Adaptive Multiscale Detection of Filamentary Structures Embedded in a Background of Uniform Random Points

Ery Arias-Castro [a], David Donoho [a], Xiaoming Huo [b]

June 2003

Abstract

We are given a set of n points that appears uniformly distributed in the unit square $[0,1]^2$. We wish to test whether the set actually is generated from a non-uniform distribution having a small fraction of points concentrated on some (*a priori* unknown) curve with C^α-norm bounded by β.

An asymptotic detection threshold exists in this problem; for a constant $T_-(\alpha,\beta) > 0$, if the number of points on the curve is smaller than $T_-(\alpha,\beta)n^{1/(1+\alpha)}$, reliable detection is not possible for large n.

We describe a Multiscale Significant-Runs Algorithm; it can reliably detect concentration of data near a smooth curve, without knowing the smoothness information α or β in advance, provided that the number of points on the curve exceeds $T_*(\alpha,\beta)n^{1/(1+\alpha)}$. This algorithm therefore has an optimal detection threshold, up to a factor T_*/T_-.

At the heart of our approach is an analysis of the data by counting membership in multiscale multianisotropic strips. The strips have an area of C/n and exhibit a variety of lengths, orientations and anisotropies. The strips are partitioned into anisotropy classes; each class is organized as a directed graph whose vertices are strips all of the same anisotropy and whose edges link such strips to their 'good continuations'. The point cloud data are reduced to counts measuring membership in strips. Each anisotropy graph is reduced to a subgraph consisting of strips with 'significant' counts. The algorithm rejects $\mathbf{H}_0$ whenever some such subgraph contains a path connecting many consecutive 'significant' counts.

Key Words and Phrases. Multiscale Geometric Analysis. Pattern Recognition. Good Continuation. Erdös-Renyi Laws. Runs Test. Beamlets.

Acknowledgements. This work was partially supported by NSF grants DMS-0077261, DMS 01-40587 (FRG), and a contract from DARPA ACMP. We would like to thank Emmanuel Candès, Hagit Hel-Or, Aapo Hyvärinen, Jean-Luc Starck and Brian Wandell for helpful discussions and references.

[a]: Department of Statistics, Stanford University. {acery,donoho}@stat.stanford.edu

[b]: School of Industrial and Systems Engineering, Georgia Institute of Technology. xiaoming@isye.gatech.edu

1 Introduction

> *We cannot help but see faces and castles in clouds, monsters in ink-blots and exotic forms in random dots. Form is so central to human perception that, I am told, it is extremely difficult to prove something random or formless.* – Mae-Wan Ho [16]

Suppose we have n data points $X_i \in [0,1]^2$ which at first glance seem uniformly distributed in the unit square. On cursory visual inspection, it seems that a suspiciously large number of the data points fall along a smooth curve. However, the curve on which these points lie has only been identified *after* inspection of the data. We know that the human visual system has the ability to 'hallucinate' curvilinear structure in truly random point clouds. We are therefore concerned about the reliability of the perceived 'pattern', and wish to follow an objective procedure for testing the existence of filamentary structure; this procedure should reliably separate 'filamentary structure' from 'random scatter' and be computationally tractable.

This is a prototype for various practical imaging problems ranging from surveillance to road and streambed tracking to particle physics [24, 27, 1]. In all cases the observer is looking for evidence of a filamentary structure in a background of heavy 'clutter'.

As a first attempt at formalizing matters, consider the problem of testing

$$\mathbf{H}_0 : X_i \overset{\text{i.i.d.}}{\sim} \text{Uniform}(0,1)^2$$

versus

$$\mathbf{H}_1(\alpha,\beta) : X_i \overset{\text{i.i.d.}}{\sim} (1-\varepsilon_n)\text{Uniform}(0,1)^2 + \varepsilon_n \text{Uniform}(\text{graph}(f)), \quad f \in \text{Hölder}(\alpha,\beta),$$

where, for $1 < \alpha \leq 2$, Hölder(α,β) is the class of functions $f : [0,1] \to [0,1]$, with continuous derivative f' obeying $|f'(x) - f'(y)| \leq \alpha\beta|x-y|^{\alpha-1}$. In words, we believe that a relatively small fraction ε_n of points lie on a smooth curve in the plane.

1.1 "Connect the Dots"

In previous work by the authors [3], it was shown that, when α and β are fixed and known, there is a detector based on the principle that, under $\mathbf{H}_0$, no Hölder(α,β) curve can pass through a very large number of points in a random point cloud. More particularly, we know that there is a threshold $T_+ = T_+(\alpha,\beta)$ so that

- If $T < T_+$, we have that, with probability tending to 1, **there exists** a Hölder(α,β) curve containing at least $T \cdot n^{1/(1+\alpha)}$ points $(X_i)_{i=1}^n$.
- If $T > T_+$, we have that with probability tending to 1, there **does not exist** a Hölder(α,β) curve, containing more than $T \cdot n^{1/(1+\alpha)}$ points $(X_i)_{i=1}^n$.

More concretely, if we deal with Lipschitz curves with $|slope| \leq 1$, we have found empirically that for moderate $n \approx 1000$ there will frequently be some Lipschitz curve containing $\sqrt{n}$ data points but rarely will there be one containing more than $3\sqrt{n}$ points.

So, if we happen to notice a curve passing through substantially more than $T_+ \cdot n^{1/(1+\alpha)}$ points, we have a strong basis for rejecting the null hypothesis of pure randomness. Moreover,

to within a constant factor, this threshold is optimal; no sequence of tests can be reliable for detecting substantially fewer than $T_- \cdot n^{1/(1+\alpha)}$ points, for a certain $T_- > 0$.

Elaborating this principle leads to a formal hypothesis test based on searching for curves containing large numbers of points. Let $N_n(A) = \#\{\{X_i\} \cap A\}$ denote the measure counting how many points lie in the set A; searching for a curve containing the maximal number of points leads to the optimization problem

$$N_n^*(\alpha, \beta) = \max\{N_n(graph(f)) : f \in \text{Hölder}(\alpha, \beta)\}$$

which searches over all Holder(α, β) graphs and rejects $\mathbf{H}_0$ for values of N^* substantially exceeding $T_+ \cdot n^{1/(1+\alpha)}$.

This approach, while very instructive, does not address the concerns of someone who might actually be interested in performing a test on real data.

- *Computational Burden.* The task of finding the largest number of points on a Hölder(α, β) curve seems to the authors computationally impractical unless $\alpha \in \{1, 2\}$.
- *Unknown α,β.* The proposed approach assumes a specific α-β combination. Instead, we desire an algorithm which works automatically regardless of the specific values of $\alpha \in (1, 2]$ and $\beta > 0$.
- *Fragments.* This approach searches only for filamentary structures which extend all the way across the square from $x = 0$ to $x = 1$. Instead, one wants an algorithm which works even for short fragments.
- *General Planar Curve .* The proposal assumes the underlying filament can be parametrized as a graph. It seems important to search for filamentary structures which are general curves rather than just graphs – for example, curves which loop around in the plane.

1.2 An Adaptive Multiscale Approach

In this paper we describe an approach which addresses the concerns just listed. Our proposal:

1. Works across for a wide range of (α, β), and only requires knowing a bound on the maximum slope of the curve;
2. Detects presence of $\mathbf{H}_1$ provided
$$\varepsilon_n > T_* \cdot n^{-\alpha/(\alpha+1)},$$
for a constant T_* which depends on α, β and other factors. In view of earlier results, this is optimally sensitive to within a factor T_*/T_-.
3. Runs in $O(\log(n) \cdot n^2)$ flops.
4. Extends naturally to detecting general planar curves which are not graphs; and, finally,
5. Extends naturally to detecting target filaments of unknown extent which, in large samples, can be very short compared to the image extent.

The detector is based on multiscale geometric analysis of the dataset, using a multiscale dictionary of parallelogram strips exhibiting a variety of lengths, locations, orientations and aspect ratios. The idea is to count membership of data points in various strips, to identify strips with significantly large counts, and search for long 'runs' of significantly large counts in collections of strips that are 'good continuations' of each other.

An interesting aspect of our approach is how simply and naturally the principle of 'good continuation' appears and leads to a solution.

1.3 Contents

The paper is organized as follows. In Section 2 we describe the underlying multiscale data structures. Section 3 describes the adaptive algorithm in general terms, and gives a statement of our main result. Section 4 describes the threshold settings underlying our algorithm, while Sections 5 and 6 analyze its behavior under $\mathbf{H}_0$ and $\mathbf{H}_1$, respectively. Section 7 finishes the paper with discussion of related work.

2 Multiscale Anisotropic Strips, and Good Continuation

Our data structures comprise a multiscale collection of anisotropic, tilted planar regions, and a sequence of directed graphs organizing them.

To define the regions, fix $\rho = 2^r$ (e.g. 2 or 4); this parameter controls overlap between the regions we will be defining. Also, fix $S > 0$ (e.g. 2 or 4) this controls the maximum $|slope|$ we shall be able to detect. We assume that n is large and let $J = \lceil \log_2(n) \rceil$.

Let $R(j, k, \ell_1, \ell_2)$ be a paralleliped with vertical sides which is $w = 2^{-j}$ 'wide' by $t = 2^{-(J-j)+r}$ 'thick'. Here j runs through the set $0, \ldots, J$. For examples, see Figure 1. The regions in question have a midline bisecting them vertically and will be 'tilted' (sheared) at a variety of angles. Notice that these regions are highly anisotropic. While implicitly the whole collection depends on n, we suppress this in our notation. Moreover, the width w and thickness t depend on j and n; but we also suppress this in our notation. Note that the degree of anisotropy is the same for all regions sharing a common value of j; we generally focus only on one anisotropy class j at a time.

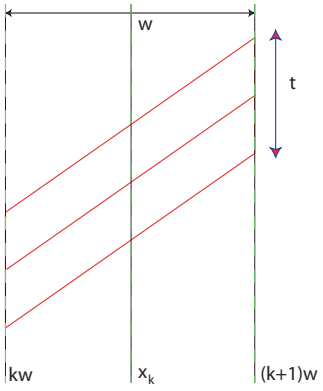


Figure 1: An Anisotropic 'Strip' R

The parameters k and ℓ_i control the horizontal location of the regions and vertical location and slope of the midline. There is an underlying assumption that we are interested only in regions whose major axis has a slope bounded in absolute value by S. The mapping between these discrete parameters is intended to insure that the regions pack together horizontally, and that they are fairly closely spaced in both vertical position and slope. We fix two scale factors, $U = 16$ and $V = 16$, which are arbitrary but should exceed 1. Let $\delta_1 = t/U$ and $\delta_2 = t/(Vw)$ (these again depend implicitly on j and n). The region $R(j, k, \ell_1, \ell_2)$ will be located in $[0, 1]^2$ with left-hand vertical side at $x = kw$, left endpoint of midline at $y = \ell_1\delta_1$, and with slope of midline $s = \ell_2\delta_2$. Here $0 \le k < w^{-1}$; ℓ_1 runs through the set $0, \ldots, \delta_1^{-1} - 1$; ℓ_2 runs through the set $-S\delta_2^{-1}, \ldots, S\delta_2^{-1}$.

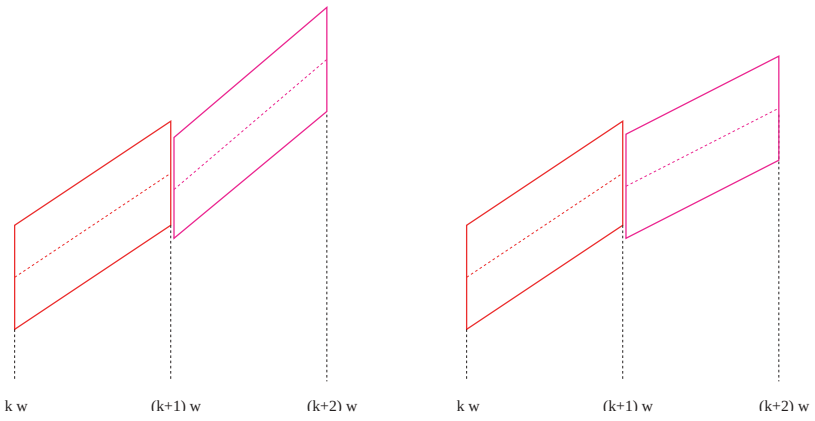
We gather all such regions at level j in $\mathcal{R}(j) = \{R(j, k, \ell_1, \ell_2) : k, \ell_1, \ell_2\}$.

To organize the regions, we define a directed graph $\mathcal{G}(j) = (\mathcal{V}(j), \mathcal{E}(j))$, with vertices $\mathcal{V}(j)$ and edges $\mathcal{E}(j)$. The vertices are simply the regions in $\mathcal{R}(j)$: $\mathcal{V}(j) \equiv \mathcal{R}(j)$. The edges connect regions to their *good continuations*, namely, regions which are horizontally adjacent, and which have altitudes and slopes which are nearly the same – less than δ_1 and δ_2 apart, respectively. Formally, we have the following directed edges in $\mathcal{E}(j)$:

$$(k, \ell_1, \ell_2) \mapsto (k + 1, \ell_1 + \ell_2 + u, \ell_2 + v)$$

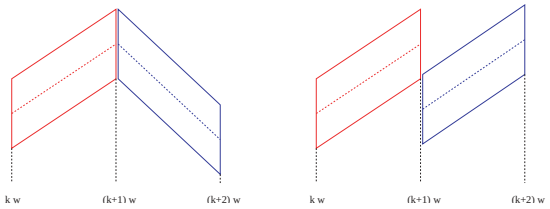
where $|u| \le U$, $|v| \le V$.

Figure 2 (resp. 3) illustrates "good continuation" (resp. "bad continuation").



(a) (b)

Figure 2: Examples of good continuations; the midlines have about the same slope and there is a substantial part of a side in common.



(a) (b)

Figure 3: Examples of bad continuations; either the midlines have very different slopes or the sides are effectively disjoint.

This graphical structure, while very simple, has a perhaps surprising property: it allows us to efficiently cover the graphs of general smooth functions exhibiting any of a range of smoothnesses. This claim is summarized in three lemmas. The first lemma associates to each Hölder class a specific anisotropy graph.

Lemma 2.1 *For each fixed (α, β) combination with $1 \le \alpha \le 2$, we have that for all sufficiently large n there is $j^* = j^*(\alpha, \beta; n)$ so that $w = w(j^*)$ and $t = t(j^*)$ obey*

$$2\beta w^\alpha \le t < 16\beta w^\alpha. \tag{2.1}$$

The next result shows that regions in the anisotropy class $\mathcal{R}(j^*(\alpha, \beta))$ are well-adapted to cover fragments of the graphs of the associated Hölder(α, β) class.

Lemma 2.2 *Let $j = j^*(\alpha, \beta)$ and suppose f is a Hölder(α, β) function with domain containing $I_k = [kw, (k+1)w)$. Set $x_k = (k + 1/2)w$, let $\ell_1\delta_1$ be the closest multiple of δ_1 to $f(x_k)$, let $\ell_2\delta_2$ be the closest multiple of δ_2 to $f'(x_k)$. We say that the region $R(j, k, \ell_1, \ell_2)$ is* **associated** *to f on I_k. This strip covers the graph of f over their common domain:*

$$graph(f, I_k) \subset R(j, k, \ell_1, \ell_2). \tag{2.2}$$

See Figure 4.

The final lemma in the sequence shows every function in the Hölder(α, β) class corresponds to a covering sequence of regions which makes a connected path in $\mathcal{G}(j)$.

Lemma 2.3 *Let $j = j^*(\alpha, \beta)$ and suppose f is a Hölder(α, β) function on $[0, 1]$. Let $\mathcal{T}_j(f)$ be the collection of all strips $R(j, k, \ell_1, \ell_2)$ associated to f by the procedure mentioned in Lemma 2.2. The sequence of strips $R_i = R(j, i, \ell_{1,i}, \ell_{2,i})$ consists of spatially adjacent regions, making a kind of 'tube'. When viewed as vertices of $\mathcal{G}(j)$, the (R_i) are neighbors in $\mathcal{G}(j)$, i.e. R_i and R_{i+1} can be connected using edges in $\mathcal{E}(j)$.*

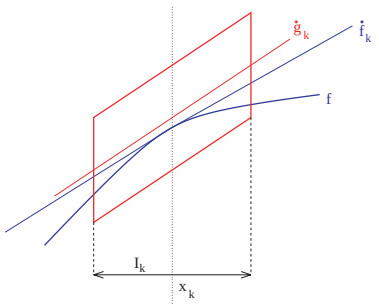


Figure 4: *graph*(f) (in blue) covered by its kth associated region at level j (in red); $\dot{f}_k$ is tangent to *graph*(f), $\dot{g}_k$ is the midline of $R(j,k,\ell_1,\ell_2)$. See the proofs in the Appendix for more discussion.

These lemmas together show that, while the graphical structure itself is based on very simple rules, it is able to associate paths in the graph with custom-fitting tubes covering the graphs of very different kinds of smooth functions. The proofs of these lemmas are given in the appendix. Figure 5 illustrates the idea.

3 The Multiscale Significant-Runs Algorithm

We now describe the complete algorithm for analysis of point cloud data (X_i) looking for suspected filaments. It depends on a counting threshold N^* and a length threshold L_n^*, both to be defined later. The algorithm has several steps:

1. *Counting Membership in Anisotropic Strips.* For every region R, in every anisotropy class, we count the number of data falling in that region:

$$N(R) = \#\{i : X_i \in R\}.$$

2. *Identifying Significant Counts.* We define a *significance indicator* which is nonzero when the counts exceed a threshold:

$$s(R) = 1_{\{N(R)>N^*\}}.$$

The significance indicator may be viewed as labels on the regions R, producing a sequence of a labelled graphs

$$\Sigma(j) = (\mathcal{V}(j), \mathcal{E}(j), \sigma(j)),$$

where $\sigma(j) = (s(R))$ gives the labels on $R \in \mathcal{R}(j)$. We call this the j-th *significance graph.*

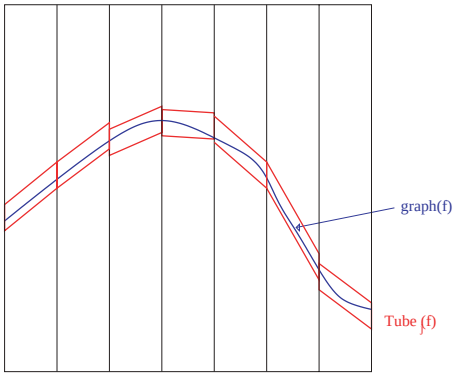


Figure 5: *graph*(f) (in blue) covered by $\mathcal{T}_j(f)$ (in red).

3. *Computing Longest Paths.* In each significance graph, we employ a depth-first search algorithm to explore all significant paths

$$\pi = (R_1, R_2, \ldots, R_m),$$

i.e. sequences of vertices which are

(a) All significant: $s(R_i) = 1$.

(b) All connected: $(R_i, R_{i+1}) \in \mathcal{E}(j)$.

We record the maximum path length in each significance graph:

$$\begin{aligned} L_{n,j}^{max} &= \max\{\text{length}(\pi) : \pi \text{ is a significant path in } \Sigma(j)\}, \\ L_n^{max} &= \max_j L_{n,j}^{max}. \end{aligned}$$

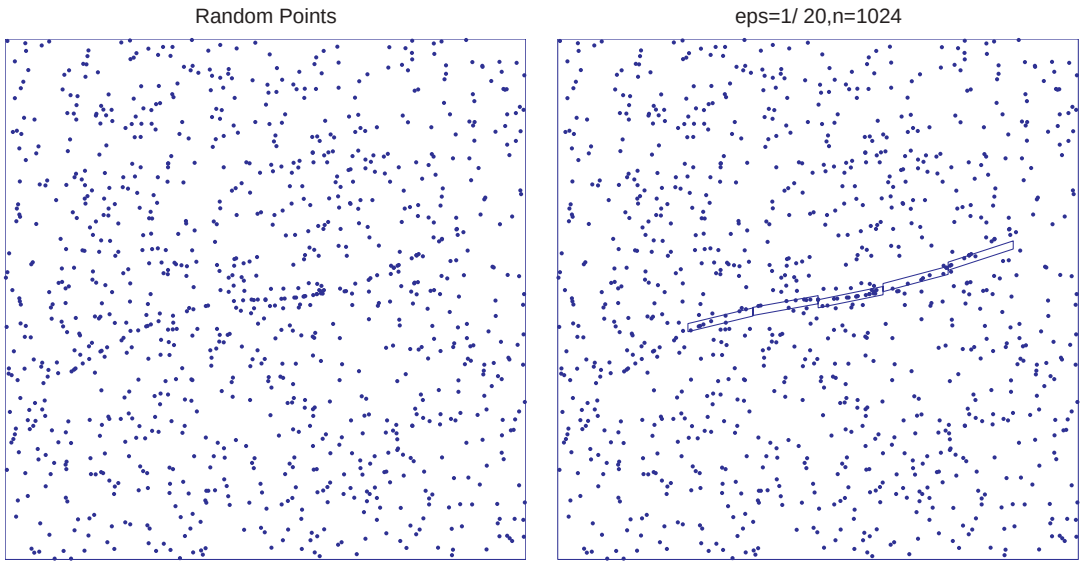
4. *Decision.* We compare L_n^{max} with a length threshold:

$$\text{If } L_n^{max} \le L_n^*, \text{ Accept } \mathbf{H}_0; \text{If } L_n^{max} > L_n^*, \text{ Reject } \mathbf{H}_0.$$

This defines the test, except for the specification of the thresholds L_n^* and N^*. Asymptotic formulas for these thresholds will be given in Section 4.

A worked example of the Multiscale Significant Runs Algorithm is illustrated in Figure 6. A pointset is mostly uniformly distributed, with about $\varepsilon = 1/20$ fraction of points on a smooth curve. The test is run with $\rho = 2$ and $S = 2$. The significance threshold was $N^* = 8$.

The longest run in this case has length 5, which in this case exceeds the run-length threshold L_n^* and leads to rejection of the null hypothesis. For this small-sample setting, the threshold



(a) (b)

Figure 6: A random pointset contaminated by $\varepsilon = 1/20$ points on a curve, together with the identified significant run (consisting of five strips).

$L_n^* = 4$ was obtained by simulation rather than asymptotic theory. For comparison, Figure 7 gives an example of a simulation in the null case $\varepsilon = 0$; the longest run has length 2 in this case. This simulation was typical; in the null case, rarely does the longest run exceed 2.

Several properties of the algorithm are immediate:

- *Complexity of Strip Counts.* The algorithm calculates all the $N(R)$ for all the anistropic strips. This takes $O(n^2 \log(n))$ flops, where n is the number of points sampled. Indeed, since each data point can belong to order $O(n \log(n))$ strips, by simply calculating which $R \ni X_i$ and incrementing a counter for those R we get all $N(R)$.

- *Complexity of Longest Path.* The algorithm calculates the longest path in each significance graph. This takes work comparable to $O(n^2 \log(n))$, based on depth-first search [2].

- *Storage Requirement.* The algorithm stores all of the $N(R)$ counts and all the significance coefficients. This requires $O(n^2 \log(n))$ storage.

To state our main result, we amend our notion of alternative hypothesis. For $\alpha > 1$, let Hölder(α, β, S) denote the collection of function $f \in$ Hölder(α, β) with $|f'|_\infty \le S$. Define

$$\mathbf{H}_1(\alpha, \beta, S, \tau): \quad X_i \overset{\text{i.i.d.}}{\sim} (1-\varepsilon_n)\text{Uniform}(0,1)^2 + \varepsilon_n \text{Uniform}(\text{graph}(f)),$$
$$f \text{ in Hölder}(\alpha, \beta; S), \quad \varepsilon_n > \tau \cdot n^{-\alpha/(1+\alpha)}.$$

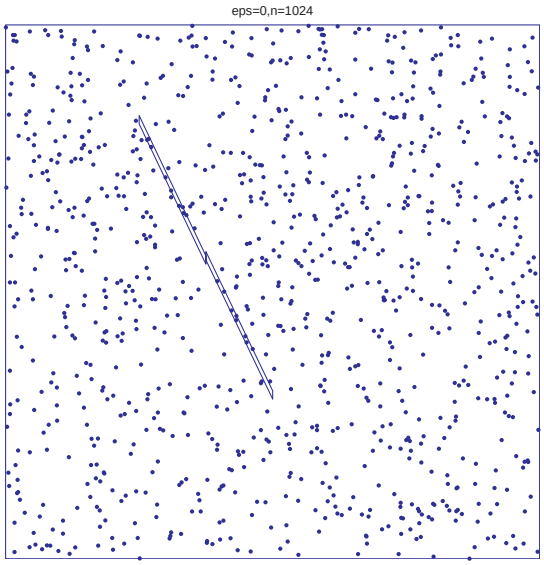


Figure 7: A random pointset, together with the identified longest run of significant strips; at length 2, it is not a significant run.

Theorem 3.1 *There is a single choice of thresholds N^* and $(L_n^*)_n$ so that for every $\alpha \in (1, 2]$ and $\beta > 0$, there is $T_*(\alpha, \beta, \rho, S)$ so that, for each $\tau > T_*$*

$$P\{ \text{ Test rejects } \mathbf{H}_0 \mid \mathbf{H}_1(\alpha, \beta, S, \tau) \} \to 1 \text{ as } n \to \infty;$$

at the same time

$$P\{ \text{ Test rejects } \mathbf{H}_0 \mid \mathbf{H}_0 \} \to 0 \text{ as } n \to \infty.$$

In words, under $\mathbf{H}_0$ the longest significant path is overwhelmingly unlikely to be substantially longer than L_n^* for large n, while under each indicated $\mathbf{H}_1$, the longest significant path is overwhelmingly likely to be substantially longer than L_n^*. The threshold T_* is within a constant factor T_*/T_- of the optimal detection threshold; this shows that, up to constants, we can adaptively test for existence of fragments of C^α graphs.

4 Asymptotics of Thresholds

For practical finite-sample application of the test just described, it is of course possible to calibrate thresholds by conducting simulation experiments. For the proof of our main result, we specify thresholds in closed form. These thresholds are very conservative and our closed-form analysis yields vastly overstated estimates of error probabilities, which are nevertheless good enough for our proofs.

4.1 Specification of N^*

Fix a probability $p_0 < 1$. Define a counting threshold $N^+(\varepsilon, \lambda)$ with the property that

$$P\{\text{Poisson}(\lambda) > N^+\} \le \varepsilon,$$

where Poisson(λ) denotes a Poisson random variable with parameter λ. Let Bin(n,p) denote a binomial random variable with parameters n and p. By Poisson approximation to the Binomial, we have

$$P\{\text{Bin}(n,\lambda/n) > N^+(\varepsilon,\lambda)\} \le 2\varepsilon, \qquad n \ge n_0.$$

Set then $N^* = N^+(p_0/(2UV),\rho)$, where U, V – introduced earlier – control the discretization of altitude and slope. Our definition of N^* gives us the key property

$$P\{s(R)=1|\mathbf{H}_0\} = P\{\text{Bin}(n,\rho/n) > N^*\} \le p_0/UV, \qquad n \ge n_0. \tag{4.1}$$

4.2 Specification of L_n^*

We use a convenient, but nonstandard, notation borrowed from Arratia and Waterman [5]. For $0 < p < 1$, $\log_p$ denotes, in this unconventional notation, the *logarithm with base* $1/p$. At the same time, we maintain the original convention that if $b > 1$, $\log_b$ means log to the base b. By our convention, $\log_2$ is the traditional logarithm base two, and $\log_{1/2}$ is actually the same quantity.

Define the length threshold

$$L_n^* = 5\log_{p_0}(n\log_{p_0} n), \tag{4.2}$$

where $p_0 \in (0,1)$ is the same as in the specification of N^*. The underlying rationale for this choice is the Erdös-Renyi law (see Arratia and Waterman [5]), which says

> In a sequence of m i.i.d. Bernoulli random variables with probability p of heads, the length of the longest run of pure heads $\sim \log_p(m)(1+o_P(1))$.

In effect, our definition makes L_n^* very substantially longer than the length of the longest run of pure heads in a linear sequence of $O(n^{1/(1+\alpha)})$ coin tosses of a p_0-coin.

4.3 Specification of T_*

Associated to the parameters N^* and L_n^* will be the threshold T_* at which $\mathbf{H}_1$ becomes detectible. To define that, set p_1 sufficiently close to 1 that, for some $n_1 > 0$,

$$\log_{p_1}(n^{1/(1+\alpha)}) > 2\cdot L_n^*, \quad n > n_1. \tag{4.3}$$

Implicitly, this choice again refers to the Erdös-Renyi law, and in some way guarantees that under $\mathbf{H}_1$ there will be very long runs.

Define an intensity threshold $\Lambda^+(\varepsilon)$ with the property that

$$P\{\text{Poisson}(\Lambda^+(\varepsilon)) < N^*\} \le \varepsilon,$$

Set $\lambda^* = \Lambda^+((1-p_1)/2)$, and set

$$T_*(\alpha,\beta,\rho,S) = 2\lambda^*\cdot(2\beta/\rho)^{1/(1+\alpha)}\cdot\sqrt{1+S^2}.$$

We will show that for

$$\varepsilon_n > T_*\cdot n^{-\alpha/(1+\alpha)}, \tag{4.4}$$

the hypothesis $\mathbf{H}_1(\alpha, \beta)$ becomes detectable by our proposal. The central point will be the property

$$n \cdot \varepsilon_n \cdot w(j^*(\alpha, \beta, n))/\sqrt{1+S^2} \geq \lambda^*. \tag{4.5}$$

To prove this, we inspect the proof of Lemma 2.1 in the appendix, and note that, by definition of $T_*(\alpha, \beta, \rho)$ above, and using notation from the Appendix,

$$n \cdot \varepsilon_n \cdot w(j^*(\alpha, \beta, n))/\sqrt{1+S^2} \geq n^{1/(1+\alpha)} \cdot T_* \cdot w(j^+(\alpha, \beta, n))/(2\sqrt{1+S^2}) = \lambda^*.$$

Incidentally, we don't claim that the algorithm fails for $\tau < T_*$, only that we can prove it succeeds for $\tau > T_*$.

5 Behavior under $\mathbf{H}_1$

Let f be the function in Hölder(α, β), generating the graph carrying the fraction ε_n of data. From Lemmas 2.1-2.3, we let $j = j^*$ and consider the tube $\mathcal{T}_j(f)$. For each region R in this tube,

$$N(R) \sim \text{Bin}(n, (1-\varepsilon_n)\text{area}(R) + \varepsilon_n \gamma(f, R)),$$

where $\text{area}(R) = \rho/n$, and γ denotes the relative arclength in the graph of f, obeying

$$\gamma(f, R) = \frac{\text{length}(graph(f_{|I}))}{\text{length}(graph(f))} \geq \frac{w}{\sqrt{1+S^2}};$$

here I denotes the projection of R on the x-axis.

By Poisson approximation to the Binomial,

$$N(R) \overset{\text{approx.}}{\sim} \text{Poisson}(\mu),$$

where, using (4.5)

$$\begin{aligned} \mu &\geq \rho/2 + n\varepsilon_n w/\sqrt{1+S^2} \\ &\geq \lambda^*. \end{aligned}$$

Hence, for every R in this tube, we have, for all sufficiently large $n > n_3$,

$$P\{N(R) > N^*\} \geq p_1. \tag{5.1}$$

Label the sequence of strips R in this tube $R_0, \ldots, R_{w^{-1}-1}$. We want to know the probable length of the longest run of the form

$$N(R_i) > N^*, \ldots, N(R_{i+L}) > N^*.$$

Then, if L_n is the length of the longest run in this special sequence, it follows that the longest-run statistic we are computing over the entire graph can only be larger:

$$L_{n,j}^{max} \geq L_n.$$

To show that the test rejects $\mathbf{H}_1$, we will show that

$$P\{L_n > L_n^*\} \to 1, \quad n \to \infty. \tag{5.2}$$

If we define

$$X_i = 1_{\{N(R_i) > N^*\}},$$

we note that each X_i is Bernoulli with probability p_i, while

$$p_i \geq p_1.$$

We let $m = w^{-1} \geq Const \cdot n^{1/(1+\alpha)}$, and $p = p_1$ and we get, by the Erdös-Rényi Law

$$L_n > \log_{p_1}(n^{1/(1+\alpha)})(1 + o_P(1)).$$

But, by hypothesis, we have chosen p_1 in such a way that

$$\log_{p_1}(n^{1/(1+\alpha)}) > 2L_n^*.$$

Hence (5.2) follows.

6 Behavior under $\mathbf{H}_0$

We need to show that with overwhelming probability under $\mathbf{H}_0$, there will be no runs in the graph exceeding L_n^*. We start by arguing that

$$P\{ \text{ significant path of length } L \text{ starts at given } R|\mathbf{H}_0\} \leq p_0^L.$$

Indeed, by choice of N^*, for each R,

$$P\{s(R) = 1|\mathbf{H}_0\} = P\{\text{Bin}(n, \rho/n) > N^*\} \leq p_0/UV.$$

Now each region R has UV neighbors in $\mathcal{G}(j)$, and so by Boole's inequality,

$$\begin{aligned} P\left\{s(R') = 1 \text{ for at least one neighbor of } R \,|\mathbf{H}_0\right\} &\leq \sum_{R' \in \text{Neighbors}(R)} P\{s(R') = 1|\mathbf{H}_0\} \\ &\leq \#\text{Neighbors} \cdot (p_0/UV) \\ &= (UV)(p_0/UV) = p_0. \end{aligned}$$

Now if we are looking for a significant path of length $> L$, we need that starting at some vertex, it has $s(R) = 1$ and is connected to a vertex with $s(R') = 1$, etc. For a given starting point R, the probability of this event is bounded by p_0^L - by negative correlation.

We now note there are at most $M_j = w^{-1} \cdot \delta_1^{-1} \cdot \delta_2^{-1} \cdot 2S$ starting points for paths in $\mathcal{G}(j)$. The probability a significant path occurs in $\mathcal{G}(j)$ under $\mathbf{H}_0$ is controllable using Boole by

$$\leq \#\text{ (Starting Points at Level } j) \cdot P\{ \text{ significant path of length L starting at } R|\mathbf{H}_0\} \leq M_j \cdot p_0^L.$$

Take $\log_2$:

$$\begin{aligned}\log_2(M_j) + \log_2(p_0) \cdot L &= \log_2(w^{-1} \cdot \delta_1^{-1} \cdot \delta_2^{-1} \cdot 2S) + \log_2(p_0) \cdot L \\ &= j + j + (J - j) + r + 9 + \log_2(S) + \log_2(p_0) \cdot L \\ &= j + J + C + \log_2(p_0) \cdot L\end{aligned}$$

Now choose L_n so that, for some $\delta > 0$,

$$L_n > 2J/\log_2(p_0) \cdot (1 + \delta),$$

for all sufficiently large n. Hence,

$$P\{\text{there is a run of length } L_n | \mathbf{H}_0\} \to 0, n \to \infty.$$

Finally, recall that $L_n^* > L_n$. □

7 Discussion

We have given only a sampling of results available by multiscale geometric analysis; much more can be done. At the same time, these results are closely related to many ideas in the literature of computer vision. In this discussion, we briefly indicate extensions and sketch a few such connections.

7.1 Extensions

7.1.1 Fragments

In the introduction, we claimed that the same testing architecture allows to detect fragments of general length, not necessarily extending across all the way across the image. The essential requirement for detectability of shorter fragments is that there be a higher normalized density of points on the curve for shorter curve lengths; slightly higher but not dramatically higher. To detect fragments in this setting, we define a sequence of thresholds N_q^* and $L_{n,q}^*$, $q = 1, \ldots$. Using these thresholds, we apply the Multiscale Significant-Runs Algorithm repeatedly, for $q = 1, 2, \ldots$. If we cannot detect a run exceeding $L_{n,1}^*$ at the initial significance threshold $N_{0,1}^*$ we look for a run exceeding $L_{n,1}^*$ with the next larger threshold, $N_{0,2}^*$, and so on.

The thresholds are defined using an auxiliary sequence $\alpha_{n,0}$ for *total error control*; this tends to zero with increasing n; we also have subordinate sequences $\alpha_{n,q} \geq 0$, $q = 1, 2, \ldots$ for *sequential error control* obeying the sum constraint

$$\sum_{q=1}^{\infty} \alpha_{n,q} \equiv \alpha_{n,0} \to 0, \qquad n \to \infty.$$

The sequence of thresholds is defined by setting $p_{0,q} = 2^{-q} p_0$ and $N_q^* = N^+(\frac{p_{0,q}}{UV}, \rho)$. $L_{n,q}^*$ is defined so that

$$n^2 \cdot \log(n) \cdot p_{0,q}^{L_{n,q}^*} \leq \alpha_q.$$

With this sequence of definitions, we have total probability of error under $\mathbf{H_0}$ tending to zero with increasing n.

On the other hand, the multistage algorithm is powerful at detecting fragments. We need a modified alternative hypothesis $\mathbf{H}_1(\alpha, \beta, S, \tau, \gamma, \eta)$, in which the parameters γ and η constrain the length of the fragment from below. This hypothesis supposes that the underlying filamentary structure has the form of a partial graph, i.e. that f has for domain an interval $I \subset [0,1]$ and the filamentary data are uniformly distributed on $graph(f|I)$. It supposes also, that for fixed $\eta, \gamma > 0$, the length of I obeys the constraint $|I| \geq \gamma \cdot n^{-1/(1+\alpha)+\eta}$. Then, by modifying our earlier analysis of behavior under $\mathbf{H}_1$, we see that, with $m = m(\gamma, \eta) \sim \gamma \cdot n^{\eta}$ there is $p_1^* = p_1^*(\gamma, \eta)$, so that for all sufficiently large q,

$$L_{n,q}^* \ll \log_{p_1}(m), \qquad n \to \infty.$$

There is a corresponding threshold T_* so that under $\tau > T_*$ we have significance probability exceeding p_1^* for all regions belonging to $\mathcal{T}_j$. It follows that the algorithm detects $\mathbf{H}_1(\alpha, \beta, S, \tau, \gamma, \eta)$, although again with threshold $T_*(\alpha, \beta, S, \gamma, \eta)$ that is perhaps suboptimal by a constant factor.

The sequential algorithm just described is less computationally demanding than it may at first seem. Note that the significance indicators $s(R)$ may be viewed as functions involving p_0 – $s(R; p_0)$ – and are monotone decreasing functions of p_0. Hence, individual runs become shorter as p_0 increases, and it is possible to avoid substantial amounts of work by using this observation.

In effect, this sequential test considers runs which would not be remarkable from the initial level $p_0 = p_{0,1}$ to be remarkable if they survive to a level $p_0 = p_{0,m}$ at which they become surprising. We leave out details.

7.1.2 General Plane Curves

In the introduction, we alluded to the fact that the approach described here can be adapted to detect general plane curves, i.e. curves which are not graphs. This adapts ideas from our work on beamlet graphs [12, 4]. We define a family of directed graphs based on regions modelled on 'dyadically thickened beamlets' with various degrees of thickening. In this directed graph structure, strips can have all orientations, including vertical and horizontal, so that the graph constraint is removed. Connections between beamlets is based once more on good continuation principles, in this case cood continuation of plane polygons rather than polygonal graphs. Otherwise the algorithms are identical.

7.1.3 Inhomogeneous Smoothness

Hölder function classes impose smoothness conditions which are spatially homogeneous – the same at each point. Other types of smoothness classes, for example, bounded variation, allow jumps at some points, and differentiability at others. To accomodate such inhomogeneous smoothness would require a different architecture than the one discussed here. We would expand the notion of good continuation to encompass connecting strips in *different anisotropy classes*, provided they are immediately adjacent.

7.1.4 Higher-Degree Smoothness

In this paper we have only considered smoothness conditions where α is intermediate between first and second order. It makes sense to ask about smoothness of higher order, for example to consider $2 < \alpha \leq 3$. The analogous set of questions make sense in this setting; for example, there will continue to be about $n^{1/(1+\alpha)}$ points on some Hölder curve, even for α in this range. To accomodate such inhomogeneous smoothness would require a different set of strips than the one discussed here – curved ones with parabolic midlines – and a notion of good continuation based on matching of sides, and matching of slopes and curvatures of midlines. Preliminary calculations suggest that analogous 'adaptivity' results hold in such a setting.

7.1.5 Higher Dimensions

The analogous detection problem in d-dimensional space – finding a curve or surface containing an unexpectedly large number of points – has been considered by the authors in [3]. That work provides nonadaptive detectors, i.e. detectors which assume knowledge of the Hölder class.

The ideas developed in this paper can be directly applied to higher dimensions for adaptive multiscale detection of filamentary structure in d-dimensional point clouds, $d > 2$. We omit details.

A more ambitious generalization – to detection of codimension-k surfaces in d-dimensional point clouds – is possible, but more involved, and messier. For example, for $d = 3$ and $k = 1$, we are attempting to find a surface in d-dimensional space containing an inordinately large number of points. The nodes of each anisotropy graph are planar slabs of volume ρ/n and the neighborhood structure in the graph is, while conceptually analogous to the case considered here, far more complex to write out. The comparable adaptive detection theorems hold in that setting, although we omit details.

7.1.6 Vector Fields

Suppose that instead of data on points X_i we had data on tangent vectors, i.e. on pairs (X_i, θ_i) naming both a position and a direction. Experiments in perceptual psychophysics (e.g. [15, 22]) suggests that this is a much more potent stimulus to 'curve finding' than simply the display of random dots. Biological evidence about early vision suggests that individual receptive fields fire when both location and orientation offer matches.

As null hypothesis, we suppose that the X_i are i.i.d. uniform$[0,1]^2$, while the θ_i are i.i.d. uniform$[-\pi/2, \pi/2]$. As alternative hypothesis, we could have that a small fraction of the X_i lie on a curve, $X_i = (x_i, f(x_i))$ and the θ_i specify angles parallel to the line with slope $f'(x_i)$.

Our paper [3] showed that such tangent vector data are substantially more powerful for identifying filamentary structure than the point data discussed so far in this paper. Namely, such data give us the ability to detect filaments containing much smaller fractions of data points. In fact, a reliable test for a Hölder$(2,1)$ filament can be based on agreement with more than $Tn^{1/4}$ tangent vectors, rather than $Tn^{1/3}$ points.

The multiscale data structures used in this paper can be applied in the tangent vector setting, where we match (X_i, θ_i) to a region based not only on $X_i \in R'$ but also on θ_i matching the slope of the midline of R. The resultant algorithm can take advantage of this more stringent matching structure to speed up the counting process, as each tangent vector will lie in only one region

in a given anisotropy class, resulting in $O(\log(n))$ flops per data point rather than $O(n\log(n))$. Searching for significant paths will be significantly faster as well, since there can only be $n\log(n)$ starting places for a path. Hence the whole algorithm can run in $O(n\log(n))$ flops. An analysis paralleling the one given here shows that a multiscale-multianisotropic significance runs algorithm can provide a detection threshold optimal to within a constant factor.

7.1.7 Pixel Imagery

In another direction, we might consider data types used to model digital imagery, for example arrays $(y(i_1, i_2) : 0 \le i_1, i_2 < n)$, where

$$y_{i_1,i_2} = \xi_f(i_1, i_2) + \sigma z(i_1, i_2), \qquad 0 \le i_1, i_2 < n$$

$(z(i_1, i_2))$ is a Gaussian white noise, σ is the noise level, and ξ_f is a pixel array with nonzero values only on pixels which intersect $graph(f)$. Despite appearances, this problem is closely related to the present problem, and analagous detection theorems are true.

In effect, we form a family of anisotropic multiscale strips, and sum pixels intersecting those strips, producing detector statistics $X(R)$ which can be significance-tested in a way paralleling the counts $N(R)$ considered in this paper, only with Gaussian rather than Poisson threshold analysis. The underlying data structures and arguments can be understood as, roughly speaking, a mixture of the ideas of this paper and those of another paper by these authors [4]. In particular, the strips considered here were called axoids in that work. In the digital array setting, there is a fast beamlet transform to rapidly compute all the required $X(R)$ detector statistics.

7.2 Relations to Other Work

7.2.1 Parametric Detection

In this paper we have considered detection of points on nonparametric curves. In certain cases, one is interested in points along lines [8] or on parabolas [1]. For an attractive non-multiscale approach to such detection problems, see [10].

In the authors' paper [4] which was just mentioned, it has been shown how to develop multiscale geometric detectors for line segments in digital imagery, for parametric forms such as circles, rectangles, and ellipses. Those ideas could be adapted to the present setting, to find situations where data have an elevated density over a blob or along some line. In the end, the underlying computations involve dyadic multiscale rectangles and strips, and the ideas are closely related to those in this paper.

7.2.2 Multiscale Geometric Analysis

The tools described here are closely related to a variety of related tools in multiscale methods; see [12, 13, 14, 17, 9, 21, 26] for discussion of related tools applied in image analysis and in mathematical analysis. We differ here in our use of a multi-scale, multi-anisotropy collection of analysing regions, which is organized and exploited in a specific way and for a specific purpose.

7.2.3 Object Grouping and Neural Architecture

In the literature of computer vision, there is extensive discussion of object grouping in perception [6]. The problem considered here – recognizing a curve against a background of random points – fits in this tradition, and there are even experiments in psychophysics that test the ability of the human visual system to accomplish similar tasks [15, 18]. In effect, what we have discussed here – (near-) optimal detection – corresponds to what psychophysicists would call 'the ideal observer' [20]. In this connection, we have exhibited a simple multiscale architecture that can provide a near-optimal detector for a very wide range of stimuli – curves of any of a wide range of degrees of smoothness.

In the Gestalt theory of perception, there arises the concept of 'good continuation' [28, 15]; experiments show that the visual system will respond better to curvilinear stimuli that follow a 'good continuation' of an initial pattern [25]. Here we have operationalized this principle by a regular connectivity pattern in a specific graphical structure. We have shown that with this implementation, we get a near-optimal detector, thus validating the significance of such a 'good continuation' principle. Note well that the connectivity pattern is invariant – i.e. it applies the same way at all nodes of the graph.

This architectural simplicity is striking when compared with the vast speculative literature proposing 'neural architectures' for visual perception. What we have shown is that, starting from a large collection of elements which are sensitive to (i.e. accumulate counts in) 'receptive fields' at a variety of lengths, widths, orientations and locations, then connecting such elements to other elements by a simple invariant rule, one very sensitively recognizes the existence of curvilinear stimuli simply by the existence of long connected paths.

Perhaps this bears comparison with biological evidence. fMRI studies [7] suggest that there are centers in primate brains which seem responsible for integrating local information into recognition of long curvilinear structures [23]. It would be interesting to know whether such integration has any resemblance to the simple multiscale connection mechanism employed here.

8 Appendix

8.1 Proof of Lemma 2.1

Extend notation so that $w(j) = 2^{-j}$ can have both real and integer arguments, and $t(j) = 2^{-(J-j)+r}$ as well. Let $j^+ = j^+(\alpha, \beta, n)$ be the real-valued solution of the system of equations

$$2\beta w(j)^\alpha = t(j),$$

i.e.

$$2\beta 2^{-j\alpha} = 2^{-(J-j)+r}.$$

Let $j^* = \lceil j^+ \rceil$ be the next larger integer, so that $w(j^+)/2 \le w(j^*) \le w(j^+)$, and $t(j^+) \le t(j^*) \le 2t(j^+)$. Then as $1 \le \alpha \le 2$, $2^\alpha \le 4$, and so

$$2\beta w(j^*)^\alpha \le 2\beta w(j^+)^\alpha = t(j^+) \le t(j^*) \le 2 \cdot t(j^+) = 4\beta w(j^+)^\alpha \le 16\beta w(j^*)^\alpha$$

Now let $w = w(j^*)$ and $t = t(j^*)$ and substitute in the last display, getting (2.1). □

8.2 Proof of Lemma 2.2

Remember that $f \in \text{Hölder}(\alpha, \beta)$ satisfies $f : [0,1] \to [0,1]$ and

$$|f'(x) - f'(y)| \leq \alpha\beta |x-y|^{\alpha-1},$$

We saw that this implies

$$|f(x) - f(y) - f'(y)(x-y)| \leq \beta |x-y|^{\alpha}, \quad x, y \in [0,1]. \tag{8.1}$$

To prove (2.2), we use the notation $\dot{f}_k(x)$ for the affine function tangent to graph(f) at x_k. Using (8.1), with I_k denoting $[kw, (k+1)w)$,

$$|f(x) - \dot{f}_k(x)| \leq \beta (w/2)^{\alpha} \leq t/4, \quad x \in I_k.$$

We also note that if $\dot{g}_k(x) = l_1\delta_1 + l_2\delta_2(x - x_k)$, then

$$\begin{aligned} |\dot{f}_k(x) - \dot{g}_k(x)| &\leq |f(x_k) - l_1\delta_1| + |f'(x_k) - l_2\delta_2||x - x_k| \\ &\leq \delta_1/2 + \delta_2/2 \times w/2 \\ &\leq t/32 + t/64. \end{aligned}$$

We conclude that

$$|f(x) - \dot{g}_k(x)| \leq t/2, \qquad x \in I_k. \tag{8.2}$$

On the other hand, the region $R(j, k, \ell_1, \ell_2)$ has $\dot{g}_k(x)$ as its mid-line, and is of half-height $t/2$. The desired relation (2.2) follows. □

Figure 4 illustrates our notation.

8.3 Proof of Lemma 2.3

It is enough to show that

$$|\dot{g}_{k+1}(x_{k+1}) - \dot{g}_k(x_{k+1})| \leq t, \tag{8.3}$$

and

$$|\dot{g}'_{k+1}(x_{k+1}) - \dot{g}'_k(x_{k+1})| \leq t/w. \tag{8.4}$$

It then follows that there is an edge in $\mathcal{E}(j)$ connecting $R(j, k, \ell_1, \ell_2)$ to $R(j, k+1, \ell'_1, \ell'_2)$ where ℓ'_1 and ℓ'_2 are the values associated to $\dot{g}_{k+1}$.

The following inequalities flow either from Hölder conditions or from simple rounding involved in quantizations:

$$\begin{aligned} |\dot{g}_{k+1}(x_{k+1}) - f(x_{k+1})| &\leq \delta_1/2 = t/32, \\ |f(x_{k+1}) - \dot{f}_k(x_{k+1})| &\leq \beta w^{\alpha} \leq t/2, \\ |\dot{f}_k(x_{k+1}) - \dot{g}_k(x_{k+1})| &\leq \delta_1/2 + \delta_1/2 \cdot w = t/16; \end{aligned}$$

combining these with the triangle inequality yields (8.3). Similarly,

$$\begin{aligned} |\dot{g}'_{k+1}(x_{k+1}) - f'(x_{k+1})| &\leq \delta_2/2 = t/(32w), \\ |f'(x_{k+1}) - \dot{f}'_k(x_{k+1})| &\leq \alpha\beta w^{\alpha-1} \leq t/(2w). \\ |\dot{f}'_k(x_{k+1}) - \dot{g}'_k(x_{k+1})| &\leq \delta_2/2 = t/(32w). \end{aligned}$$

which combine to give (8.4).

References

[1] H. Abramowicz, D. Horn,U. Naftali, C. Sahar-Pikielny (1997). An orientation-selective neural network for pattern identification in particle detectors. In *Advances in Neural Information Processing Systems, 9* Eds. M. C. Mozer, M. J. Jordan and T. Petsche, MIT Press, pp. 925–931.

[2] A.V. Aho, J.E. Hopcroft, and J.D. Ullman (1983). *Data Structures and Algorithms.* Addison-Wesley, Reading, Mass.

[3] E. Arias-Castro, D.L. Donoho, X. Huo, C. Tovey. Connect-The-Dots: How many random points can a regular curve pass through? (Work in progress.)

[4] E. Arias-Castro, D.L. Donoho, X. Huo. Near-Optimal Detection of Geometric Objects by Fast Multiscale Methods. (Work in progress.)

[5] R. Arratia and M. S. Waterman (1989). The Erdös-Rényi strong law for pattern matching with a given proportion of mismatches. *Ann. Probab.* **17**, 1152–1169.

[6] J. M. Buhmann, J. Malik, and P. Perona (1999). Image recognition: Visual grouping, recognition, and learning. *Proc. Natl. Acad. Sci. USA*, **96** 14203-14204.

[7] S.M. Courtney and L.G. Ungerleider (1997). What fMRI has taught us about human vision. *Current Opinion in Neurobiology*, **7**, 554-561.

[8] A. C. Copeland, G. Ravichandran, and M.H. Trivedi (1995). Localized Radon transform-based detection of ship wakes in SAR image. *IEEE Trans. Geosci. and Remote Sens.* **33**, 35-45.

[9] G. David and S. Semmes (1993). *Analysis of and on uniformly rectifiable sets*, *Math. Surveys and Monographs*, volume 38. Amer. Math. Soc.

[10] A. Desolneux, L. Moisan, and J.-M. Morel (2000). Meaningful Alignments. *International Journal of Computer Vision* **40**, 7–23.

[11] D. Donoho and X. Huo (2000). Beamlet pyramids: A new form of multiresolution analysis, suited for extracting lines, curves, and objects from very noisy image data. In *Proceedings of SPIE*, volume 4119, July.

[12] D. Donoho and X. Huo (2001). Beamlets and Multiscale Image Analysis. In: *Multiscale and Multiresolution Methods.* T.J. Barth, T. Chan, and R. Haimes (Eds.), Springer Lecture Notes in Computational Science and Engineering **20**, 149-196.

[13] D. L. Donoho (1999). Wedgelets: nearly minimax estimation of edges. *Annals of Statistics* **27**, 859–897.

[14] D. L. Donoho and O. Levi (2002). Fast X-ray and beamlet transforms for three-dimensional data. in *Modern Signal Processing*, Dan Rockmore and Dennis Healy, eds. Mathematical Science Research Institute Publications, **46**, Cambridge University Press, Cambridge and New York.

[15] D. Field, A. Hayes, R. Hess (1993). Contour integration by the human visual system: Evidence for a local "association field". *Vision Research* **33**, 173–193.

[16] Mae-Wan Ho, In Search of the Sublime, at Institute for Science in Society Website, `http://www.i-sis.org.uk/sublime.php`

[17] P. W. Jones (1990). Rectifiable sets and the traveling salesman problem. *Inventiones Mathematicae*, **102**, 1–15.

[18] I. Kovacs, B. Julesz (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segementation. *Proc. Natl. Acad. Sci. USA*, **90**, 7495-7497.

[19] M. R. Leadbetter, G. Lindgren, and H. Rootzén (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.

[20] G. E. Legge, D. Kersten, and F A. E. Burgess (1987). Contrast discrimination in noise *J. Opt. Soc. Am. A* 391-404, 1987.

[21] G. Lerman (2000). *Geometric Transcriptions of Sets and Their Applications to Data Analysis.* Ph.D. Thesis, Yale University, Department of Mathematics.

[22] Dennis M. Levi, Stanley A. Klein (1999). Seeing circles: what limits shape perception? *Vision Research*,**40** (240), 2329-2339.

[23] J. D. Mendola, A. M. Dale, B. Fischl, A. K. Liu, and R. B. H. Tootell (1999). The Representation of Illusory and Real Contours in Human Cortical Visual Areas Revealed by Functional Magnetic Resonance Imaging. *J. Neuroscience*, **19**, 8560-8572.

[24] N. Qaddoumi, E. Ranu, J. D. McColskey, R. Mirshahi, and R. Zoughi (2000). Microwave detection of stress-induced fatigue cracks in steel and potential for crack opening determination. *Research in Nondestructive Evaluation,* **12**, 87-103.

[25] Z. Pizlo, M Salach-Golyska, A. Rosenfeld (1997). Curve Detection in a Noisy Image, *Vision Research*, **37** (9), 1217-1241.

[26] E. Sharon, A. Brandt, and R Basri (2000). Fast multiscale image segmentation. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition* **1**, 70–77.

[27] F. Tupin, H. Maitre, J. F. Mangin, J. M. Nicolas, and E. Pechersky (1998). Detection of linear features in SAR images: application to road network extraction. *IEEE Trans. on Geoscience and Remote Sensing* **36**, 434-453, March.

[28] Wertheimer, M. (1938). *Laws of Organization in Perceptual Forms*. London: Harcourt Brace Jovanovich.