Diffusion Runs Low on Persistence Fast*

Chao Chen [†]　　Herbert Edelsbrunner [‡]

Abstract

Interpreting an image as a function on a compact subset of the Euclidean plane, we get its scale-space by diffusion, spreading the image over the entire plane. This generates a 1*-parameter family of functions alternatively defined as convolutions with a progressively wider Gaussian kernel. We prove that the corresponding* 1*-parameter family of persistence diagrams have norms that go rapidly to zero as time goes to infinity. This result rationalizes experimental observations about scale-space. We hope this will lead to targeted improvements of related computer vision methods.*

1. Introduction

The work described in this paper is motivated by experimental observations about scale-space, which is defined by progressive diffusion of an image. Here, we think of the image as a real-valued function on a compact subset of the Euclidean plane, and we idealize the effect of diffusion by convolving the function with an isotropic Gaussian kernel. This generates a 1-parameter family of functions on the entire $\mathbb{R}^2$, which we refer to as the *scale-space* of the image. The construction is popular for the extraction of local to global features [2, 17], which are useful for image registration, camera calibration, and object recognition [18], among other computer vision tasks. Scale-spaces can be computed for the image itself, or for derivatives including the Laplacian and the determinant of the Hessian [19, 20]. Scale-space has been defined by Iijima [10], and was rediscovered by Witkin [27] and by Koenderink [11]; but see also [16].

In this paper, we study the evolution of the structural information contained in the functions of scale-space. As a general tendency, diffusion washes out details, so we can expect the number of critical points to decrease. However, there are cases in which the critical points grow in number. Here we discuss one classic example [14, 16]. Making the construction symmetric, we connect two mountains by a relatively narrow arched bridge that reaches its highest point in the middle, so that the function has its sole maximum halfway between the mountains; see Figure 1, left. Even a small amount of diffusion suffices to erode the arch, while the two mountains remain relatively unaffected. We thus get two maxima separated by a saddle; see Figure 1, middle. Further diffusion also erodes the two mountains and leaves one relatively shallow hill, so we are back to a single maximum; see Figure 1, right. Drawing the trajectories of the critical points in scale-space [12], we see the initial maximum split into two maxima and a saddle which later merge back into a single maximum.

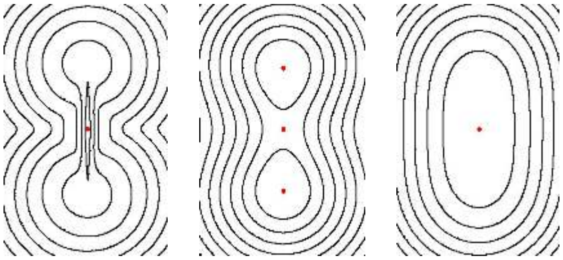

Figure 1: From left to right: the collapse of the narrow bridge connecting two mountains. The functions are illustrated by their level sets and critical points (red dots).

While critical points are sometimes created by diffusion, the experimental evidence suggests that this rarely happens. Nevertheless, it is known that the creation of critical points is a generic event; see Damon [6] and Rieger [23]. More surprising than this creation is perhaps the possibility of diffusing a finite number of point masses and getting more maxima than point masses during an open time interval; see [4]. The created critical points tend to be fragile, existing only for a short time, but there are again counterexamples: we can design our arched bridge to make the gorge that opens up between the two mountains as deep as we like.

In this paper, we give a quantification of the diffusion process, in terms of the persistent homology of the functions, that explains the experimental evidence. In a nutshell, we sweep out the function by gradually increasing the cut-off value, and we observe topology changes of the subset of points where the function value lies below the cut-off. New features are acquired and old features are lost. Calling

---

*The authors acknowledge partial support by the FWF under grant P20134-N13 and the NFS under grant DBI-0820624.

[†] IST Austria, Klosterneuburg & PRIP, Vienna Univ. Techn., Vienna, Austria. http://www.ist.ac.at/~cchen

[‡] IST Austria, Klosterneuburg, Austria & Dept. Comput. Sci., Duke Univ., Durham, North Carolina & Geomagic, North Carolina, USA.

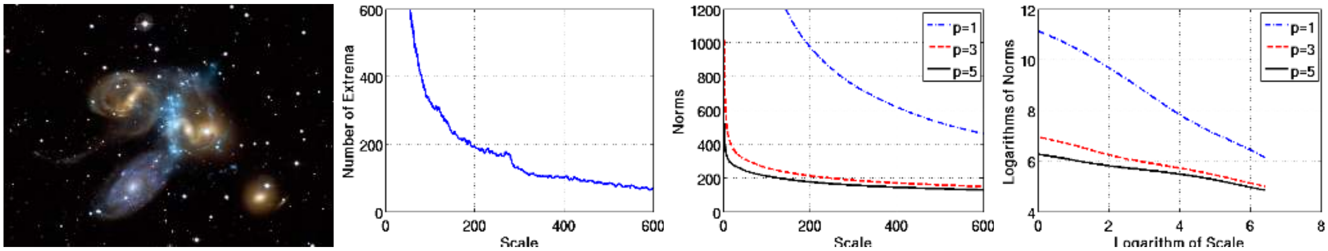

Figure 2: From left to right: an image of galaxies (from Flickr), the number of extrema as a function of the scale t, the norms of the persistence diagram for $p = 1, 3, 5$, and same norms in log-log scale. In contrast to the number of extrema, we observe steadily decreasing norms, whose logarithms are roughly linear in the logarithm of the scale. We have a proof of const/t as an upper bound only for $p > 4.562$.

these events *births* and *deaths*, we pair them up and define the difference between their values as the *persistence* of the feature. Consider the middle function in Figure 1 as an example. As explained in Section 2, the global minimum is paired with the global maximum and the saddle is paired with the minor maximum. The persistence of the first pair is the height of the main mountain, while the persistence of the second pair is the depth of the gorge between the mountains. Since the gorge can be as deep as we like, we see that diffusion can create features with arbitrarily large persistence. In contrast, its ability to affect the average persistence is more limited.

To make this precise, we introduce the p-norm of the persistence diagram, which takes the p-th root of the sum of the p-th powers of all persistences, and we prove bounds on this measure. We state our result for a compact subset Ω of the Euclidean plane, leaving the formulation of the n-dimensional result to the technical sections of this paper. Let $f : \Omega \to \mathbb{R}$ be a function, and define $f_t : \mathbb{R}^2 \to \mathbb{R}$ by convolving f with the isotropic Gaussian kernel with scale $t > 0$. Then for every real number $p > \frac{1}{2}(5 + \sqrt{17})$, the p-norm of the persistence diagram of f_t satisfies

$$\|\mathrm{Dgm}(f_t)\|_p \quad \leq \quad \mathrm{const}/t; \tag{1}$$

see Figure 2. This upper bound is tight. We mention that measuring the information contained in a smoothed function is not a new idea. Other such measures studied in the literature include the generalized entropy [25], the method noise [3], and the L_2-norms of functions and residues [9]. The norm of the persistence diagram proposed in this paper is different as it captures more of the high-level structural information contained in a function. Persistence is related to Morse theory [21], couched in the algebraic language of homology, and blessed with efficient combinatorial as well as algebraic algorithms [7].

Most directly related to our result is the work by Lindeberg [15]. He considers n-dimensional images and proves, both theoretically and experimentally, that for $n = 1$ and for random noise, the expected number of critical points is $\mathrm{const}/t^{n/2}$, conjecturing the same for $n > 1$. The restriction of our Main Theorem to $n = 1$ gives a similar result with weaker assumptions on the image, which extends to higher dimensions as expected. It thus rationalizes an intuition that has been used implicitly for decades. We interpret our result as evidence that persistent homology can gain insight into popular techniques, including the extraction of keypoints [20]. The additional insight may lead to refinements of these techniques, e.g. by weighting the critical points with their persistence. In contrast to many other studies of keypoint extraction, persistence has solid mathematical foundations, while being intuitive and applying to real-valued functions of any dimension. It therefore gives hope for extensions to global image structures [26] and features in 3- and 4-dimensional images [13].

Outline. Section 2 introduces the necessary background in analysis and algebraic topology. Section 3 establishes a connection between the amplitude and the persistence diagram of a function. Section 4 analyzes the convolutions of a function and proves our main result. Section 5 concludes the paper.

2. Background

In this section, we introduce the background we will need in Sections 3 and 4. Beginning with topics in analysis and algebraic topology, we conclude with persistent homology, which forms a bridge between the two mathematical disciplines.

Convolution and diffusion. The normal distribution plays a special role in probability. The associated *normal density* is the function $g_t : \mathbb{R}^n \to \mathbb{R}$ defined by

$$g_t(x) \quad = \quad \frac{1}{(2\pi t)^{\frac{n}{2}}} \cdot e^{-\frac{|x|^2}{2t}}, \tag{2}$$

where $|x|$ is the Euclidean norm of $x \in \mathbb{R}^n$. In $\mathbb{R}^1$, this density function has a symmetric, bell-shaped graph with exponentially decaying tails at both sides. We refer to g_t as the *Gaussian kernel* with *mean* zero and *scale* t. It is a mathematical model of many physically important phenomena, including the stochastic location of a randomly moving particle with initial position at the origin. If instead of a fixed position, we begin with an initial density function,

$f : \mathbb{R}^n \to \mathbb{R}$, we get the density at time $t > 0$ by *convolution* with the Gaussian kernel: $f_t = f * g_t : \mathbb{R}^n \to \mathbb{R}$, defined by

$$f_t(x) \quad = \quad \int_{y \in \mathbb{R}^n} f(y) g_t(x - y) \, \mathrm{d}y. \tag{3}$$

An important property of the Gaussian kernel is its closure under convolution: $g_s * g_t = g_{s+t}$ for all $s, t > 0$. In words: instead of repeatedly convolving, we can convolve f once with a Gaussian kernel of appropriately chosen larger scale. Interpreting f as an initial distribution of heat, we get f_t as the distribution at time $t > 0$. The 1-parameter family of functions f_t is thus the solution to the *heat equation*:

$$\nabla^2 f_t \quad = \quad \mathrm{const} \cdot \frac{\partial f_t}{\partial t}, \tag{4}$$

for some positive constant and initial condition $f_0 = f$. Here $\nabla^2 = \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_n^2}$ is the *Laplace operator*. If the initial condition is a unit amount of heat at the origin, then the solution to the heat equation is $f_t = g_t$.

Amplitudes and norms. Letting $f : \mathbb{R}^n \to \mathbb{R}$ be a function, and B a subset of $\mathbb{R}^n$, we define the *amplitude* of f over B as the supremum difference between two values:

$$\mathrm{amp}_B(f) \quad = \quad \sup_{x,y \in B} |f(x) - f(y)|. \tag{5}$$

Note that the amplitude is related to the *infinity-norm* of f, defined as $\|f|_B\|_\infty = \sup_{x \in B} |f(x)|$, but it is not the same. We have $\mathrm{amp}_B(f) \leq 2\|f|_B\|_\infty$. Assuming f is smooth, we can take its gradient, $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$, and the magnitude of the gradient, $|\nabla f| : \mathbb{R}^n \to \mathbb{R}$. The latter is a real-valued function, so we can define its *p-norm* over B as the p-th root of the integral of the p-th power of the magnitude:

$$\|\nabla f|_B\|_p \quad = \quad \left(\int_{x \in B} |\nabla f(x)|^p \, \mathrm{d}x \right)^{\frac{1}{p}}. \tag{6}$$

Letting p grow, we get the infinity-norm of $|\nabla f|$ in the limit as $\|\nabla f|_B\|_\infty = \sup_{x \in B} |\nabla f(x)|$.

We will be interested in bounding the amplitude of f in terms of the norm of the gradient, which is achieved by the Sobolev inequalities. A version that is particularly useful for our purposes is Theorem 1.4.2 in [24, page 22]:

Sobolev Inequality. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be smooth, $\mathbb{B} \subseteq \mathbb{R}^n$ a closed Euclidean ball, and $p > n$. Then*

$$|f(x) - f(y)| \quad \leq \quad \mathrm{const} \cdot \mathrm{vol}(\mathbb{B})^{\frac{1}{n} - \frac{1}{p}} \|\nabla f|_{\mathbb{B}}\|_p, \tag{7}$$

for all points $x, y \in \mathbb{B}$, where $\mathrm{vol}(\mathbb{B})$ is the n-dimensional volume, and the constant factor depends on n and p.

It is not very difficult to adjust the proof to get the same inequality (with a different constant) for a cube instead of a Euclidean ball. This gives the following easy consequence of the Sobolev Inequality.

Corollary. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be smooth, $B \subseteq \mathbb{R}^n$ an n-dimensional cube, and $p > n$. Then*

$$\mathrm{amp}_B(f) \quad \leq \quad \mathrm{const} \cdot \mathrm{vol}(B)^{\frac{1}{n} - \frac{1}{p}} \|\nabla f|_B\|_p, \tag{8}$$

where the constant depends on n and p.

Homology. Given a topological space, we use homology groups to characterize how the space is connected. There are a number of different but equivalent theories to construct these groups, and we will sketch the simplicial homology since we will need triangulations to prove our main result. Within each theory, we can use different coefficient groups, leading to potentially different homology groups, but the differences are well understood. There are many textbooks in algebraic topology that cover homology groups in detail, and we recommend [22] as one of them.

We now give a formal introduction of simplicial complexes and the related homology theory. Recall that a *j-simplex* in $\mathbb{R}^n$ is the convex hull of $j + 1$ affinely independent points. A subset of $i + 1$ of the $j + 1$ points defines an i-simplex that is a *face* of the j-simplex. A *simplicial complex* is a finite collection of simplices, K, that is closed under the face relation such that any two simplices are either disjoint or they intersect in a common face. The *underlying space* of K, denoted as $|K|$, is the union of the simplices in K together with the topology inherited from $\mathbb{R}^n$. Note that $|K|$ is a topological space, while K is a combinatorial representation of the same. A *triangulation* of a topological space $\mathbb{X}$ is a simplicial complex K together with a homeomorphism $h : |K| \to \mathbb{X}$. We will now construct the homology groups of K and consider them as the groups of $\mathbb{X}$, which makes sense because different triangulations of the same space give isomorphic groups. We get one group for each dimension, j, which we denote as $\mathsf{H}_j(\mathbb{X})$. Assuming the binary coefficient group, $\mathsf{U} = \mathbb{Z}/2\mathbb{Z}$, with addition modulo 2, each group is a vector space of the form $\mathsf{H}_j(\mathbb{X}) \simeq \mathsf{U}^{\beta_j}$. We call $\beta_j = \beta_j(\mathbb{X})$ the *rank* of $\mathsf{H}_j(\mathbb{X})$ and the *j-th Betti number* of $\mathbb{X}$. To construct the homology group, we call a set c of j-simplices in K a *j-chain*. The *boundary* of c is the set ∂c of $(j-1)$-simplices that belong to an odd number of j-simplices in c. A *j-cycle* is a j-chain with empty boundary. Two j-cycles are *homologous* if their symmetric difference is the boundary of a $(j + 1)$-chain. Finally, a *j-dimensional homology class* is a maximal set of homologous j-cycles, and $\mathsf{H}_j(\mathbb{X})$ is the group of these classes, with addition defined by symmetric difference of representative cycles.

To give an example, let Ω be a rectangular subset of $\mathbb{R}^2$, and let $f : \Omega \to \mathbb{R}$ be the function whose level sets are shown in Figure 1, middle. Let a be halfway between the function value of the saddle and the shared function value of the two maxima. Define $\mathbb{X} = f^{-1}(-\infty, a]$ and note that it is a rectangle with two holes. It has two non-trivial homology groups, namely $\mathsf{H}_0(\mathbb{X}) \simeq \mathsf{U}$ and $\mathsf{H}_1(\mathbb{X}) \simeq \mathsf{U}^2$. The rank of the former, $\beta_0(\mathbb{X}) = 1$, is the number of components, and the rank of the latter, $\beta_1(\mathbb{X}) = 2$, is the number of holes in the rectangle. It will often be convenient to suppress the homological dimension, which we do by writing $\mathsf{H}(\mathbb{X})$ for the direct sum of all homology groups of the space $\mathbb{X}$.

Persistent homology. Instead of a single space, we now assume a nested sequence of spaces. We point out that this sequence is not related to scale-space (or not yet) but is used to study a fixed function, e.g. by considering the sequence of sublevel sets: $\mathbb{X}_a \subseteq \mathbb{X}_b$ for $a \leq b$, where $\mathbb{X}_a = f^{-1}(-\infty, a]$ and similarly for b. Since every cycle in $\mathbb{X}_a$ is also contained in $\mathbb{X}_b$, we have a homomorphism from $\mathsf{H}(\mathbb{X}_a)$ to $\mathsf{H}(\mathbb{X}_b)$, which is induced by the inclusion. In other words, we have a 1-parameter family of homology groups with homomorphisms connecting them from left to right. We call this a *filtration*, and reading it from beginning to end, we can observe when homology classes are *born* and when they *die*. The key insight is the existence of a canonical pairing between births and deaths, which is used to define the *persistence* of a homology class as the absolute difference between the values at its birth and at its death. We refer the reader to [7] for further background on persistent homology.

A convenient representation of the information in a filtration is the *persistence diagram*, which we denote by $\mathrm{Dgm}(f)$. We also assume that f is *tame*, by which we mean that every sublevel set has finite rank homology groups, and that there are only finitely many values at which the group changes non-isomorphically. The persistence diagram is a multiset of dots in the plane, in which each dot represents a birth-death pair, marking the two events with its two coordinates. To avoid the complications caused by classes of $\mathbb{X}$ that are born but never die, we assume that $\mathbb{X} = \mathbb{S}^n$ (the n-dimensional Euclidean space, compactified by adding a point at infinity). In this case, we have only two unpaired births, namely the first component, which is born at the global minimum, and the n-dimensional class, which is born at the global maximum. Pairing up these two events, we thus have finitely many dots, each with finite coordinates. Take the function shown in the middle of Figure 1 as an example. We have four critical points: a minimum at infinity, w, a saddle, x, and two maxima, y and z. In the evolution of the sublevel set, we first have the birth of a 0-dimensional class at $f(w)$, second the birth of a 1-dimensional class at $f(x)$, third the death of the 1-dimensional class at $\min\{f(y), f(z)\}$, and finally the birth of a 2-dimensional class at $\max\{f(y), f(z)\}$. Assuming $f(y) < f(z)$, the two dots in the diagram are $(f(w), f(z))$ and $(f(x), f(y))$. We note an ambiguity when $f(y) = f(z)$, but in this case, we get the same diagram for the alternative pairing of w with y and x with z.

For each dot $u \in \mathrm{Dgm}(f)$, we write $\mathrm{pers}(u)$ for the absolute difference between the two coordinates, and we note that this is the persistence of every class represented by this dot. Finally, we define the *p-norm* of the diagram as

$$\|\mathrm{Dgm}(f)\|_p = \left[\sum_{u \in \mathrm{Dgm}(f)} \mathrm{pers}(u)^p \right]^{\frac{1}{p}}. \tag{9}$$

As proved in [5], the p-norm is a stable measure provided $\mathbb{X}$ is compact with polynomially growing mesh, f is Lipschitz, and $p > n$. In this paper, we consider non-Lipschitz functions on non-compact spaces, and we focus on convergence properties as opposed to stability. Nevertheless, we will get results which again hold for all $p > n$.

The rest of this paper focuses on the proof of the Main Theorem. More experimental results and details of the algorithm for computing persistence will be made available in a technical report available at the author's homepage. As a proof of concept, we show experimental results for a single image in Figure 2.

3. Regions and Cycles

Given a smooth function, we show how to subdivide the domain so that the amplitude within each region is bounded. Using the Corollary of the Sobolev Inequality, we bound the number of regions in terms of the norm of the gradient.

Subdivision. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. For a radius $r \geq 0$, let $\delta(f, r)$ be the amplitude of f outside the n-dimensional cube $B_r = [-r, r]^n$, that is, $\delta(f, r) = \mathrm{amp}_{\bar{B}_r}(f)$, where $\bar{B}_r = \mathbb{R}^n - B_r$. Fixing f, this defines a non-increasing function in r. We say that f has a *flat tail* if $\delta(f, r)$ goes to zero when r goes to infinity. We invert the relationship by defining $r(f, \delta)$ as the infimum radius r for which $\mathrm{amp}_{\bar{B}_r}(f) \leq \delta$. To prepare the next step, we fix a bound $\delta > 0$, let $r = r(f, \delta)$, and consider the n-dimensional cube B_r. Writing $\frac{1}{m} = \frac{1}{n} - \frac{1}{p}$, we define

$$F(B_r) = \mathrm{vol}(B_r)^{\frac{1}{m}} \cdot \|\nabla f|_{B_r}\|_p, \tag{10}$$

noting that $\mathrm{const} \cdot F(B_r)$ is equal to the right-hand side of the Corollary of the Sobolev Inequality. When we subdivide B_r, the volume of a region is predictably smaller while the norm of the gradient over the region may be as large as over the entire B_r. For example, if $s = \frac{r}{k}$, for $k \geq 1$, then

$$\begin{aligned} F(B_s) &= \mathrm{vol}(B_s)^{\frac{1}{m}} \cdot \|\nabla f|_{B_s}\|_p & (11)\\ &\leq (2s)^{\frac{n}{m}} \cdot \|\nabla f|_{B_r}\|_p & (12)\\ &= F(B_r)/k^{\frac{n}{m}}. & (13) \end{aligned}$$

The same inequality holds for every n-dimensional cube of radius $s = \frac{r}{k}$ inside B_r. Assuming k is a positive integer, we can therefore subdivide B_r into k^n cubes B of radius

$s = \frac{r}{k}$, such that $F(B) \leq F(B_r)/k^{\frac{n}{m}}$ for each cube. Our goal is to choose k as small as possible under the constraint

$$F(B) \quad \leq \quad \delta / C_0, \tag{14}$$

where C_0 is the constant in the Corollary of the Sobolev Inequality. By (13), it suffices to choose k such that $F(B_r)/k^{\frac{n}{m}} \leq \delta/C_0$. If $F(B_r) \leq \delta/C_0$ then we pick $k = 1$ and we are done without subdividing. Else, $X = C_0^{\frac{m}{n}} F(B_r)^{\frac{m}{n}}/\delta^{\frac{m}{n}}$ exceeds 1, and we pick $k = 2\lfloor X \rfloor > X$. The number of n-dimensional cubes in the subdivision is therefore

$$\begin{aligned} k^n \quad & \leq \quad 2^n C_0^m \cdot F(B_r)^m / \delta^m & (15) \\ & = \quad \frac{2^n C_0^m}{\delta^m} \cdot \mathrm{vol}(B_r) \cdot \|\nabla f|_{B_r}\|_p{}^m & (16) \\ & \leq \quad (4r)^n \left(C_0 \cdot \|\nabla f\|_p / \delta \right)^m . & (17) \end{aligned}$$

Including the outside region, $\bar{B}_r$, we have generally at most 1 more region than stated on the right-hand side in (17), and at most 2 more if $k = 1$. We therefore define

$$M_f(\delta) \quad = \quad 2 + 4^n r(f,\delta)^n \left(C_0 \frac{\|\nabla f\|_p}{\delta} \right)^m . \tag{18}$$

Finally, we show that the amplitude of f within each region is bounded. Using the Corollary of the Sobolev Inequality stated in Section 2, we get

$$\begin{aligned} \mathrm{amp}_B(f) \quad & \leq \quad C_0 \cdot \mathrm{vol}(B_s)^{\frac{1}{m}} \cdot \|\nabla f|_B\|_p & (19) \\ & = \quad C_0 \cdot F(B), & (20) \end{aligned}$$

which, by (14), is at most δ for each cube B. We have $\mathrm{amp}_{\bar{B}_r}(f) \leq \delta$ by definition of $r = r(f,\delta)$. It follows that the amplitude is bounded from above by δ in every region of the subdivision.

Flat tail compactness. Later in the proof of the Main Theorem, we will use results that have only been established for functions on compact domains. To finesse the difficulties caused by the non-compactness of $\mathbb{R}^n$, we transform f to a function on the n-dimensional unit sphere, $g : \mathbb{S}^n \to \mathbb{R}$. Assuming f has a flat tail, we have the same limit no matter in which direction we go to infinity: $a = \lim_{|x| \to \infty} f(x)$. Let $N = (0, \ldots, 0, 1)$ be the north-pole of $\mathbb{S}^n \subseteq \mathbb{R}^{n+1}$, write $\mathbb{R}^n$ for the n-dimensional plane spanned by the first n coordinate axes, and let $\varpi : \mathbb{S}^n - \{N\} \to \mathbb{R}^n$ be the stereographic projection that maps every point $y \in \mathbb{S}^n$ different from N to

$$\varpi(y) \quad = \quad N + \frac{2(y - N)}{|y - N|^2} \tag{21}$$

in $\mathbb{R}^n$. Accordingly, define $g(\varpi^{-1}(x)) = f(x)$, for all points $x \in \mathbb{R}^n$, and complete the construction by defining $g(N) = a$. Clearly, g has compact support, and if f is continuous with flat tail, then g is continuous.

We now prepare a connection between the norm of the gradient of f and the ranks of the homology groups of the sublevel sets of f. For this purpose, we use a triangulation of $\mathbb{S}^n$, which we recall is a simplicial complex K together with a homeomorphism $h : |K| \to \mathbb{S}^n$. For every $\delta > 0$, we are interested in a triangulation with as few simplices as possible such that the amplitude of f restricted to the image of every simplex, $\varpi(h(\xi))$, is bounded from above by δ.

Subdivision Lemma. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be smooth with flat tail. Then for every $p > n$ and every $\delta > 0$, there is a triangulation of $\mathbb{S}^n$ with at most $2^n n! M_f(\delta)$ simplices, such that the amplitude of f within the image of each simplex is at most δ.*

The extra factor, $2^n n!$, allows for the decomposition of every n-dimensional cube into $n!$ n-simplices, see e.g. [8], and for counting all 2^n faces of each n-simplex.

Counting cycles. The reason for our interest in the number of simplices needed to guarantee a bound on the amplitude within each simplex is its connection to the number of high-persistence dots in the persistence diagram of f. Recall that $\mathrm{Dgm}(f)$ records the events (births and deaths, and their correspondence) during a sweep through the sublevel sets, $f^{-1}(-\infty, a]$, in which a goes from $-\infty$ to ∞. To develop an intuition, let $u, v \in \mathrm{Dgm}(f)$ be two dots representing i-dimensional homology classes of persistence larger than δ. Because of the small amplitude within each $(i+1)$-simplex, each of these classes has a representative in the i-skeleton of the triangulation. Having the same representation in the i-skeleton would contradict both having large persistence, but if they are different then we are limited to the classes generated by the i-skeleton. This suggests that the number of simplices gives an upper bound on the number of dots:

Persistent Cycle Lemma. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be tame and with flat tail. Then the number of dots in $\mathrm{Dgm}(f)$ with persistence larger than δ is at most $2^n n! M_f(\delta)$.*

A formal proof of this lemma can be found in [5]. More specifically, the lemma with the same name in this reference is formulated in terms of an upper bound on the Lipschitz constant of the function f. This difference is not essential and the proof extends virtually unchanged. Similar to [5], we need an integrated version of the bound. To that end, we consider the p-norm of the diagram after removing all dots with persistence at most δ:

$$\|\mathrm{Dgm}(f,\delta)\|_p \quad = \quad \left[\sum_{\mathrm{pers}(u) > \delta} \mathrm{pers}(u)^p \right]^{\frac{1}{p}} , \tag{22}$$

Clearly, $\|\mathrm{Dgm}(f,0)\|_p = \|\mathrm{Dgm}(f)\|_p$, the p-norm of the diagram as defined in Section 2. As proved in [5], we can get a bound on $\|\mathrm{Dgm}(f,\delta)\|_p$ by integrating the bound provided by the Persistent Cycle Lemma:

Diagram Norm Lemma. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be tame and with flat tail. Then $\|\mathrm{Dgm}(f, \delta)\|_p$ is at most*

$$\mathrm{const} \cdot \left[\delta^p M_f(\delta) + \int_{\varepsilon=\delta}^{\mathrm{amp}(f)} M_f(\varepsilon)\varepsilon^{p-1}\,\mathrm{d}\varepsilon \right]^{\frac{1}{p}},$$

for every $\delta > 0$.

We delay the computation of the integral until later, when we will know how $M_f(\varepsilon)$ depends on ε. We will see that this dependence is only polynomial so that it will be easy to evaluate the integral. Letting δ go to zero, we will then obtain a bound on the p-norm of the diagram.

4. Convergence

In this section, we prove the main result of this paper in two steps, first establishing how fast the gradient diffuses to zero and second how the norm of the persistence diagram follows this trend.

Two results on Gaussian kernels. We begin with two exercises in the analysis of a Gaussian kernel, which we will use to prove properties of general functions with compact support. Specifically, we consider $g_t : \mathbb{R}^n \to \mathbb{R}$ and the magnitude of its gradient, defined by

$$|\nabla g_t(x)| \quad = \quad \frac{|x|}{t} \cdot g_t(x). \tag{23}$$

It is easy to compute the maximum value as a function of t, which is $g_t(0) = 1/(2\pi t)^{\frac{n}{2}}$. It decreases monotonically with increasing t. This is different from the value at a point $x \neq 0$, where it first increases to a maximum and then increases. Determining the maximum value at x thus amounts to computing when the derivative with respect to t vanishes. We get

$$\frac{\partial g_t(x)}{\partial t} \quad = \quad \left[\frac{|x|^2}{2t^2} - \frac{n}{2t} \right] g_t(x), \tag{24}$$

which vanishes at $t_0 = |x|^2/n$. To see that this indeed corresponds to a maximum, we may compute the second derivative and verify that it is negative at $t = t_0$. We get the upper bound by plugging t_0 into the formula for $g_t(x)$:

First Kernel Lemma. *Let g_t be the Gaussian kernel with $t > 0$ in $\mathbb{R}^n$. Then $g_t(x) \leq \left(\frac{n}{2e\pi|x|^2} \right)^{\frac{n}{2}}$ for any $|x| > 0$.*

Second, we are interested in the p-norm of the gradient. It is intuitively clear that this norm will go to zero when t goes to infinity, but it will be important to determine how fast it vanishes. Computing the p-norm of $|\nabla g_t(x)|$ reduces to a standard exercise in integration, and results can be found in standard mathematical handbooks, including [1].

Second Kernel Lemma. *Let g_t be the Gaussian kernel with $t > 0$ in $\mathbb{R}^n$, and $p \geq 1$. Then $\|\nabla g_t\|_p \leq \mathrm{const}/t^{\frac{n+1}{2} - \frac{n}{2p}}$.*

For example, the 1-norm goes to zero like $t^{\frac{1}{2}}$, which does not depend on n. For $p > 1$, the convergence is faster in higher than in lower dimensions.

Quantifying flatness. Let Ω be a compact subset of $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}$ a function, and $f_0 : \mathbb{R}^n \to \mathbb{R}$ defined by $f_0(x) = f(x)$ if $x \in \Omega$ and $f_0(x) = 0$ otherwise. We are interested in the behavior of the convolution, $f_t = f_0 * g_t$, which is defined by

$$f_t(x) \quad = \quad \int_{y \in \mathbb{R}^n} f_0(y) g_t(x - y)\,\mathrm{d}y. \tag{25}$$

It will be convenient to side-step the definition of f_0 and write $f_t = f * g_t$ for $t \geq 0$. When f is the Dirac delta function, with unit mass concentrated at the origin, then $f_t = g_t$ for all $t > 0$. Hence, its amplitude is equal to the maximum value: $\mathrm{amp}(f_t) = 1/(2\pi t)^{\frac{n}{2}}$. Even for general f, the amplitude cannot be much larger than that, which we prove using $\mathrm{amp}(f_t) \leq 2\|f_t\|_\infty$. The maximum absolute value is

$$|f_t(x)| \quad \leq \quad \int_{y \in \Omega} |f(y)| \cdot \frac{1}{(2\pi t)^{\frac{n}{2}}} \cdot e^{-\frac{|x-y|^2}{2t}}\,\mathrm{d}y, \tag{26}$$

for all $t > 0$. Noting that the exponential term is at most 1, we get the desired upper bound.

Amplitude Lemma. *Let Ω be a compact subset of $\mathbb{R}^n$ and $f : \Omega \to \mathbb{R}$. Then the amplitude of $f_t : \mathbb{R}^n \to \mathbb{R}$ satisfies $\mathrm{amp}(f_t) \leq \mathrm{const}/t^{\frac{n}{2}}$, for all $t > 0$.*

The constant in this bound is $2\|f\|_1/(2\pi)^{\frac{n}{2}}$. We are also interested in the amplitude of f_t outside a sufficiently large cube. Specifically, we use the First Kernel Lemma to bound how fast the radius $r(f_t, \delta)$ grows when δ goes to zero. We content ourselves with a bound that applies uniformly, for all t. To state the result, we assume a radius r_0 large enough so that the cube $[-r_0, r_0]^n$ contains $\Omega \subseteq \mathbb{R}^n$, and we define

$$R(\delta) \quad = \quad r_0 + \sqrt{\frac{n}{2e\pi}} \cdot \left(\frac{2\|f\|_1}{\delta} \right)^{\frac{1}{n}}. \tag{27}$$

Furthermore, we write $B_{R(\delta)} = [-R(\delta), R(\delta)]^n$ for the cube this radius defines, as before.

Flat Tail Bound. *Let $f : \Omega \to \mathbb{R}$ be with compact support $\Omega \subseteq [-r_0, r_0]^n$. Then the amplitude of $f_t : \mathbb{R}^n \to \mathbb{R}$ outside $B_{R(\delta)}$ is at most δ uniformly for all $t \geq 0$.*

PROOF. Let x be a point outside $B_{R(\delta)}$ and note that it is further than $R(\delta) - r_0$ from every point in Ω. By the First Kernel Lemma, the value of g_t at a point y with distance $R(\delta) - r_0$ from the origin is at most

$$X_\delta \quad = \quad \left(\frac{n}{2e\pi(R(\delta) - r_0)^2} \right)^{\frac{n}{2}}. \tag{28}$$

We can therefore bound $f_t(x)$ by concentrating all the mass at a point at distance $R(\delta) - r_0$ from x. More formally, we have

$$
\begin{aligned}
|f_t(x)| &\leq \int_{y \in \Omega} |f(y)| g_t(x-y) \, \mathrm{d}y & (29) \\
&\leq X_\delta \cdot \|f\|_1, & (30)
\end{aligned}
$$

where the 1-norm is finite because Ω is compact. Solving for $X_\delta \|f\|_1 = \delta/2$, we get $R(\delta)$ as defined in (27). The claim follows because $|f_t(x)| \leq \delta/2$ for all points outside $B_{R(\delta)}$ implies that the amplitude of f_t restricted to $\bar{B}_{R(\delta)}$ is at most δ. $\square$

We may read (27) as saying that there is a constant that depends on f and n such that $R(\delta) = \text{const} \cdot [1 + 1/\delta^{\frac{1}{n}}]$. Hence, the infimum radius $r(f_t, \delta)$, which is bounded from above by $R(\delta)$, grows at most like the n-th root of $1/\delta$.

Extending to compact support. Using the Second Kernel Lemma, it is not too difficult to bound the norm of the gradient of a more general diffusing function. A crucial tool in this analysis is the Hölder Inequality, which we now recall. Given two functions $\Phi, \Psi : \mathbb{R}^n \to \mathbb{R}$, it says that

$$
\int \Phi(x) \Psi(x) \, \mathrm{d}x \leq \left(\int \Phi(x)^p \, \mathrm{d}x \right)^{\frac{1}{p}} \left(\int \Psi(x)^q \, \mathrm{d}x \right)^{\frac{1}{q}},
$$

whenever $1 \leq p, q \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$. To illustrate its use, suppose that Ψ has compact support $\Omega \subseteq \mathbb{R}^n$ and set $\Phi(x) = 1$ for all $x \in \Omega$ to get

$$
\int_{x \in \Omega} \Psi(x)^a \, \mathrm{d}x \leq \mathrm{vol}(\Omega)^{\frac{1}{p}} \left(\int_{x \in \Omega} \Psi(x)^{aq} \, \mathrm{d}x \right)^{\frac{1}{q}}.
$$

Assuming $\mathrm{vol}(\Omega) = 1$, we get $\|\Psi\|_a \leq \|\Psi\|_{aq}$ for all $q \geq 1$. In words, the norms are non-decreasing for non-decreasing index. We are now ready to relate the norm of a general diffusing function with the norm of the gradient of the diffusing Gaussian kernel.

Compact Gradient Lemma. *Let $\Omega \subseteq \mathbb{R}^n$ be compact, $f : \Omega \to \mathbb{R}$ a function, $f_t = f * g_t$ for $t > 0$, and $p \geq 1$. Then $\|\nabla f_t\|_p \leq \text{const} \cdot \|\nabla g_t\|_p$.*

PROOF. In a first step, we write the gradient of f_t in terms of the gradient of the Gaussian kernel and get an upper bound by integrating magnitudes:

$$
\begin{aligned}
|\nabla f_t(x)| &\leq \int_{y \in \Omega} |f(y)| \cdot |\nabla g_t(x-y)| \, \mathrm{d}y & (31) \\
&\leq \|f|_\Omega\|_q \left(\int_{y \in \Omega} |\nabla g_t(x-y)|^p \, \mathrm{d}y \right)^{\frac{1}{p}} & (32)
\end{aligned}
$$

where we apply the Hölder Inequality with $\frac{1}{q} + \frac{1}{p} = 1$. Note that $\|f|_\Omega\|_q$ is no more than the q-norm of f, while the second integral is *not* the p-norm of $|\nabla g_t|$. Indeed, the size of the second integral depends on the relative position of x and Ω. In the second step, we bound the p-norm of $|\nabla f_t|$ by integrating the above bound on the magnitude:

$$
\begin{aligned}
\|\nabla f_t\|_p &= \left(\int_{x \in \mathbb{R}^n} |\nabla f_t(x)|^p \, \mathrm{d}x \right)^{\frac{1}{p}} & (33) \\
&= \|f\|_q \left(\int_{\mathbb{R}^n} \int_\Omega |\nabla g_t(x-y)|^p \, \mathrm{d}y \, \mathrm{d}x \right)^{\frac{1}{p}} & (34) \\
&= \|f\|_q \left(\int_{y \in \Omega} \|\nabla g_t\|_p{}^p \, \mathrm{d}x \right)^{\frac{1}{p}}, & (35)
\end{aligned}
$$

where we get the last line by exchanging the integrals. The claimed inequality follows by noticing that the integral is equal to $\mathrm{vol}(\Omega) \|\nabla g_t\|_p{}^p$. Simplifying the resulting inequality by absorbing $\|f\|_q$ and the p-th root of $\mathrm{vol}(\Omega)$ into the constant gives the claimed inequality. $\square$

Bounding the persistence. We are now ready to prove the main result of this paper, which states that the p-norm of the persistence diagram goes to zero like $1/t^{\frac{n}{2}}$.

Main Theorem. *Let $\Omega \subseteq \mathbb{R}^n$ be compact and $f : \Omega \to \mathbb{R}$ a function such that $f_t = f * g_t$ is tame for all $t \geq 0$. Then*

$$
\|\mathrm{Dgm}(f_t)\|_p \leq \text{const}/t^{\frac{n}{2}}, \tag{36}
$$

for all $p > \frac{1}{2}(2n + 1 + \sqrt{4n^2 + 1})$, and the exponent on the right hand side of the inequality is best possible.

PROOF. The tightness of the bound follows from the existence of f with $\mathrm{amp}(f_t) = \text{const}/t^{\frac{n}{2}}$. We thus get a dot whose persistence is this amplitude. Taking the p-th root of the p-th power implies the claim.

To prove the upper bound, we simplify the relevant inequalities by focusing on the terms that depend on δ or on t. For example, for (18) and (27), we get

$$
\begin{aligned}
M_{f_t}(\delta) &\leq \text{const} \cdot \left[1 + \frac{r(f_t, \delta)^n \cdot X_p}{\delta^m} \right], & (37) \\
R(\delta) &\leq \text{const} \cdot \left[1 + \frac{1}{\delta^{\frac{1}{n}}} \right]. & (38)
\end{aligned}
$$

where $X_p = \|\nabla f_t\|_p{}^m$. By the Flat Tail Bound, we have $r(f_t, \delta) \leq R(\delta)$, so we can plug (38) into (37) to get

$$
M_{f_t}(\delta) \leq \text{const} \cdot \left[1 + \frac{X_p}{\delta^{1+m}} \right]. \tag{39}
$$

Plugging (39) into the Diagram Norm Lemma gives a first term $\delta^p + \delta^\ell X_p$ within the outer pair of brackets, where $\ell = p - 1 - m = p - 1 - \frac{np}{p-n}$. The lower bound on p given in the statement implies $\ell > 0$, so the first term can be neglected as δ goes to zero. The second term within the brackets is

$$
X \leq \text{const} \cdot \int_{\varepsilon = \delta}^{\mathrm{amp}(f_t)} \left(\varepsilon^{p-1} + X_p \varepsilon^{\ell - 1} \right) \mathrm{d}\varepsilon. \tag{40}
$$

Substituting $\|\nabla g_t\|_p$ for $\|\nabla f_t\|_p$ using the Compact Gradient Lemma and applying the Second Kernel Lemma, we get $X_p \leq \mathrm{const}/t^{m(\frac{n+1}{2}-\frac{n}{2p})}$. Plugging this bound into (40) and integrating gives

$$X \quad \leq \quad \mathrm{const} \cdot \left[\mathrm{amp}(f_t)^p + \frac{\mathrm{amp}(f_t)^\ell}{t^{m(\frac{n+1}{2}-\frac{n}{2p})}} \right]. \qquad (41)$$

Using $\mathrm{amp}(f_t) \leq \mathrm{const}/t^{\frac{n}{2}}$ from the Amplitude Lemma, we note that the first term in (41) is dominated by the second term. We therefore get $X \leq \mathrm{const}/t^{\frac{\ell n}{2}+m(\frac{n+1}{2}-\frac{n}{2p})}$. Substituting $p-1-m$ for ℓ and $\frac{pn}{p-n}$ for m finally gives $X \leq \mathrm{const}/t^{\frac{np}{2}}$, and taking the p-th root gives the claimed inequality in (36). □

5. Discussion

The main contribution of this paper is a bound on the norm of the persistence diagram of a diffusing function on $\mathbb{R}^n$. Its proof uses a Sobolev inequality to establish a connection between the gradient and the amplitude of the function. Indeed, we may think of the Sobolev inequality as the technical means needed to finesse the difficulties caused by the non-compactness of Euclidean spaces.

We close by formulating questions related to the work presented in this paper. Can the assumptions under which our bounds hold be weakened? In particular, does the upper bound hold for value of p that are smaller than allowed in the Main Theorem? How do the upper bounds extend to non-Euclidean spaces? What are the characteristics of spaces with particularly fast or particularly slowly vanishing norm of the persistence? Can the analysis of the continuous case presented in this paper be extended to discretized versions of convolution, as considered in [15]?

Acknowledgment

The first author thanks Christoph Lampert, Gehua Yang and Ganesh Sundaramoorthi for helpful discussions.

References

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, NY, 1964. 6

[2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *ECCV*, pages 404–417, 2006. 1

[3] A. Buades, B. Coll, and J. Morel. A non-local algorithm for image denoising. In *CVPR*, pages 60–65, 2005. 2

[4] 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. 1

[5] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, and Y. Mileyko. Lipschitz functions have L_p-stable persistence. *Found. Comput. Math.*, 10(2):127–139, 2010. 4, 5

[6] J. Damon. Local Morse theory for solutions to the heat equation and Gaussian blurring. *J. Diff. Eq.*, 115(2):368–401, 1995. 1

[7] H. Edelsbrunner and J. Harer. *Computational Topology. An Introduction*. Amer. Math. Soc., Providence, RI, 2010. 2, 4

[8] H. Freudenthal. Simplizialzerlegung von beschränkter Flachheit. *Ann. of Math.*, 43:580–582, 1942. 5

[9] D. Gustavsson, K. Pedersen, F. Lauze, and M. Nielsen. On the rate of structural change in scale spaces. In *SSVM*, pages 832–843, 2009. 2

[10] T. Iijima. Basic theory on normalization of a pattern (in case of typical one-dimensional pattern), (in Japanese). *Bull. Electrotechn. Lab.*, 26:368–388, 1962. 1

[11] J. Koenderink. The structure of images. *Bio. Cyber.*, 50(5):363–370, 1984. 1

[12] A. Kuijper and L. Florack. The relevance of non-generic events in scale space models. *IJCV*, 57(1):67–84, 2004. 1

[13] I. Laptev. On space-time interest points. *IJCV*, 64(2):107–123, 2005. 2

[14] L. M. Lifshitz and S. M. Pizer. A multiresolution hierarchical approach to image segmentation based on intensity extrema. *PAMI*, 12(6):529–540, 1990. 1

[15] T. Lindeberg. Effective scale: a natural unit for measuring scale-space lifetime. *PAMI*, 15(10):1068–1074, 1993. 2, 8

[16] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer, Dortrecht, the Netherlands, 1993. 1

[17] T. Lindeberg. Feature detection with automatic scale selection. *IJCV*, 30(2):79–116, 1998. 1

[18] D. Lowe. Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157, 1999. 1

[19] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004. 1

[20] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Gool. A comparison of affine region detectors. *IJCV*, 65(1):43–72, 2005. 1, 2

[21] J. Milnor. *Morse Theory*. Princeton Univ. Press, NJ, 1963. 2

[22] J. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, Reading, MA, 1984. 3

[23] J. Rieger. Generic evolutions of edges on families of diffused greyvalue surfaces. *J. Math. Imaging Vis.*, 5(3):207–217, 1995. 1

[24] L. Saloff-Coste. *Aspects of Sobolev-Type Inequalities*. Cambridge Univ. Press, England, 2002. 3

[25] J. Sporring and J. Weickert. Information measures in scale-spaces. *IEEE Trans. Inform. Theory*, 45(3):1051–1058, 1999. 2

[26] G. Sundaramoorthi, P. Petersen, V. S. Varadarajan, and S. Soatto. On the set of images modulo viewpoint and contrast changes. In *CVPR*, pages 832–839, 2009. 2

[27] A. Witkin. Scale space filtering. In *Proc. 8th Int. Joint Conf. Art. Intell.*, pages 1019–1022, 1983. 1