

Ghosts in Gaussian Mixture Models

Sphere Packing Workshop, Fields Institute, Toronto

Brittany Terese Fasy

Duke University, Computer Science Department
IST Austria (Institute of Science and Technology Austria)

15 November 2011

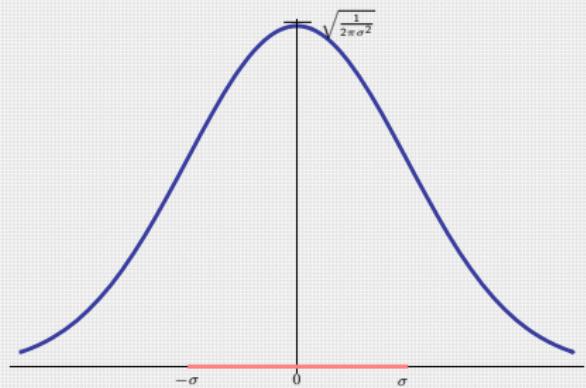
Motivating Goal

To characterize the critical values of an isotropic Gaussian Mixture Model.

Gaussian Kernel

Definition

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z)^2}{2\sigma^2}}$$

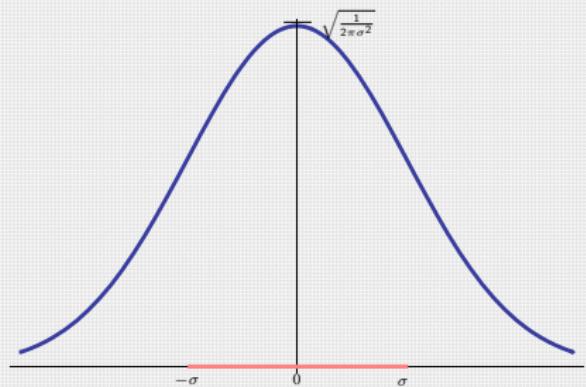


Gaussian Kernel

Definition

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z)^2}{2\sigma^2}}$$

Standard Deviation: σ



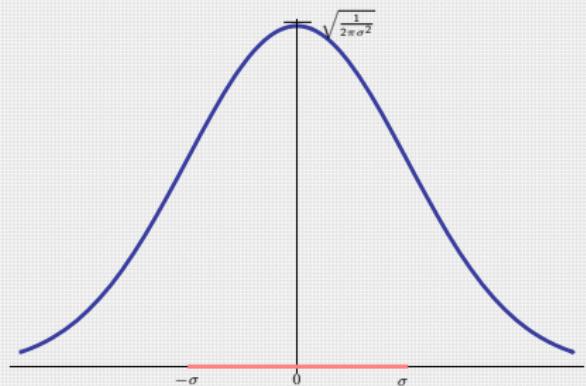
Gaussian Kernel

Definition

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z)^2}{2\sigma^2}}$$

Standard Deviation: σ

Height: $\frac{1}{\sqrt{2\pi\sigma^2}}$



Gaussian Kernel

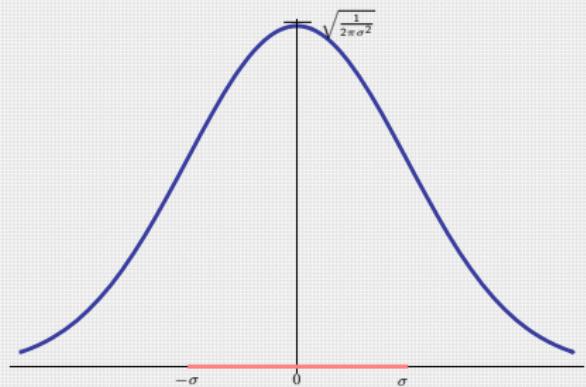
Definition

$$g(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-z)^2}{2\sigma^2}}$$

Standard Deviation: σ

Height: $\frac{1}{\sqrt{2\pi\sigma^2}}$

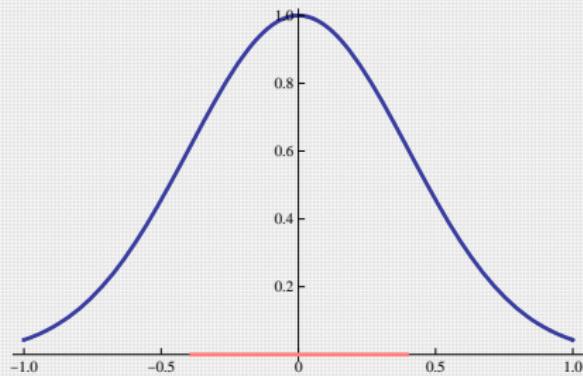
Center: z



Standardized Gaussian Kernel

Definition

$$g(x) = e^{-\pi(x-z)^2}$$

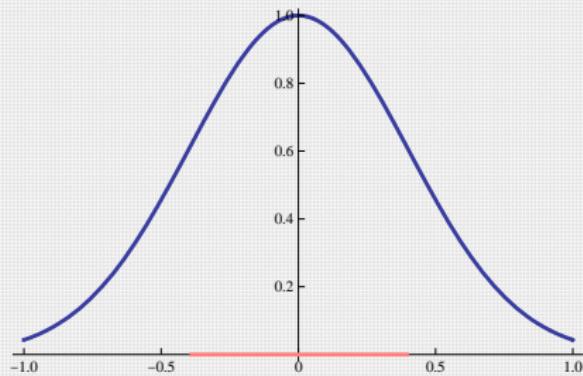


Standardized Gaussian Kernel

Definition

$$g(x) = e^{-\pi(x-z)^2}$$

Standard Deviation: $\sigma_0 = \frac{1}{\sqrt{2\pi}}$



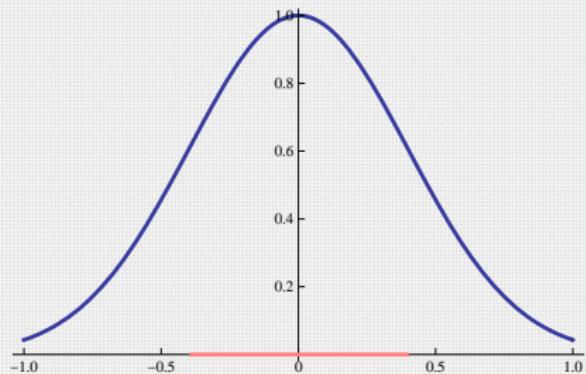
Standardized Gaussian Kernel

Definition

$$g(x) = e^{-\pi(x-z)^2}$$

Standard Deviation: $\sigma_0 = \frac{1}{\sqrt{2\pi}}$

Height: 1



Standardized Gaussian Kernel

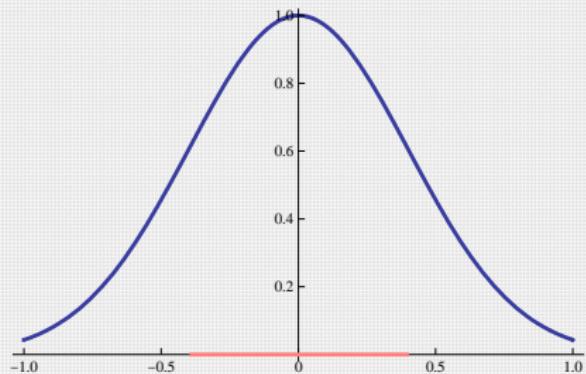
Definition

$$g(x) = e^{-\pi(x-z)^2}$$

Standard Deviation: $\sigma_0 = \frac{1}{\sqrt{2\pi}}$

Height: 1

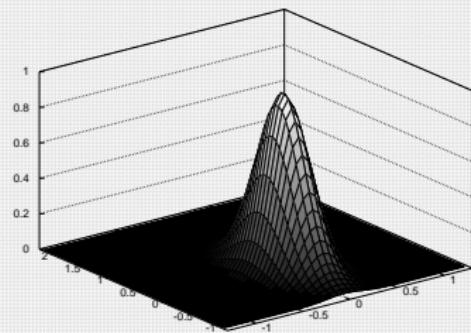
Center: z



n -Dimensional Gaussian Kernel

Definition

$$g(x) = e^{-\sum_{j=0}^n \frac{(x_j - z_j)^2}{2\sigma_j^2}}$$

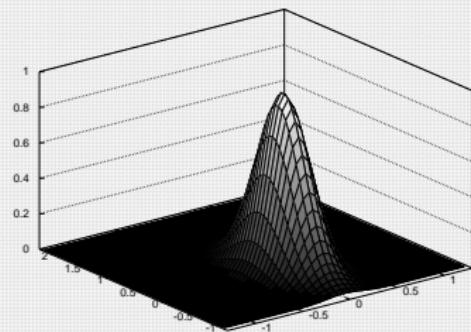


n -Dimensional Gaussian Kernel

Definition

$$g(x) = e^{-\sum_{j=0}^n \frac{(x_j - z_j)^2}{2\sigma_j^2}}$$

Width in j^{th} direction: σ_j



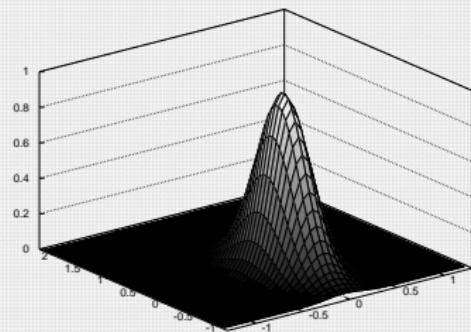
n -Dimensional Gaussian Kernel

Definition

$$g(x) = e^{-\sum_{j=0}^n \frac{(x_j - z_j)^2}{2\sigma_j^2}}$$

Width in j^{th} direction: σ_j

Height: 1



n -Dimensional Gaussian Kernel

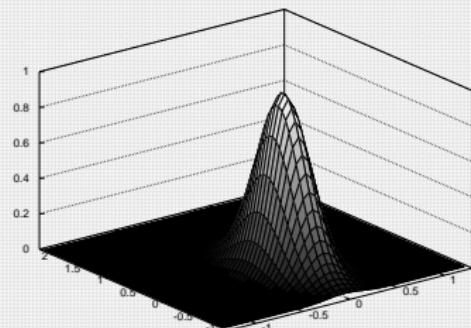
Definition

$$g(x) = e^{-\sum_{j=0}^n \frac{(x_j - z_j)^2}{2\sigma_j^2}}$$

Width in j^{th} direction: σ_j

Height: 1

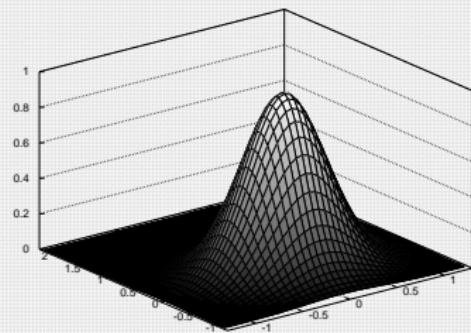
Center: z



n -Dimensional Isotropic Gaussian Kernel

Definition

$$g(x) = e^{-\pi||x-z||^2}$$

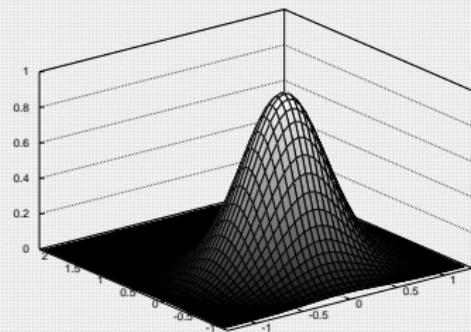


n -Dimensional Isotropic Gaussian Kernel

Definition

$$g(x) = e^{-\pi||x-z||^2}$$

Width: $\frac{1}{\sqrt{2\pi}}$

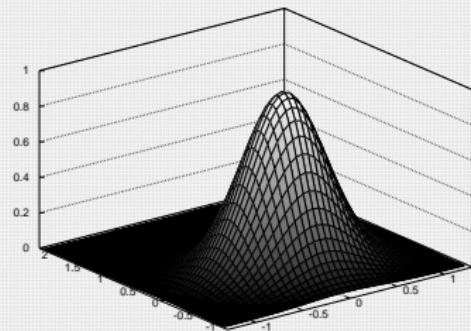


n -Dimensional Isotropic Gaussian Kernel

Definition

$$g(x) = e^{-\pi||x-z||^2}$$

Width: $\frac{1}{\sqrt{2\pi}}$
Height: 1



n -Dimensional Isotropic Gaussian Kernel

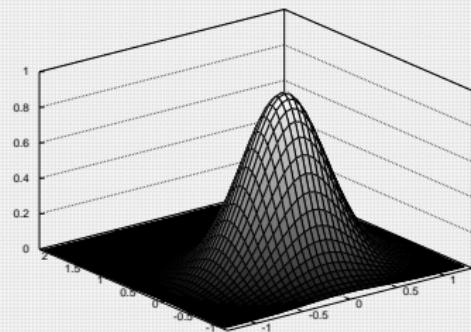
Definition

$$g(x) = e^{-\pi||x-z||^2}$$

Width: $\frac{1}{\sqrt{2\pi}}$

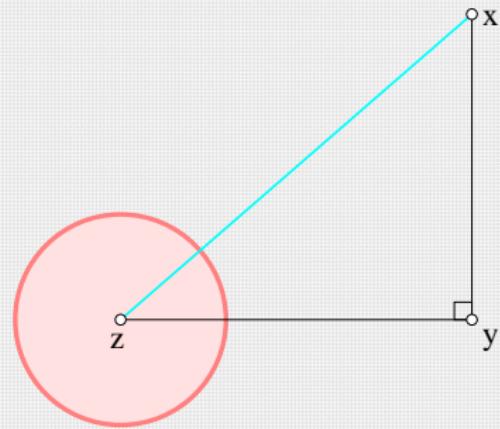
Height: 1

Center: z



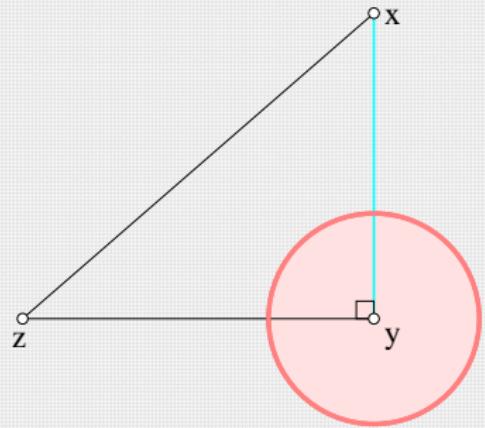
Separability of the Gaussian Kernel

$$e^{-\pi||x-z||^2}$$



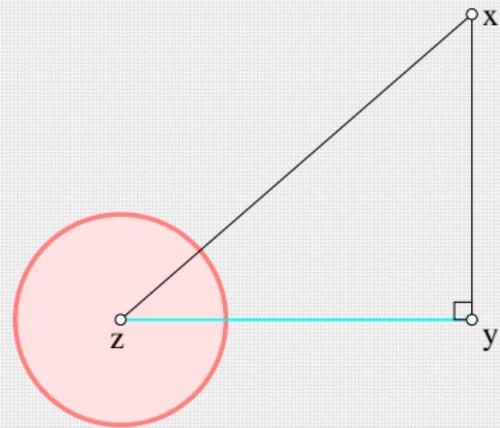
Separability of the Gaussian Kernel

$$e^{-\pi||x-z||^2} = e^{-\pi||x-y||^2}$$



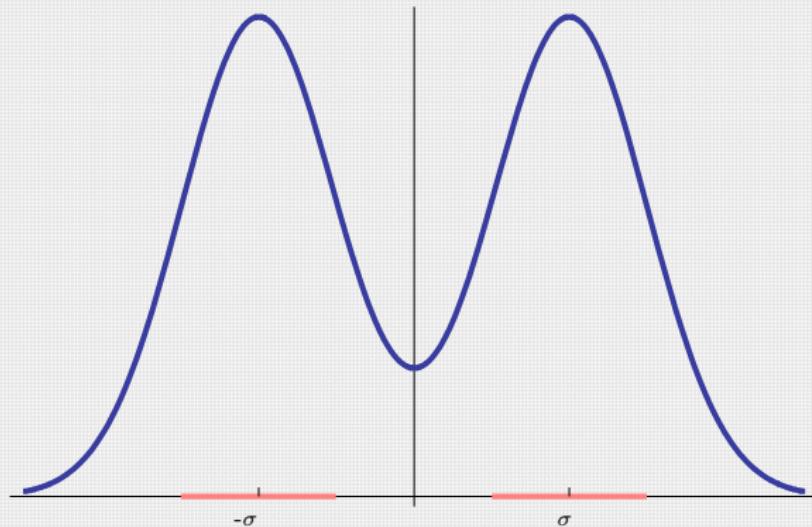
Separability of the Gaussian Kernel

$$e^{-\pi||x-z||^2} = e^{-\pi||x-y||^2} e^{-\pi||y-z||^2}$$



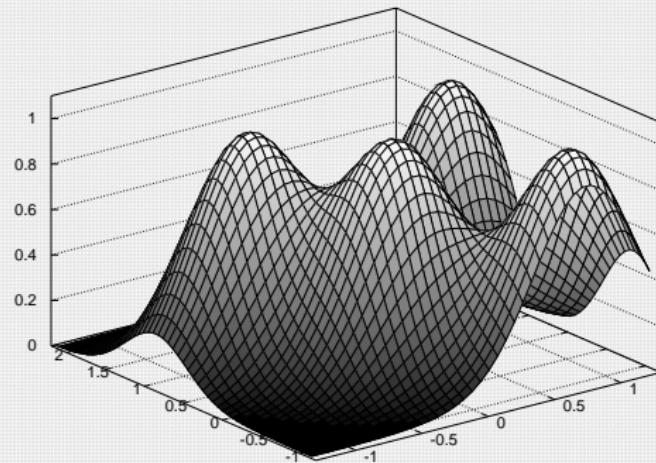
Gaussian Mixture Model (GMM)

A *Gaussian Mixture Model* is the sum of Gaussian Kernels.



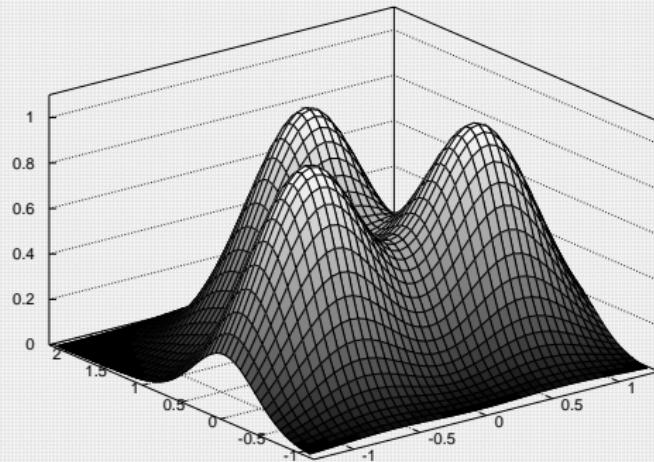
Gaussian Mixture Model (GMM)

A *Gaussian Mixture Model* is the sum of Gaussian Kernels.



Gaussian Mixture Model (GMM)

A *Gaussian Mixture Model* is the sum of Gaussian Kernels.

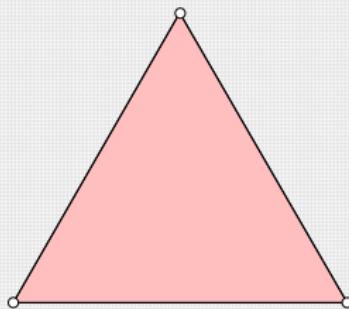


Motivating Goal

To characterize the critical values of an isotropic Gaussian Mixture Model.

Standard n -Simplex, Δ^n

A simplex is the convex hull of $n + 1$ vertices.

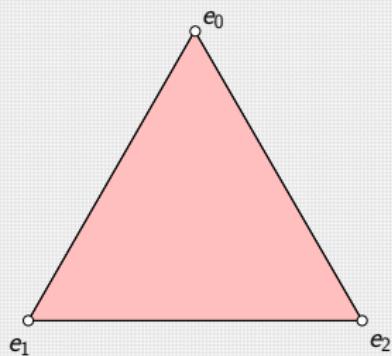


Standard n -Simplex, Δ^n

A simplex is the convex hull of $n + 1$ vertices.

The *standard simplex* has the standard basis elements as the vertices:

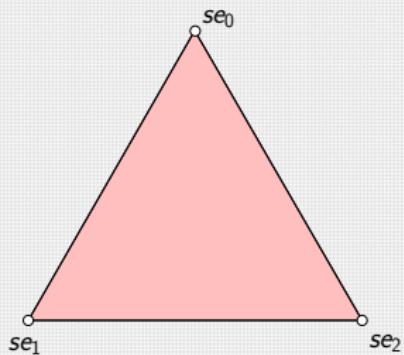
$$e_0, e_1, \dots, e_n.$$



Scaled n -Simplex, $s\Delta^n$

The *Scaled Standard n -Simplex* in \mathbb{R}^{n+1} is defined by the $n + 1$ standard basis elements, scaled by a factor s

$$se_0, se_1, \dots, se_n.$$



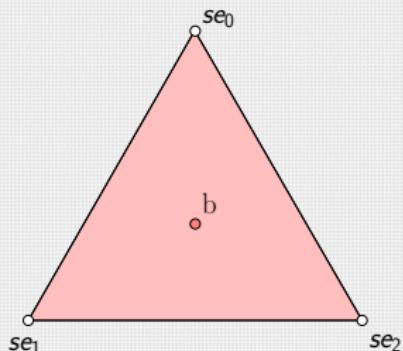
Scaled n -Simplex, $s\Delta^n$

The *Scaled Standard n -Simplex* in \mathbb{R}^{n+1} is defined by the $n + 1$ standard basis elements, scaled by a factor s

$$se_0, se_1, \dots, se_n.$$

The *barycenter* is the average vertex position:

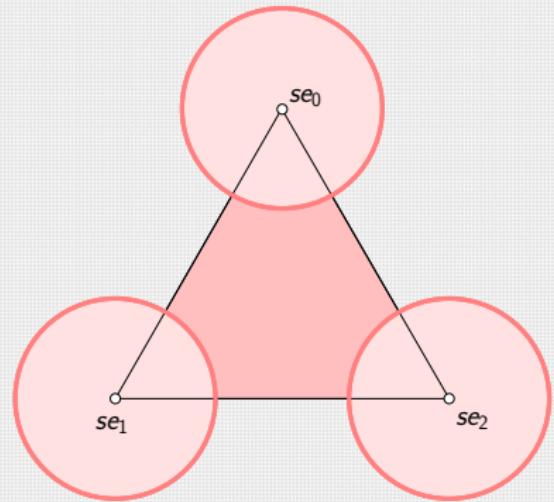
$$\left(\frac{s}{n+1}, \frac{s}{n+1}, \dots, \frac{s}{n+1} \right).$$



Standard n -Design

Definition

The *Standard n -Design* is the GMM with centers at the $n + 1$ vertices of the scaled n -simplex.



Motivating Goal

To characterize the critical values of an isotropic Gaussian Mixture Model.

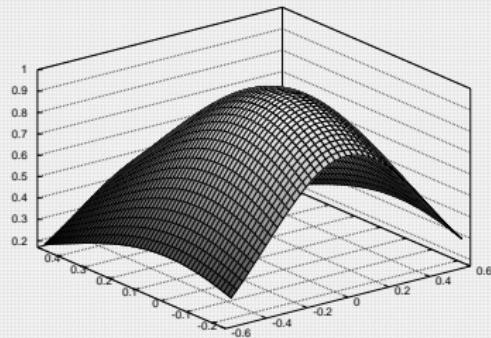
Motivating Goal

To characterize the modes of the standard n -Design.

How many modes?

A *mode* of a GMM is a local maximum.

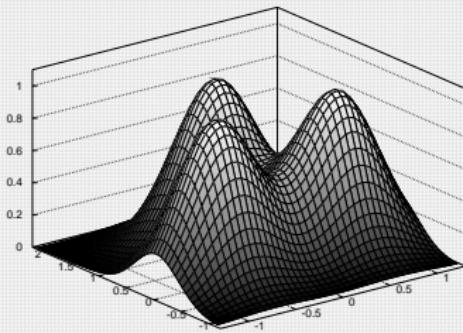
With the standard 3-design, we can see 1 mode.



How many modes?

A *mode* of a GMM is a local maximum.

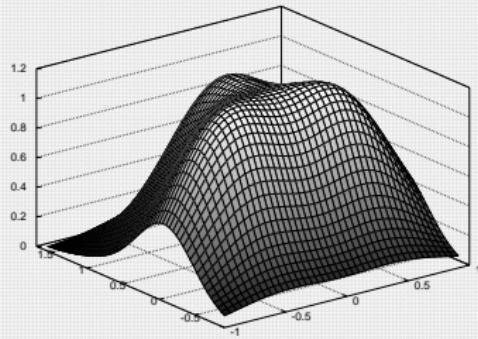
With the standard 3-design, we can see 3 modes.



How many modes?

A *mode* of a GMM is a local maximum.

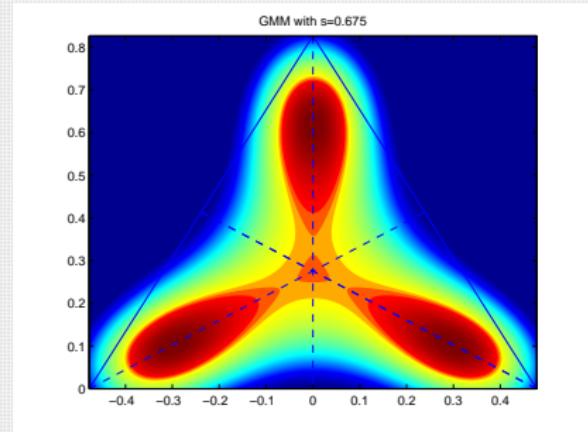
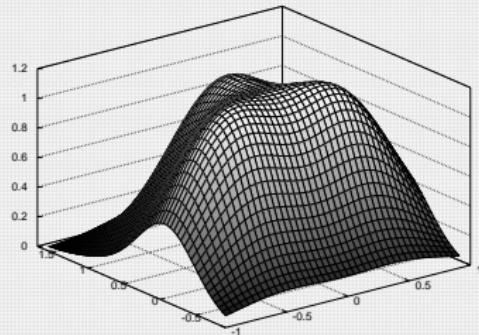
With the standard 3-design, we can see 4 modes.



How many modes?

A *mode* of a GMM is a local maximum.

With the standard 3-design, we can see 4 modes.



Ghost Interval

Question

For which values of s do we see $n + 2$ modes?

The Barycenter

Lemma

The barycenter of $s\Delta^n$ is a mode for $s < U_n = \sqrt{\frac{n+1}{2\pi}}$, and a saddle of index 1 for $s > U_n$.

The Barycenter

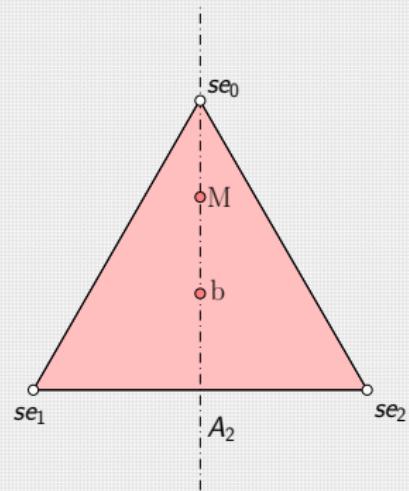
Lemma

The barycenter of $s\Delta^n$ is a mode for $s < U_n = \sqrt{\frac{n+1}{2\pi}}$, and a saddle of index 1 for $s > U_n$.

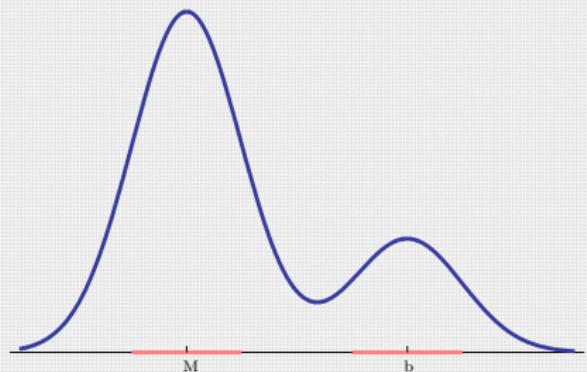
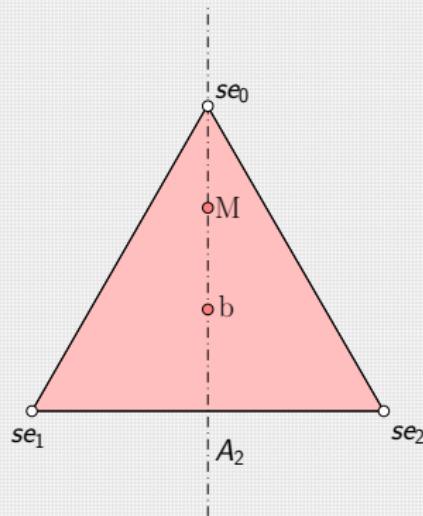
Proof: Look at the Hessian with respect to the standard basis elements and compute the eigenvalues.

Axis

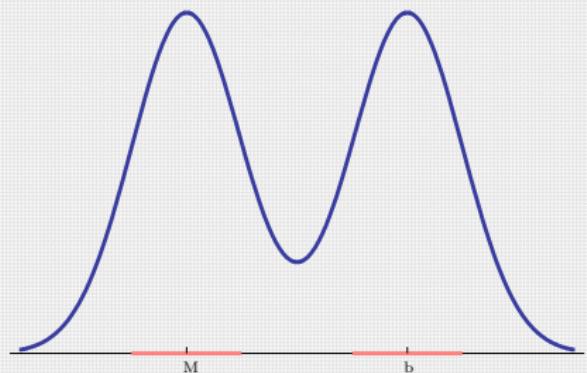
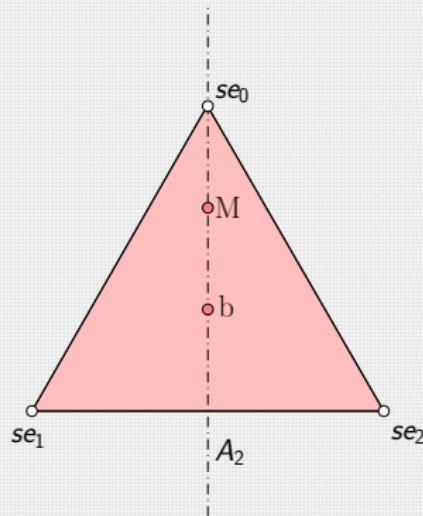
The axis A_n connects a vertex with the barycenter of the complementary $n - 1$ -face.



Restriction to the Axis



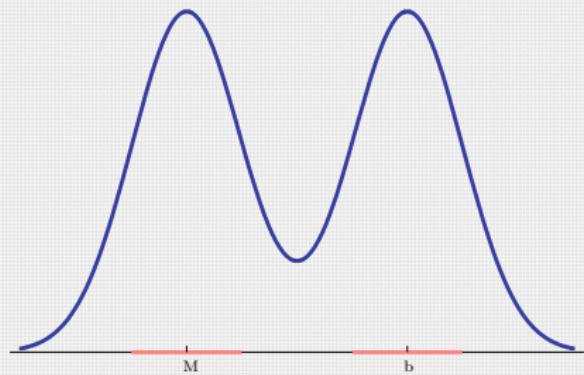
Restriction to the Axis



Balancing Scale Factor

Lemma

$$B_n = \sqrt{\frac{n \ln(n)}{(n-1)\pi}}.$$



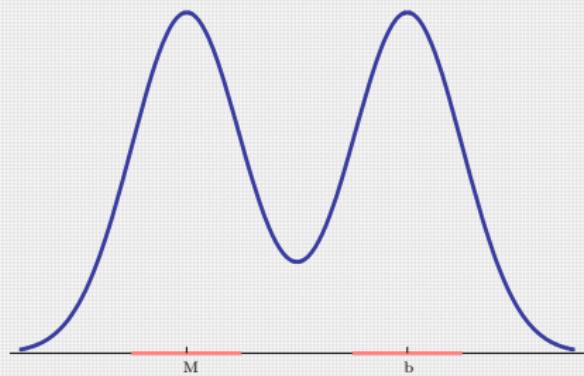
Balancing Scale Factor

Lemma

$$B_n = \sqrt{\frac{n \ln(n)}{(n-1)\pi}}.$$

Lemma

$$B_n < U_n = \sqrt{\frac{n+1}{2\pi}}$$



Balancing Scale Factor

Lemma

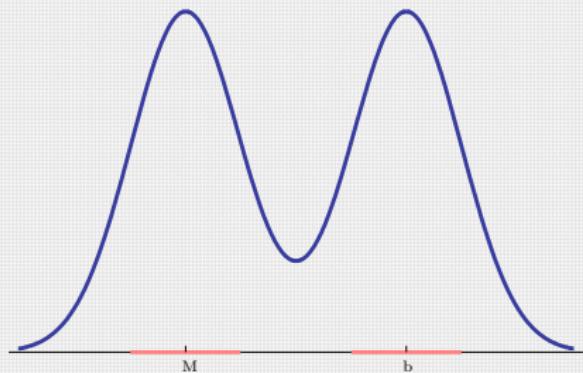
$$B_n = \sqrt{\frac{n \ln(n)}{(n-1)\pi}}.$$

Lemma

$$B_n < U_n = \sqrt{\frac{n+1}{2\pi}}$$

Corollary

If $s = B_n$, then there is a mode at the barycenter.

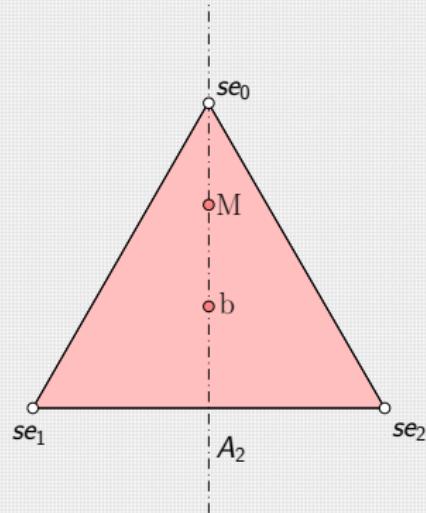


Witnessing the Modes

$$G(b) = e^{-\pi h_\ell} G_\ell(b_\ell) + g_0(b)$$

Theorem

If $s < U_n$, then A_n witnesses two modes.

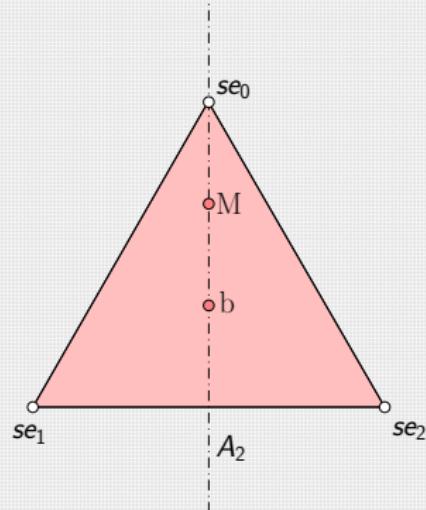


Witnessing the Modes

$$\begin{aligned} G(b) &= e^{-\pi h_\ell} G_\ell(b_\ell) + g_0(b) \\ e^{-\pi h_\ell} &> e^{-\pi(h_\ell + \delta)} \end{aligned}$$

Theorem

If $s < U_n$, then A_n witnesses two modes.



Witnessing the Modes

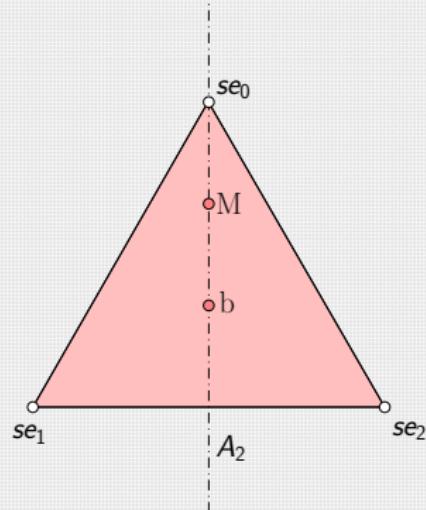
$$G(b) = e^{-\pi h_\ell} G_\ell(b_\ell) + g_0(b)$$

$$e^{-\pi h_\ell} > e^{-\pi(h_\ell + \delta)}$$

$$g_0(b) < g_0(M)$$

Theorem

If $s < U_n$, then A_n witnesses two modes.



Witnessing the Modes

$$G(b) = e^{-\pi h_\ell} G_\ell(b_\ell) + g_0(b)$$

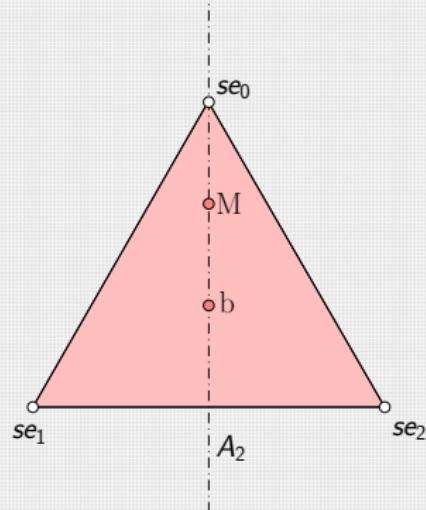
$$e^{-\pi h_\ell} > e^{-\pi(h_\ell + \delta)}$$

$$g_0(b) < g_0(M)$$

$$G(M) = e^{-\pi(h_\ell + \delta)} G_\ell(b_\ell) + g_0(M).$$

Theorem

If $s < U_n$, then A_n witnesses two modes.



Ghost Interval

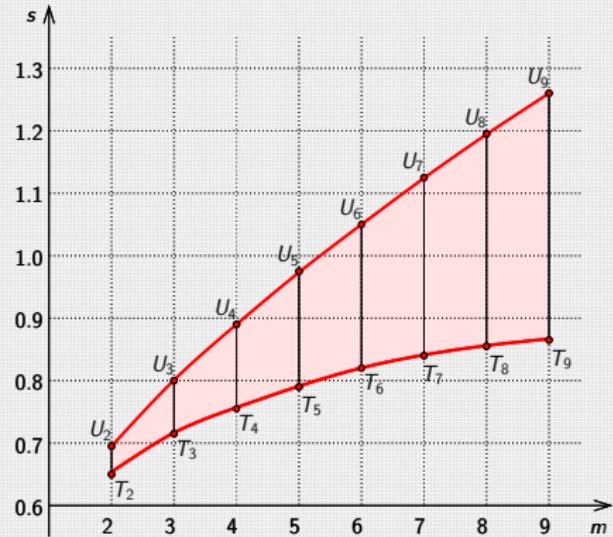
Question

For which values of s do we see $n + 2$ modes?

Resilience of Ghost

Theorem

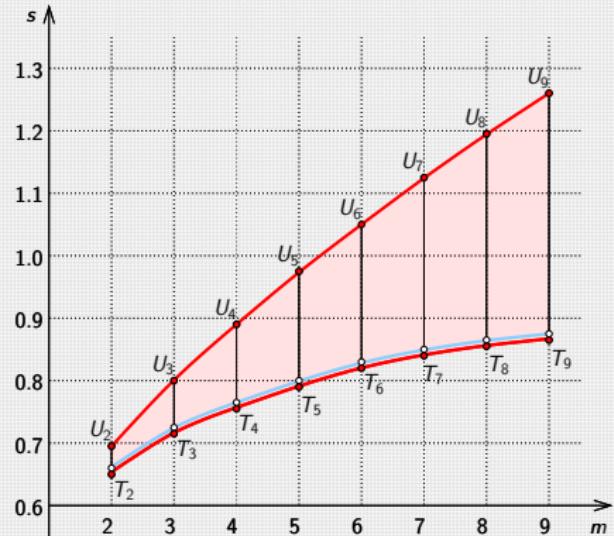
There are $n + 2$ modes for $s \in (T_n, U_n)$, where T_n is the smallest s such that A_n has two maxima.



Resilience of Ghost

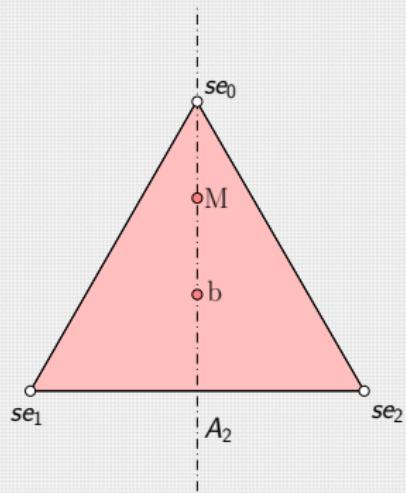
Theorem

There are $n + 2$ modes for $s \in (T_n, U_n)$, where T_n is the smallest s such that A_n has two maxima.



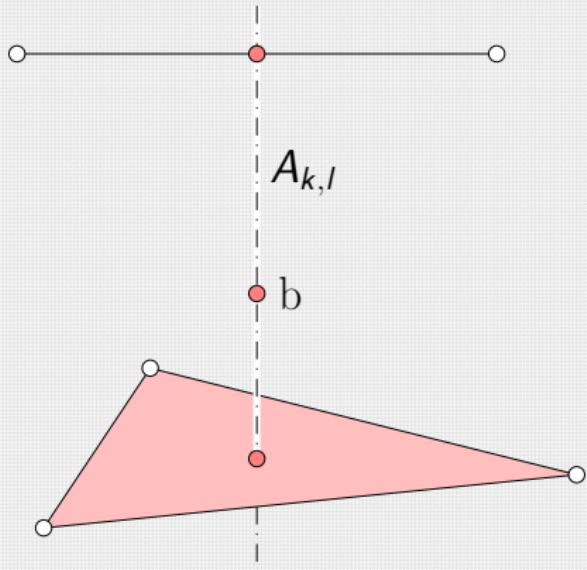
Axis

The axis A_n connects a vertex with the barycenter of the complementary $n - 1$ -face.



Axis

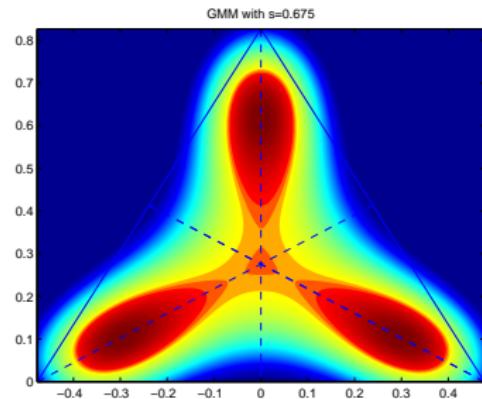
The axis $A_{k,\ell}$ connects the barycenter of a k -face with the barycenter of the complementary ℓ -face.



Critical Points

Lemma

All critical points of the standard n -design lies on an axis of $s\Delta^n$.

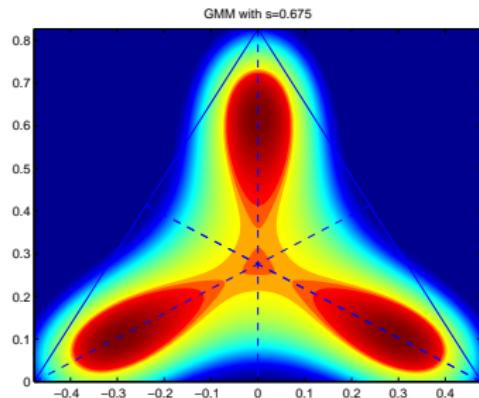


Critical Points

Lemma

All critical points of the standard n -design lies on an axis of $s\Delta^n$.

Proof: Relies on the strict convexity of e^{-x} .



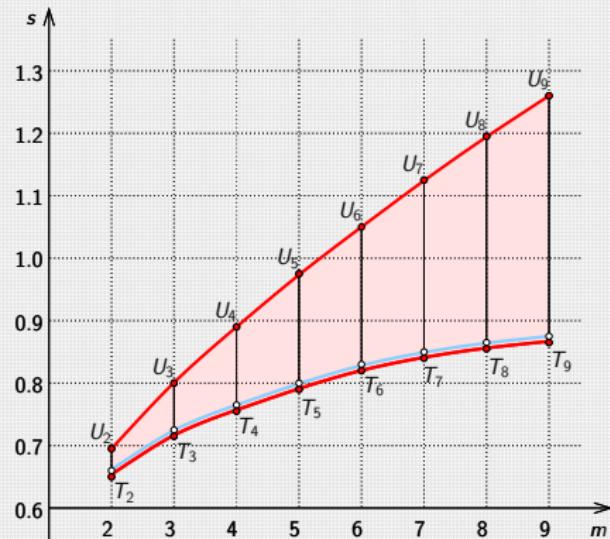
Motivating Goal

To characterize the critical values of an isotropic Gaussian Mixture Model.

Witnessing Critical Values

Theorem

If $s \in (T_n, U_n)$, then A_n witnesses two modes.



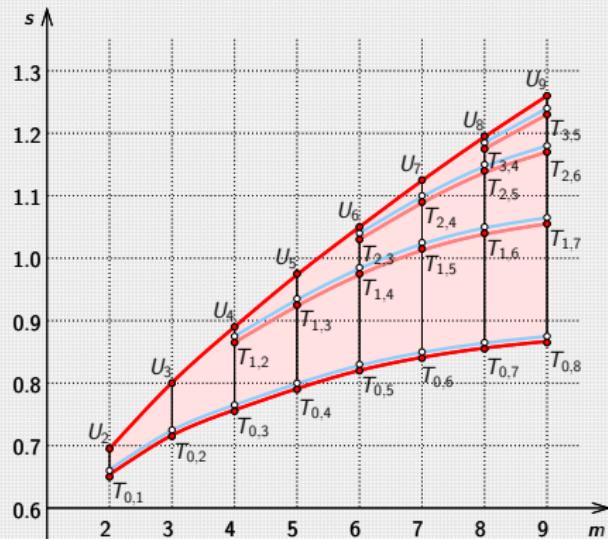
Witnessing Critical Values

Theorem

If $s \in (T_n, U_n)$, then A_n witnesses two modes.

Theorem

If $s \in (T_{k,l}, U_n)$, then $A_{k,l}$ witnesses three critical values.



Witnessing Critical Values

Theorem

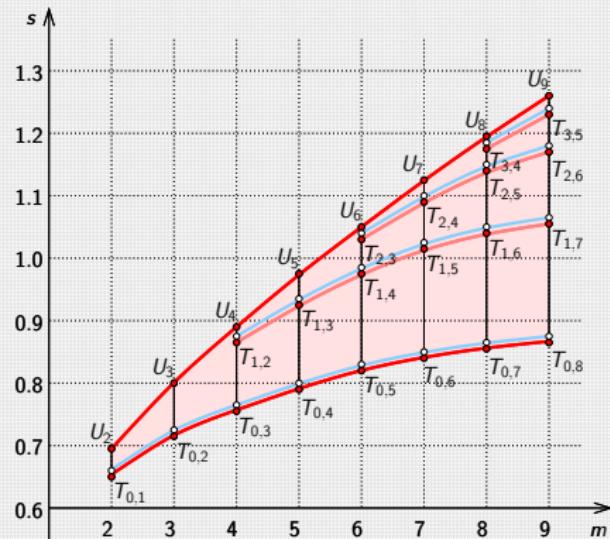
If $s \in (T_n, U_n)$, then A_n witnesses two modes.

Theorem

If $s \in (T_{k,l}, U_n)$, then $A_{k,l}$ witnesses three critical values.

Question

Can these critical values be modes?



No More Ghosts!

These extra critical points are not modes.

Conjecture

For $s \in (T_{k,\ell}, U_n)$, the critical points on $A_{k,l}$ have index $n+1$, $\ell+1$, and $\ell+2$.

No More Ghosts!

These extra critical points are not modes.

Conjecture

For $s \in (T_{k,\ell}, U_n)$, the critical points on $A_{k,l}$ have index $n+1$, $\ell+1$, and $\ell+2$.

Why? Strong numerical evidence.

Summary

A GMM can have:

- more modes than components
- an exponential number of critical points.

Thank You

- Herbert Edelsbrunner
- Günter Rote
- You!

Questions?

Contact Information

Brittany Terese Fasy
brittany@cs.duke.edu
www.fasy.us