Add Isotropic Gaussian Kernels at Own Risk

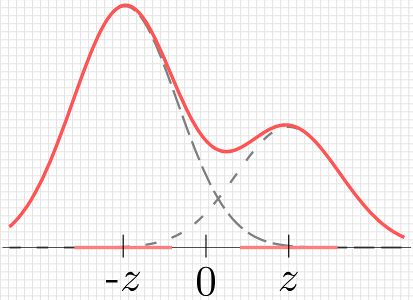
More and More Resilient Modes in Higher Dimensions

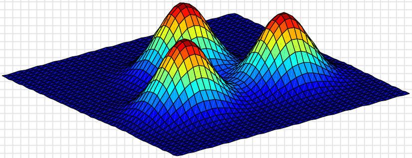
Herbert Edelsbrunner, BRITTANY TERESE FASY, and Günter Rote

Symposium on Computational Geometry 2012
Chapel Hill, North Carolina

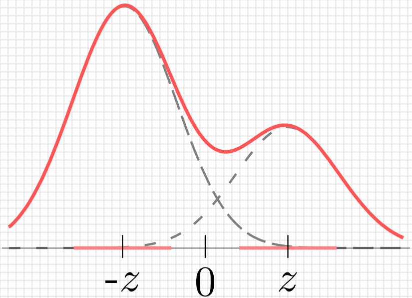
18 June 2012

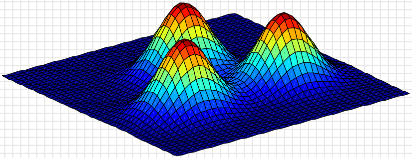
Counting Modes and Critical Points





Counting Modes and Critical Points

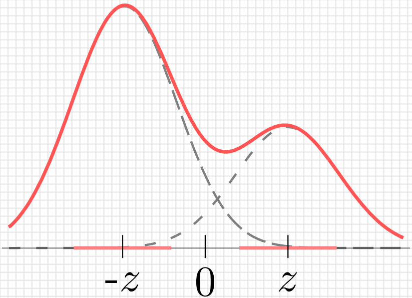


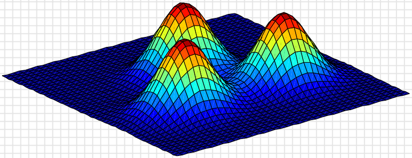


Definition

A *critical point* is a point with a zero gradient.

Counting Modes and Critical Points





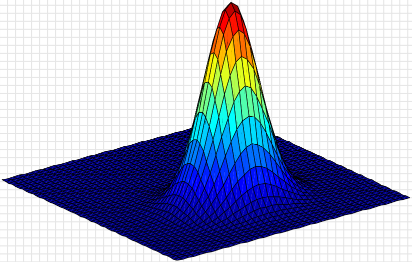
Definition

A *critical point* is a point with a zero gradient.

Definition

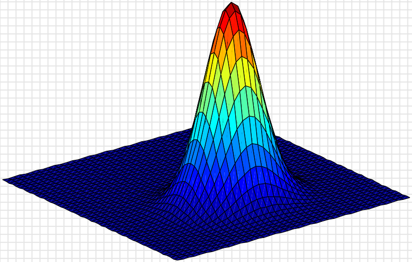
A *mode* is a local maximum.

How Many Modes (Local Maxima)?



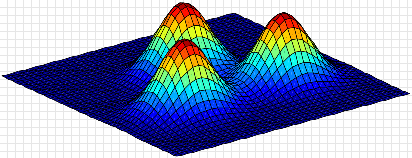
How Many Modes (Local Maxima)?

In the begining, we see 1 local maximum.



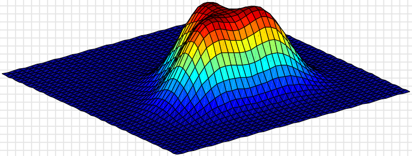
How Many Modes (Local Maxima)?

At the end, we see 3 local maxima.



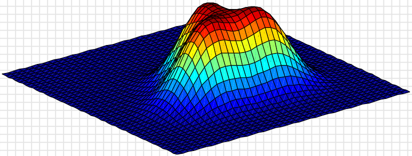
How Many Modes (Local Maxima)?

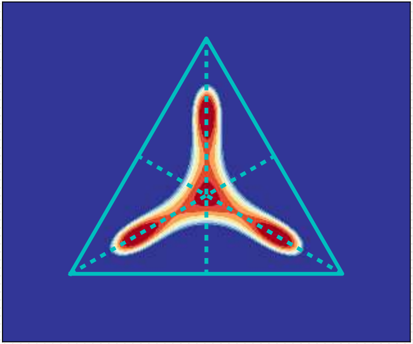
In the middle, we see 4 local maxima.



How Many Modes (Local Maxima)?

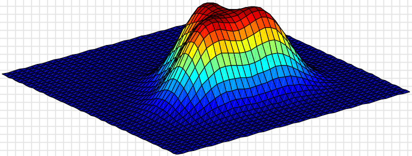
In the middle, we see 4 local maxima.

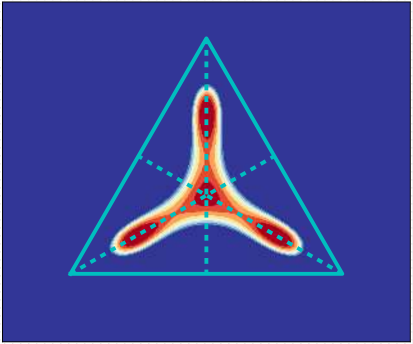




How Many Modes (Local Maxima)?

In the middle, we see 4 local maxima.





Existence proven in [M. Carreira-Perpiñán and C. Williams, Scotland 2003].

Brief Overview

- Define Gaussian kernel and mixture.

Brief Overview

- Define Gaussian kernel and mixture.
- Analyze 1-dimensional mixtures.

Brief Overview

- Define Gaussian kernel and mixture.
- Analyze 1-dimensional mixtures.
- Locate and count all critical points of an n-dimensional mixture.

Brief Overview

- Define Gaussian kernel and mixture.
- Analyze 1-dimensional mixtures.
- Locate and count all critical points of an n-dimensional mixture.
- Locate and count all modes of an n-dimensional mixtures.

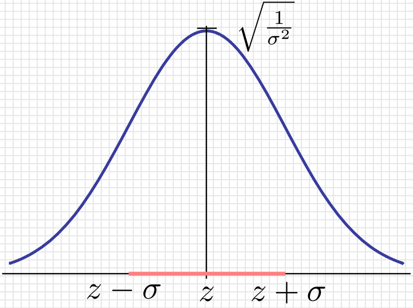
Brief Overview

- Define Gaussian kernel and mixture.
- Analyze 1-dimensional mixtures.
- Locate and count all critical points of an n-dimensional mixture.
- Locate and count all modes of an n-dimensional mixtures.
- (Describe the resilience of the ghost mode.)

Gaussian Kernel

Definition

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

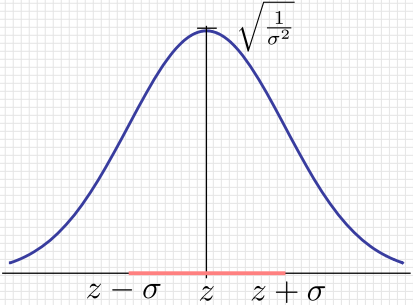


Gaussian Kernel

Definition

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

Center: z



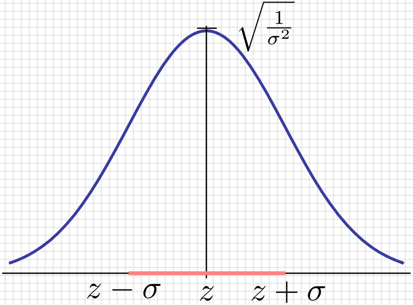
Gaussian Kernel

Definition

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

Center: z

Standard Deviation: σ



Gaussian Kernel

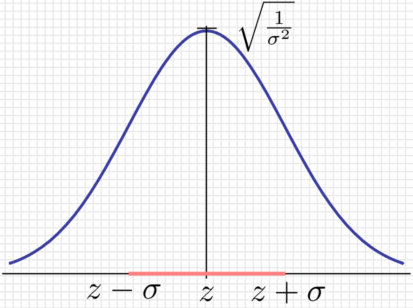
Definition

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

Center: z

Standard Deviation: σ

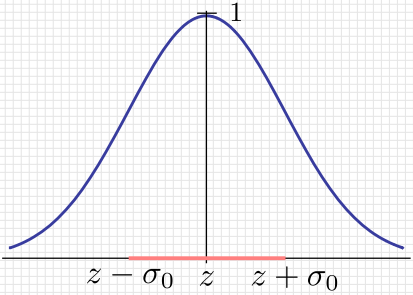
Height: $\frac{1}{\sqrt{2\pi\sigma^2}}$



Standardized Gaussian Kernel

Definition

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

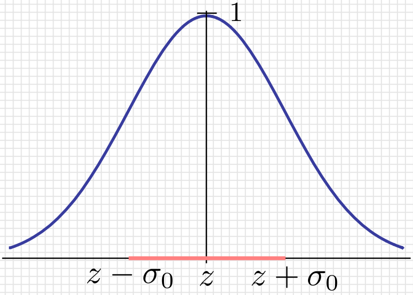


Standardized Gaussian Kernel

Definition

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

Center: z

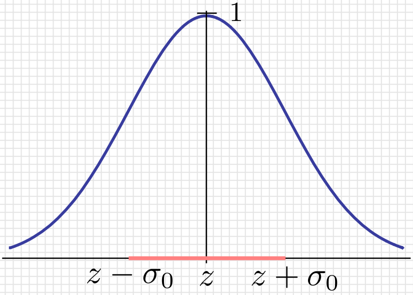


Standardized Gaussian Kernel

Definition

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

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



Standardized Gaussian Kernel

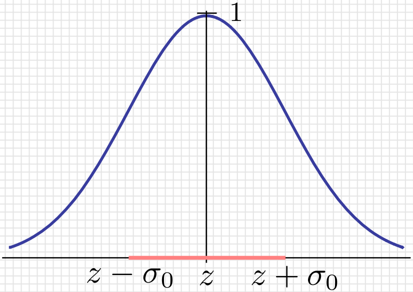
Definition

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

Center: z

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

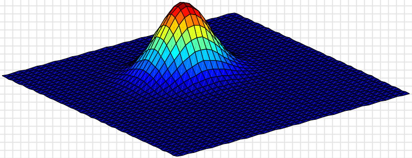
Height: 1



n-Dimensional Isotropic Gaussian Kernel

Definition

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

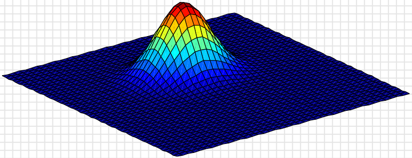


n-Dimensional Isotropic Gaussian Kernel

Definition

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

Center: z

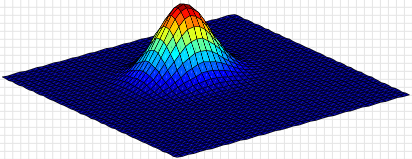


n-Dimensional Isotropic Gaussian Kernel

Definition

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

Center: z
Width: $\sigma_0 = \frac{1}{\sqrt{2\pi}}$

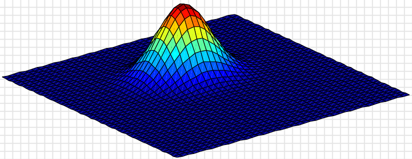


n-Dimensional Isotropic Gaussian Kernel

Definition

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

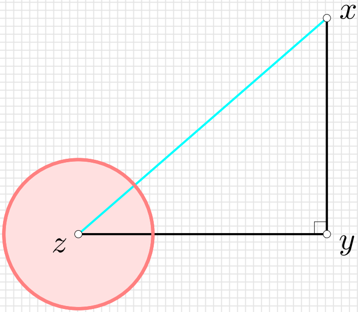
Center: z
Width: $\sigma_0 = \frac{1}{\sqrt{2\pi}}$
Height: 1



Separability of the Gaussian Kernel

Separability Lemma

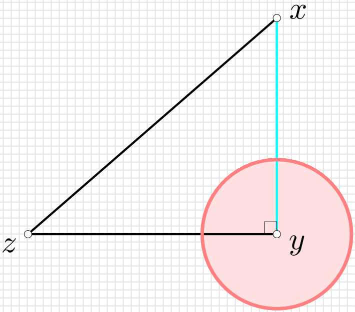
$e^{-\pi||x-z||^2}$



Separability of the Gaussian Kernel

Separability Lemma

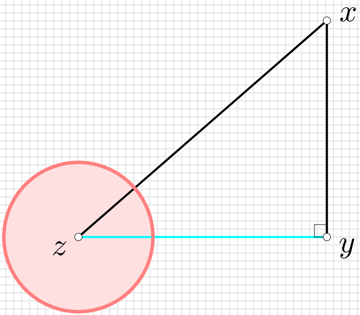
$$e^{-\pi||x-z||^2} = e^{-\pi||x-y||^2}$$



Separability of the Gaussian Kernel

Separability Lemma

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

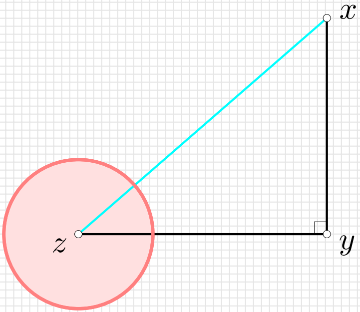


Separability of the Gaussian Kernel

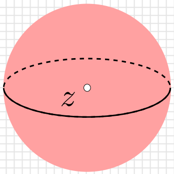
Separability Lemma

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

$$||x - z||^2 = ||x - y||^2 + ||y - z||^2$$



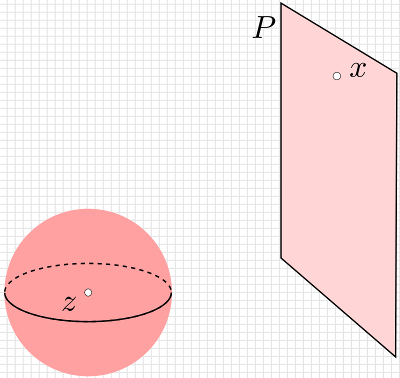
Restrictions of Kernels



Restrictions of Kernels

Definition

A *restriction* of g_z is the evaluation of the function on a lower-dimensional plane P.

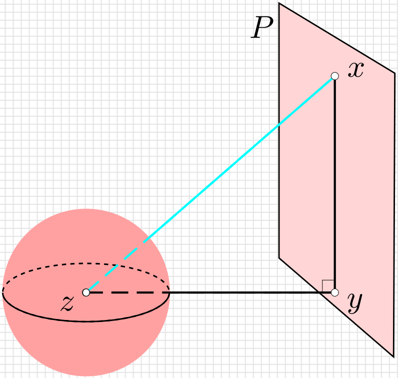


Restrictions of Kernels

Definition

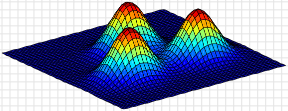
A *restriction* of g_z is the evaluation of the function on a lower-dimensional plane P.

$$g_z|_P(x) = c_z g_y(x).$$



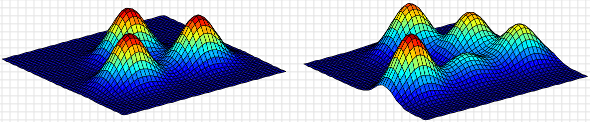
Gaussian Mixture

A *Gaussian mixture* is the sum of Gaussian kernels.



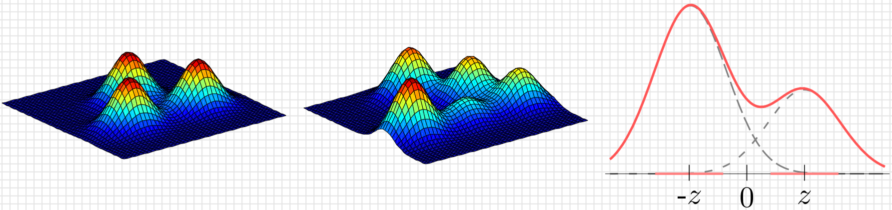
Gaussian Mixture

A *Gaussian mixture* is the sum of Gaussian kernels.

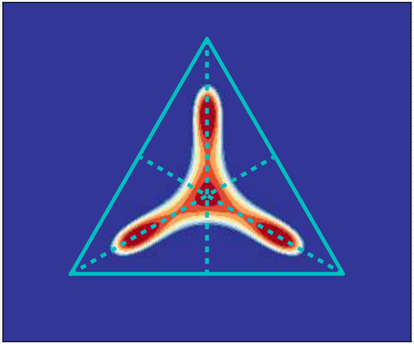


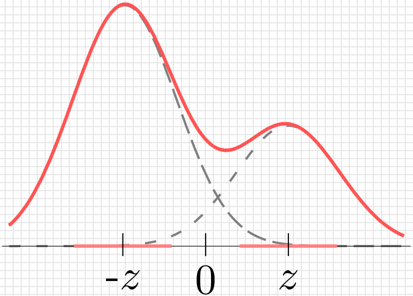
Gaussian Mixture

A *Gaussian mixture* is the sum of Gaussian kernels.



Restrictions of Mixtures





No Ghost Modes

Theorem

In $\mathbb{R}^1$, the number of modes is at most the number of components.

No Ghost Modes

Theorem

In $\mathbb{R}^1$, the number of modes is at most the number of components.

- Balanced sum of two kernels: [Burke, 1956].

No Ghost Modes

Theorem

In $\mathbb{R}^1$, the number of modes is at most the number of components.

- Balanced sum of two kernels: [Burke, 1956].
- Weighted sum of two kernels: [Behboodian, 1970].

No Ghost Modes

Theorem

In $\mathbb{R}^1$, the number of modes is at most the number of components.

- Balanced sum of two kernels: [Burke, 1956].
- Weighted sum of two kernels: [Behboodian, 1970].
- General sum: [M. Carreira-Perpiñán and C. Williams, LNCS 2003] relies heavily on [Silverman, 1981].

No Ghost Modes

Theorem

In $\mathbb{R}^1$, the number of modes is at most the number of components.

- Balanced sum of two kernels: [Burke, 1956].
- Weighted sum of two kernels: [Behboodian, 1970].
- General sum: [M. Carreira-Perpiñán and C. Williams, LNCS 2003] relies heavily on [Silverman, 1981].

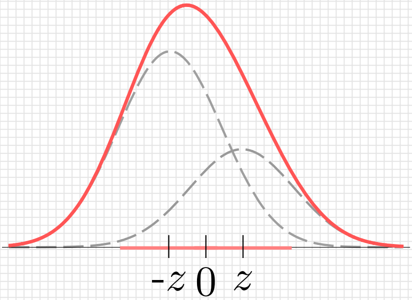
Question

When is the transition between having one mode and two?

Weighted Gaussian Mixture

$$G_w(x) = c_k g_{-z}(x) + c_\ell g_z(x).$$

Weighted Gaussian Mixture

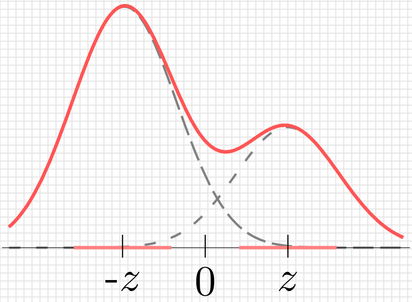


$$G_w(x) = c_k g_{-z}(x) + c_\ell g_z(x).$$

The Weighted Mixture

1. If z is small enough, then G_w has one critical point.

Weighted Gaussian Mixture

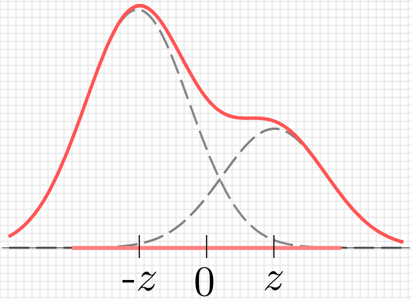


$$G_w(x) = c_k g_{-z}(x) + c_\ell g_z(x).$$

The Weighted Mixture

1. If z is small enough, then G_w has one critical point.
2. If z is large enough, then G_w has three critical points.

Weighted Gaussian Mixture



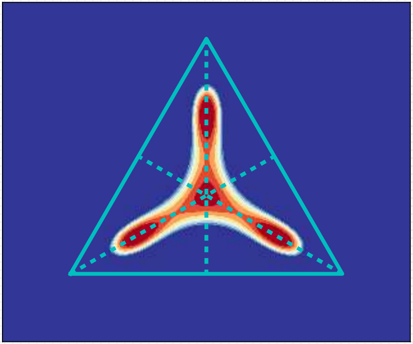
$$G_w(x) = c_k g_{-z}(x) + c_\ell g_z(x).$$

The Weighted Mixture

1. If z is small enough, then G_w has one critical point.
2. If z is large enough, then G_w has three critical points.
3. G_w has exactly 2 critical points when $\frac{c_k}{c_\ell} = r(x) + 1$.

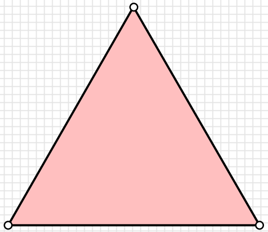
Counting Modes in $\mathbb{R}^n$

For $n \geq 2$, there can be more modes than components of a Gaussian mixture in $\mathbb{R}^n$.



Standard n-Simplex, Δ^n

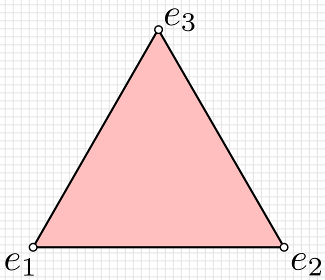
An n-simplex is the convex hull of $n + 1$ vertices.



Standard n-Simplex, Δ^n

An n-simplex is the convex hull of $n+1$ vertices.
The *standard n-simplex* has the standard basis elements as the vertices:

$$e_1, e_2, \ldots, e_{n+1}.$$

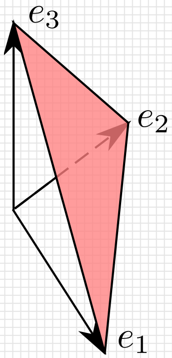


Standard n-Simplex, Δ^n

$\mathbb{R}^3$

An *n*-simplex is the convex hull of $n+1$ vertices.
The *standard n-simplex* has the standard basis elements as the vertices:

$$e_1, e_2, \ldots, e_{n+1}.$$

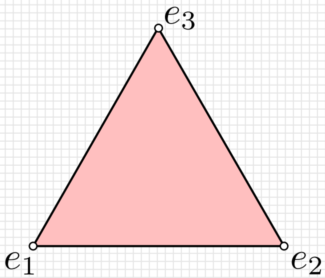


Standard n-Simplex, Δ^n

$\mathbb{R}^3$

An n-simplex is the convex hull of $n+1$ vertices.
The *standard n-simplex* has the standard basis elements as the vertices:

$$e_1, e_2, \ldots, e_{n+1}.$$

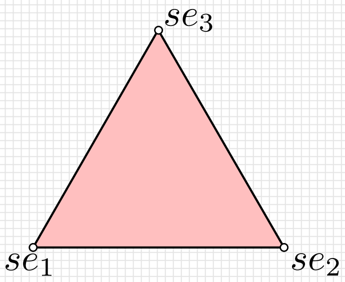


Scaled n-Simplex, $s\Delta^n$

$\mathbb{R}^3$

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

$$se_1, se_2, \ldots, se_{n+1}.$$



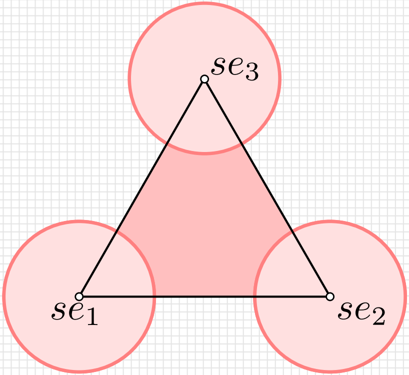
Scaled n-Design

$\mathbb{R}^3$

Definition

The *Scaled n-Design* is the Gaussian mixture with centers at the $n+1$ vertices of the scaled n-simplex:

$$G_s(x) = \sum_{i=1}^{n+1} g_{se_i}(x)$$

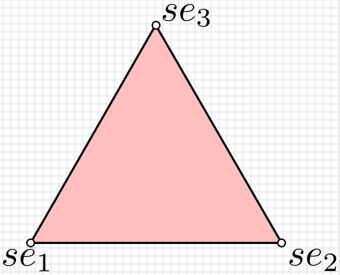


Scaled n-Simplex, $s\Delta^n$

$\mathbb{R}^3$

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

$$se_1, se_2, \ldots, se_{n+1}.$$



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

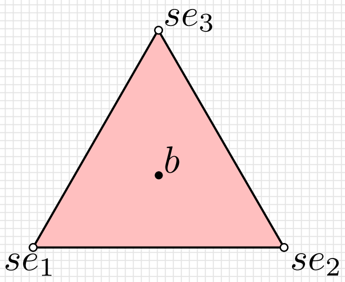
$\mathbb{R}^3$

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

$$se_1, se_2, \ldots, se_{n+1}.$$

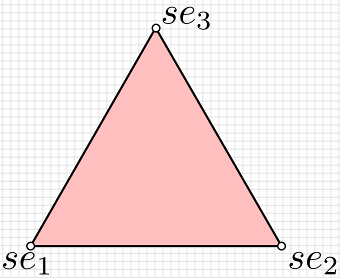
The *barycenter* is the average vertex position:

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



Complementary Faces

$\mathbb{R}^3$



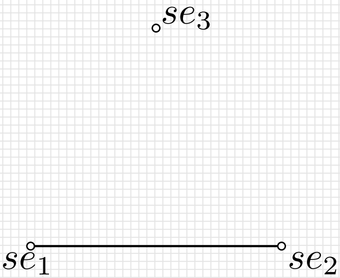
Complementary Faces

$\mathbb{R}^3$

We partition the vertices of the scaled n-simplex into two sets:

$$K = \{se_3\},$$
$$L = \{se_1, se_2\}.$$

Let $k = |K| - 1$ and $\ell = |L| - 1$.



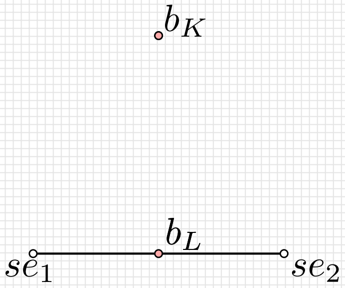
Complementary Faces

$\mathbb{R}^3$

We partition the vertices of the scaled n-simplex into two sets:

$$K = \{se_3\},$$
$$L = \{se_1, se_2\}.$$

Let $k = |K| - 1$ and $\ell = |L| - 1$.

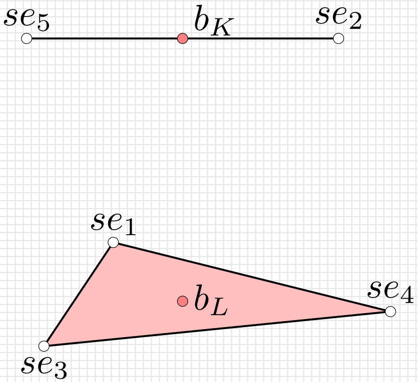


Complementary Faces

$\mathbb{R}^5$

We partition the vertices of the scaled *n*-simplex into two sets:

$$K = \{se_2, se_5\},$$
$$L = \{se_1, se_3, se_4\}.$$

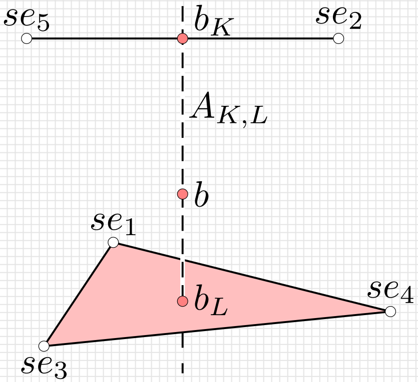


Let $k = |K| - 1$ and $\ell = |L| - 1$.

Location of Critical Points

$\mathbb{R}^5$

The axis $A_{K,L}$ is the line defined by b_K and b_L.



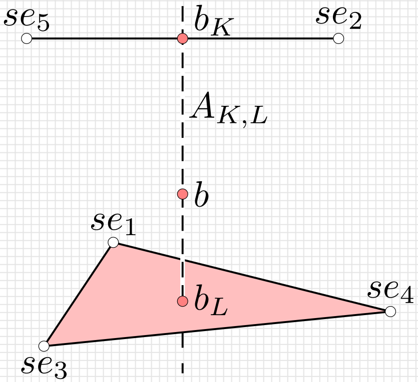
Location of Critical Points

$\mathbb{R}^5$

The axis $A_{K,L}$ is the line defined by b_K and b_L.

Location of Critical Values

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



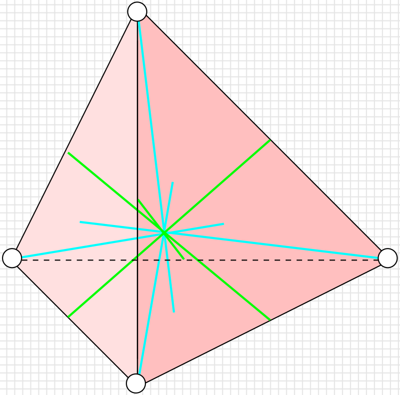
Location of Critical Points

$\mathbb{R}^4$

The axis $A_{K,L}$ is the line defined by b_K and b_L.

Location of Critical Values

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



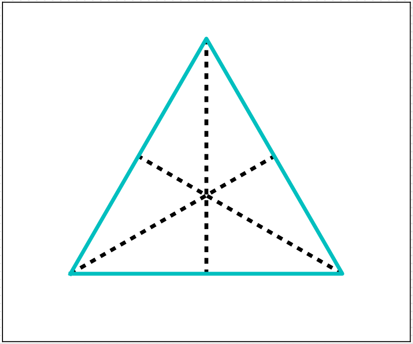
Location of Critical Points

$\mathbb{R}^3$

The axis $A_{K,L}$ is the line defined by b_K and b_L.

Location of Critical Values

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



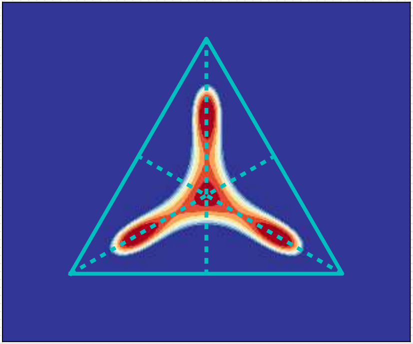
Location of Critical Points

$\mathbb{R}^3$

The axis $A_{K,L}$ is the line defined by b_K and b_L.

Location of Critical Values

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



Location of Critical Points

$\mathbb{R}^3$

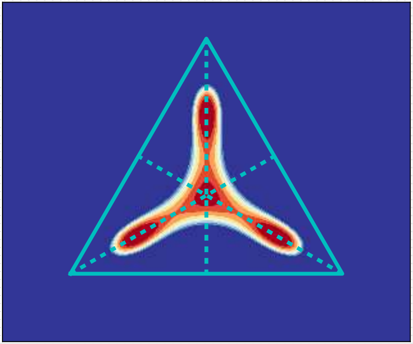
The axis $A_{K,L}$ is the line defined by b_K and b_L.

Location of Critical Values

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

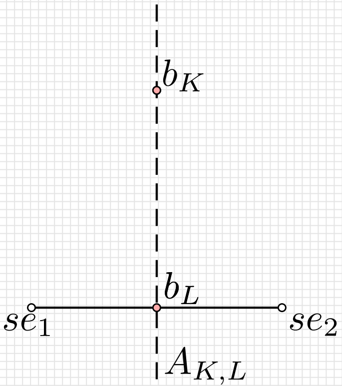
Proof

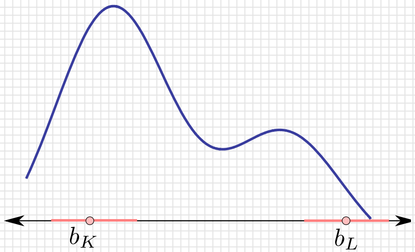
Assume a critical point x is not on an axis ...



Restriction to an Axis

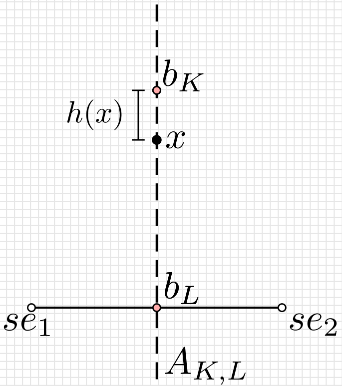
$\mathbb{R}^3$

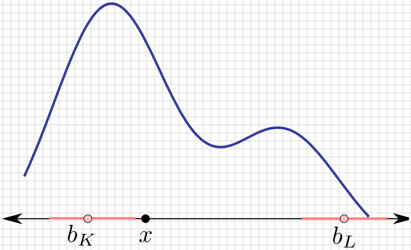




Restriction to an Axis

$\mathbb{R}^3$



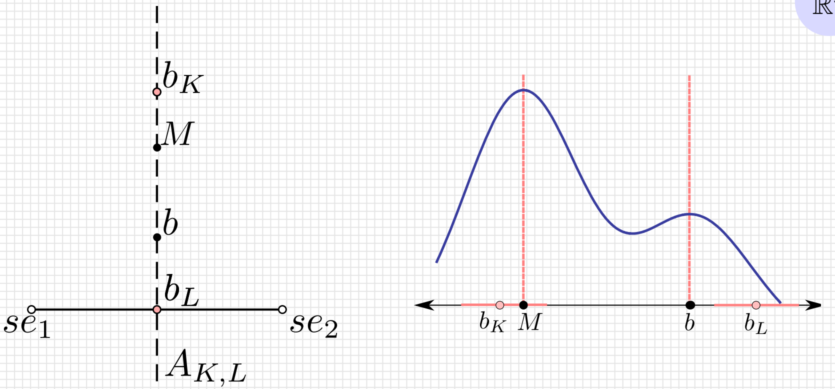


$$G_s|_A(x) = c_k e^{-\pi h(x)} + c_\ell e^{-\pi(D_{k,\ell} - h(x))},$$

where $c_k = (k+1) g_{se_i}(b_L), \quad c_\ell = (\ell+1) g_{se_j}(b_K).$

Restriction to an Axis

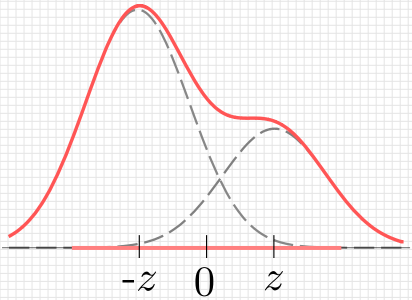
$\mathbb{R}^3$



$$G_s|_A(x) = c_k e^{-\pi h(x)} + c_\ell e^{-\pi(D_{k,\ell} - h(x))},$$

$$\text{where } c_k = (k+1) g_{se_i}(b_L), \quad c_\ell = (\ell+1) g_{se_j}(b_K).$$

Weighted Gaussian Mixture



$$G_w(x) = c_k g_{-z}(x) + c_\ell g_z(x).$$

The Weighted Mixture

1. If z is small enough, then G_w has one critical point.
2. If z is large enough, then G_w has three critical points.
3. G_w has exactly 2 critical points when $\frac{c_k}{c_\ell} = r(x) + 1$.

Lower Transition Scale Factor $T_{k,\ell}$

Definition

$T_{k,\ell}$ is the scale factor for which

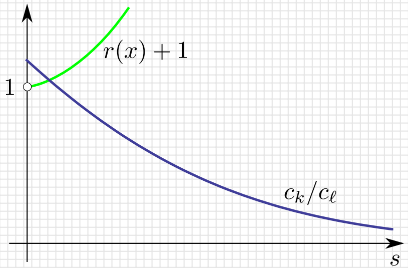
$$\frac{c_k}{c_\ell} = r(x) + 1.$$

Lower Transition Scale Factor $T_{k,\ell}$

Definition

$T_{k,\ell}$ is the scale factor for which

$$\frac{c_k}{c_\ell} = r(x) + 1.$$

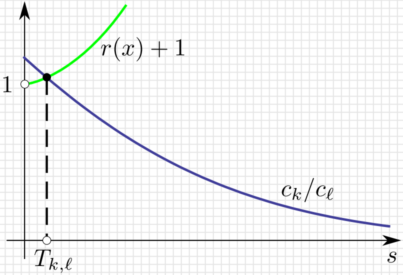


Lower Transition Scale Factor $T_{k,\ell}$

Definition

$T_{k,\ell}$ is the scale factor for which

$$\frac{c_k}{c_\ell} = r(x) + 1.$$



Lower Transition Scale Factor $T_{k,\ell}$

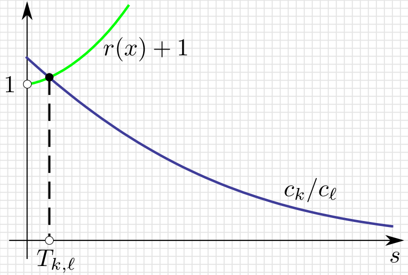
Definition

$T_{k,\ell}$ is the scale factor for which

$$\frac{c_k}{c_\ell} = r(x) + 1.$$

1-Dimensional Maxima Lemma

For all $s > T_{k,\ell}$, the axis $A_{K,L}$ witnesses two one-dimensional maxima.



Upper Transition Scale Factor U_n

Definition

$T_{k,\ell}$ is the scale factor for which:

$$\frac{c_k}{c_\ell} = r(x) + 1.$$

1-Dimensional Maxima Lemma

For all $s > T_{k,\ell}$, the axis $A_{K,L}$ witnesses two one-dimensional maxima.

Definition

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

Upper Transition Scale Factor U_n

Definition

$T_{k,\ell}$ is the scale factor for which:

$$\frac{c_k}{c_\ell} = r(x) + 1.$$

Definition

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

1-Dimensional Maxima Lemma

For all $s > T_{k,\ell}$, the axis $A_{K,L}$ witnesses two one-dimensional maxima.

Barycenter Lemma

The barycenter of $s\Delta^n$ is a mode for $s < U_n$, and a saddle of index 1 for $s > U_n$.

One-Dimensional Maxima

Definition

$T_{k,\ell}$ is the scale factor for which:

$$\frac{c_k}{c_\ell} = r(x) + 1.$$

Definition

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

1-Dimensional Maxima Lemma

For all $s > T_{k,\ell}$, the axis $A_{K,L}$ witnesses two one-dimensional maxima.

Barycenter Lemma

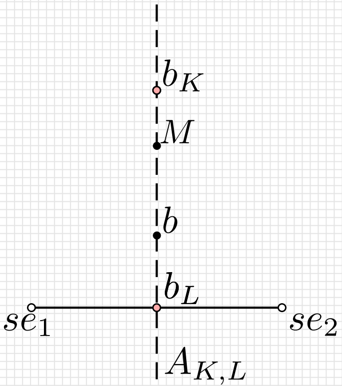
The barycenter of $s\Delta^n$ is a mode for $s < U_n$, and a saddle of index 1 for $s > U_n$.

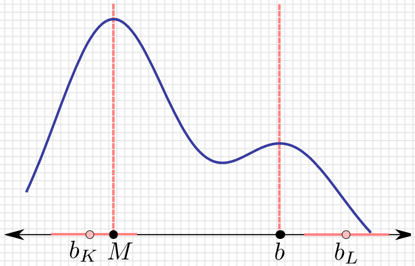
Theorem

If $s \in (T_{k,\ell}, U_n)$, then $A_{K,L}$ witnesses two one-dimensional maxima, one of which is at the barycenter.

Restriction to an Axis

$\mathbb{R}^5$

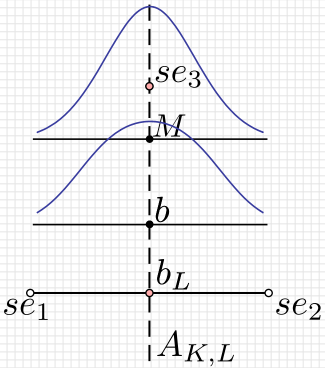




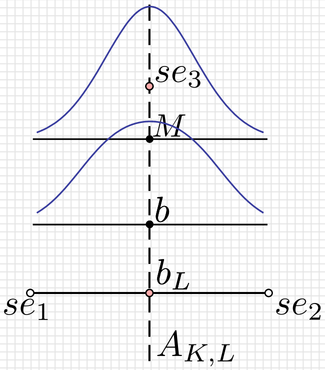
$$G_s|_A(x) = e^{-\pi h(x)} + c_\ell e^{-\pi(D_{0,n-1} - h(x))},$$

$$\text{where } c_\ell = (\ell + 1) g_{se_j}(b_L).$$

Witnessing the Modes



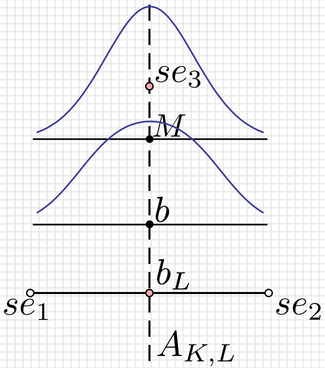
Witnessing the Modes



Witnessing Modes

If $|K| = 1$, then $A_{K,L}$ witnesses two modes for $s \in (T_{0,n-1}, U_n)$.

Witnessing the Modes



Witnessing Modes

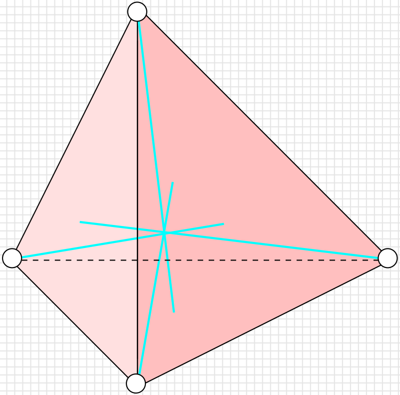
If $|K| = 1$, then $A_{K,L}$ witnesses two modes for $s \in (T_{0,n-1}, U_n)$.

Witnessing Critical Points

If $|K| > 1$, then M is a critical point, not a mode.

Many Axes

$\mathbb{R}^4$

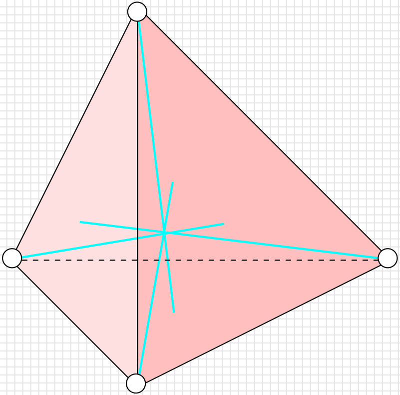


Many Axes

Number of Axes with $k = 0$:

$$n + 1.$$

$\mathbb{R}^4$



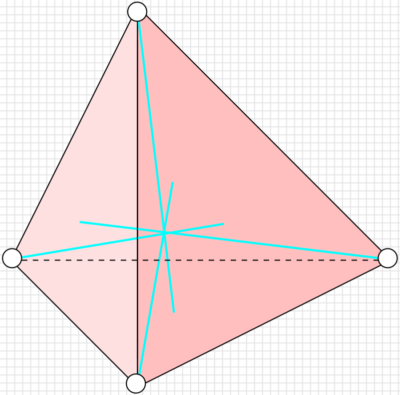
Many Axes

$\mathbb{R}^4$

Number of Axes with $k = 0$:

$$n + 1.$$

The scaled design has $n + 2$ modes.



Many Axes

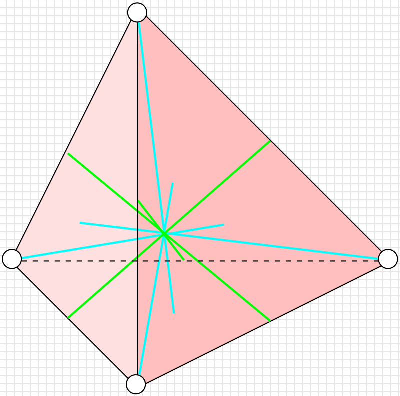
Number of Axes with $k = 0$:

$$n + 1.$$

The scaled design has $n + 2$ modes.

Total Number of Axes:

$$\frac{1}{2} \sum_{k=1}^{n+1} \binom{n+1}{k} = 2^n - 1.$$



Many Axes

Number of Axes with $k = 0$:

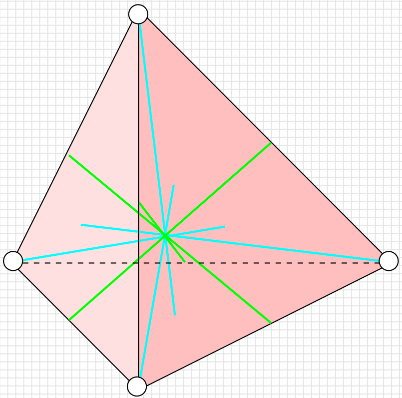
$$n + 1.$$

The scaled design has $n + 2$ modes.

Total Number of Axes:

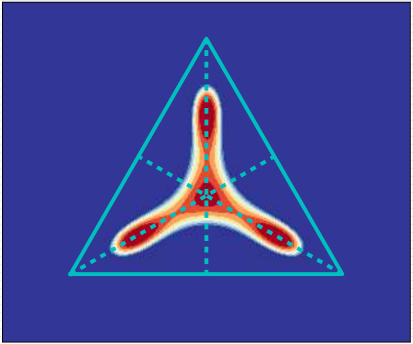
$$\frac{1}{2} \sum_{k=1}^{n+1} \binom{n+1}{k} = 2^n - 1.$$

The scaled design has $\Theta(2^n)$ critical points.



$\mathbb{R}^4$

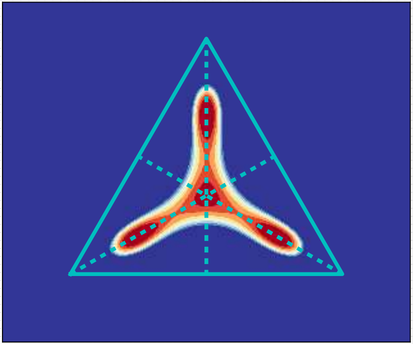
Summary of Results



Summary of Results

The *n*-design has:

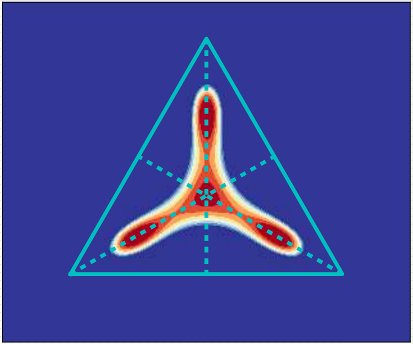
1. at most ONE ghost mode.



Summary of Results

The *n*-design has:

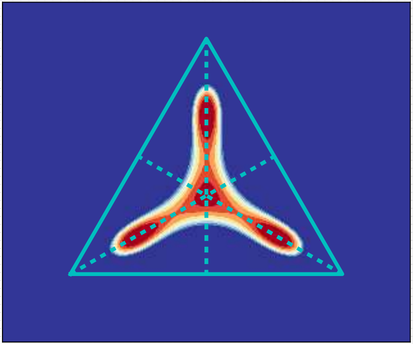
1. at most ONE ghost mode.
2. an exponential number of critical points.



Summary of Results

The *n*-design has:

1. at most ONE ghost mode.
2. an exponential number of critical points.
3. all critical points on axes.



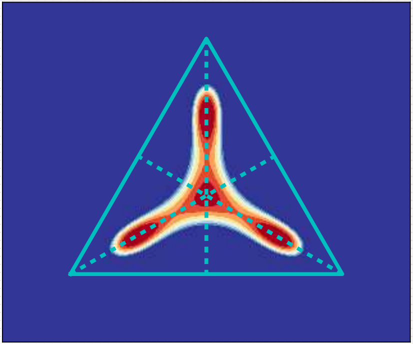
Summary of Results

The *n*-design has:

1. at most ONE ghost mode.
2. an exponential number of critical points.
3. all critical points on axes.

Wednesday at 2:50 in SN011:

1. How does $U_n - T_{0,n-1}$ (the resilience) grow with dimension?



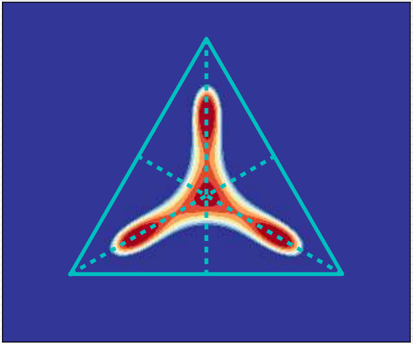
Summary of Results

The *n*-design has:

1. at most ONE ghost mode.
2. an exponential number of critical points.
3. all critical points on axes.

Wednesday at 2:50 in SN011:

1. How does $U_n - T_{0,n-1}$ (the resilience) grow with dimension?
2. What is the persistence of the ghost mode?



Add Isotropic Gaussian Kernels at Own Risk

More and More Resilient Modes in Higher Dimensions

Herbert Edelsbrunner, BRITTANY TERESE FASY, and Günter Rote

Symposium on Computational Geometry 2012
Chapel Hill, North Carolina

18 June 2012

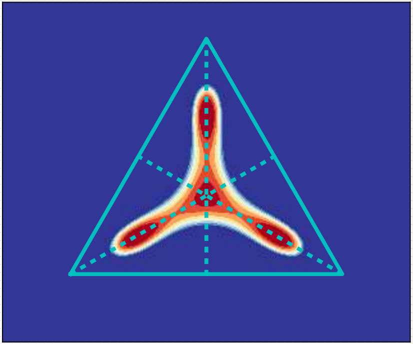
Summary of Results

The *n*-design has:

1. at most ONE ghost mode.
2. an exponential number of critical points.
3. all critical points on axes.

Wednesday at 2:50 in SN011:

1. How does $U_n - T_{0,n-1}$ (the resilience) grow with dimension?
2. What is the persistence of the ghost mode?



References

BEHBOODIAN, J.
On the modes of a mixture of two normal distributions.
Technometrics 12, 1 (1970), 131–139.

BURKE, P. J.
Solution of problem 4616 [1954, 718], proposed by A. C. Cohen, Jr.
Amer. Math. Monthly 63, 2 (Feb. 1956), 129.

CARREIRA-PERPINÁN, M., AND WILLIAMS, C.
On the number of modes of a Gaussian mixture.
Scale Space Methods in Computer Vision, Lecture Notes in Computer Science 2695 (2003), 625–640.

References

Carreira-Perpiñán, M., and Williams, C.
An isotropic Gaussian mixture can have more modes than components.
Informatics Research Report EDI-INF-RR-0185, Institute for Adaptive and Neural Computation, University of Edinbugh, Dec. 2003.

Fasy, B. T.
Modes of Gaussian mixtures and an inequality for the distance between curves in space, June 2012.
PhD Dissertation, Duke University Comput. Sci. Dept.

Silverman, B. W.
Using kernel density estimates to investigate multimodality.
J. R. Stat. Soc. Ser. B. Stat. Methodol. 43 (1981), 97–99.