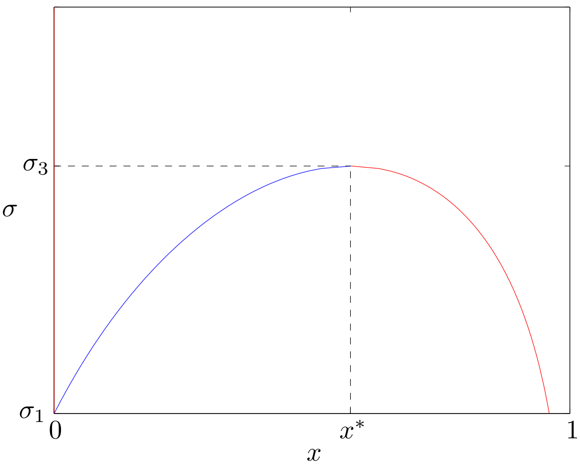


Figure 4: Evolution of the critical points for the segment $x \in [0,1]$ for $D = 4$ for $\sigma \geq \sigma_1 = \frac{1}{2}$. There are three critical points: a maximum at the origin $x = 0$ (red vertical line); a saddle travelling rightwards from the origin (blue curve); and a maximum travelling leftwards from near the centroid at $x = 1$ (red curve). The latter two annihilate at the second catastrophe $\sigma_3 \approx 0.592183$ at $x^* \approx 0.574862$. The function $F(x, \sigma, D) = 0$ on these critical point curves.

Numerically, $x^*$ can be obtained by the following fixed-point iterative scheme with initial value $x = \frac{1}{2}$:

$$x \leftarrow 1 - \frac{(D+1)x}{(1+Dx)\log\frac{1+Dx}{1-x}}.$$

Figure 5 shows the dependence on the dimension $D$ of the interval

$$\Delta\sigma = \sigma_3 - \sigma_1 = \sqrt{\frac{(1-x^*)(1+Dx^*)}{D}} - \frac{1}{\sqrt{D}} \tag{10}$$

where the extra mode at the origin exists. This interval is zero for $D = 1$, peaks at about 0.2833 for $D = 698$, and decays thereafter. Although $\sigma_1$ and $\sigma_3$ decrease monotonically with $D$, they do so at different rates, hence the maximum for $\Delta\sigma(D)$.

### 3.1 The $M = 3$ case: the equilateral triangle

The equilateral triangle is also a counterexample in any dimension $d > 2$. This is because $p(\mathbf{t})$ becomes the product of two terms $p(t_1, t_2)p(t_3, \dots, t_d)$ (where we take the basic triangle construction to lie in 2D). $p(t_1, t_2)$ is the mixture of 3 Gaussians in 2D we have analysed above, while $p(t_3, \dots, t_d)$ is a $\mathcal{N}(0, \sigma^2 I)$ density on the remaining dimensions. Thus, the centre remains a critical point and the Hessian at the centre is block diagonal. There is a $t_1, t_2$ block, and the remaining block equals a negative constant times the identity. Hence, if the 2D Hessian is negative definite, the origin is a maximum; otherwise it is a saddle. Finally, the density value $p(x)$ along the centre-vertex axis is proportional to the 2D case. Thus, the critical intervals for $\sigma$ are still given by the 2D case. In summary, everything remains as in the 2D case, except that for $\sigma < \sqrt{2}/2$ the origin is not a minimum but a saddle.

## 4 Discussion

Adding a very small amount of noise to the centroids, mixing proportions or covariances in the triangle construction still preserves the appearance of a fourth mode at the centre. However, adding a larger (but still very small) amount of noise prevents it. Thus, the mode creation at the origin is generic but fragile.

The simplex construction creates $M + 1$ modes from $M$ Gaussians. It is interesting to ask if it is possible to get a greater productivity. For example we can consider a triangular lattice, with Gaussians at every vertex. In fact in this case numerical experiments show that the phenomenon disappears and we do not get additional

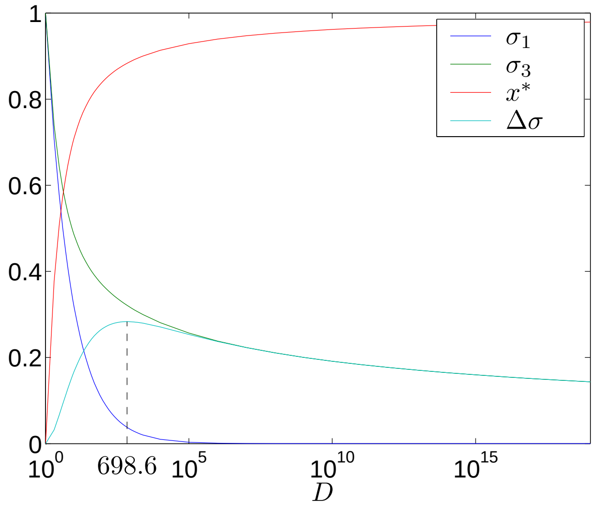


Figure 5: Dependence on the dimension $D$ of the interval $\Delta\sigma = \sigma_3 - \sigma_1$ in which the extra mode at the origin exists.

modes. However, after being told of the simplex construction David MacKay constructed a "Kekule lattice"[3] and has shown that asymptotically this has 5/3 times as many modes as Gaussians as the size of the lattice goes to infinity, see `http://www.inference.phy.cam.ac.uk/mackay/gaussians/`.

## Acknowledgements

CW thanks Dr. Luc Florack for the suggesting to contact Prof. Duistermaat about this problem.

## A Gradient and Hessian of a Gaussian mixture

For a Gaussian mixture density

$$p(\mathbf{t}) = \sum_{m=1}^{M} \pi_m \left|2\pi\boldsymbol{\Sigma}_m\right|^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{t}-\boldsymbol{\mu}_m)^T\boldsymbol{\Sigma}_m^{-1}(\mathbf{t}-\boldsymbol{\mu}_m)} \qquad \mathbf{t} \in \mathbb{R}^D$$

the gradient $\mathbf{g}(\mathbf{t})$ and Hessian $\mathbf{H}(\mathbf{t})$ are as follows:

$$\mathbf{g}(\mathbf{t}) = \sum_{m=1}^{M} p(\mathbf{t}, m)\boldsymbol{\Sigma}_m^{-1}(\boldsymbol{\mu}_m - \mathbf{t}),$$

$$\mathbf{H}(\mathbf{t}) = \sum_{m=1}^{M} p(\mathbf{t}, m)\boldsymbol{\Sigma}_m^{-1}\left((\boldsymbol{\mu}_m - \mathbf{t})(\boldsymbol{\mu}_m - \mathbf{t})^T - \boldsymbol{\Sigma}_m\right)\boldsymbol{\Sigma}_m^{-1}.$$

## References

J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda. Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 8(1):26–33, Jan. 1986.

M. Á. Carreira-Perpiñán. Mode-finding for mixtures of Gaussian distributions. Technical Report CS–99–03, Dept. of Computer Science, University of Sheffield, UK, Mar. 1999. Revised August 4, 2000. Available online at `http://www.dcs.shef.ac.uk/~miguel/papers/cs-99-03.html`.

[3] A triangular lattice with some vertices unfilled so that a pattern of connected hexagons is obtained.

M. Á. Carreira-Perpiñán and C. K. I. Williams. On the number of modes of a Gaussian mixture. In L. Griffin and M. Lillholm, editors, *Proc. of the 4th International Conference on Scale-Space theories in Computer Vision*, pages 625–640, Isle of Skye, Scotland, UK, June 10–12 2003. Springer-Verlag, Berlin.

J. Damon. Local Morse theory for solutions to the heat equation and Gaussian blurring. *J. Diff. Equations*, 115 (2):368–401, Jan. 20 1995.

J. J. Duistermaat. Bifurcations of periodic solutions near equilibrium points of Hamiltonian systems. In L. Salvadori, editor, *Bifurcation Theory and Applications*, number 1057 in Lecture Notes in Mathematics, pages 57–105. Springer-Verlag, Berlin, 1984.

J. J. Koenderink. The structure of images. *Biol. Cybern.*, 50:363–370, 1984.

L. M. Lifshitz and S. M. Pizer. A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 12(6):529–540, June 1990.

T. Lindeberg. *Scale-Space Theory in Computer Vision.* Kluwer Academic Publishers Group, Dordrecht, The Netherlands, 1994.

A. L. Yuille and T. A. Poggio. Scaling theorems for zero crossings. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 8(1):15–25, Jan. 1986.