ADD ISOTROPIC GAUSSIAN KERNELS AT OWN RISK:
MORE AND MORE RESILIENT MODES IN HIGHER DIMENSIONS

HERBERT EDELSBRUNNER, BRITTANY TERESE FASY, AND GÜNTER ROTE

ABSTRACT. The fact that the sum of isotropic Gaussian kernels can have more modes than kernels is surprising. Extra (ghost) modes do not exist in $\mathbb{R}^1$ and are generally not well studied in higher dimensions. We study a configuration of $n+1$ Gaussian kernels for which there are exactly $n+2$ modes. We show that all modes lie on a finite set of lines, which we call axes, and study the restriction of the Gaussian mixture to these axes in order to discover that there are an exponential number of critical points in this configuration. Although the existence of ghost modes remained unknown due to the difficulty of finding examples in $\mathbb{R}^2$, we show that the resilience of ghost modes grows like the square root of the dimension. In addition, we exhibit finite configurations of isotropic Gaussian kernels with superlinearly many modes.

Keywords: Unit Gaussian kernels, diffusion, convolution, standard simplex, critical points, modes, resilience.

ACM Computing Reviews Subject Classification: I.4.7 (Image Processing and Computer Vision: Feature measurement) and G.3 (Probability and Statistics: Distribution Functions) and I.6 (Simulation and Modeling)

1. INTRODUCTION

The diffusion of chemical substances, such as hormones, and of physical quantities, such as temperature, is a general phenomenon. Assuming a uniform medium, the process is described by the solution to the heat equation. In Euclidean space, solving this equation is synonymous to convolving with a Gaussian kernel. This is also a popular computational method, in particular in computer vision, where the 1-parameter family of convolutions of a given image is known as its *scale space*; see [13, 18]. A one-dimensional Gaussian kernel is known as a *normal density function* in probability [8].

We are interested in the quantitative analysis of diffusion and Gaussian convolution. In particular, we study the evolution of the critical points of a function that is convolved with a progressively wider Gaussian kernel. If the function is 1-dimensional, from $\mathbb{R}$ to $\mathbb{R}$, then Gaussian convolution does not create new critical points [1, 9, 16, 19]. As a consequence, the diffusion of m point masses (a sum of m Dirac delta functions) cannot have more than m modes (local maxima); see [2, 4, 17]. For two- or higher-dimensional functions, this is no longer true; see [12] for a two-dimensional function for which diffusion temporarily increase the number of modes and [7, 15] for a mathematical analysis of the unfolding events that cause this effect. It has been observed that these events are rare in practice [10, 11] and it has been confirmed that the ability to create critical points with non-negligible persistence deteriorates rapidly [6]. It is also known that $n+1$ point masses can be arranged in $\mathbb{R}^n$ so that diffusion creates $n+2$ modes during a non-empty time interval; see [5].

The contribution of this paper is a strengthening of the cautionary voice on using Gaussian convolution in dimensions beyond 1. In particular, we give a detailed analysis of the
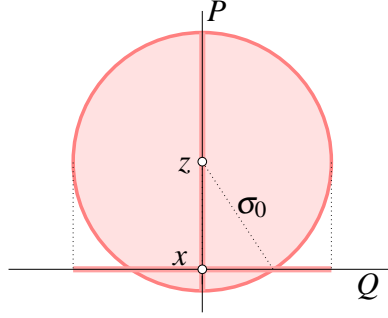

FIGURE 1. A unit Gaussian kernel with center z in $\mathbb{R}^2$ represented by the disk with radius $\sigma_0 = 1/\sqrt{2\pi}$. The lines P and Q define 1-dimensional sections.

sum of $n+1$ identical isotropic Gaussian kernels placed at the vertices of a regular n-simplex in $\mathbb{R}^n$. We prove that all critical points lie on the symmetry axes of the n-simplex, and we characterize their indices, confirming the $(n+2)$-nd mode at the barycenter as the sole extra mode. While the extra mode seems fragile, we show that the interval of widths during which it exists grows like the square root of the dimension. It thus seems likely that the phenomenon of extra modes is more prevalent in higher dimensions. Providing additional evidence, we construct finite configurations of isotropic Gaussian kernels with superlinearly many modes.

Outline. Section 2 provides background on Gaussian kernels and the geometry of regular simplices. Section 3 analyzes the sum of kernels placed at the vertices of a regular simplex, characterizes its critical points, estimates the resilience of the extra mode, and exhibits configurations with superlinearly many modes. Section 4 concludes this paper.

2. BACKGROUND

Our results depend on 1-dimensional Gaussian kernels and n-dimensional regular simplices. We study these two topics in two subsections.

2.1. **Curve Analysis.** In this subsection, we introduce Gaussian kernels and discuss some of their fundamental properties.

Gaussian kernels and derivatives. We call a real-valued function of the form $g(x) = We^{-C\|x-z\|^2}$ an $(n+1)$-*dimensional Gaussian kernel*, where W and C are real constants, $z \in \mathbb{R}^{n+1}$ is a point, and $\|x-z\|$ denotes the Euclidean distance between the two points. We call the kernel *normalized* if it integrates to 1, in which case it can be written as:

$$g_z(x) \;=\; \frac{1}{(2\pi\sigma^2)^{\frac{n+1}{2}}} \cdot e^{-\frac{\|x-z\|^2}{2\sigma^2}}.$$

We call $z \in \mathbb{R}^{n+1}$ the *center* or *mean* and $\sigma > 0$ the *width* (or *standard deviation* if $n = 0$). Finally, g_z is *unit* if it is normalized with height 1. Independent of the dimension, the width and the formula of the unit Gaussian kernel are:

$$\sigma_0 \;=\; 1/\sqrt{2\pi},$$
$$g_z(x) \;=\; e^{-\pi\|x-z\|^2},$$

which will simplify our computations. It can be transformed into every other Gaussian kernel by a translation (to change the center), a scaling of the domain (to change the width), and a scaling of the range (to change the height). For the case $n = 0$, we can write the
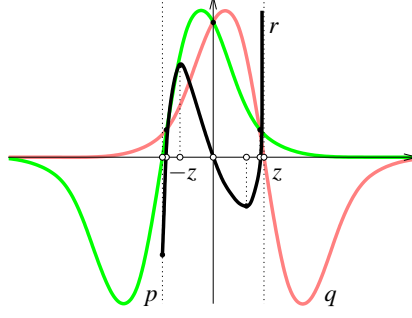
FIGURE 2. The graphs of the derivatives of g_{-z} and of g_z intersect above the roots of the (black) ratio function.

formulas for the first two derivatives of the Gaussian kernel centered at the origin:

$$(1) \qquad g_0'(x) \quad = \quad [-2\pi x] \cdot g_0(x),$$

$$(2) \qquad g_0''(x) \quad = \quad [4\pi^2 x^2 - 2\pi] \cdot g_0(x).$$

Note that g_0'' is negative in the interior of $[-\sigma_0, \sigma_0]$ and positive outside this interval. A Gaussian kernel is concave at every point in the interior of the closed ball with center z and radius σ, and it fails to be concave at every point outside this ball. This implies that the ball is a convenient illustration of the kernel; see Figure 1.

Balanced sums. Consider the sum of two unit Gaussian kernels. For symmetry, we choose their centers at distance $z \geq 0$ to the left and right of the origin. As proven in [3], $G = g_{-z} + g_z$ has either 1 or 3 critical points and no other number is possible. More specifically, G has 1 maximum iff $z \leq \sigma_0$ and G has 2 maxima and 1 minimum iff $z > \sigma_0$. We present our own proof of this result, as we need the concepts it uses.

Since $G' = g_{-z}' + g_z'$, a point $x \in \mathbb{R}$ is a critical point of G iff the graphs of $p = -g_{-z}'$ and $q = g_z'$ intersect above x. Since g' is an odd function, the number of intersections (counted with multiplicities) must be odd, and it is visually plausible that the number can only be 1 or 3. To be sure, we introduce the *ratio function*, $r = \frac{p}{q} - 1$, with formula

$$(3) \qquad r(x) \quad = \quad \frac{z+x}{z-x} \cdot e^{-4\pi z \cdot x} - 1;$$

see Figure 2. Setting $r(x) = 0$ gives us the intersections of p and q and thus the critical points of G. The roots of r are necessarily in $[-z, z]$. Independent of z, we have $r(0) = 0$. To see whether there are additional roots, we take the derivative:

$$r'(x) \quad = \quad \frac{2z - 4\pi z(z^2 - x^2)}{(z-x)^2} \cdot e^{-4\pi z \cdot x}.$$

Setting $r'(x) = 0$, we have $x^2 = z^2 - \frac{1}{2\pi}$, which has two real solutions if $z > \sigma_0$ and no solution if $z < \sigma_0$. Consider first the case that $r'(x)$ has no real solution. Observing that $r(-z)$ is negative and that the limit of $r(x)$ as x approaches z from the right is positive infinity, we conclude that there is exactly one root of $r(x)$. In the case where $r'(x) = 0$ has two solutions, a similar argument shows that $r(x)$ has 3 roots. As anticipated, G has 1 critical point if $z \leq \sigma_0$ and it has 3 non-degenerate critical points if $z > \sigma_0$.

Unbalanced sums. Next, we study sums $G_w = g_{-z} + w g_z$, where $w \geq 0$ is the *weight* of the second term. The number of critical points of G_w is at most 3, but in contrast to the balanced case, it can also be 2, as we will see. More specifically, [2] gives necessary and sufficient conditions for all three cases (1, 2, or 3 critical points), but they are not as easy to state as in the balanced case. As before, we present our own proof since we need the concepts it uses.

Recalling that the critical points of G_w correspond to the intersections between the graphs of $p = -g'_{-z}$ and $wq = wg'_z$, we consider their ratio function, which is

$$r_w \;=\; \frac{p}{wq} - 1 \;=\; \frac{r}{w} - \left(1 - \frac{1}{w}\right).$$

Its derivative is $r'_w = \frac{r'}{w}$, which has at most 2 roots. Similarly, $r_w(-z) = -1$ and $r_w(z)$ goes to infinity as x approaches z, so we see that r_w has at most 3 roots. It follows that G_w has at most 3 critical points, just like G.

A new phenomenon is the possibility of 2 critical points. To see when this case arises, we set $w = r(x) + 1$ and note that for this choice of weight $r_w(x) = 0$. In words, p and wq intersect above x and, equivalently, x is a critical point of G_w. If x has the additional property of being critical for r, then the intersection between p and wq is degenerate. As computed above, the critical points of r_w are given by $x^2 = z^2 - \sigma_0^2$. Let $x_1 = -\sqrt{z^2 - \sigma_0^2}$ and $x_2 = \sqrt{z^2 - \sigma_0^2}$ be the two solutions, and note that x_1 gives a weight $w_1 = r(x_1) + 1$ that is larger than 1, while x_2 gives a weight $w_2 = r(x_2) + 1$ between 0 and 1. We call w_1 and w_2 the *transition weights* for z, remembering that they exist iff $z \geq \sigma_0$.

We illustrate the case analysis by moving the centers and observing how the shape of G_w changes. We get qualitatively the same behavior for every positive weight. Fixing the weight to $w = \frac{1}{2}$, let 2ζ be the distance between the two centers for which w is a transition weight. Starting with $-z = z = 0$, we see the evolution sketched in Figure 3. For $z < \zeta$, the function G_w has only one maximum. At $z = \zeta$, we have a degenerate critical point forming a shoulder on the right. For $z > \zeta$, this shoulder turns into a min-max pair. This is a generic event in the 1-parameter evolution of a Morse function, known as an anti-cancellation.

2.2. Simplex Design.

2.2. **Simplex Design.** In this subsection, we design a sum of Gaussian kernels in $\mathbb{R}^{n+1}$ that has the symmetry group of the regular n-simplex. We begin with a geometric study of the simplex, whose shape properties will play a central role in our design.
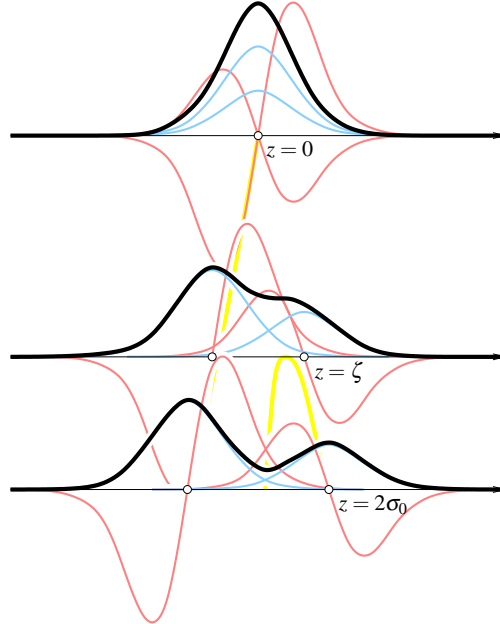

FIGURE 3. From top to bottom: the sum of a unit kernel and half a unit kernel. The kernels are blue, their sum is black, their derivatives (plus-minus) are pink, and the trajectories of the critical points, drawn over the interval from 0 to $2\sigma_0$, are yellow.
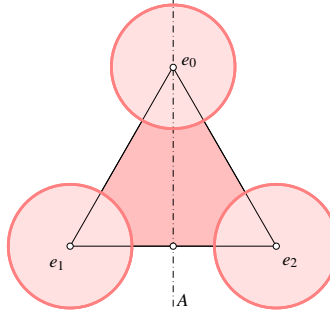
FIGURE 4. A 2-design, which is the sum of 3 unit Gaussian kernels placed at the vertices of an equilateral triangle.

Standard simplex. A convenient model is the *standard n-simplex*, defined as the convex hull of the $n+1$ unit coordinate vectors in $\mathbb{R}^{n+1}$: $\Delta^n = \text{conv}\{e_0, e_1, \ldots, e_n\}$. Each subset of $k+1$ vectors defines a *k-face* of Δ^n, which is itself a standard k-simplex. The *barycenter* of Δ^n is the point whose $n+1$ coordinates are all equal to $\frac{1}{n+1}$. Let $0 \leq k \leq \ell$ with $k + \ell = n - 1$ and consider a k-face and the unique disjoint ℓ-face. Their barycenters use complementary subsets of the $n+1$ coordinates, which makes it easy to compute the distance between them as

$$(4) \qquad D_{k,\ell} \quad = \quad \sqrt{\frac{1}{k+1} + \frac{1}{\ell+1}}.$$

For example, the height of the n-simplex, which is defined as the distance between a vertex to the opposite $(n-1)$-face, is $D_{0,n-1} = \sqrt{(n+1)/n}$. Similarly, we can compute the radius of the circumsphere:

$$(5) \qquad R_n \quad = \quad \sqrt{\frac{n}{n+1}}.$$

Sections. A *k-section* of a Gaussian kernel is the restriction to a k-dimensional plane P. Assuming a unit Gaussian kernel, we can write this as $g_z|_P(x) = g_z(y) \cdot g_y(x)$, where y is the orthogonal projection of z onto P and g_y is the k-dimensional unit Gaussian kernel with center $y \in P$. We call $g_z(y)$ the *weight*, noting that it is equal to the integral of g_z over P and we call $g_z|_P$ a *weighted* unit Gaussian kernel. Importantly, we see that the section has the same width as the original kernel. It is a unit kernel itself iff P passes through z. In this case, $g_z|_P(x)$ is the weight of the $(n-k)$-section defined by the plane Q that intersects P orthogonally at x; see Figure 1. Iterating this construction, we can write the $(n+1)$-dimensional unit Gaussian kernel as a product of 1-dimensional kernels:

$$(6) \qquad g_z(x) \quad = \quad \prod_{i=0}^{n} g(x_i - z_i),$$

where the x_i and z_i are the Cartesian coordinates of x and z. In words, the high-dimensional unit kernel can be *separated* into mutually orthogonal one-dimensional unit kernels.

Standard design. We turn Δ^n into a function by placing a unit Gaussian kernel at every vertex. Writing g_i for the kernel with center e_i, we get $G = g_0 + g_1 + \ldots + g_n$, which we call an *n-design*; see Figure 4. Its symmetry group is that of the n-simplex with the additional reflection across the n-dimensional plane spanned by the n-simplex. It is therefore isomorphic to $\Sigma_{n+1} \oplus \mathbb{S}^0$, where Σ_{n+1} is the symmetric group on $n+1$ elements. To argue about the symmetries, we use lines that connect barycenters of complementary faces of the n-simplex. We call these lines *axes*. Suppose for example that $0 \leq k \leq \ell$ are integers with $k + \ell = n - 1$, and that A is the axis that connects the barycenter of the k-face spanned by e_0 to e_k with the barycenter of the ℓ-face spanned by e_{k+1} to e_n. Except for the barycenter of Δ^n, every point x of A has two distinct barycentric coordinates, one with multiplicity

$k+1$ and the other with multiplicity $\ell+1$. The orbit of x has therefore size $\binom{n+1}{k+1}$. Recall that the section defined by the axis is the restriction of the function to the line: $G|_A$. The restrictions of $g_0, g_1, \ldots, g_k$ to A are all identical, namely a weighted 1-dimensional unit Gaussian kernel whose center is the barycenter of the k-face. Similarly, the restrictions of $g_{k+1}, \ldots, g_{n-1}, g_n$ are all identical, and we can write the 1-section as the sum of two kernels:

$$(7) \qquad G|_A \;=\; (k+1) \cdot g_0|_A + (\ell+1) \cdot g_n|_A,$$

which are one-dimensional weighted unit Gaussian kernels, with weights $(k+1)g(R_k)$ and $(\ell+1)g(R_\ell)$, and distance $D_{k,\ell}$ between their centers.

We are interested in changing the widths of the $(n+1)$-dimensional kernels uniformly. Equivalently, we scale the n-simplex by moving the centers of the unit Gaussian kernels closer to or further from each other, without changing their widths and heights. To do this, we introduce the *scaled n-design*, $G_s = g_{se_0} + g_{se_1} + \ldots + g_{se_n}$. Here, we call s the *scale factor*, and we write $s\Delta^n$ for the *scaled n-simplex* whose vertices are the se_i. We are interested in the evolution of the critical points in the 1-parameter family of scaled n-design $G_s : \mathbb{R}^{n+1} \to \mathbb{R}$, as s goes from zero to infinity.

3. ANALYSIS

We begin this section by proving that all critical points lie on the axes of the n-simplex. Thereafter, we analyze each axis, characterizing for which scales we see 1, 2, or 3 critical points. To decide which of the one-dimensional maxima are modes, we analyze the n-sections orthogonal to the axes. As it turns out, all modes lie on axes that pass through vertices of the n-simplex. Most interesting is the critical point at the barycenter, which changes from unique mode during an initial interval of scales, to $(n+2)$-nd mode during a non-empty intermediate interval, to a saddle of index one during a final interval. We call the length of the intermediate interval the resilience of the extra mode and show that it grows like the square root of the dimension. Finally, we construct sums of isotropic Gaussian kernels with a superlinear number of modes.

3.1. **Lines of Critical Points.** In this subsection, we note that all critical points of a scaled n-design lie on the axes of the scaled n-simplex. We begin by introducing coordinates that are more natural for the n-design, and we show how they relate to the barycentric coordinates.

Distance coordinates. Write $v_i = se_i$, for $0 \le i \le n$, and let x be a point of the corresponding scaled n-simplex $s\Delta^n$. Setting $r_i = \|x - v_i\|$, we note that x is uniquely defined by the vector of $n+1$ distances since x lies on the hyperplane spanned by $\{v_i\}$. We express this by writing $x = (r_0, r_1, \ldots, r_n)_D$, and by calling the r_i the *distance coordinates* of x. Recall that $(x_0, x_1, \ldots, x_n)_B$ is the representation of the same point in barycentric coordinates. We are interested in computing the barycentric from the distance coordinates via the Coordinate Transformation below.

Coordinate Transformation. *For $0 \le i \le n$, the ith barycentric coordinate is given by:*

$$(8) \qquad x_i \;=\; \frac{1}{n+1} + \frac{1}{2(n+1)s^2} \left(\sum_{j=0}^{n} r_j^2 - (n+1)r_i^2 \right).$$

Proof. Let $(r_0, r_1, \ldots, r_n)_D$ be the distance coordinates of a point x in the scaled n-simplex $s\Delta^n$. Let $i \ne j$ and consider the edge connecting v_i with v_j, recalling that v_i and v_j are two vertices of $s\Delta^n$. The length of the edge is $s\sqrt{2}$. Let x_{ij} be the distance between v_j and the orthogonal projection of x onto the edge, normalized by dividing with $s\sqrt{2}$; see Figure 5. We first show that

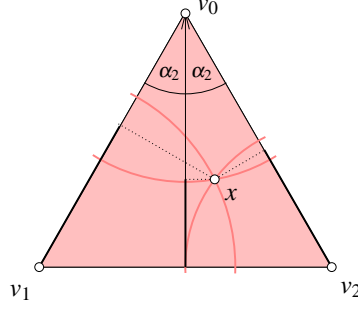$$(9) \qquad x_{ij} \;=\; \frac{1}{2} + \frac{1}{4s^2}(r_j^2 - r_i^2).$$

FIGURE 5. The radii of the three circles are the distance coordinates of the point x. The orthogonal projections onto the vertical axis and the left and right edges of the triangle give x_0, x_{01}, and x_{02}.

Indeed, if $x = (1-t)v_j + tv_i$, then we have $r_j = s\sqrt{2}t$ and $r_i = s\sqrt{2}(1-t)$. Furthermore, $x_{ij} = t$, which agrees with the equation we get by plugging the values of r_i and r_j into (9). Realizing that $r_j^2 - r_i^2$ is constant along hyperplanes orthogonal to the edge, we get (9) for all points of the n-simplex.

For the next step, let b_i be the barycenter of the $(n-1)$-face complementary to v_i and y be the orthogonal projection of x onto the edge between v_i and v_j. Set α_n to the angle between the edges that connect v_i to v_j and to b_i. Because $s\Delta^n$ is regular, this angle does not depend on the choice of i and j. Suppose x lies on the latter edge, which connects v_i and b_i. Then $x = (1-x_i)b_i + x_i v_i$ and we have two expressions for $\cos\alpha_n$. Setting these two expression equal we arrive at

$$\frac{D_{0,n-1}}{\sqrt{2}} = \frac{\sqrt{2}(1-x_{ij})}{D_{0,n-1}(1-x_i)},$$

for every $0 \le j \le n$ and $j \ne i$. Adding the n equations gives

$$(10) \qquad nD_{0,n-1}^2(1-x_i) = 2n - 2\sum_{j\ne i} x_{ij}.$$

Similar to before, we notice that the two sides of the equation are constant along hyperplanes orthogonal to the axis defined by v_i and b_i. Hence, (10) holds for all points x of the scaled n-simplex. It remains to plug (4) and (9) into (10), which gives

$$(n+1)(1-x_i) = n + \frac{1}{2s^2}\left(nr_i^2 - \sum_{j\ne i} r_j^2\right).$$

The equation simplifies to the claimed equation. $\qquad\qquad\square$

Non-zero gradients. Recall that $G_s : \mathbb{R}^{n+1} \to \mathbb{R}$ is the scaled n-design formed by taking the sum of the $n+1$ unit Gaussian kernels whose centers are the vertices of $s\Delta^n$. We use the Coordinate Transformation to show that G_s has no critical points away from the axes of the scaled n-simplex:

Lemma (Axes Lemma). *Every critical point of G_s lies on an axis of the scaled n-simplex $s\Delta^n$.*

Proof. Recall that a point x belongs to an axis of $s\Delta^n$ iff it has at most two distinct barycentric coordinates. We will show that if x has three distinct barycentric coordinates, then the gradient of G_s at x is non-zero. Writing $f_i = g_{se_i}$, we obtain

$$\nabla G_s(x) = -2\pi G_s(x)\cdot x + 2\pi \sum_{i=0}^n f_i(x)\cdot v_i.$$

Setting the gradient to zero, we solve for x

$$(11) \qquad x \;=\; \sum_{i=0}^{n} \frac{f_i(x)}{G_s(x)} \cdot v_i.$$

We will show that Equation (11) can hold only if x has at most two distinct barycentric coordinates. To this end, we write x in distance coordinates: $x = (r_0, r_1, \ldots, r_n)_D$. Similar to the barycentric coordinates, x lies on an axis of $s\Delta^n$ iff there are at most two distinct distance coordinates. Transforming x into barycentric coordinates, we have $x = (x_0, x_1, \ldots, x_n)_B$, in which

$$x_i \;=\; \frac{1}{n+1} + \frac{1}{2(n+1)s^2}\left(\sum_{j=0}^{n} r_j^2 - (n+1)r_i^2 \right),$$

for $0 \le i \le n$. Assume now that x does not lie on any of the axes. If follows there are three distinct distance coordinates: $r_k < r_\ell < r_m$. Subtracting the m-th barycentric coordinate from the other two , we obtain

$$(12) \qquad x_k - x_m \;=\; \frac{1}{2s^2}(r_m^2 - r_k^2),$$

$$(13) \qquad x_\ell - x_m \;=\; \frac{1}{2s^2}(r_m^2 - r_\ell^2).$$

Assuming a zero gradient, the barycentric coordinates of x have the form given in (11). Hence, $x_k - x_m$ and $x_\ell - x_m$ are equal to

$$(14) \qquad \frac{f_k(x) - f_m(x)}{G_s(x)} \;=\; \frac{e^{-\pi r_k^2} - e^{-\pi r_m^2}}{G_s(x)},$$

$$(15) \qquad \frac{f_\ell(x) - f_m(x)}{G_s(x)} \;=\; \frac{e^{-\pi r_\ell^2} - e^{-\pi r_m^2}}{G_s(x)},$$

respectively. Since the right hand sides of (12) and (14) are equal, as well as the right hand sides of (13) and (15), we have

$$\frac{r_m^2 - r_k^2}{e^{-r_m^2} - e^{-r_k^2}} \;=\; \frac{r_m^2 - r_\ell^2}{e^{-r_m^2} - e^{-r_\ell^2}}.$$

But this is impossible because $r_m^2 - r_k^2 > r_m^2 - r_\ell^2$, by assumption, and because the function $f(t) = e^{-t}$ is strictly convex. $\qquad\qquad \square$

3.2. One-Dimensional Sections.
The restriction of G_s to an axis of $s\Delta^n$ is a sum of two weighted Gaussian kernels. This sum has two maxima for a range of scale factors, which we now analyze.

Transitions. Recall that the n-design consists of $n+1$ unit Gaussian kernels placed at the vertices of the standard n-simplex. Consider the 1-section defined by the line that connects the barycenter of a k-face with the barycenter of the complementary ℓ-face, with $k + \ell = n - 1$, and vary the construction by scaling the design with $s \ge 0$. We call a value a *transition* if the number of critical points of the 1-section changes as s passes the value. It is easy to compute the transition for $k = \ell = \frac{n-1}{2}$ because the corresponding 1-section is balanced for all scale factors s. The distance between the two centers is $sD_{\ell,\ell} = 2s/\sqrt{n+1}$ and we find the transition by setting the distance equal to $2\sigma_0$, which gives s equal to

$$(16) \qquad U_n \;=\; \sqrt{\frac{n+1}{2\pi}}.$$

Consider next the case $k < \ell$. Equation (7) gives the weights of the two kernels in the decomposition of the 1-section as $(k+1)g(sR_k)$ and $(\ell+1)g(sR_\ell)$. Using (5) and taking
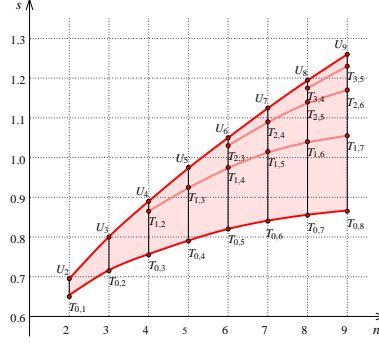
FIGURE 6. The vertical interval bounded from below by $T_{0,\ell}$ and above by U_n contains the scale factors for which an extra mode appears.

the ratio, the *weight function* is computed by the following function:

$$
(17) \qquad \omega_{k,\ell}(s) \quad = \quad \frac{\ell+1}{k+1} \cdot e^{-\pi s^2 \left(\frac{1}{k+1} - \frac{1}{\ell+1} \right)}.
$$

We compare this with the two *transition functions*, which we get by setting $z = \frac{s}{2} D_{k,\ell}$ and plugging the two solutions of $x^2 = z^2 - \sigma_0^2$ into the formula for $r(x) + 1$, which we get from (3). This gives

$$
(18) \qquad \tau_{k,\ell}(s) \quad = \quad \frac{z - \sqrt{z^2 - \sigma_0^2}}{z + \sqrt{z^2 - \sigma_0^2}} \cdot e^{4\pi z \sqrt{z^2 - \sigma_0^2}},
$$

$$
(19) \qquad \upsilon_{k,\ell}(s) \quad = \quad \frac{z + \sqrt{z^2 - \sigma_0^2}}{z - \sqrt{z^2 - \sigma_0^2}} \cdot e^{-4\pi z \sqrt{z^2 - \sigma_0^2}}.
$$

Note that $\upsilon_{k,\ell}(s) = 1/\tau_{k,\ell}(s)$. We find the first transition, $T_{k,\ell}$, by solving $\omega_{k,\ell}(s) = \tau_{k,\ell}(s)$, and the second transition, U_n, by solving $\omega_{k,\ell}(s) = \upsilon_{k,\ell}(s)$. Appendix A will prove that both transitions are well defined, also showing that the second transition depends on n but not on k and ℓ and is given by (16) in all cases. While we have no analytic expression for $T_{k,\ell}$, we will derive one for an upper bound in Section 3.4.

Section evolution. We follow the 1-section defined by an axis of the n-simplex as the scale factor, s, goes from 0 to infinity. By construction, we have qualitative changes at the transitions, which we now summarize.

Lemma (1-Section Lemma). *Let $0 \leq k \leq \ell$ with $k + \ell = n - 1$, and let A be the axis passing through the barycenters of a k-face and its complementary ℓ-face of $s\Delta^n$. Then $G_s|_A$ has one maximum whenever $s < T_{k,\ell}$, and two maxima whenever $T_{k,\ell} < s$ and $s \neq U_n$.*

Indeed, the double intersection is responsible for the special evolution of the 1-section. In particular, we go from one maximum for $s < T_{k,\ell}$ to two maxima for $T_{k,\ell} < s < U_n$, of which one is the barycenter of the n-simplex. After the second transition at the double intersection, we still have two maxima, but now the separating minimum is the barycenter of the n-simplex. Figure 6 shows all transition scale factors in a single picture for small values of k and ℓ. First, we look at $T_{k,\ell}$ for a fixed value of n. We observe that $T_{k,\ell}$ increases with growing k. This implies that for constant n, the axes defined for small values of k spawn a second maximum earlier than do axes defined by large values of k. For $k = \ell$, the two transitions coincide and the corresponding 1-section does not witness the extra maximum at all. Second, we fix k and observe that $T_{k,\ell}$ increases with growing ℓ. This implies that the k-faces of a low-dimensional simplex spawn second maxima earlier than do the k-faces in high-dimensional simplices.

Next, we look at the second transition, U_n. As we have observed, it depends only on n. This implies that all 1-sections lose the maximum at the barycenter at the same scale factor. For $k = \ell$, the two transitions coincide, so the interval collapses.

3.3. n-Dimensional Sections.

In this subsection, we show that most maxima of the 1-sections are not modes. We begin with the analysis of the barycenter of $s\Delta^n$, which belongs to every axis of the scaled n-simplex.

Barycenter of n-simplex. The n-design has the symmetry group of the n-simplex, which implies that the barycenter, $b_G \in s\Delta^n$, is a critical point of G_s. Indeed, if b_G is not a critical point, then it has a non-zero gradient, which contradicts the symmetry. More specifically, b_G is either a maximum or a minimum of the n-section defined by the n-simplex, and it is a maximum of the orthogonal 1-section defined by the diagonal line of $\mathbb{R}^{n+1}$.

Lemma (Barycenter Lemma). *Let $n \geq 1$. Then the barycenter of $s\Delta^n$ is a mode of G_s for $s < U_n$ and it is a saddle of index 1 for $s > U_n$.*

Proof. We compute the Hessian of G_s at b_G by taking partial derivatives with respect to the Cartesian basis of $\mathbb{R}^{n+1}$. Because of the symmetry, we have

$$d \;=\; \frac{\partial^2 G_s}{\partial x_0^2}(b_G) \;=\; \frac{\partial^2 G_s}{\partial x_i^2}(b_G),$$

$$c \;=\; \frac{\partial^2 G_s}{\partial x_0 \partial x_1}(b_G) \;=\; \frac{\partial^2 G_s}{\partial x_i \partial x_j}(b_G),$$

for all $0 \leq i \leq n$ and all $j \neq i$. The characteristic polynomial of the Hessian is therefore

$$(20) \qquad \det \begin{bmatrix} d-\xi & c & \ldots & c \\ c & d-\xi & \ldots & c \\ \vdots & \vdots & \ddots & \vdots \\ c & c & \ldots & d-\xi \end{bmatrix}.$$

Its roots are the eigenvalues of the Hessian, which are $\xi = d + nc$, with multiplicity one, and $\xi = d - c$, with multiplicity n. For simple geometric reasons, $d + nc$ is negative and corresponds to the eigenvector in the diagonal direction of $\mathbb{R}^{n+1}$. Since all Gaussian kernels live in a common plane, any point in that plane will be a maximum value in the direction orthogonal to that plane. To compute $d - c$, it suffices to consider just one 1-section through the barycenter, and we choose the line, B, that passes through the vertex $v_0 = se_0$ and the barycenter of the complementary $(n-1)$-face, which we denote as b_0. The distances from the barycenter of the n-simplex are $\|b_G - v_0\| = sR_n$ and $\|b_G - b_0\| = \frac{s}{n}R_n$. Furthermore, the common distance of the vertices $v_i = se^i$ from B is $\|v_i - b_0\| = sR_{n-1}$, for $1 \leq i \leq n$. Plugging these distances into the second derivative of the one-dimensional section given by (2), we compute the second derivatives at b_G of the 1-sections defined by the $n+1$ unit Gaussian kernels as

$$(21) \qquad\qquad [4\pi^2 s^2 R_n^2 - 2\pi] \cdot e^{-\pi s^2 R_n^2},$$

$$(22) \qquad\qquad \left[4\pi^2 \frac{s^2}{n^2} R_n^2 - 2\pi\right] \cdot e^{-\pi s^2 (R_{n-1}^2 + R_n^2/n^2)},$$

where the first line applies for $i = 0$ and the second line for $1 \leq i \leq n$. Note that $R_{n-1}^2 + R_n^2/n^2 = R_n^2$ and $R_n^2(1 + \frac{1}{n}) = 1$. Adding (21) and n times (22), we obtain the second derivative of the sum of $n+1$ one-dimensional Gaussian kernels as

$$d - c \;=\; [4\pi^2 s^2 - 2\pi(n+1)] \cdot e^{-\pi s^2 R_n^2},$$

which has the same sign as $s^2 - \frac{n+1}{2\pi}$. Thus, b_G is a maximum of G_s for $s < U_n$ and a saddle of index one for $s > U_n$ as claimed.

$\square$

We note here that the barycenter is an index-1 saddle for $s > U_n$, as opposed to a minimum, because we place the n-simplex in $\mathbb{R}^{n+1}$. At the transition, when $s = U_n$, the barycenter of the n-simplex is a degenerate critical point.

Orthogonal sections. We generalize the analysis of the barycenter. Let $1 \le k \le \ell$ with $k + \ell = n - 1$, and consider a k-face of the n-simplex as well as the complementary ℓ-face. Writing G_s as the sum of the $f_i = g_{se_i}$, for $0 \le i \le n$, we assume that the centers of f_0 to f_k span the k-face, and that the centers of f_{k+1} to f_n span the ℓ-face. Hence, $G_s = K_s + L_s$, where $K_s = \sum_{i=0}^{k} f_i$ and $L_s = \sum_{i=k+1}^{n} f_i$. Writing b_K and b_L for the barycenters of the two faces, we let A be the axis defined by $A(t) = (1-t)b_K + tb_L$. We are interested in the Hessian of G_s at $x = A(t)$. For symmetry reasons, it has at most four distinct eigenvalues, each a second derivative along pairwise orthogonal lines. One line is the axis, another is orthogonal to the n-simplex, a third line is parallel to the k-face, and a fourth line is parallel to the ℓ-face. The latter two eigenvalues have multiplicity k and ℓ. We write κ for the length parameter along the third line and λ for the length parameter along the fourth line.

Lemma (n-Section Lemma). *Let $1 \le k \le \ell$ with $k + \ell = n - 1$. The second derivatives of G_s at $x = A(t)$ along lines parallel to the complementary k- and ℓ-faces of $s\Delta^n$ are*

$$(23) \qquad \frac{\partial^2 G_s}{\partial \kappa^2}(x) \;=\; -2\pi G_s(x) + 4\pi^2 s^2 f_0(x),$$

$$(24) \qquad \frac{\partial^2 G_s}{\partial \lambda^2}(x) \;=\; -2\pi G_s(x) + 4\pi^2 s^2 f_n(x).$$

Proof. Recall $G_s = \sum_{i=0}^{n} f_i$ and $f_i(x) = e^{-\pi \|x - se_i\|^2}$. The derivative with respect to the i-th coordinate direction is

$$\frac{\partial G_s}{\partial x_i}(x) \;=\; -2\pi x_i G_s(x) + 2\pi s f_i(x).$$

Deriving again, with respect to the same and a different coordinate direction, we have

$$\frac{\partial^2 G_s}{\partial x_i^2}(x) \;=\; [-2\pi + 4\pi^2 x_i^2] G_s(x)$$
$$(25) \qquad\qquad\qquad\qquad - 4\pi^2 (2sx_i - s^2) f_i(x),$$
$$\frac{\partial^2 G_s}{\partial x_i \partial x_j}(x) \;=\; 4\pi^2 [x_i x_j G_s(x)$$
$$(26) \qquad\qquad\qquad\qquad - sx_i f_j(x) - sx_j f_i(x)].$$

The point at which we take the second derivative has only two distinct coordinates, $\frac{(1-t)s}{k+1}$, repeated $k+1$ times, and $\frac{ts}{\ell+1}$, repeated $\ell+1$ times. We can therefore substitute x_0 and x_1 for any two among the first $k + 1 \ge 2$ coordinate directions, and we can substitute x_n and x_{n-1} for any two among the last $\ell + 1 \ge 2$ coordinate directions. The Hessian at the point x is

$$H(x) \;=\; \begin{bmatrix} d & \cdots & c & \gamma & \cdots & \gamma \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ c & \cdots & d & \gamma & \cdots & \gamma \\ \gamma & \cdots & \gamma & D & \cdots & C \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \gamma & \cdots & \gamma & C & \cdots & D \end{bmatrix},$$

where

$$(27) \qquad d \;=\; \frac{\partial^2 G_s}{\partial x_0^2}(x), \; c \;=\; \frac{\partial^2 G_s}{\partial x_0 \partial x_1}(x),$$

$$(28) \qquad D \;=\; \frac{\partial^2 G_s}{\partial x_n^2}(x), \; C \;=\; \frac{\partial^2 G_s}{\partial x_n \partial x_{n-1}}(x),$$

$$(29) \qquad \gamma \;=\; \frac{\partial^2 G_s}{\partial x_0 \partial x_n}(x).$$

We get the eigenvalues as the roots of the characteristic polynomial, which we find by subtracting the variable ξ from each diagonal element and taking the determinant, as in (20). In particular, $d - c$ is the k-fold eigenvalue that corresponds to the k-face, and $D - C$ is the ℓ-fold eigenvalue that corresponds to the ℓ-face. Plugging (25) and (26) into (27) and (28), we arrive at

$$\begin{aligned} d - c &= -2\pi G_s(x) + 4\pi^2 s^2 f_0(x), \\ D - C &= -2\pi G_s(x) + 4\pi^2 s^2 f_n(x). \end{aligned}$$

These are the two claimed second derivatives of (23) and (24). $\qquad\qquad\qquad\qquad \boxdot$

Sign change. A point $x = A(t)$ is a mode of $G_s : \mathbb{R}^{n+1} \to \mathbb{R}$ iff it is a maximum of the 1-section defined by A as well as of the n-section defined by H_t. Focusing on the latter, we compute the values of the parameter t at which the second derivatives with respect to κ and with respect to λ vanish. Beginning with κ, we set (23) to zero and find

$$(30) \qquad [4\pi^2 s^2 - 2\pi(k+1)]f_0(x) \;=\; 2\pi(\ell+1)f_n(x).$$

We note that the natural logarithm of $f_n(x)/f_0(x)$ is $-\pi$ times the following difference of squared distances:

$$(31) \qquad \|x - se_n\|^2 - \|x - se_0\|^2 = 2s^2 \cdot \frac{(\ell+1) - t(n+1)}{(k+1)(\ell+1)}.$$

Plugging (31) into (30) gives us

$$\frac{2\pi s^2 - (k+1)}{\ell+1} \;=\; e^{-2\pi s^2 \frac{(\ell+1)-t(n+1)}{(k+1)(\ell+1)}}.$$

Solving this equation, we get t as a function of the scale parameter. We call this function t_K. Doing the symmetric computations for λ, we find a second function $t_L : \mathbb{R} \to \mathbb{R}$, both defined by

$$(32) \qquad t_K(s) \;=\; \frac{k\ell+n}{2\pi s^2(n+1)} \cdot \ln \frac{2\pi s^2 - k - 1}{\ell+1} + \frac{\ell+1}{n+1},$$

$$(33) \qquad t_L(s) \;=\; \frac{k\ell+n}{2\pi s^2(n+1)} \cdot \ln \frac{k+1}{2\pi s^2 - \ell - 1} + \frac{\ell+1}{n+1}.$$

For example, for $s = U_n$, we get $t_K = t_L = \frac{\ell+1}{n+1}$, which is consistent with the Barycenter Lemma, where $s = \frac{\ell+1}{n+1}$ is identified as the scale factor at which the barycenter of the n-simplex changes from a maximum to a minimum. Note also that t_K is undefined for $s = U_k$, and t_L is undefined for $s = U_\ell$.

Chandelier. To get a feeling for the situation, we draw the trajectories of the critical points of G_s, and in particular those of the modes. We call this set in $\mathbb{R}^{n+1} \times \mathbb{R}$ the *chandelier* of the 1-parameter family of functions. Letting s increase from bottom to top, Figure 7 sketches the chandelier for $n = 1, 2$. The most prominent feature is the *base point*, which we use to decompose the chandelier into curves. Two of these curves are vertical, both swept out by the barycenter of the n-simplex, which changes from index $n + 1$ to index 1 when it passes through the base point. For each curve, we consider the *height function* defined by mapping $(x, s) \in \mathbb{R}^{n+1} \times \mathbb{R}$ to s, and we further subdivide so that the height
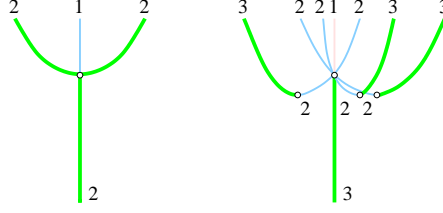
FIGURE 7. The chandelier for $n = 1$ on the the left and for $n = 2$ on the right. Each curve is labeled by the index of its critical points.

function is injective. In other words, we cut each curve at the local minima and maxima of the height function. The benefit of this subdivision is that now each curve is swept out by a critical point of G_s with constant index. While the total number of curves in the chandelier grows exponentially with the dimension, the number of curves that correspond to modes grows only by one for each dimension. To count the curves, we compute the number of complementary face pairs of the n-simplex:

$$p_n = \frac{1}{2} \sum_{k=0}^{n-1} \binom{n+1}{k+1} = 2^n - 1.$$

For each pair, two branches emanate from the base point. Adding the vertical line, we count $2p_n + 2 = 2^{n+1}$ branches. For each complementary face pair with $0 \leq k < \ell$, the height function of one of the two corresponding branches has a local minimum and is therefore subdivided into two curves. The number of local minima is

$$l_n = \begin{cases} p_n & \text{if } n \text{ is even,} \\ p_n - \frac{1}{2}\binom{n+1}{(n+1)/2} & \text{if } n \text{ is odd.} \end{cases}$$

The total number of curves is therefore $2p_n + 2 + l_n$. Of these, only $n + 2$ correspond to modes.

Indices. The index swept out by a curve in the chandelier is easy to determine numerically, but at this time, we lack analytic proofs. We fist state the result and second explain the numerical evidence that supports it.

$0 \leq k < \ell$**::** There are $\binom{n+1}{k+1}$ complementary face pairs of k- and ℓ-faces. Besides the barycenter, the corresponding axes witness critical points of index $\ell + 2$ and $\ell + 1$ for $s \in (T_{k,\ell}, U_n)$ and two critical points of index $k + 2$ for $s > U_n$.

$k = \ell = \frac{n-1}{2}$**::** There are $\frac{1}{2}\binom{n+1}{k+1}$ complementary pairs of k-faces. Besides the barycenter, the corresponding axes witness two critical points of index $k + 2$ for $s > T_{k,k} = U_n$.

To explain the numerical evidence, we consider $t_K(s)$ and $t_L(s)$, which are given by (32) and (33). We make t_K injective by restricting it to the range $[0, \frac{\ell+1}{n+1}]$ and we make t_L injective by restricting it to the range $[\frac{\ell+1}{n+1}, 1]$; see Figure 8, which plots the inverses of the restricted functions. Drawing the horizontal line for a value of s, we note that the portion below the graphs of t_K and t_L consists of the points x at which the n-section orthogonal to the axis has a maximum at x. We see these graphs for even and odd values of n in Figure 8. For each scale factor s, there is either one or two modes witnessed by $A_{k,\ell}$, drawn in cyan in Figure 8. We notice empirically that the mode at the barycenter, given by $t = \frac{\ell+1}{n+1}$, is the only mode under the piecewise defined curve for $0 < k < \ell$. This means that the only mode is at the barycenter and the other critical points are saddles of index $\ell + 1$ and $\ell + 2$.

3.4. **Resilient Modes.** We have seen that the sum of Gaussian kernels can have extra modes. In this subsection, we study their significance, showing that they last for an interval of scale factors whose length increases with the dimension.

(a) $k = 2$ and $\ell = 4$.   (b) $k = 2$ and $\ell = 5$.

(c) $k = 1$ and $\ell = 5$.   (d) $k = 3$ and $\ell = 4$.
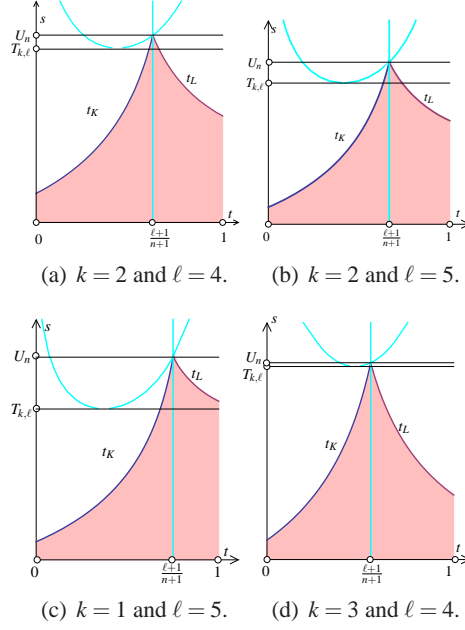
FIGURE 8. Graphs of t_K and t_L for select values of k and ℓ.

Balancing scales. To get started, we need more information on the transition at which the extra maxima appear. We get an upper bound on $T_{k,\ell}$ by studying the scale factor at which the weights of the two 1-dimensional kernels in the decomposition of G_s restricted to a relevant axis are balanced. For $k = \ell$, the two 1-dimensional kernels in the decomposition are always balanced. For $k < \ell$, the *balancing scale factor* is

$$(34) \qquad B_{k,\ell} \;\;=\;\; \sqrt{\frac{\ln(\ell+1) - \ln(k+1)}{\pi \left(\frac{1}{k+1} - \frac{1}{\ell+1} \right)}}.$$

Indeed, recomputing the weights gives $(k+1)g(B_{k,\ell}R_k) = (\ell+1)g(B_{k,\ell}R_\ell)$. Similar to $T_{k,\ell}$ and U_n, the balancing scale factor increases with respect to k and ℓ. Numerically, we observed that $B_{k,\ell}$ is not very different, but consistently larger than $T_{k,\ell}$. We prove that this relationship is not accidental.

Lemma (Transition Lemma). *We have $T_{k,\ell} < B_{k,\ell} < U_n$ for all integers $0 \le k < \ell$ with $k + \ell = n - 1$.*

Proof. We prove the claim indirectly, by showing that $s = B_{k,\ell}$ gives two maxima in the 1-section along any axis connecting the barycenter of a k-face with the barycenter of the complementary ℓ-face. For balanced weights, we have two maxima iff the centers of the two 1-dimensional kernels are further apart than twice the width; see Section 2. To prove the latter property, we compute

$$(35) \qquad \frac{B_{k,\ell}D_{k,\ell}}{2\sigma_0} \;\;=\;\; \sqrt{\frac{n+1}{2(\ell-k)} \cdot \ln \frac{\ell+1}{k+1}},$$

using Equations (4), (1), and (34). Recall the logarithmic inequality:

$$\frac{x}{1 + \frac{x}{2}} \;\;<\;\; \ln(1+x),$$

for $x > 0$. Setting $x = \frac{\ell-k}{k+1}$, we see that the right hand side of (35) exceeds 1 for all choices of $0 \le k < \ell$. This implies that we have two maxima along the axis, which implies that the balancing scale factor lies between the first and second transitions, as claimed.   ⌑

Resilience. We define the *resilience* of a mode as the length of the interval of scale values at which it exists. This definition is not satisfactory for a general 1-parameter family of smooth functions; however, it will suffice in our context, in which we know enough about the modes to follow them through the family parameterized by the scale s. Specifically, we have a single mode for $0 \leq s \leq T_{0,n-1}$, and we have $n+1$ modes for $U_n \leq s$. The picture is more interesting in the interval $T_{0,n-1} < s < U_n$, in which we have $n+2$ modes. One of these modes is the barycenter of the n-simplex, and we study the resilience of this *extra* mode. The upper endpoint of the interval is defined in (16), and an upper bound for the lower endpoint is given in the Transition Lemma, with the definition of the bound in (34):

$$T_{0,n-1} \quad < \quad \sqrt{\frac{\ln n}{\pi(1 - 1/n)}}.$$

As n goes to infinity, U_n grows roughly like the square root of n, and $T_{0,n-1}$ grows roughly like the square root of the logarithm of n. The gap between the two widens, so that the resilience of the mode at the barycenter of the n-simplex grows roughly like $\sqrt{n}$; see Figure 6.

Summary. We are now ready to summarize the findings in regard to the critical points and the modes of the 1-parameter family of functions $G_s : \mathbb{R}^{n+1} \to \mathbb{R}$. For values $s < T_{0,n-1}$, we have a single critical point with index $n+1$. Thereafter, we pick up $2\binom{n+1}{k+1}$ critical points at every $T_{k,\ell}$, for $0 \leq k < \ell$, until we accumulate $2l_n + 1$ critical points right before reaching U_n. The barycenter has index $n+1$, and the other critical points come in pairs, with indices $\ell + 2$ and $\ell + 1$, for $\frac{n-1}{2} < \ell \leq n-1$. For $U_n < s$, we have $2p_n + 1$ critical points. The barycenter has index 1, and the other critical points come in pairs with indices $\ell + 2$ and $k + 2$, for $\frac{n-1}{2} \leq \ell \leq n-1$. As a sanity check, we consider the Euler-Poincaré formula, which states that the alternating sum of critical points is equal to the Euler characteristic of $\mathbb{R}^{n+1}$:

$$(36) \qquad \sum_{i=0}^{n+1} (-1)^i c_i \quad = \quad (-1)^{n+1},$$

where c_i counts the critical points with index i. We also write $c = \sum_{i=0}^{n+1} c_i$. Trivially, (36) holds in the first case. Thereafter, we pick up the critical points in pairs whose contribution to the alternating sum cancel, so (36) is maintained. Finally, for $U_n < s$, we have a bijection between the critical points and the faces of the n-simplex such that the index is $n+1$ minus the dimension of the face. Since the n-simplex is a closed ball, its Euler characteristic is 1, which again implies (36). We thus have a complete description of the critical points of the n-design as the scale factor increases from zero to infinity.

Main Theorem. *Let $n \geq 1$ and consider the sum of $n+1$ unit Gaussian kernels placed at the vertices of the scaled standard n-simplex, $s\Delta^n$.*

(1) *For $s < T_{0,n-1}$, we have 1 critical point which is also a mode.*
(2) *For $T_{0,n-1} < s < U_n$, we have gradually more critical points after passing each $T_{k,\ell}$, until we accumulate $2l_n + 1$ critical points right before U_n. Of these critical points, $n+2$ are modes, and they exist during the entire interval.*
(3) *For $U_n < s$, we have $2p_n + 1$ critical points, of which $n+1$ are modes.*

The resilience of the extra mode in Case (2) is $U_n - T_{0,n-1}$, which grows like $\sqrt{n}$.

3.5. **Many Modes.** In this subsection, we construct a finite configuration of isotropic Gaussian kernels with a superlinear number of modes. While there is a family of such constructions, it will suffice to explain one.

Products of simplices. The basic building block of our construction is the standard 2-simplex. Let the dimension be $3n$ and write the $3n$-dimensional Euclidean space as the

Cartesian product of n 3-dimensional planes: $\mathbb{R}^{3n} = H_1 \times H_2 \times \ldots \times H_n$, in which H_i is spanned by the three coordinate vectors e_{3i-2}, e_{3i-1}, e_{3i}, for $1 \leq i \leq n$. Let Δ_i^2 be the standard 2-simplex in H_i, with vertices $v_{i0} = e_{3i-2}$, $v_{i1} = e_{3i-1}$, $v_{i2} = e_{3i}$. Correspondingly, we write $g_{ij} : H_i \to \mathbb{R}$ for the 3-dimensional unit Gaussian kernel with center v_{ij}, for $0 \leq j \leq 2$, and $G_i : H_i \to \mathbb{R}$ defined by

$$G_i(x) \quad = \quad g_{i0}(x) + g_{i1}(x) + g_{i2}(x)$$

for their sum. Next, we construct a $3n$-dimensional sum of Gaussian kernels by taking products. To begin, we let $P \subseteq \mathbb{R}^{3n}$ be the largest subset of points whose orthogonal projection to H_i is $\{v_{i0}, v_{i1}, v_{i2}\}$, for $1 \leq i \leq n$. This is the set of 3^n points formed by taking the Cartesian product of the n triplets of points. For each point $p \in P$, let $f_p : \mathbb{R}^{3n} \to \mathbb{R}$ be the unit Gaussian kernel with center p. Adding these kernels, we get $F : \mathbb{R}^{3n} \to \mathbb{R}$, defined by

$$F(x) \quad = \quad \sum_{p \in P} f_p(x).$$

To understand F, we recall that f_p can be written as the product of $3n$ 1-dimensional unit Gaussian kernels; see (6). Collecting the terms in sets of three, we can write

$$f_p(x) \quad = \quad \prod_{i=1}^{n} g_{ij}(x),$$

where j is chosen such that v_{ij} is the orthogonal projection of p onto H_i. Substituting the sum of the three kernels for the singletons, we obtain

$$F(x) \quad = \quad \prod_{i=1}^{n} G_i(x).$$

In words, the sum of the 3^n $3n$-dimensional unit Gaussian kernels is the product of n sums of three 3-dimensional unit Gaussian kernels.

Counting modes. We arrive at the final construction by reintroducing the scale factor, writing $F_s : \mathbb{R}^{3n} \to \mathbb{R}$ for the product of the $G_{is} : H_i \to \mathbb{R}$, where G_{is} is of course the sum of the three unit Gaussian kernels with centers se_{3i-2}, se_{3i-1}, se_{3i}. We have seen in Section 3 that s can be chosen such that G_{is} has 4 modes. Since F_s is the product of the G_{is}, its sets of modes is the largest subset of $\mathbb{R}^{3n}$ whose orthogonal projection to H_i is the set of four modes of G_{is}, for $1 \leq i \leq n$. Its size is $4^n = 3^{(1+\log_3 \frac{4}{3})n} > 3^{1.261n}$. This shows that the number of modes is roughly the number of kernels to the power 1.261.

There is an entire family of similar constructions. The one presented here neither maximizes the number nor the resilience of the extra modes. Indeed, we can increase the exponent by improving the ratio of modes over kernels in each H_i, and we can improve the resilience by using higher-dimensional simplices.

4. DISCUSSION

The main contribution of this paper is a cautionary message about the sum of Gaussian kernels. Giving a detailed analysis of the construction studied in [5], we show that there is indeed only one extra mode, but that its resilience increases like the square root of the dimension. We also exhibit configurations of finitely many identical isotropic Gaussian kernels whose sums have superlinearly many modes. We thus give precisely quantified contradictions to our intuition that diffusion erodes and eliminates local density maxima.

The results in this paper raise a number of questions. How stable are the extra maxima? Our analysis in Section 3.4 answers the question when the perturbation is the diffusion of density. How robust are they under moving individual kernels or changing their weights? Related to this question, we ask about the probability of extra modes for randomly placed Gaussian kernels in Euclidean or other spaces. Carreira-Perpiñán and Williams report that their computerized searches in $\mathbb{R}^2$ did not turn up any extra modes [5], but what if we did similar experiments in three and higher dimensions? Finally, it would be interesting to

determine the persistence of the extra modes; see [6] for a recent related study. In other words, how large is the difference in function value between an extra mode and the highest saddle? Understanding the persistence, as well as the basin of attraction for each mode would complement the analysis provided in this paper.

REFERENCES

[1] J. BABAUD, A. P. WITKIN, W. BAUDIN AND R. O. DUDA. Uniqueness of the Gaussian kernel for scale-space filtering. *IEEE Trans. Pattern Anal. Mach. Intel.* **8** (1986), 26–33.

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

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

[4] M. CARREIRA-PERPIÑÁN AND C. WILLIAMS. On the number of modes of a Gaussian mixture. In *Scale-Space Methods in Computer Vision*, Lecture Notes in Computer Science **2695**, 2003, 625–640.

[5] M. CARREIRA-PERPIÑÁN AND C. WILLIAMS. An isotropic Gaussian mixture can have more modes than components. Rept. EDI-INF-RR-0185, School of Informatics, Univ. Edinburgh, Scotland, 2003.

[6] C. CHEN AND H. EDELSBRUNNER. Diffusion runs low on persistence fast. In "Proc. 13th Internat. Conf. Comput. Vision, 2011", 423–430.

[7] J. DAMON. Local Morse theory for solutions to the heat equation and Gaussian blurring. *J. Diff. Equations* **115** (1995), 368–401.

[8] W. FELLER. *An Introduction to Probability Theory and Its Applications, Vol. I.* Wiley, New York, 1950.

[9] J. J. KOENDERINK. The structure of images. *Biol. Cybern.* **50** (1984), 363–370.

[10] A. KUIJPER AND L. M. J. FLORACK. The application of catastrophe theory to image analysis. Rept. UU-CS-2001-23, Dept. Comput. Sci., Utrecht Univ., Utrecht, the Netherlands, 2001.

[11] A. KUIJPER AND L. M. J. FLORACK. The relevance of non-generic events in scale space models. *Internat. J. Comput. Vision* **57** (2004), 67–84.

[12] L. M. LIFSHITZ AND S. M. PIZER. A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Trans. Pattern Anal. Mach. Intell.* **12** (1990), 529–540.

[13] T. LINDEBERG. *Scale-Space Theory in Computer Vision.* Kluwer, Dortrecht, the Netherlands, 1994.

[14] J. MILNOR. *Morse Theory.* Princeton Univ. Press, New Jersey, 1963.

[15] J. RIEGER. Generic evolutions of edges on families of diffused greyvalue surfaces. *J. Math. Imaging Vision* **5** (1995), 207–217.

[16] S. J. ROBERTS. Parametric and non-parametric unsupervised cluster analysis. *Pattern Recognition* **30** (1997), 261–272.

[17] B. W. SILVERMAN. Using kernel density estimates to investigate multimodality. *J. Royal Stat. Soc., B* **43** (1981), 97–99.

[18] A. P. WITKIN. Scale-space filtering. *In* "Proc. 8th Internat. Joint Conf. Artif. Intell., 1983", 1019-1022.

[19] A. L. YUILLE AND T. A. POGGIO. Scaling theorems for zero crossings. *IEEE Trans. Pattern Anal. Mach. Intel.* **8** (1986), 15–25.

APPENDIX A

In this appendix, we give a detailed analysis of the intersections between the weight function, $\omega_{k,\ell}$, and the two transition functions, $\tau_{k,\ell}$ and $\upsilon_{k,\ell}$, all introduced in Section 3.2. We recall that the two transitions, $T_{k,\ell}$ and U_n, are the solutions to $\omega_{k,\ell}(s) = \tau_{k,\ell}(s)$ and to $\omega_{k,\ell}(s) = \upsilon_{k,\ell}(s)$ respectively. We will find that both transitions are well defined, and the second transition depends on n but not on the choice of k and ℓ, as claimed in Section 3.2.

Curve analysis. We discuss the graphs of the three functions to convey a feeling for how they intersect. To begin, we note that $s_0 = 2\sigma_0/D_{k,\ell}$ is the smallest scale factor for which the transition functions are defined. Writing $z = \frac{s}{2}D_{k,\ell}$, we have $z^2 - \sigma_0^2 = 0$ for $s = s_0$. It follows that $\tau_{k,\ell}(s_0) = \upsilon_{k,\ell}(s_0) = 1$. To compare this with the weight function at the same value, we compute

$$\omega_{k,\ell}(s_0) = \frac{\ell+1}{k+1} \cdot e^{-\frac{2(\ell-k)}{n+1}},$$

(37)
$$\ln \omega_{k,\ell}(s_0) = \ln \frac{\ell+1}{k+1} - \frac{2(\ell-k)}{n+1}.$$

We interpret the right hand side of (37) as the difference between the area below the graph of $\frac{1}{x}$, for $k+1 \leq x \leq \ell+1$, and the area below the line that touches the graph at the midpoint
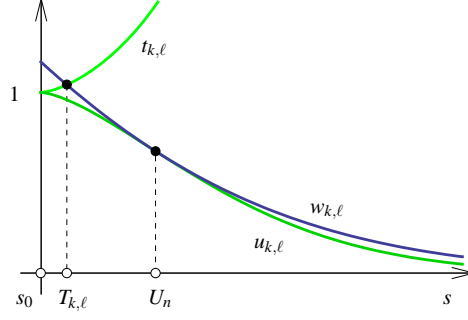
FIGURE 9. The graphs of the weight function and the two transition functions. The intersections define the transitions.

of the interval, again for $k+1 \leq x \leq \ell+1$. Since $\frac{1}{x}$ is a convex function, the second area is smaller, which implies $\ln \omega_{k,\ell}(s_0) > 0$ and therefore $\omega_{k,\ell}(s_0) > 1$. This implies (38), which is the first of the two pairs of inequalities that describe the relation between the three functions on the left and the right:

$$\tag{38} \tau_{k,\ell}(s_0) \;=\; \upsilon_{k,\ell}(s_0) \;<\; \omega_{k,\ell}(s_0),$$

$$\tag{39} \upsilon_{k,\ell}(s) \;<\; \omega_{k,\ell}(s) \;<\; \tau_{k,\ell}(s),$$

as in Figure 9. To compare the functions on the right, for sufficiently large s, we look at the exponents. Writing $z = \frac{s}{2}D_{k,\ell}$, as before, the exponent of $\omega_{k,\ell}(s)$ is $-4\pi z^2 \frac{\ell-k}{n+1}$, which is clearly smaller than the exponent of $\tau_{k,\ell}(s)$, which is $4\pi z \sqrt{z^2 - \sigma_0^2}$. Less obvious is the comparison with the exponent of $\upsilon_{k,\ell}(s)$, which is $-4\pi z\sqrt{z^2 - \sigma_0^2}$. After dividing by $-4\pi z$ and squaring, we obtain $z^2(\ell-k)^2/(n+1)^2 < z^2 - \sigma_0^2$. It follows that the exponent of $\omega_{k,\ell}(s)$ is larger than that of $\upsilon_{k,\ell}(s)$, for s large which implies (39).

From (38) and (39), we conclude that the number of intersections between the graphs of $\omega_{k,\ell}$ and $\tau_{k,\ell}$ (counting with multiplicity) is odd. Since both functions are monotonic, with slopes of opposite signs, we have exactly one intersection. It follows that $T_{k,\ell}$ is well defined.

Double intersection. Similarly, we use (38) and (39) to conclude that the number of intersections between the graphs of $\omega_{k,\ell}$ and $\upsilon_{k,\ell}$ (again counting with multiplicity) is even. We will establish that there is only one double intersection, namely at $s = U_n$. To prove that U_n is a solution to the equation $\omega_{k,\ell}(s) = \upsilon_{k,\ell}(s)$, we write $A = (\ell+1) + (k+1)$ and $B = (\ell+1) - (k+1)$, noting that $A^2 - B^2 = 4(k+1)(\ell+1)$. Setting

$$s \;=\; \sqrt{\frac{n+1}{2\pi} + \frac{y}{2\pi(n+1)}},$$

and recalling that $A = n+1$, we have

$$z \;=\; \sqrt{\frac{A^2+y}{2\pi(A^2-B^2)}},$$

$$\sqrt{z^2 - \sigma_0^2} \;=\; \sqrt{\frac{B^2+y}{2\pi(A^2-B^2)}},$$

$$\frac{z + \sqrt{z^2-\sigma_0^2}}{z - \sqrt{z^2-\sigma_0^2}} \;=\; \frac{\sqrt{A^2+y} + \sqrt{B^2+y}}{\sqrt{A^2+y} - \sqrt{B^2+y}},$$

$$-4\pi z\sqrt{z^2-\sigma_0^2} \;=\; -\frac{2\sqrt{A^2+y}\sqrt{B^2+y}}{A^2-B^2}.$$

Using the definitions of the weight function in (17) and the second transition function in (19), we arrive at

$$\omega_{k,\ell}(s) = \frac{(A+B)^2}{A^2-B^2} \cdot e^{-\frac{2AB+2\frac{B}{A}y}{A^2-B^2}},$$

$$\upsilon_{k,\ell}(s) = \frac{[\sqrt{A^2+y}+\sqrt{B^2+y}]^2}{A^2-B^2} \cdot e^{-\frac{2\sqrt{A^2+y}\sqrt{B^2+y}}{A^2-B^2}}.$$

Clearly, $\omega_{k,\ell}(s) = \upsilon_{k,\ell}(s)$ if $y = 0$, which shows that $s = U_n$ is indeed a solution to the equation. We continue by showing that for small but non-zero y, we have $\omega_{k,\ell}(s) > \upsilon_{k,\ell}(s)$. This is equivalent to showing that the natural logarithm of $\upsilon_{k,\ell}(s)$ over $\omega_{k,\ell}(s)$ is smaller than 0. Equivalently, LHS < RHS, where

$$\text{LHS} = \ln \frac{\sqrt{A^2+y}+\sqrt{B^2+y}}{A+B},$$

$$\text{RHS} = \frac{\sqrt{A^2+y}\sqrt{B^2+y}-AB-\frac{B}{A}y}{A^2-B^2}.$$

To prove this inequality for small values of y, we use the Taylor expansions of the square root and the natural logarithm functions:

$$\sqrt{A^2+y} = A+\frac{y}{2A}-\frac{y^2}{8A^3}+\frac{y^3}{16A^5}-\cdots,$$

$$\ln(1+x) = x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots.$$

With this, we can re-write the two sides of the inequality: LHS $= l_1 y + l_2 y^2 + \ldots$, and RHS $= r_1 y + r_2 y^2 + \ldots$. Computing the coefficients, in turn, we find

$$l_1 = \frac{1}{2AB} = r_1,$$

$$l_2 = -\frac{A^2+B^2}{8A^3B^3} < -\frac{A^2-B^2}{8A^3B^3} = r_2.$$

In words, $s = U_n$ is a double solution of the equation, and $\omega_{k,\ell}(s) > \upsilon_{k,\ell}(s)$ for values of s chosen in a small neighborhood but different from U_n.

H. EDELSBRUNNER, IST AUSTRIA, KLOSTERNEUBURG, AUSTRIA,
E-mail address: edels@ist.ac.at

B. FASY, IST AUSTRIA, KLOSTERNEUBURG, AUSTRIA, DEPARTMENT OF COMPUTER SCIENCE, DUKE UNIVERSITY, DURHAM, NORTH CAROLINA, *Present address:* COMPUTER SCIENCE DEPARTMENT, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PENNSYLVANIA.
E-mail address: brittany@cs.duke.edu

G. ROTE, INSTITUT FÜR INFORMATIK, FREIE UNIVERSITÄT BERLIN, GERMANY,
E-mail address: rote@inf.fu-berlin.de