Networks of Polynomial Pieces with Application to the Analysis of Point Clouds and Images

Ery Arias-Castro[1]*, Boris Efros[2] and Ofer Levi[2]

[1] University of California, San Diego, USA
[2] Ben Gurion University of the Negev, Be'er Sheva, Israel

Abstract

We consider Hölder smoothness classes of surfaces for which we construct piecewise polynomial approximation networks, which are graphs with polynomial pieces as nodes and edges between polynomial pieces that are in 'good continuation' of each other. Little known to the community, a similar construction was used by Kolmogorov and Tikhomirov in their proof of their celebrated entropy results for Hölder classes.

We show how to use such networks in the context of detecting geometric objects buried in noise to approximate the scan statistic, yielding an optimization problem akin to the Traveling Salesman. In the same context, we describe an alternative approach based on computing the longest path in the network after appropriate thresholding.

For the special case of curves, we also formalize the notion of 'good continuation' between beamlets in any dimension, obtaining more economical piecewise linear approximation networks for curves.

We include some numerical experiments illustrating the use of the beamlet network in characterizing the filamentarity content of 3D datasets, and show that even a rudimentary notion of good continuity may bring substantial improvement.

Keywords and Phrases: Detection of filaments; Multiscale analysis; Beamlets; Extracting information from graphs; Hölder smoothness classes; Piecewise polynomials.

1 Introduction

1.1 From function approximation to set approximation

An important trend in Approximation Theory and Harmonic Analysis focuses on designing dictionaries (e.g. orthonormal bases) $\{\phi_n\}$ well-adapted to a given function class $\mathcal{F}$, in the

---

*Corresponding author: `eariasca@math.ucsd.edu`

sense that any $f \in \mathcal{F}$ is well-approximated by a linear combination of a few functions from the dictionary:

$$f \doteq \sum_{n \in N(f)} a_n \phi_n. \tag{1}$$

Examples of dictionaries include the Fourier basis, polynomials, splines [53], radial basis functions, wavelets [44] and others such as wedgelets [22], platelets [55], bandelets [39], rigelets [13], curvelets [16], chirplets [14].

In the context of function estimation in additive white noise [32, 24, 21], approximations by sums of atoms as in (1) are particularly suitable. Consider instead a setting where a geometric object (i.e. a set) is buried in noise or clutter, a setting considered e.g. in [23, 6]. This is for example relevant in target tracking, where the object of interest is the target's trajectory often modeled as a curve. In this setting, set approximation plays the role of function approximation, and the aim of the present paper is to develop strategies to compute approximations of sets by unions of simple building blocks akin to how beamlets are used to approximate curves in [23].

Since a set may be equivalently represented by its indicator function, approximation of sets may appear to be a special case of approximation of functions. Though indeed closely related, function approximation does not directly translate into set approximation. In part, this comes from the fact that the image of a sum of functions is in general not the union of the functions' images. However, when a parametrization of the set is available, approximating the parametrization in piecewise fashion (i.e. when the supports of the functions involved in the sum do not overlap) does result in a proper approximation of the set itself. This is for example the case when the set is the graph of a function, considered in [38, 22] in the context of image processing.

The selection of building blocks in set approximation is also not necessarily parallel to that of atoms in function approximation. The main difference is in the fact that overlapping building blocks create redundancy, while atoms with overlapping supports may cancel each other in appropriate ways to fit the target function, e.g. in trigonometric, polynomial or wavelet expansions. Typically, the approximation (1) is constructed by first computing the coefficients a_n, often as the inner product of f and ϕ_n, and then keeping the largest ones in absolute value, which is in effect equivalent to thresholding [21]. A similar strategy may be implemented for sets, where building blocks with the most overlap with the set (as a fraction of their size) are selected; this corresponds to the simplest beamlet-based algorithm presented in [23]. However, the result is often redundant, as nearby building blocks tend to have a similar overlap with a given set. This is avoided in the beamlet-based coder for curves introduced in [30], by choosing at most one beamlet per dyadic square in a a given recursive dyadic partitioning. The strategy we adopt here consists of looking for a union of building blocks that obey some sort of 'good continuation'. In the approximation of curves, this corresponds to the most complex algorithm presented in [23] which amounts to chaining beamlets together that share one endpoint and have similar orientations. This is formalized in [7] for curves that are graphs of Hölder functions and in [15] for chirps (highly oscillating functions), chaining chirplets in good continuation. Good-continuation principles are also considered in [18, 19] inspired by the Gestalt theory of vision, with applications to the detection of (parametric) geometric object in images.

2

1.2 Networks of polynomial pieces

We define a network of polynomial pieces to be a graph with nodes indexing polynomial pieces and edges between polynomial pieces in good-continuation. (We use the word 'network' instead of 'graph' so as to avoid confusion with the notion of 'graph of a function'.) We construct such networks to approximate surfaces of varying dimension and smoothness. Just as beamlets are line-segments spanning a wide range of location, orientations and scale, so do the polynomial pieces. The networks are akin to that of [7] in that the notion of good-continuation is explicit and the scales are not mixed together, thus better adapted to surfaces with homogeneous smoothness, e.g. graphs of Hölder functions. We do, however, suggest ways to go multiscale.

In the 1950's, Kolmogorov and Tikhomirov used piecewise polynomial approximations together with a similar kind of good-continuation notion to bound the ϵ-entropy of Hölder function classes [35]. Our construction may be seen as a formalization of their approach, with emphasis on the approximation of sets instead of functions. Note that this ancestry was discovered after the fact; the present perspective was indeed independently suggested in [5] with the intention of generalizing the system used in [7].

1.3 The special case of beamlets

Beamlets were introduced in [23] for the explicit purpose of approximating curves in 2D, with a 3D version latter developed in [25]. A variety of algorithms are proposed in [23], where the more elaborate ones are based on the chaining of beamlets in good continuation. However, the notion of good-continuation remains implicit and only palpable through numerical experiments.

We formalize here this notion of good-continuation for beamlets. This was previously done in [7] for a beamlet-like system built to detect graphs of Hölder functions. In the resulting beamlet network, small beamlets have a large number of neighbors; this seems unavoidable if the network is to enable accurate approximation of curves. To address this issue, we develop an alternative network of line-segments, with emphasis on developing an economical system both in terms of number of nodes and connectivity.

1.4 Application to the analysis of point clouds and images

1.4.1 Detection of geometric objects

Consider a simple model for detection, where we observe a point cloud (i.e. a spatial point process) and the goal is to decide whether the points were generated uniformly at random or a fraction of them were sampled from a geometric object (i.e. a surface) belonging to a given class.

This is a standard setting where the scan statistic is used [28, 27], which consists in computing the largest number of points belonging to one of the objects of interest. Though the scan statistic achieves the best known detection rates, both for parametric [6] and nonparametric [8] classes of objects, it is not computationally friendly as it involves an optimization over a large (function) class of objects, though there are exceptions [31]. Instead, we propose to replace the optimization over the class of objects with an optimization over paths in an

approximating network. The idea of replacing an optimization over a function space with an optimization over a carefully constructed graph is quite natural, and in fact appears in other situations, e.g. in the computation of minimal surfaces [12, 37, 34, 29]. In our particular case, the resulting optimization problem is akin to the Budget-Reward problem [17]; though still NP-hard, this problem admits polynomial time approximations. Moreover, we show that this approximation achieves the best known detection rates.

Note that the optimization above is computationally more tractable (e.g. using dynamic programming ideas) when the network is direct and acyclic, which is the case for example in multiframe target tracking; see e.g. [41], where a beamlet network is used to track the time-space primitives.

We also consider an alternative approach based on computing the size of the longest path in the network after thresholding, which is suggested in [23]. Dynamic programming ideas may also be implemented here, as done in [7]. We show that this method also achieves the best known detection rates.

We mention that the same approaches may be implemented in the setting of an image, where a geometric object is buried in white Gaussian noise.

1.4.2   Characterization of spatial distributions

In Astrophysics, the study of the galaxy distribution involves quantifying the content in filaments, sheets and blobs in 3D galaxy catalogs [45, 1]. Ultimately, scientists would like to know which of the many cosmological models is best (and well) supported by observations. Practically, the task is to meaningfully compare simulated galaxy distributions from various cosmological models with the observed galaxy distribution.

With the presence of highly anisotropic features such as filaments, traditional tools for analyzing spatial data become irrelevant, among them the classical two-point correlation function. Instead, a method based on beamlets is very attractive, as beamlets provide good approximations for filaments. And indeed, [20] presents evidence that beamlets are useful at separating various cosmological models, even though the algorithm implemented in [20] is of the simplest kind and in particular does not involve chaining (i.e. good-continuation). We perform a number of numerical experiments on simulated data that show that chaining may bring substantial improvement.

Note that such tools are in demand in other scientific fields, such as Medical Imaging, for example in the examination of vascular networks [50] or cancer cells [43].

1.5   Contents

The contents are organized as follow. In Section 2, we introduce piecewise polynomial networks designed to approximate surfaces of any intrinsic dimension and (Hölder) smoothness. In Section 3, we consider the detection of geometric objects buried in noise and develop methods based on these networks. In Section 4, we formalize a notion of good-continuation for beamlets in arbitrary dimension and show that the resulting network has desirable approximation properties for curves. In Section 5, we perform some numerical experiments showing that the notion of good-continuation may bring practical improvement. Some of the proofs and technical arguments are gathered in the Appendix.

2 Networks of Polynomial Pieces

We build an explicit, algorithmically friendly approximating network for Hölder surfaces, i.e. graphs or images of Hölder functions. As a Hölder function is well-approximated locally by a polynomial, in fact its Taylor expansion, it is natural to construct piecewise polynomial approximations. The idea is to partition the unit hypercube into smaller hypercubes and in each smaller hypercube provide a choice of approximation by polynomials; since the functions we consider are uniformly smooth, approximations in nearby hypercubes should be close, which we formalize as a neighboring condition. We thus form a network with nodes indexing polynomial pieces and edges linking those in good-continuation, restricting the possible combinations to those useful in approximating functions of given smoothness, in such a way that functions and certain connected components in this network are in correspondence. Though we build a different network for each smoothness class, it is possible to discretize the range of parameters resulting in a dyadic organization of this family of networks by scale.

Little known to the community, a similar construction was used by Kolmogorov and Tikhomirov in their seminal work on ϵ-entropy of Hölder function classes (and others) [36, 35]. Note that the present construction was independently suggested in [5], as a generalization of the beamlet-like system used in [7].

We first introduce some notation. For $i \in \{1, \ldots, k\}$, let $\mathbf{e}_i$ denote the ith canonical vector in $\mathbb{R}^k$. For a vector $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$, its supnorm is defined as $\|\mathbf{x}\| = \max\{|x_i| : i = 1, \ldots, k\}$. For $\mathbf{s} = (s_1, \ldots, s_k) \in \mathbb{N}^k$, let $\mathbf{s}! = s_1! \cdots s_k!$ and $|\mathbf{s}| = s_1 + \cdots + s_k$. For a function f and $\mathbf{s} = (s_1, \ldots, s_k) \in \mathbb{N}^k$, $f^{(\mathbf{s})} = \partial_{x_1}^{s_1} \cdots \partial_{x_k}^{s_k} f$.

We define the following constants:

$$c_1 = \sum_{|\mathbf{s}|=\lfloor\alpha\rfloor} \frac{1}{\mathbf{s}!}, \qquad c_2 = \sum_{|\mathbf{s}|\leq\lfloor\alpha\rfloor} \frac{2^{-|\mathbf{s}|}}{\mathbf{s}!}. \tag{2}$$

Note that $c_1 \leq \exp(k)$ and $c_2 \leq \exp(k/2)$.

2.1 Hölder smoothness classes

For $\alpha, \beta > 0$, define $\mathcal{H}^k(\alpha, \beta)$ as the Hölder smoothness class of $\lfloor\alpha\rfloor$-times differentiable functions functions $f : [0,1]^k \to [0,1]$ satisfying:

$$|f^{(\mathbf{s})}(\mathbf{x})| \leq \beta, \quad \forall \mathbf{x} \in [0,1]^k, \forall \mathbf{s} \in \mathbb{N}^k, |\mathbf{s}| \leq \lfloor\alpha\rfloor; \tag{3}$$

$$|f^{(\mathbf{s})}(\mathbf{y}) - f^{(\mathbf{s})}(\mathbf{x})| \leq \beta\|\mathbf{y} - \mathbf{x}\|^{\alpha-\lfloor\alpha\rfloor}, \quad \forall \mathbf{x}, \mathbf{y} \in [0,1]^k, \forall \mathbf{s} \in \mathbb{N}^k, |\mathbf{s}| = \lfloor\alpha\rfloor. \tag{4}$$

Example. ($k = 1$, $\alpha \in (1,2]$)

$$|f'(x)| \leq \beta, \quad \forall x \in [0,1];$$

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

Functions in $\mathcal{H}^k(\alpha, \beta)$ are uniformly well-approximated locally by polynomials, specifically their Taylor expansions. For $f \in \mathcal{H}^k(\alpha, \beta)$ and $\mathbf{x} \in [0,1]^k$, the Taylor expansion of f

at $\mathbf{x}$ of degree $\lfloor\alpha\rfloor$ is defined as follows:

$$\dot{f}_{\mathbf{x}}(\mathbf{y}) = \sum_{|\mathbf{s}|\leq\lfloor\alpha\rfloor} f^{(\mathbf{s})}(\mathbf{x}) \prod_{i=1}^{d} \frac{(y_i - x_i)^{s_i}}{s_i!}.$$

Example. ($k = 1$, $\alpha \in (1, 2]$)

$$\dot{f}_x(y) = f(x) + f'(x)(y - x).$$

Lemma 2.1. *For any $f \in \mathcal{H}^k(\alpha, \beta)$,*

$$|f(\mathbf{y}) - \dot{f}_{\mathbf{x}}(\mathbf{y})| \leq c_1\beta\|\mathbf{y} - \mathbf{x}\|^{\alpha}, \quad \forall \mathbf{x}, \mathbf{y} \in [0, 1]^k.$$

Proof. A Taylor approximation of degree $\lfloor\alpha\rfloor$ gives:

$$f(\mathbf{y}) = \dot{f}_{\mathbf{x}}(\mathbf{y}) + \sum_{|\mathbf{s}|=\lfloor\alpha\rfloor} (f^{(\mathbf{s})}(\mathbf{z}) - f^{(\mathbf{s})}(\mathbf{x})) \prod_{i=1}^{d} \frac{(y_i - x_i)^{s_i}}{s_i!},$$

for some $\mathbf{z}$ on the segment joining $\mathbf{x}$ and $\mathbf{y}$. Hence,

$$|f(\mathbf{y}) - \dot{f}_{\mathbf{x}}(\mathbf{y})| \leq c_1\|\mathbf{y} - \mathbf{x}\|^{\lfloor\alpha\rfloor} \max_{|\mathbf{s}|=\lfloor\alpha\rfloor} |f^{(\mathbf{s})}(\mathbf{z}) - f^{(\mathbf{s})}(\mathbf{x})|.$$

Now apply (4) and the fact that $\|\mathbf{z} - \mathbf{x}\| \leq \|\mathbf{y} - \mathbf{x}\|$ to get

$$|f^{(\mathbf{s})}(\mathbf{z}) - f^{(\mathbf{s})}(\mathbf{x})| \leq \beta\|\mathbf{y} - \mathbf{x}\|^{\alpha-\lfloor\alpha\rfloor}, \quad \forall \mathbf{s} \in \mathbb{N}^k, |\mathbf{s}| = \lfloor\alpha\rfloor.$$

$\square$

2.2   Nets of piecewise polynomials

We now build a family of nets for $\mathcal{H}^k(\alpha, \beta)$ by dividing $[0, 1]^k$ into hypercubes and then offering a choice of approximation by polynomials within each hypercube, which is most relevant in view of Lemma 2.1. Fix $\Delta \in (0, 1)$ and $\delta > 0$, and define $\delta_s = \Delta^{-s}\delta, s = 0, \ldots, \lfloor\alpha\rfloor$. In the construction that follows, the parameter Δ quantizes the variable space, while each parameter δ_s quantizes the range of values of derivatives of order s of functions in $\mathcal{H}^k(\alpha, \beta)$. Note that the quantization is coarser for higher order derivatives, and specifically chosen so that the approximation result in Lemma 2.2 below holds.

Divide $[0, 1]^k$ into hypercubes indexed by $\mathbf{m} \in \{1, \ldots, \Delta^{-1}\}^k$ of the form:

$$I_{\mathbf{m}} = \prod_{i=1}^{k} [(m_i - 1)\Delta, m_i\Delta].$$

Let $\mathbf{x_m} = (x_{\mathbf{m},1}, \ldots, x_{\mathbf{m},k})$ denote the center of $I_{\mathbf{m}}$, i.e. $x_{\mathbf{m},i} = (m_i - 1/2)\Delta$. On each hypercube $I_{\mathbf{m}}$, consider polynomials of the form

$$p_{\mathbf{m},\mathbf{h}}(\mathbf{x}) = \sum_{|\mathbf{s}|\leq\lfloor\alpha\rfloor} h^{(\mathbf{s})}\delta_{|\mathbf{s}|} \prod_{i=1}^{d} \frac{(x_i - x_{\mathbf{m},i})^{s_i}}{s_i!}, \tag{5}$$

where $\mathbf{h} = (h^{(\mathbf{s})} : |\mathbf{s}| \leq \lfloor\alpha\rfloor)$, with $h^{(\mathbf{s})} \in \mathbb{Z}$ and $|h^{(\mathbf{s})}\delta_{|\mathbf{s}|}| \leq \beta$.
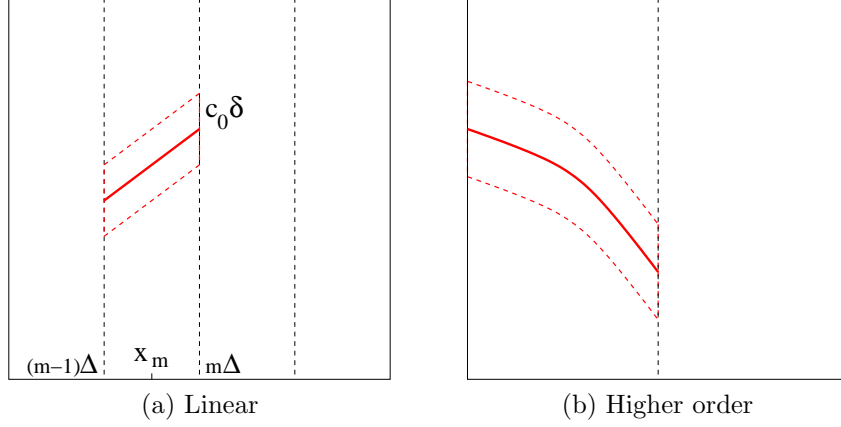
6

| (a) Linear | (b) Higher order |

Figure 1: Examples of polynomial pieces for different choices of Δ and δ, together with their associated region as defined in (8).

Example. $(k = 1,\ \alpha \in (1, 2])$ For $\mathbf{h} = (h^{(0)}, h^{(1)})$,

$$p_{m,\mathbf{h}}(x) = h^{(0)}\delta + h^{(1)}\delta_1(x - x_m).$$

For $f \in \mathcal{H}^k(\alpha, \beta)$, let $h^{(\mathbf{s})}(\mathbf{m}, f)$ denote the closest integer to $f^{(\mathbf{s})}(\mathbf{x_m})/\delta_{|\mathbf{s}|}$.

Lemma 2.2. *For $f \in \mathcal{H}^k(\alpha, \beta)$ and $\mathbf{m} \in \{1, \ldots, \Delta^{-1}\}^k$,*

$$|f(\mathbf{x}) - p_{\mathbf{m},\mathbf{h}(\mathbf{m},f)}(\mathbf{x})| \leq (c_1/2^\alpha)\beta\Delta^\alpha + (c_2/2)\delta, \quad \forall \mathbf{x} \in I_\mathbf{m}.$$

Proof. The proof of Lemma 2.2 is in the Appendix. □

Partly justified by Lemma 2.2, we now assume that

$$\delta = c_1\beta\Delta^\alpha. \tag{6}$$

The system $\{p_{\mathbf{m},\mathbf{h}}\}$ is therefore rich enough to provide a certain degree of approximation locally. The same degree of approximation may be achieved globally by simply considering functions that coincide with the polynomials above within each hypercube, namely

$$g_\mathbf{h}(\mathbf{x}) = \sum_\mathbf{m} \mathbf{1}_{\{\mathbf{x} \in I_\mathbf{m}\}}\, p_{\mathbf{m},\mathbf{h}(\mathbf{m})}(\mathbf{x}),$$

where this time $\mathbf{h}$ also depends on $\mathbf{m}$, representing a different choice for each hypercube.

As Kolmogorov and Tikhomirov realized, generating a $g_\mathbf{h}$ by simply picking a polynomial in each hypercube independently would result in a wasteful system, for in fact polynomials in neighboring hypercubes may be restricted to have similar coefficients. This comes from the fact that the derivatives up to order $\lfloor \alpha \rfloor$ of a function in $f \in \mathcal{H}^k(\alpha, \beta)$ have a certain degree of smoothness.

Lemma 2.3. *For $f \in \mathcal{H}^k(\alpha, \beta)$ and $\mathbf{s}$ such that $|\mathbf{s}| \leq \lfloor \alpha \rfloor$,*

$$\left| f^{(\mathbf{s})}(\mathbf{x} + \eta\mathbf{e}_i) - \sum_{t \leq \lfloor \alpha \rfloor - |\mathbf{s}|} f^{(\mathbf{s}+t\mathbf{e}_i)}(\mathbf{x})\, \frac{\eta^t}{t!} \right| \leq c_1\beta\eta^{\alpha - |\mathbf{s}|},$$

for all $\mathbf{x} \in [0, 1]^k$, $i = 1, \ldots, k$ and η such that $\mathbf{x} + \eta\mathbf{e}_i \in [0, 1]^k$.

7

Proof. As with Lemma 2.1, perform a Taylor approximation of degree $\lfloor \alpha \rfloor - |\mathbf{s}|$ along $\mathbf{e}_i$ and apply (4). In fact, β may be replaced by $\beta/(\lfloor \alpha \rfloor - |\mathbf{s}|)!$. $\qquad\square$

2.3 Approximating networks for Hölder graphs

For a function $f : [0,1]^k \to [0,1]$, define its graph as

$$\mathrm{graph}(f) = \{(\mathbf{x}, f(\mathbf{x})) : \mathbf{x} \in [0,1]^k\}.$$

Assuming (6), we define a network of polynomial pieces $\mathbb{G}_\delta^k(\alpha, \beta)$ with the property that a certain kind of connected components index piecewise polynomial approximations for graphs of Hölder functions. The network $\mathbb{G}_\delta^k(\alpha, \beta)$ has nodes of the form $(\mathbf{m}, \mathbf{h})$ indexing the polynomials $p_{\mathbf{m},\mathbf{h}}$ defined in (5). Two nodes in the network $(\mathbf{m}, \mathbf{h})$ and $(\mathbf{m}_\star, \mathbf{h}_\star)$ are neighbors if the corresponding hypercubes, $I_{\mathbf{m}}$ and $I_{\mathbf{m}_\star}$, are adjacent, and if the corresponding polynomials, $p_{\mathbf{m},\mathbf{h}}$ and $p_{\mathbf{m}_\star,\mathbf{h}_\star}$, and their derivatives assume nearby values both at $\mathbf{x}_{\mathbf{m}}$ and $\mathbf{x}_{\mathbf{m}_\star}$. Formally, this corresponds to $\mathbf{m}_\star = \mathbf{m} + \xi \mathbf{e}_i$ for some $i \in \{1, \ldots, k\}$ and $\xi \in \{-1, +1\}$, and for all $\mathbf{s} \in \mathbb{N}^k, |\mathbf{s}| \le \lfloor \alpha \rfloor$,

$$\left| h_\star^{(\mathbf{s})} - \sum_{t \le \lfloor \alpha \rfloor - |\mathbf{s}|} \frac{\xi^t}{t!} h^{(\mathbf{s}+t\mathbf{e}_i)} \right| < 3 \quad \text{and} \quad \left| h^{(\mathbf{s})} - \sum_{t \le \lfloor \alpha \rfloor - |\mathbf{s}|} \frac{(-\xi)^t}{t!} h_\star^{(\mathbf{s}+t\mathbf{e}_i)} \right| < 3. \tag{7}$$

This last property is a discrete version of Lemma 2.3.

Example. $(k = 1, \alpha \in (1,2])$ The nodes $(m, h^{(0)}, h^{(1)})$ and $(m_\star, h_\star^{(0)}, h_\star^{(1)})$ are neighbors if $|m_\star - m| = 1$ and

$$|h_\star^{(1)} - h^{(1)}| < 3, \quad |h_\star^{(0)} - h^{(0)} - (m_\star - m)h^{(1)}| < 3, \quad |h^{(0)} - h_\star^{(0)} + (m_\star - m)h_\star^{(1)}| < 3.$$

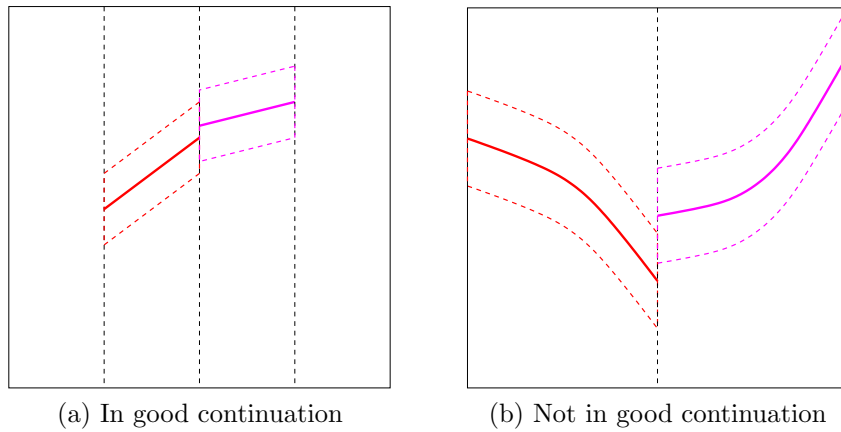(a) In good continuation   (b) Not in good continuation

Figure 2: Examples of polynomial pieces in good continuation (left) and not in good continuation (right).

The family of networks $\mathbb{G}_\delta^k(\alpha, \beta)$ effectively generalizes the system presented in [7] for the case $k = 1, \alpha \in (1, 2]$.

Define constants

$$c_3 = \sum_{s=0}^{\lfloor \alpha \rfloor} \binom{s+k-1}{k-1}, \qquad c_4 = \sum_{s=0}^{\lfloor \alpha \rfloor} s \binom{s+k-1}{k-1}.$$

Lemma 2.4. $\mathbb{G}_\delta^k(\alpha, \beta)$ *has* $O(\beta^{c_3}\delta^{-c_3}\Delta^{c_4-k}) = O(\beta^{c_3-(c_4-k)/\alpha}\delta^{-c_3+(c_4-k)/\alpha})$ *nodes and each node has at most* $2k6^{c_3}$ *neighbors.*

Proof. The proof of Lemma 2.4 is in the Appendix. $\qquad\qquad\square$

Example. $(k = 1, \alpha \in (1, 2])$ $\mathbb{G}_\delta^1(\alpha, \beta)$ has $O(\beta\delta^{-2})$ nodes, each with degree at most 72.

To each vertex $(\mathbf{m}, \mathbf{h})$ we associate a region around the graph of $p_{\mathbf{m},\mathbf{h}}$ restricted to $I_{\mathbf{m}}$, of thickness given by the error bound of Lemma 2.2:

$$R(\mathbf{m}, \mathbf{h}) = \{(\mathbf{x}, z) \in I_{\mathbf{m}} \times [0, 1] : |z - p_{\mathbf{m},\mathbf{h}}(\mathbf{x})| \leq c_0\delta\}, \tag{8}$$

where $c_0 = (2^{-\alpha} + c_2/2)$. See Figure 1.

Example. $(k = 1, \alpha \in (1, 2])$ The regions are in this case parallelograms.

For a subset of nodes π, define $R(\pi) = \bigcup_{(\mathbf{m},\mathbf{h}) \in \pi} R(\mathbf{m}, \mathbf{h})$.

Let $\Pi_\delta^k(\alpha, \beta)$ denote the set of connected components of $\mathbb{G}_\delta^k(\alpha, \beta)$, homeomorphic to the square grid $\{1, \ldots, \Delta^{-1}\}^k$.

Theorem 2.5. *For each* $f \in \mathcal{H}^k(\alpha, \beta)$*, there is a connected component* $\pi \in \Pi_\delta^k(\alpha, \beta)$ *such that* $\mathrm{graph}(f) \subset R(\pi)$.

Proof. The proof of Theorem 2.5 is in the Appendix. $\qquad\qquad\square$
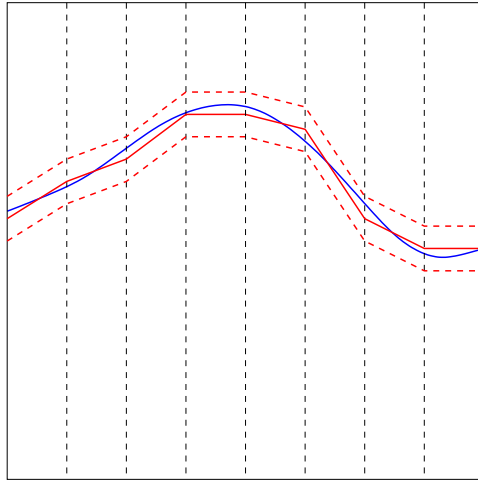

Figure 3: Example of covering of an Hölder graph (blue) with polynomial (linear) pieces in good continuation (red).

Theorem 2.5 implies that the system of functions

$$\sum_{(\mathbf{m},\mathbf{h}) \in \pi} \mathbf{1}_{\{\mathbf{x} \in I_{\mathbf{m}}\}} \, p_{\mathbf{m},\mathbf{h}}, \quad \pi \in \Pi_\delta^k(\alpha, \beta),$$

is an ε-net for $\mathcal{H}^k(\alpha, \beta)$ with $\varepsilon = c_0\delta$. From Lemma 2.4 and Theorem 2.5, we see that the system has entropy of order $O(\varepsilon^{-k/\alpha})$, which is essentially the smallest possible [36].

9

2.4 Approximating networks for Hölder immersions

Define $\mathcal{H}^{k,d}(\alpha, \beta)$ as the class of functions $f : [0,1]^k \to [0,1]^d$ with coordinates in $\mathcal{H}^k(\alpha, \beta)$, i.e. $f = (f_1, \ldots, f_d)$ with $f_r \in \mathcal{H}^k(\alpha, \beta)$ for $r = 1, \ldots, d$.

For a function $f : [0,1]^k \to [0,1]^d$, define its image as

$$\text{im}(f) = \{f(\mathbf{x}) : \mathbf{x} \in [0,1]^k\}.$$

An approximation network for images of Hölder functions is simply built out of a tensor product of copies of the network built in the previous section. Specifically, define the network $\mathbb{G}_\delta^{k,d}(\alpha, \beta)$, with nodes of the form $(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_d)$, indexing the multivariate polynomial $(p_{\mathbf{m},\mathbf{h}_1}, \ldots, p_{\mathbf{m},\mathbf{h}_d})$; and edges between nodes indexing polynomial pieces on adjacent hypercubes and satisfying (7) coordinate-wise.

Example. ($k = 1$, $\alpha \in (1,2]$) The polynomial pieces are of the form:

$$(h_1^{(0)}\delta + h_1^{(1)}\delta_1(x - x_m), h_2^{(0)}\delta + h_2^{(1)}\delta_1(x - x_m)), \quad |x - x_m| \leq \Delta/2.$$

Lemma 2.6. $\mathbb{G}_\delta^{k,d}(\alpha, \beta)$ *has* $O(\beta^{dc_3}\delta^{-dc_3}\Delta^{dc_4-k}) = O(\beta^{dc_3-(dc_4-k)/\alpha}\delta^{-dc_3+(dc_4-k)/\alpha})$ *nodes and each node has at most* $2k6^{dc_3}$ *neighbors.*

Proof. The proof follows that of Lemma 2.4. Details are omitted. □

Example. ($k = 1$, $\alpha \in (1,2]$) $\mathbb{G}_\delta^{1,2}(\alpha, \beta)$ has $O(\beta^{2-1/\alpha}\delta^{-4+1/\alpha})$ nodes, each with degree at most 2592.

As before, to each vertex $(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_d)$ we associate a region around the image of $(p_{\mathbf{m},\mathbf{h}_1}, \ldots, p_{\mathbf{m},\mathbf{h}_d})$ restricted to $I_{\mathbf{m}}$, of thickness given by the error bound of Lemma 2.2:

$$R(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_d) = \{\mathbf{z} \in [0,1]^d : \exists \mathbf{x} \in [0,1]^k, \forall r = 1, \ldots, d, \ |z_r - p_{\mathbf{m},\mathbf{h}_r}(\mathbf{x})| \leq c_0\delta\}.$$

Let $\Pi_\delta^{k,d}(\alpha, \beta)$ denote the set of connected components of $\mathbb{G}_\delta^{k,d}(\alpha, \beta)$, homeomorphic to the square grid $\{1, \ldots, \Delta^{-1}\}^k$.

Theorem 2.7. *For each* $f \in \mathcal{H}^{k,d}(\alpha, \beta)$, *there is a connected component* $\pi \in \Pi_\delta^{k,d}(\alpha, \beta)$ *such that* $\text{im}(f) \subset R(\pi)$.

Proof. The proof follows that of Theorem 2.5. Details are omitted. □

2.5 Networks organized by scale

In practice, the parameters α and β are often unknown, so that it becomes necessary and/or useful to look through a discrete set. We introduce a scale parameter and another parameter indexing the approximation order, and organize the graphs accordingly. Fix $J \in \mathbb{N}$ as the maximum scale and a sequence $a_J \to \infty$ as $J \to \infty$. At scale $j \in \{0, \ldots, J\}$ and approximation order $\iota \in \mathbb{N}$, let $\mathbb{G}_{j,J}^k(\iota)$ (resp. $\mathbb{G}_{j,J}^{k,d}(\iota)$) be defined as $\mathbb{G}_\delta^k(\alpha, \beta)$ (resp. $\mathbb{G}_\delta^{k,d}(\alpha, \beta)$) with $\lfloor\alpha\rfloor = \iota$, $\Delta = 2^{-j}$, $\delta = 2^{j-J}$ and $|h^{(\mathbf{s})}\delta_{|\mathbf{s}|}| \leq a_J^{|\mathbf{s}|+1}$.

We have the following corollary of Theorems 2.5 and 2.7.

Proposition 2.8. *For* $\alpha, \beta > 0$, *let* $j = j(\alpha, \beta) = \lceil\frac{J+\log_2(c_1\beta)}{1+\alpha}\rceil$. *Assume J is large enough that* $j(\alpha, \beta) \leq J$ *and* $a_J \geq \beta$. *Also, take* $\iota = \lfloor\alpha\rfloor$. *Then the covering result of Theorem 2.5 (resp. Theorem 2.7) applies with* $\mathbb{G}_{j,J}^k(\iota)$ *(resp.* $\mathbb{G}_{j,J}^{k,d}(\iota)$*).*

Proof. This is a simple corollary of Theorems 2.5 and 2.7. □

2.5.1 A single multiscale network

The networks $\mathbb{G}_{j,J}^k(\iota)$, $j = 1, \ldots, J$, $\iota \in \mathbb{N}$, constitute a family of monoscale networks. Instead, one may want to mix scales together (and possibly mix approximation orders too, though not done here) so as to better approximate functions with varying smoothness, akin to how Besov functions are decomposed into a sum of wavelets at various scales [46]. Specifically, consider $\mathcal{F}^k(\iota)$ to be the class of ι-times continuously differentiable functions of the form $f = \sum_{p \in P} f_p \chi_p$, where P is a finite partition of $[0,1]^k$ into regions with boundaries of finite length and $f_p \in \mathcal{H}^k(\alpha_p, \beta_p)$ with $\alpha_p \in (\iota, \iota + 1]$ and $\beta_p > 0$. For such a function class, we may want to involve a number of scales, each adapted to a different smoothness degree.

In particular, such a multiscale approximating network may be built out of the union of $\mathbb{G}_{j,J}^k(\iota)$, $j = 1, \ldots, J$, with additional edges between nodes at different scales. The neighboring condition across scales may be chosen to be identical to that defined in Section 2.3, namely that $(\mathbf{m}, \mathbf{h}) \in \mathbb{G}_{j,J}^k$ and $(\mathbf{m}_\star, \mathbf{h}_\star) \in \mathbb{G}_{j_\star,J}^k$ are neighbors if $I_{\mathbf{m}}$ and $I_{\mathbf{m}_\star}$ are adjacent, and if $p_{\mathbf{m},\mathbf{h}}$ and $p_{\mathbf{m}_\star,\mathbf{h}_\star}$ together with their derivatives assume nearby values both at $\mathbf{x}_{\mathbf{m}}$ and $\mathbf{x}_{\mathbf{m}_\star}$. (The condition translates into a precise statement involving $\mathbf{h}$ and $\mathbf{h}_\star$ akin to (7), yet more cumbersome. We omit details.) Let $\mathbb{G}_J^k(\iota)$ denote this multiscale network.

Given a function $f \in \mathcal{F}^k(\iota)$, $f = \sum_{p \in P} f_p \chi_p$, we use a recursive dyadic partitioning (RDP), the cornerstone of many multiscale algorithms [44, 23, 22, 39], to approximate the partition P. Specifically, we start at $j = 0$ and then recursively subdivide each dyadic hypercube S at scale j until $j \geq j(\alpha_p, \beta_p)$ for all $p \in P$ with $|p \cap S| > 0$. We then use a polynomial piece from $\mathbb{G}_{j,J}^k(\iota)$ within each RDP cell at scale j. See Figure 4.

For a subset of nodes π in $\mathbb{G}_J^k(\iota)$, define $R(\pi) = \bigcup_{(\mathbf{m},\mathbf{h}) \in \pi} R(\mathbf{m}, \mathbf{h})$, where each region is defined with the appropriate scale.

Proposition 2.9. *For each function $f \in \mathcal{F}^k(\iota)$, there is a connected component π within $\mathbb{G}_J^k(\iota)$ in correspondence with an RDP such that $\mathrm{graph}(f) \subset R(\pi)$.*

Proof. This is essentially a corollary of Proposition 2.8. Details are omitted. $\qquad\square$

Note however that the typical degree of a node in $\mathbb{G}_J^k(\iota)$ increases with J, as a result of connecting nodes across scales.

3   Detection of Objects in Point Clouds and Images

We consider a simple model of detection of objects in point clouds and design some algorithms based on the networks introduced in Section 2. The objects are assumed to be of Hölder type. Fix an Hölder class $\mathcal{H}^{k,d-k}(\alpha, \beta)$. For $f \in \mathcal{H}^{k,d-k}(\alpha, \beta)$ and $\eta \geq 0$, define

$$\mathrm{graph}_\eta(f) = \{(\mathbf{x}, \mathbf{z}) \in [0,1]^d : \|\mathbf{z} - f(\mathbf{x})\| \leq \eta\},$$

which is a region centered around the graph of f and of thickness 2η.

Suppose we observe a point cloud $X_1, \ldots, X_n \in [0,1]^d$, and want to decide between the following two hypotheses (generative models):

$$
\begin{aligned}
H_0: \quad & X_1, \ldots, X_n \sim^{\mathrm{iid}} \quad \mathrm{Uniform}[0,1]^d; \\
H_1: \quad & X_1, \ldots, X_n \sim^{\mathrm{iid}} \quad (1 - \varepsilon_n)\mathrm{Uniform}[0,1]^d + \varepsilon_n \mathrm{Uniform}(\mathrm{graph}_\eta(f^*)), \\
& \qquad\qquad \text{for some (unknown) } f^* \in \mathcal{H}^{k,d-k}(\alpha, \beta).
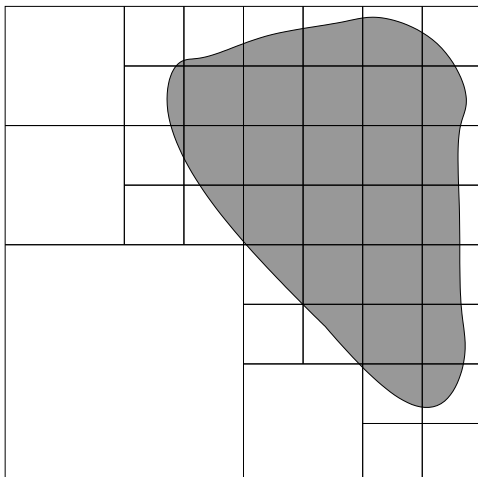\end{aligned}
$$

Figure 4: A partition into two regions p_1 (white) and p_2 (grey) associated with some function $f = f_{p_1}\chi_{p_1} + f_{p_2}\chi_{p_2}$, with corresponding scales $j(\alpha_{p_1}, \beta_{p_1}) = 1$ and $j(\alpha_{p_2}, \beta_{p_2}) = 3$.

The same situation was considered in [7, 5, 8].

For a (measurable) set $S \subset [0,1]^d$, let $N(S)$ denote the number of data points belonging to S, which under the null has the binomial distribution with parameters n and $|S|_d$, the d-dimensional Lebesgue measure of S:

$$N(S) = \#\{i : X_i \in S\} \sim^{H_0} \text{Bin}(n, |S|_d).$$

3.1 Generalized Likelihood Ratio Test

If f^* were known, the most powerful test would be the likelihood ratio (i.e. Neyman-Pearson) test, which rejects for large values of $N(\text{graph}_\eta(f^*))$. The scan statistic is the maximum over those statistics:

$$M_\eta^{k,d}(\alpha, \beta) = \max_{f \in \mathcal{H}^{k,d}(\alpha,\beta)} N(\text{graph}_\eta(f)).$$

The generalized likelihood ratio test (GLRT) rejects when $M_\eta^{k,d}(\alpha, \beta)$ is large.

Define

$$\rho(k, d, \alpha) = \frac{k}{k + \alpha(d - k)}.$$

Theorem 3.1. *There are constants $A, B > 0$ not depending on n such that*

$$\mathbf{P}\left\{A(n^\rho \vee \eta^{d-k}n) \le M_\eta^{k,d}(\alpha, \beta) \le B(n^\rho \vee \eta^{d-k}n)|H_0\right\} \to 1, \quad n \to \infty.$$

Proof. The lower bound is obtained by interpolation of carefully selected points, while the upper bound is obtained using a precise-enough net for $\mathcal{H}^{k,d-k}(\alpha, \beta)$ of near-optimal entropy (such as introduced in Section 2). We refer the reader to Section 2.3 in [5] for more details. □

Since $M_\eta^{k,d}(\alpha, \beta) \ge N(\text{graph}_\eta(f^*))$ and $\text{graph}_\eta(f^*)$ contains at least an ε_n proportion of the point cloud (roughly), the GLRT asymptotically separates H_0 and H_1 if, for some fixed

$B' > B$, $\varepsilon_n \geq B'(n^\rho \vee \eta^{d-k}n)$, meaning that both the probabilities of false alarm (type I error) and missed detection (type II error) tend to zero as n increases.

The GLRT as defined above is challenging, if not impossible to compute exactly. We use instead an approximation based on the coverings constructed in Section 2, turning an optimization over a functional space into an optimization over paths in a network, for which a large number of algorithms have been developed. See also [34, 12], where variational problems related to computing minimal surfaces are turned into combinatorial optimizations over paths and other structures within networks.

Recall the notation used in Section 2, and assume again that Δ and δ are related according to (6). We use the network $\mathbb{G}_\delta^{k,d-k}(\alpha, \beta)$ to build appropriate coverings, this time with slightly enlarged regions:

$$R_\delta(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) = \{(\mathbf{x}, \mathbf{z}) : \mathbf{x} \in I_\mathbf{m}, \forall r = 1, \ldots, d-k, \ |z_r - p_{\mathbf{m}, \mathbf{h}_r}(\mathbf{x})| \leq (c_0 + 1)\delta\},$$

where $c_0 = (2^{-\alpha} + c_2/2)$ as in (8).

Fix a path $\{\mathbf{m}_{\mathrm{zz}}(t) : t = 1, \ldots, \Delta^{-k}\}$ in the square grid $\{1, \ldots, \Delta^{-1}\}^k$ covering the whole grid in zig-zag fashion, and define $\mathcal{P}_\delta^{k,d-k}(\alpha, \beta)$ as the set of paths in $\mathbb{G}_\delta^{k,d-k}(\alpha, \beta)$ of the form

$$\{(\mathbf{m}_{\mathrm{zz}}(t), \mathbf{h}_1(t), \ldots, \mathbf{h}_{d-k}(t)) : t = 1, \ldots, \Delta^{-k}\}.$$

Also, recall the definition of $\Pi_\delta^{k,d-k}(\alpha, \beta)$ in Section 2. Now consider the following alternative statistics:

$$M_{\Pi,\delta}^{k,d}(\alpha, \beta) = \max_{\pi \in \Pi_\delta^{k,d-k}(\alpha,\beta)} N(R_\delta(\pi)), \qquad M_{\mathcal{P},\delta}^{k,d}(\alpha, \beta) = \max_{P \in \mathcal{P}_\delta^{k,d-k}(\alpha,\beta)} N(R_\delta(P)).$$

In that case,
$$M_\eta^{k,d}(\alpha, \beta) \leq M_{\Pi,\delta}^{k,d}(\alpha, \beta) \leq M_{\mathcal{P},\delta}^{k,d}(\alpha, \beta).$$

The first inequality comes from Theorem 2.5 and the triangle inequality; the second from the fact that $\Pi_\delta^{k,d-k}(\alpha, \beta) \subset \mathcal{P}_\delta^{k,d-k}(\alpha, \beta)$. Actually, by Theorem 3.2 below, with a proper choice for δ all three are of same order of magnitude with high probability.

Theorem 3.2. *With $\delta = n^{-\alpha/(k+\alpha(d-k))} \vee \eta$, there is a constant $C = C(k, d, \alpha, \beta)$ such that*

$$\mathbf{P}\left\{M_{\mathcal{P},\delta}^{k,d}(\alpha, \beta) \leq C(n^\rho \vee \eta^{d-k}n)|H_0\right\} \to 1, \quad n \to \infty.$$

Proof. The proof of Theorem 3.2 is in the Appendix. $\qquad\square$

Computing $M_{\mathcal{P},\delta}^{k,d}(\alpha, \beta)$ may be done efficiently using dynamic programming ideas, for example as implemented in [15]; this is due to the fact that the paths in $\mathcal{P}_\delta^{k,d-k}(\alpha, \beta)$ are oriented and with no loops. Though $M_{\Pi,\delta}^{k,d}(\alpha, \beta)$ provides a better approximation to the scan statistic, we do not know of an efficient way to compute it directly.

The results above hold for Hölder immersions as well, though in that case the computations are much more challenging. This comes from the fact that Hölder immersions may self-intersect, so that dynamic programming approaches do not apply. In fact, if the optimization is over all paths of length Δ^{-k} instead, the setting is equivalent to the Budget-Reward Problem [17] (also called Bank Robber Problem), closely related to the Prize-Collecting

Traveling Salesman problem [11]; though those problems are known to be NP-hard, there are polynomial-time approximations [17]. Other approaches have been suggested in this situation, for example in [23], where ratios of additive criteria are used to recover chains of beamlets; or algorithms implemented to extract curves from saliency networks [51, 42].

3.2   Longest Significant Run

We propose an alternative approach based on the size of the longest path after discarding nodes with low counts within their associated region, which in effect generalizes the algorithm introduced in [7].

For a threshold $\tau > 0$, define

$$S(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) = \begin{cases} 1, & N(R_\delta(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k})) > \tau n \delta^{k/\alpha + d - k}, \\ 0, & \text{otherwise.} \end{cases}$$

Then define $L_\tau^{k,d}(\alpha, \beta)$ as the length of the longest path of the form

$$\{(\mathbf{m}_{\mathrm{zz}}(t), \mathbf{h}_1(t), \ldots, \mathbf{h}_{d-k}(t)) : t = t_0, \ldots, t_0 + \ell - 1\},$$

such that $S(\mathbf{m}_{\mathrm{zz}}(t), \mathbf{h}_1(t), \ldots, \mathbf{h}_{d-k}(t)) = 1$ for all $t = t_0, \ldots, t_0 + \ell - 1$. If we see each $S(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k})$ as a test at node $(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k})$ and say that it is significant if it equals 1, then $L_\tau^{k,d}(\alpha, \beta)$ is the length of the longest significant run (LSR).

Theorem 3.3. *With $\delta = n^{-\alpha/(k+\alpha(d-k))} \vee \eta$, for τ large enough,*

$$\mathbf{P}\left\{L_\tau^{k,d}(\alpha, \beta) \leq \log(1/\delta)|H_0\right\} \to 1, \quad n \to \infty.$$

Also, there is a constant $C = C(k, d, \alpha, \beta)$ such that, if $\varepsilon > Cn^\rho \vee \eta^{d-k}n$,

$$\mathbf{P}\left\{L_\tau^{k,d}(\alpha, \beta) > \log(1/\delta)|H_1\right\} \to 1, \quad n \to \infty.$$

Proof. The proof of Theorem 3.3 is in the Appendix. □

Therefore, the LSRT achieves the detection rate established for the GLRT (Theorem 3.1) and its approximation (Theorem 3.2). Moreover, it can be computed using dynamic programming ideas since again the paths considered are oriented and without loops. This is done in [7].

The same approach applies essentially unchanged to the case of Hölder immersions, though with pathological cases obscuring the exposition, so that we omit details.

A similar approach is advocated in [48], where a network of streams is monitored for pollution levels and large connected components of areas marked as problematic (polluted) are of interest. Two of the same researchers suggest a hybrid method in [49] applied to the identification of regions of interest (hot spots) in raster maps.

3.3 Detection in Grey-Level Images

The results we developed for point clouds can be obtained (in slightly different form) for digitized images, which is the context for the experiments of Section 5.

Suppose we observe a d-dimensional pixel array Y, with a total of n pixels, of the form:

$$Y = \mu \, \xi_{\mathrm{graph}_\eta(f^*)} + \sigma Z,$$

where Z is white Gaussian noise, with independent, standard normal entries; $\sigma > 0$ is the noise level; μ is the signal level; and for a subset $S \subset [0,1]^d$, ξ_S is the array with ℓ^2-norm 1 identifying the pixels that S intersects, namely $\xi_S(\mathbf{i}) \propto \mathbf{1}\{S \cap \mathrm{pixel}(\mathbf{i}) \neq \emptyset\}$.

We observe Y and want to decide between H_0 and H_1 below:

$$
\begin{aligned}
H_0: \quad & \mu = 0; \\
H_1: \quad & \mu > 0, \text{ and } f^* \in \mathcal{H}^{k,d-k}(\alpha, \beta) \text{ is unknown.}
\end{aligned}
$$

This setting is considered in [6] (with a slightly different definition for ξ_S) in the context of parametric objects.

Following the same arguments as for point clouds, we find that the GLRT asymptotically separates H_0 and H_1 if $\mu \geq C(n^{k/(2\alpha d)} \vee \eta^{-k/(2\alpha)})$ with C large enough, and that a detection threshold of same order of magnitude is achieved both by the approximate GLRT and the LSRT, with $\delta = \eta \vee n^{-1/d}$. We omit details.

4 Beamlet and Beamlet-Like Networks

We now focus on curves ($k = 1$) of Hölder smoothness with $\alpha \in (1, 2]$. The corresponding net described in Section 2 is made of piecewise linear functions. In this section, we show that such a net may be obtained by carefully chaining beamlets as suggested in [23], yielding a more economical net at a comparable degree of approximation.

The case of curves is special in the sense that it is the simplest, and in particular allows us to use a parametrization by arclength. Higher dimensional surfaces, even though smooth, may exhibit strange behavior, for example a very thin 2D surface in 3D.

4.1 Hölder Curves

We adopt here a slightly more intrinsic definition for curves. For $\alpha \in (1, 2]$ and $\lambda, \kappa > 0$, let $\Gamma(\alpha, \lambda, \kappa)$ be the set of curves $\gamma \subset [0,1]^d$ with $\mathrm{length}(\gamma) \leq \lambda$ and parametrized by arclength with

$$\|\gamma(t) - \gamma(s) - (t-s)\gamma'(s)\| \leq \kappa \, |t-s|^\alpha, \quad \forall s, t \in [0, \mathrm{length}(\gamma)]. \tag{9}$$

$\Gamma(\alpha, \lambda, \kappa)$ is in close correspondence with $\mathcal{H}^{1,d}(\alpha, \beta)$ as defined in Section 2. Note that the case $\alpha = 2$ includes all twice differentiable curves with curvature bounded by 2κ.

Curves in $\Gamma(\alpha, \lambda, \kappa)$ satisfy the following properties.

Lemma 4.1. *For all $\gamma \in \Gamma(\alpha, \lambda, \kappa)$,*

$$\|\gamma'(t) - \gamma'(s)\| \leq 2\kappa |t-s|^{\alpha-1}, \quad \forall s, t.$$

Proof. Fix $0 \leq s < t \leq \text{length}(\gamma)$. The triangle inequality and (9) give

$$
\begin{aligned}
(t-s)\|\gamma'(t) - \gamma'(s)\| &\leq \|\gamma(t) - \gamma(s) - \gamma'(s)(t-s)\| + \|\gamma(s) - \gamma(t) - \gamma'(t)(s-t)\| \\
&\leq 2\kappa(t-s)^\alpha.
\end{aligned}
$$

$\square$

Lemma 4.2. *Let $\gamma \in \Gamma(\alpha, \lambda, \kappa)$. For all arclengths $r < s < t$,*

$$
\left\| \gamma(s) - \gamma(r) - \frac{s-r}{t-r}\left(\gamma(t) - \gamma(r)\right) \right\| \leq 2\kappa(t-r)^\alpha.
$$

Proof. Applying (9) twice yields

$$
\|\gamma(s) - \gamma(r) - (s-r)\gamma'(r)\| \leq \kappa(s-r)^\alpha \leq \kappa(t-r)^\alpha
$$

and

$$
\left\| \frac{s-r}{t-r}\left(\gamma(t) - \gamma(r)\right) - (s-r)\gamma'(r) \right\| \leq \kappa\,(s-r)(t-r)^{\alpha-1} \leq \kappa(t-r)^\alpha.
$$

Then apply the triangle inequality and conclude. $\square$

4.2 Beamlets

Beamlets were introduced in 2D by Donoho and Huo [23], and then in 3D by Donoho and Levi [25, 20]. We define them here in any dimension $d \geq 2$. Fix a maximum scale $J \in \mathbb{N}$. Define $\delta_0 = 2^{-J}$ and at scale $j \in \{0, \ldots, J\}$, define $\Delta = 2^{-j}$. For a given coordinate $r = 1, \ldots, d$, hyperplanes of the form

$$
\{(x_1, \ldots, x_d) : x_r = h\Delta\},
$$

where $h \in \{0, \ldots, \Delta^{-1}\}$, are called r-hyperplanes. Such hyperplanes will be called Δ-hyperplanes; they partition the unit hypercube $[0,1]^d$ into smaller hypercubes of sidelength Δ that we call Δ-hypercubes. On each Δ-hyperplane, we consider a regular square grid with spacing δ_0. Formally, we consider gridpoints of the form $(h_1\delta_0, \ldots, h_d\delta_0)$, $h_r = 0, 1, \ldots, \delta_0^{-1}$, with at least one coordinate an integer multiple of $\Delta\delta_0^{-1}$; if this happens at the rth coordinate, we speak of an r-gridpoint, which by definition belongs to an r-hyperplane. A beamlet is simply a line-segment joining two gridpoints belonging to the same Δ-hypercube. See Figures 5.

The beamlet graph in [23] refers to the graph with nodes the gridpoints and edges the beamlets. Our interest instead is in building a beamlet good-continuation graph (i.e. network), with nodes the beamlets themselves and edges between beamlets in good-continuation, such that paths in that network provide a useful net for smooth curves. Such a graph is used in some experiments performed in [23], yet never formally defined there or elsewhere, though a related construction is presented in [7] for curves that are graphs of Hölder functions. Formally, two beamlets are neighbors if they are of the form $[b_1, b_2]$ and $[b_2, b_3]$ and satisfy

$$
\begin{aligned}
\|(b_{r,2} - b_{r,1})(b_3 - b_2) - (b_{r,3} - b_{r,2})(b_2 - b_1)\| & \\
\leq 2^{j-J}(\|b_3 - b_2\| + \|b_2 - b_1\|), \quad \forall r = 1, \ldots, d. \quad (10)
\end{aligned}
$$

(a) Coarsest scale, j=0   (b) Next finer scale, j=1

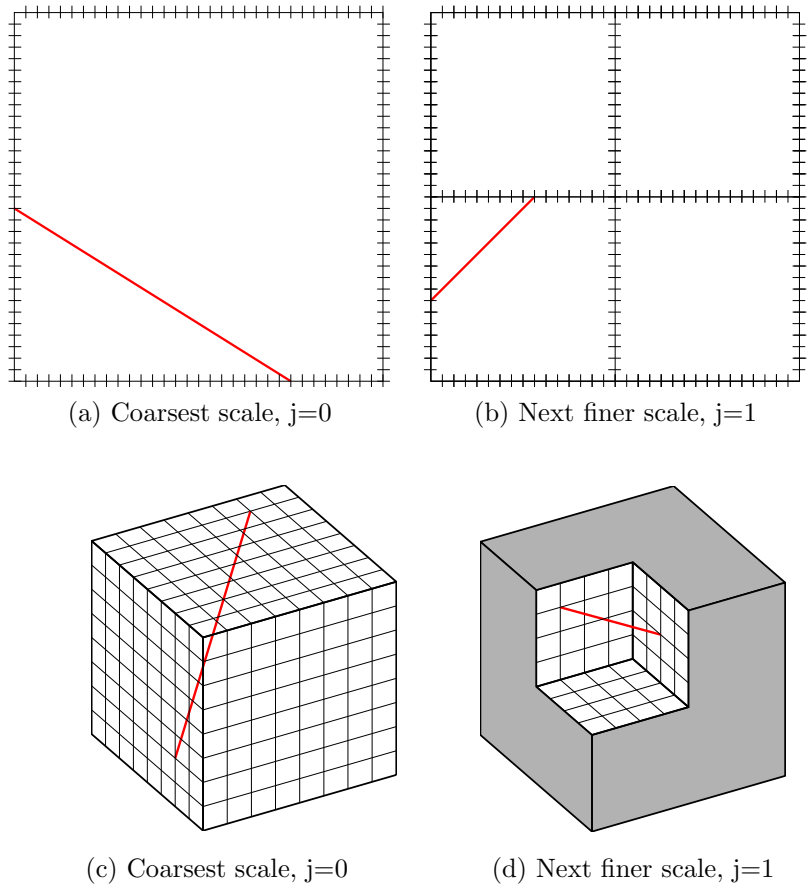(c) Coarsest scale, j=0   (d) Next finer scale, j=1

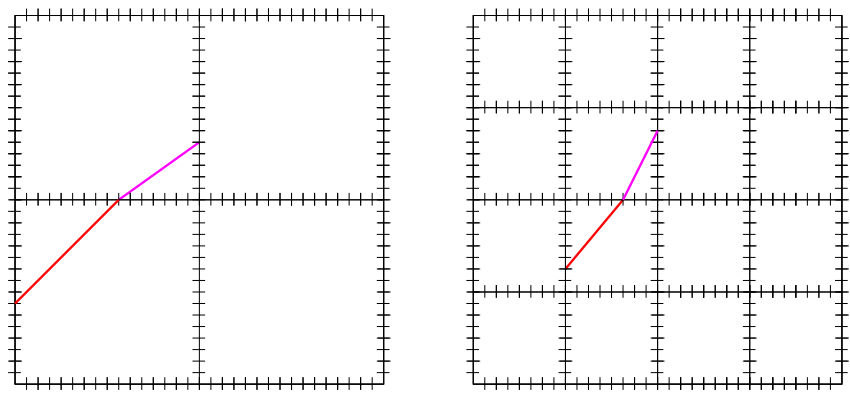Figure 5: Examples of beamlets in 2D and 3D.

Figure 6: Examples of 2D beamlets in good-continuation.

This is basically a constraint on the angle formed by $[b_1, b_2]$ and $[b_2, b_3]$; see Figure 6.

Let $\mathbb{B}^d_{j,J}$ denote the resulting beamlet good-continuation network at scale j.

Lemma 4.3. *The number of beamlets at scale j is of order $\mathbb{B}^d_{j,J} = O(d^2 2^{2(d-1)J-(d-2)j})$ and a given beamlet $B \in \mathbb{B}^d_{j,J}$ of length $|B|$ has $O(d|B|^{-(d-1)})$ neighbors.*

Proof. The proof of Lemma 4.3 is in the Appendix. □

Remark. Short beamlets in the network $\mathbb{B}^d_{j,J}$ are highly connected. This is an undesirable feature and we were not able to avoid it, other than defining a closely related system in Section 4.3.

To each beamlet B we associate a tubular region:

$$R(B) = \{\mathbf{x} \in [0,1]^d : \min_{\mathbf{y} \in B} \|\mathbf{x} - \mathbf{y}\| \leq 2^{j-J}\}.$$

We extend this definition to subsets of beamlets $R(\pi) = \bigcup_{B \in \pi} R(B)$.

Theorem 4.4. *Fix $\alpha \in (1,2]$ and $\lambda, \kappa > 0$. Let $j(\alpha, \kappa) = \left\lceil \frac{\log_2(\kappa) + J}{\alpha+1} \right\rceil$. There is a universal constant K such that, when J is large enough and $j \geq j(\alpha, \beta) + K$, to each curve $\gamma \in \Gamma(\alpha, \lambda, \kappa)$ corresponds a path π_j in $\mathbb{B}^d_{j,J}$ chaining at most $(2d)(\lambda 2^j + 2)$ beamlets such that γ is included in $R(\pi_j)$.*

Proof. Theorem 4.4 is a consequence of Theorem 4.6; see the comments following Theorem 4.6. □

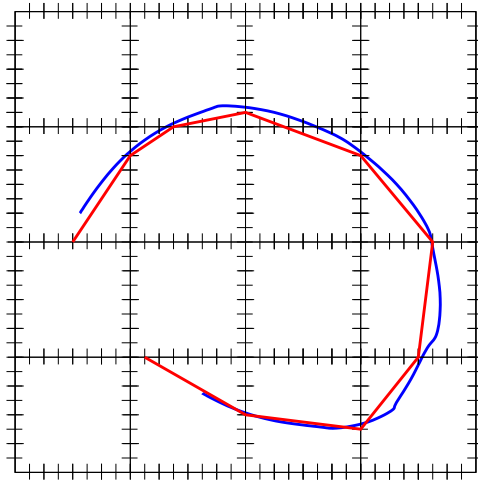Such a covering is illustrated in Figure 7.

Figure 7: Example of a curve (blue) approximation by a chain of 2D beamlets (red) in good continuation.

4.2.1   Multiscale beamlet good-continuation network

The chaining algorithms proposed in [23] involve linking beamlets at different scales. Such a multiscale network may simply be constructed by taking the union of all $\mathbb{B}^d_{j,J}, j = 0, \ldots, J$ and adding edges between beamlets at different scale satisfying (10) with j replaced by the maximum of the two scales involved; see Figure 8. As in Section 2.5.1, an approximation is built using an RDP, this time of the ambient space $[0, 1]^d$. Such a network provides more precise approximations to curves with varying smoothness, however at the cost of (substantially) increasing connectivity.
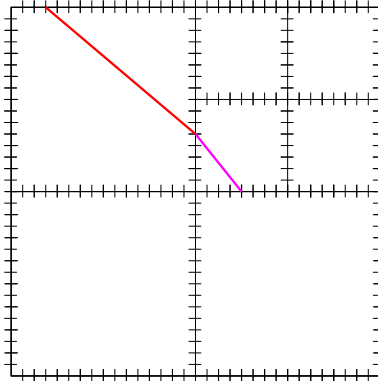

Figure 8: Example of 2D beamlets at different scales in good continuation.

4.3   Beams

We found in Lemma 4.3 that defining a notion of good continuation directly between beamlets is problematic in that small corner beamlets become hubs, connected to a large number of other beamlets. The construction presented here uses other line-segments akin to beamlets, that we call beams. (Note that the term 'beam' refers to something else in [23].) Beams at a given scale are of similar length, thus avoiding the problem just mentioned, and each beam can be well approximated by a short chain of beamlets at that same scale.

Another drawback of beamlets as defined in Section 4.2 is that the quantization by δ_0 does not change with the scale as Δ does; this results in a system that is more rich than needed, therefore wasteful. Just as in Section 2.5, at scale j we define $\delta = 2^{j-J}$, and $\Delta = 2^{-j}$ as before; we assume that $j \leq J/2$, so that Δ is an integer multiple of δ.

We define two kinds of beams. For $r = 1, \ldots, d$, an r-beam is a line segment joining two r-gridpoints b_1 and b_2 belonging to the same Δ-hypercube on opposite sides, and such that

$$\|b_1 - b_2\| \leq \Delta + \delta. \tag{11}$$

Therefore, an r-beam makes an angle of about 45 degrees or less with the rth axis.

For $r_1, r_2 = 1, \ldots, d$, with $r_1 \neq r_2$, an $r_1 r_2$-beam is a line segment joining an r_1-gridpoint

b_1 and a r_2-gridpoint b_2 such that

$$\Delta \leq |b_{r_1,1} - b_{r_1,2}| \vee |b_{r_2,1} - b_{r_2,2}| < 2\Delta, \tag{12}$$

$$| \, |b_{r_1,1} - b_{r_1,2}| - |b_{r_2,1} - b_{r_2,2}| \, | \leq \delta, \tag{13}$$

$$(|b_{r_1,1} - b_{r_1,2}| \vee |b_{r_2,1} - b_{r_2,2}|) - |b_{r_3,1} - b_{r_3,2}| \geq -\delta, \quad \forall r_3 = 1, \ldots, d. \tag{14}$$

Note that b_1 and b_2 do not belong to the same Δ-hypercube and that (14) is void in dimension $d = 2$. Hence, an $r_1 r_2$-beam connects an r_1-hyperplane and a r_2-hyperplane, making angles of about 45 degrees at the intersection with those hyperplanes. The reason $r_1 r_2$-beams are so restricted is that they are only used to connect r_1-beams and r_2-beams.

(a) $r2$-beam ($j = 0$)   (b) $r1$-beam ($j = 1$)   (c) $r1r2$-beam ($j = 2$)
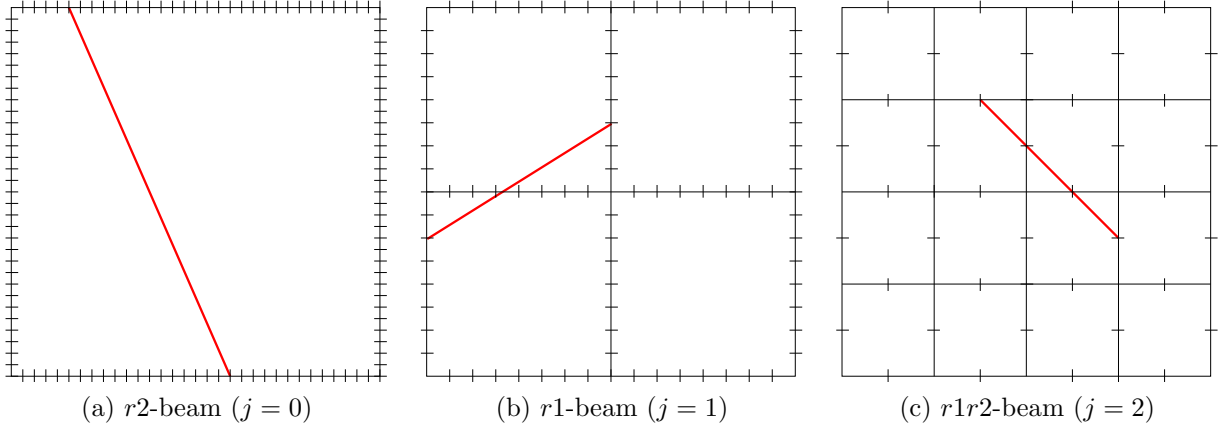
Figure 9: Example of 2D beams.

We define neighborhood relationships between beams as we did for beamlets in (10); see Figure 10. Specifically, two beams are neighbors if they are of the form $[b_1, b_2]$ and $[b_2, b_3]$ and satisfy

$$\|(b_{r,2} - b_{r,1})(b_3 - b_2) - (b_{r,3} - b_{r,2})(b_2 - b_1)\|$$
$$\leq (11\delta/20)(\|b_3 - b_2\| + \|b_2 - b_1\|), \quad \forall r = 1, \ldots, d. \tag{15}$$

For example, suppose $[b_1, b_2]$ and $[b_2, b_3]$ are both r_1-beams. If b_3^* is the intersection of the line (b_1, b_2) with the r_1-hyperplane b_3 belongs to, then condition (15) implies $\|b_3 - b_3^*\| < (5/2)\delta$, so that b_3 is among the 5^{d-1} r_1-gridpoints closest to b_3^*.

Let $\overline{\mathbb{B}}_{j,J}^d$ denote the resulting beam good-continuation network at scale j.

Lemma 4.5. *The number of nodes in $\overline{\mathbb{B}}_{j,J}^d$ is of order $O(d2^{2(d-1)J-(3d-4)j}(1+d2^{2j-J}))$. Moreover, all nodes have at most $2d \cdot 7^{d-1}$ neighbors, with most nodes having at most $2 \cdot 5^{d-1}$ neighbors.*

Proof. The proof of Lemma 4.5 is in the Appendix. □

Just as we did for beamlets, to each beam B we associate a tubular region:

$$R(B) = \{\mathbf{x} \in [0,1]^d : \min_{\mathbf{y} \in B} \|\mathbf{x} - \mathbf{y}\| \leq \delta\}.$$
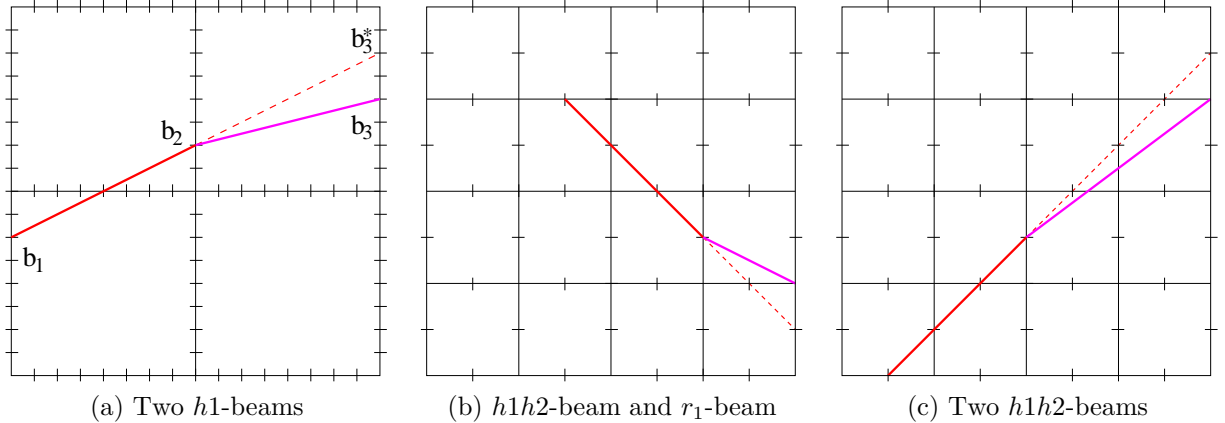
(a) Two $h1$-beams　　　(b) $h1h2$-beam and r_1-beam　　　(c) Two $h1h2$-beams

Figure 10: Example of 2D beams in good continuation.

Theorem 4.6. *Fix $\alpha \in (1,2]$ and $\lambda, \kappa > 0$. There is a universal constant K such that, when J is large enough and $j \geq j(\alpha, \beta) + K$, to each curve $\gamma \in \Gamma(\alpha, \lambda, \kappa)$ corresponds a path $\pi_j(\gamma)$ in $\overline{\mathbb{B}}^d_{j,J}$ chaining at most $\lambda 2^j + 2$ beams such that γ is included in $R(\pi_j)$.*

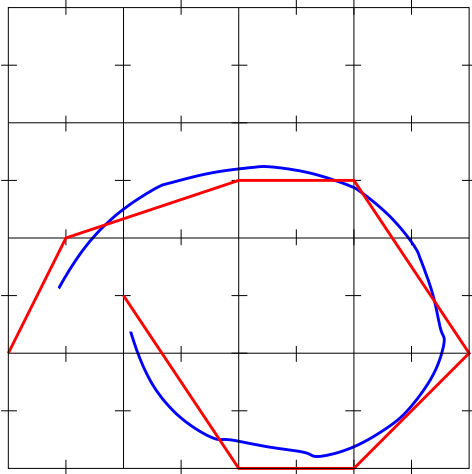*Proof.* The proof of Theorem 4.6 is in the Appendix.　　　□

Figure 11: Example of a curve (blue) approximation by a chain of 2D beams (red) in good continuation.

We now explain how Theorem 4.6 implies Theorem 4.4. First, any beam may be approximated within distance δ_0 by a chain of at most $2d$ beamlets at the same scale. Indeed, a beam touches at most $2d$ Δ-hyperplanes (at most $d+1$ for a typical r-beam); on each one, select a (Δ, δ_0)-gridpoint closest to the beam. By successively connecting those gridpoints with line-segments, a chain of beamlets is born. Therefore, a chain of beams of length at most $\lambda 2^j + 2$ may be approximated by a chain of beamlets of length at most $(2d)(\lambda 2^j + 2)$. That the successive beamlets in such a chain are in good continuation according to (10) comes from the fact that the beams themselves are in good continuation according to (15),

how beamlets are chained to approximate a beam, and the triangle inequality, in parallel to how Claim 3 is established in the proof of Theorem 4.6.

Remark. Both $\overline{\mathbb{B}}_{j,J}^{d}$ and $\mathbb{G}_{j,J}^{1,d}$ provide ε-nets for Hölder curves embedded in the d-dimensional unit hypercube, with ε of order 2^{j-J}. Which one is more economical? In terms of number of nodes, $\log_2(|\mathbb{G}_{j,J}^{1,3}|) \sim 2dJ - (3d-1)j$ (Lemma 2.4), while $\log_2(|\overline{\mathbb{B}}_{j,J}^{d}|) \sim 2(d-1)J - (3d-4)j$ (Lemma 4.5); the latter is smaller for all relevant scales $j \leq J/2$. Perhaps more importantly, $\mathbb{G}_{j,J}^{1,d}$ is substantially more connected, with most nodes with $2 \cdot 36^d$ neighbors (2592 in 2D; 93312 in 3D), compared to $\overline{\mathbb{B}}_{j,J}^{d}$, with most nodes having $2 \cdot 5^d$ neighbors (50 in 2D; 250 in 3D).

5 Numerical Experiments

We perform some numerical experiments showing the power of approximation networks built on good-continuation principles. Specifically, we compare the filamentary content of simulated 3D datasets with a variety of beamlet-based algorithms of our own creation. Using software developed in [40] and a basic notion of good-continuity as introduced in Section 4.2 is enough to outperform simpler algorithms disregarding any spatial information, such as introduced in [20].

We consider three situations illustrated in Figure 12, where each column corresponds to a different case, and for each case the goal is to distinguish between the top and bottom situations. In setting (a), we compare a random point cloud (top) with a set of random filaments of different lengths, orientations and curvatures (bottom). In setting (b), we compare a set of random filaments (top) with a set of random filaments constrained to pass through a small number of hubs (bottom). In setting (c), we compare a set of short random filaments (top) with a set of long random filaments (bottom), all filaments oriented in the direction of the first coordinate.

Each dataset is a pixel array of size 64^3. The images are corrupted with a certain amount of additive white Gaussian noise (see Figure 13), calibrated so the algorithm introduced in [20] (and described in Section 5.4.1) is powerless, i.e. essentially useless at distinguishing between the top and bottom settings. All filaments in these datasets are synthesized using trigonometric functions, each with randomly selected location, amplitude, frequency and phase shift.

Though we introduce a variety of statistics, the workflow is the same:

1. *Compute the beamlet transform.* This amounts to computing all the beamlet coefficients, where a beamlet coefficient is defined as the line-integral of the image along the beamlet [23, 25].

2. *Thresholding of the beamlet coefficients.* Beamlets with a large coefficient are suspected to be intersecting with a filament, and therefore informative since our objective involves detecting filaments. We therefore focus on those beamlets with large coefficients by discarding those with low coefficients [23, 25]. The choice of threshold is not determined in advance, rather a variety of thresholds are considered within some range.
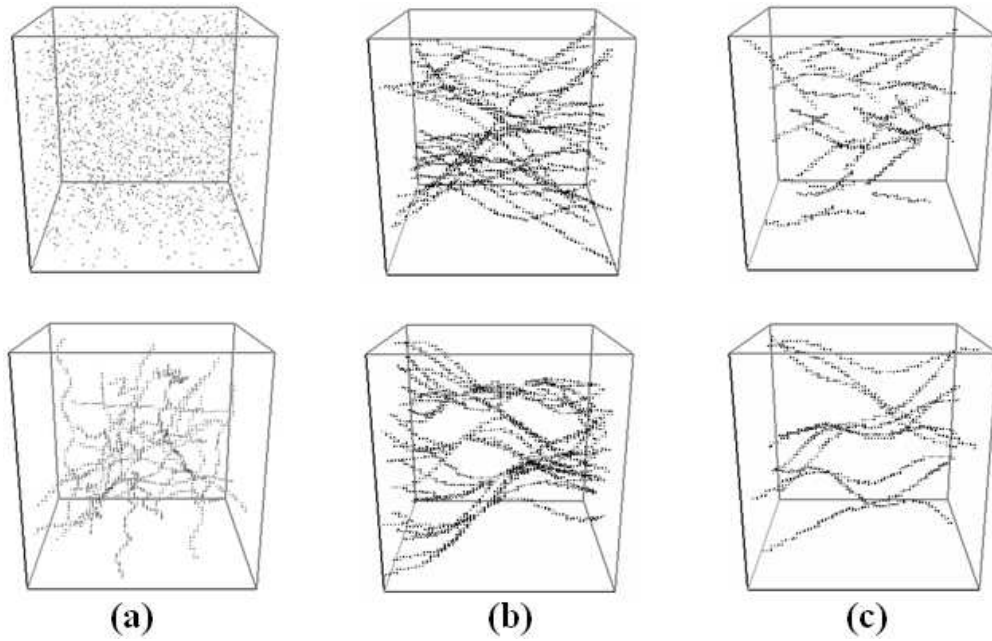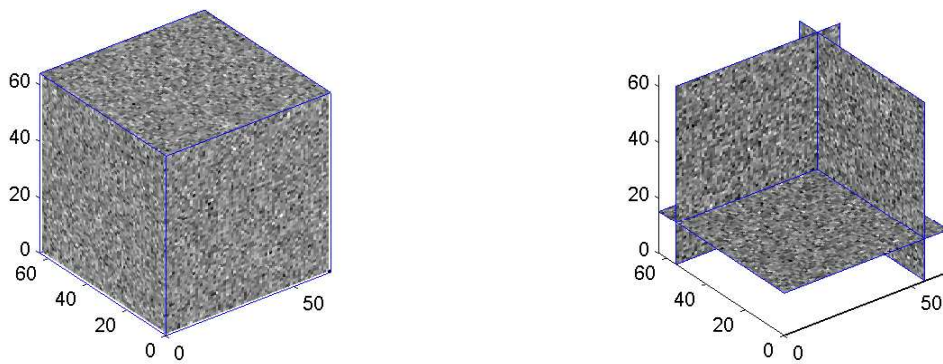
22

Figure 12: Simulated datasets: (a) random point cloud (top) vs. random filaments (bottom); (b) random filaments (top) vs. random filaments with hubs (bottom); (c) random short filaments vs. random long filaments.

(a) Surface of the noisy data volume   (b) Orthogonal slices through the noisy data

Figure 13: Noisy version of the simulated data in the lower panel of column (a) in Figure 12.

3. *Construction of a good-continuation network (GCN) for each scale.* Based only on the beamlets surviving thresholding, neighboring relationships as introduced in Section 4.2 are considered.

4. *Extraction of some relevant network statistics.* We extract a number of statistics from the beamlet network that are sensitive to different degrees and kinds of filamentarity such as node and edge cardinality or connected components counts, centrality measures or even more sophisticated statistics based on path search in the network.

Our software is based on the beamlet transform as implemented in [40] and graph algorithms from the LEDA library [2]. Our code is available online [3].

In what follows, we will say that a test or method is 'powerful' if it faithfully (more than 95% of the time) distinguishes between the top and bottom settings.

5.1   Number of Edges vs. Number of Nodes

Our simplest algorithm looks at how the number of edges and the number of nodes vary together as functions of the threshold used.

Consider column (a) of Figure 12, where we compare a random point cloud (top) with a set of filaments with random lengths, orientations and curvatures (bottom). In Figure 14, the number of edges is plotted as a function of the number of nodes for both top and bottom datasets. As expected, the curve corresponding to the random filaments rests above the curve corresponding to the random point cloud, since the GCN is more connected in the former setting. From our simulation studies, we found that this statistic is not powerful at per pixel signal-to-noise ratios (SNRs) below 0.8 for this specific situation. The dataset with filaments contains 20 of them with random lengths in the range $[10, 64]$ (in number of pixels).
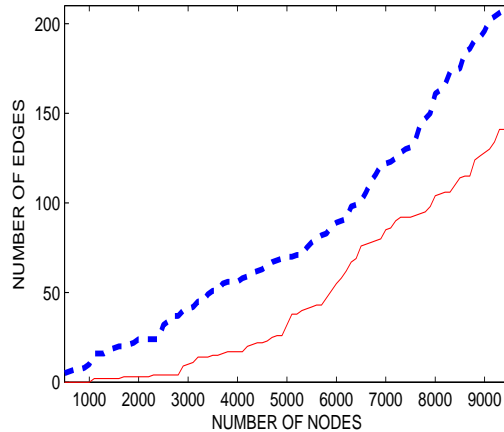

Figure 14: Number of edges vs. number of vertices for the random point cloud (red, solid curve) and the dataset of random filaments (blue, dashed curve) from column (a) of Figure 12.
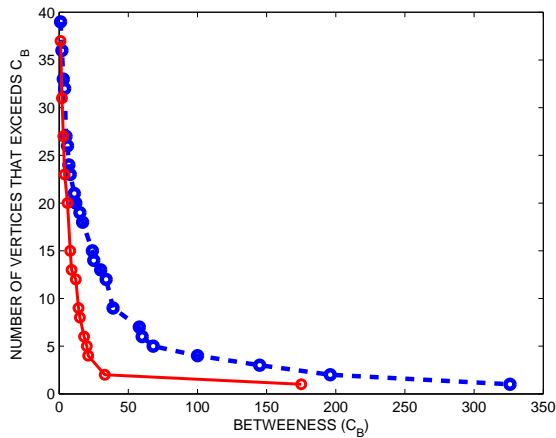
Figure 15: Number of vertices whose betweeness exceeds a given threshold for the dataset of random filaments without hubs (red, solid) and the dataset of random filaments with hubs (blue, dashed) from column (b) of Figure 12.

5.2 Vertex Betweeness

Centrality attributes are well-known from network analysis [47] and used for example in social networks studies [54]. The *betweeness* of vertex v is defined as

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\rho_{st}(v)}{\rho_{st}},$$

where s, t run through all vertices except v, ρ_{st} is the number of longest paths from s to t, and $\rho_{st}(v)$ the number of longest paths from s to t that pass through a vertex v. (Note that shortest paths are sometimes used instead.)

Consider column (b) in Figure 12, where the top image contains randomly distributed filaments and the bottom image also contains random filaments but with hubs. The number of filaments and their individual characteristics are the same in both cases. In Figure 15 we plot the number of vertices with betweeness exceeding a given threshold for both datasets. As expected, the curve corresponding to the random filaments with hubs rests above the curve corresponding to the random filaments without hubs, since in the former case the hubs translate into vertices in the GCN with large betweeness. From our simulation studies, we also found this statistic not powerful at SNRs below 0.8 for the type of filaments chosen here. The data set without hubs (top) contains 20 filaments with horizontal orientations and random lengths in range [60, 63]. The data set with hubs contains 5 groups of 4 filaments with horizontal orientations and length 64, then all 4 filaments in the group have common filamentarity region of length 3 (hub).

Now, consider a slightly different setting with just one large hub, illustrated in the left column of Figure 16. In the right column of the Figure 16 we compare the maximal betweeness of both datasets for different noise levels, which is significantly larger for the dataset with a hub. Again, we found that the maximal betweeness statistic is not effective when the voxel-level SNR is less than 0.8. The data set without hub (top) contains 18 filaments with

horizontal orientations and random lengths in range [48, 64]. The data set with hub contains 18 filaments with horizontal orientations and length 64, then all filaments in the group have common filamentarity region of length 5 (hub).
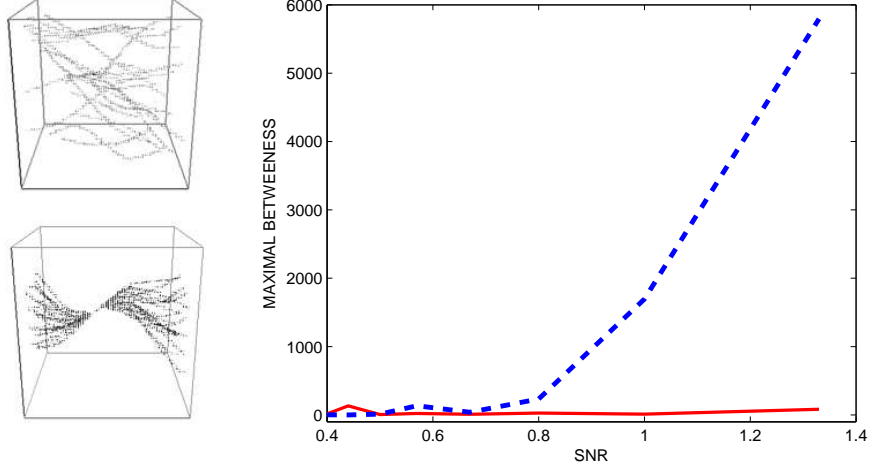

Figure 16: Maximal betweeness as a function of SNR, for the dataset of random filaments without hub (red, solid) and the dataset of random filaments with a hub (blue, dashed) from the left column.

5.3   Filamentarity Survival Index

For each beamlet v in the network, let $\omega(v)$ denote its coefficient (or weight), i.e. the line-integral of the image along this beamlet. The weight of a beamlet path p is then defined as the sum of the weights of the beamlets it passes through:

$$\omega(p) = \sum_{v \in p} \omega(v).$$

Let P be the partitioning of $\mathbb{G}$ into disjoint paths defined recursively as follows:

$$\begin{cases} p_1 = \arg\max_{p \in \mathbb{G}} \omega(p); \\ p_i = \arg\max_{p \in \mathbb{G} \setminus \{p_1, \dots, p_{i-1}\}} \omega(p), & i > 1. \end{cases} \tag{16}$$

We define the *Filamentarity Survival Index* $D(t)$ as the fraction of paths in P with weight greater than t. The computation of the piecewise constant function $D(t)$ is iterative, where at each iteration the *Longest Weighted Path* (LWP) is found and removed from the network. The algorithm terminates when the network is empty.

We next define the *Filamentarity Survival Ratio* (FSR) which compares the filamentary survival index of a given dataset I with the filamentary survival index a random point cloud with same energy, denoted $\tilde{I}$

$$R(t) = \frac{(D_I(t) + \epsilon)}{(D_{\tilde{I}}(t) + \epsilon)},$$

26

where ϵ is a small positive number, used to avoid divide-by-zero situations.

Consider column (c) in Figure 12, where we compare a dataset of random short filaments (top) with a dataset of random long filaments (bottom), all filaments oriented in the direction of the first coordinate. Figure 17 presents the FSR curves for these two datasets. Additionally, this figure contains the FSR curve for the case of purely random points, as in the top image in column (a) in Figure 12. As expected, the FSR curve for the purely random dataset (black, dotted line) remains close to 1. For the dataset containing long filaments, however, the FSR (blue, dashed line) is significantly higher than 1 for certain values of t. For the dataset containing short filaments, the FSR curve (red, solid line) is found between the other two curves. We found that the FSR statistic is powerless at distinguishing between short and long filaments (in this setting) for SNRs below 1 in this specific situation. The data set with short filaments (top) contains 30 filaments with horizontal orientations and lengths 20. The data set with long filaments (bottom) contains 10 filaments with horizontal orientations and lengths 60.
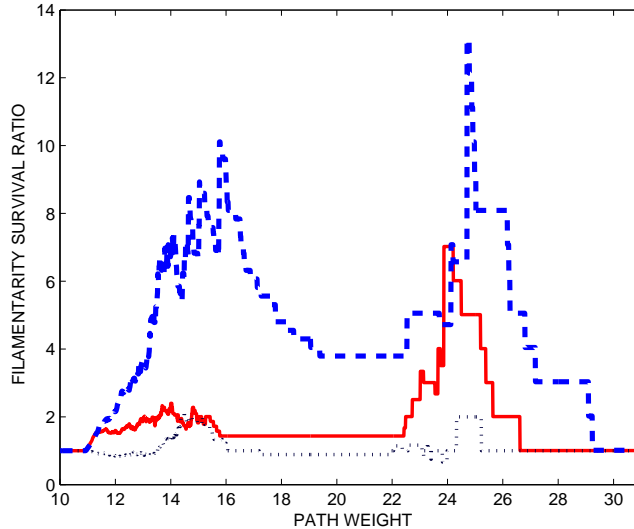
Figure 17: Filamentarity Survival Ratios for the dataset of random short filaments (red, solid) and for the dataset of random long filaments (blue, dashed) in column (c), and for the random point cloud (black, dotted) in column (a) of Figure 12.

5.4 Discussion

5.4.1 Comparison with statistics based on beamlet coefficients only

The numerical experiments we presented provide evidence that using the spatial relationship between beamlets allows for the design of algorithms that have the ability to perform well even in the case of very low SNR, where statistics based only on the beamlets coefficients fail. More precisely, we compared our different algorithms described above with the *Log-Survival*

Index (LSI) introduced in [20]; see Table 1. The LSI is defined as

$$S_j(t) = \frac{\log(1 + N_j(t))}{\log(1 + N_j)},$$

where $N_j(t)$ is the number of beamlet coefficients at the j-th scale that exceed t, and N_j is the total number of beamlet coefficients at scale j.

Case	(a)	(b)	(c)
Statistic	Graph connectivity	Vertex betweeness	FSR
Lowest SNR	0.8 (1.33)	0.8 (2)	1 (1.33)

Table 1: Summary of the experiments. The numbers (resp. numbers in parentheses) in the last row correspond to the lowest SNR at which the statistic used for that particular case (resp. the LSI statistic) is still powerful.

Figure 18 shows the behavior of this statistic for situation (a) of Figure 12. The LSI curves are clearly disjoint at SNR = 4; however, they merge at SNR = 0.8, resulting in the LSI being powerless, while the statistic presented in Section 5.1 is still powerful.
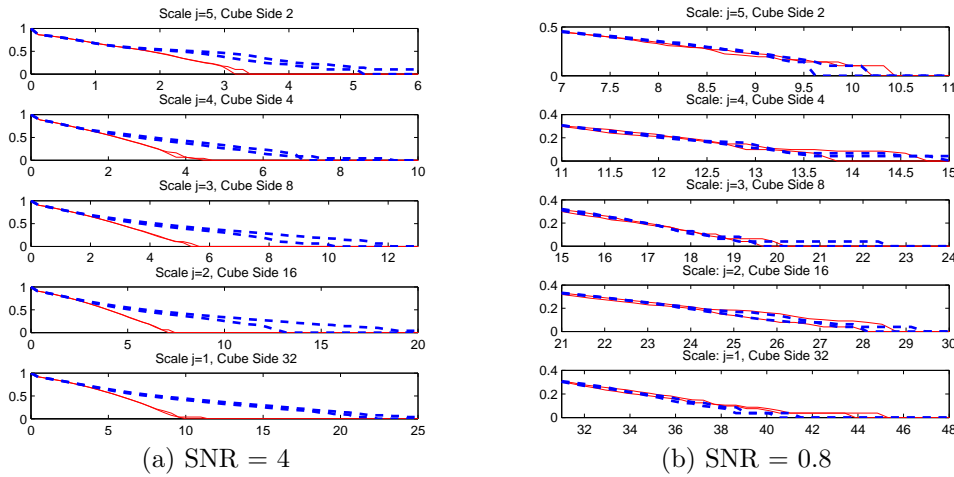
Figure 18: LSI curves at various scales for the random point cloud (red, solid) and for the dataset with random filaments (blue, dashed) in column (a) of Figure 12.

5.4.2 Computational Complexity

The computational burden comes from the beamlet transform. In these experiments, we used the version developed in [40], which runs in order $O(n^5)$ flops for an n^3 pixel array. The version based on the Fast Slant Stack [25, 10] is in theory faster, $O(n^4 \log n)$ flops, yet in practice the implementation in [40] is more precise and faster on smaller arrays as considered here.

We mention that beamlets have only been developed for 2D and 3D datasets, partly because the method suffers from the curse of dimensionality, since the number of beamlets increases exponentially with the ambient dimension; see Lemma 4.3. Note also that the implementation in [23] runs in $O(n^2 \log n)$ for an n^2 pixel array, so that the 3D implementations are comparatively heavier.

5.4.3   Computation of the Longest Weighted Paths

Since finding the LWP is an NP-Hard problem [33], in the experiments described above we restricted ourselves to acyclic networks. We assume that the foreground signal consists of a set of curves of the form $\gamma(s) = (s, \gamma_y(s), \gamma_z(s))$. Therefore, it is possible to approximate such curves with beamlets oriented along the fist axis only, so their chaining will never induce a cycle. This rather artificial assumption was made only for computational reasons; for general datasets one would be forced to use approximations [4] or other strategies [26, 23]. We tried statistics based on connected components, which are computationally tractable, but were not able to improve on the LSI.

5.4.4   Full multiscale analysis

We built a good-continuation network for each scale separately along the lines presented in Section 4.2. Based on Theorem 4.6, this is fine if we expect filaments of homogenous smoothness. If we want to test for filaments of varying smoothness, building a single good-continuation network that includes neighboring relationships across scales may be more useful.

Appendix

Proof of Lemma 2.4

We first count the number of nodes $(\mathbf{m}, \mathbf{h})$. There are Δ^{-k} choices for $\mathbf{m}$ and $2\beta\delta_{|\mathbf{s}|}^{-1}$ for each $h^{(\mathbf{s})}$, so for $\mathbf{h}$ a total of

$$\prod_{|\mathbf{s}| \leq \lfloor \alpha \rfloor} 2\beta\delta_{|\mathbf{s}|}^{-1} = (2\beta)^{c_3} \delta^{-c_3} \Delta^{c_4}.$$

For a given node $(\mathbf{m}, \mathbf{h})$, $\mathbf{m}$ has at most $2k$ neighbors $\mathbf{m}_\star$'s in the square grid. And for each one of them, there are at most 6^{c_3} $\mathbf{h}_\star$'s satisfying the left part of (7), since at each $\mathbf{s}$ there are at most 6 choices. □

Proof of Theorem 2.5

In view of Lemma 2.2, it is enough to show that, for $\mathbf{m}, \mathbf{m}_\star \in \{1, \ldots, \Delta^{-1}\}^k$ with $\mathbf{m}_\star = \mathbf{m} + \mathbf{e}_i$,

$$\left| h^{(\mathbf{s})}(\mathbf{m}_\star, f) - \sum_{t \leq \lfloor \alpha \rfloor - |\mathbf{s}|} \frac{h^{(\mathbf{s}+t\mathbf{e}_i)}(\mathbf{m}, f)}{t!} \right| < 3.$$

By the triangle inequality, we have

$$\left| h^{(\mathbf{s})}(\mathbf{m}_\star, f) - \sum_{t \leq \lfloor \alpha \rfloor - |\mathbf{s}|} \frac{h^{(\mathbf{s}+t\mathbf{e}_i)}(\mathbf{m}, f)}{t!} \right| \leq \left| h^{(\mathbf{s})}(\mathbf{m}_\star, f) - \delta_{|\mathbf{s}|}^{-1} f^{(\mathbf{s})}(x_{\mathbf{m}_\star}) \right| +$$

$$\left| \delta_{|\mathbf{s}|}^{-1} f^{(\mathbf{s})}(x_{\mathbf{m}_\star}) - \sum_{t \leq \lfloor \alpha \rfloor - |\mathbf{s}|} \frac{\delta_{|\mathbf{s}|+t}^{-1} f^{(\mathbf{s}+t\mathbf{e}_i)}(x_{\mathbf{m}})}{t!} \right| +$$

$$\sum_{t \leq \lfloor \alpha \rfloor - |\mathbf{s}|} \frac{\left| \delta_{|\mathbf{s}|+t}^{-1} f^{(\mathbf{s}+t\mathbf{e}_i)}(x_{\mathbf{m}}) - h^{(\mathbf{s}+t\mathbf{e}_i)}(\mathbf{m}, f) \right|}{t!}.$$

By definition of $h^{(\mathbf{s})}(\mathbf{m}, f)$, the first term on the right handside is bounded by $1/2$ while the third term is bounded by $1/2 \sum_t 1/t! \leq \exp(1)/2 < 3/2$. So we are left with showing that the second term is bounded by 1. To do that, we use Lemma 2.3 and the fact that $x_{\mathbf{m}_\star,i} - x_{\mathbf{m},i} = \Delta$ to get

$$\left| f^{(\mathbf{s})}(x_{\mathbf{m}_\star}) - \sum_t f^{(\mathbf{s}+t\mathbf{e}_i)}(x_{\mathbf{m}}) \frac{\Delta^t}{t!} \right| \leq c_1 \beta \Delta^{\alpha - |\mathbf{s}|}.$$

Since $\Delta^t = \delta_s \delta_{s+t}^{-1}$ for all s, t, we further get

$$\left| \delta_{|\mathbf{s}|}^{-1} f^{(\mathbf{s})}(x_{\mathbf{m}_\star}) - \sum_t \frac{\delta_{|\mathbf{s}|+t}^{-1} f^{(\mathbf{s}+t\mathbf{e}_i)}(x_{\mathbf{m}})}{t!} \right| \leq c_1 \beta \Delta^\alpha / \delta.$$

This concludes the proof of Theorem 2.5. $\qquad\qquad\square$

Proof of Theorem 3.2

We use simplified notation for clarity. By Boole's Inequality, we have

$$\mathbf{P}\left\{ M_{\mathcal{P}} > C\delta^{d-k} n | H_0 \right\} \leq |\mathcal{P}| \cdot \max_{P \in \mathcal{P}} \mathbf{P}\left\{ N(R_\delta(P)) > C\delta^{d-k} n | H_0 \right\}.$$

For the number of paths,

$$|\mathcal{P}| = O(\delta^{-(d-k)c_3} \Delta^{(d-k)c_4} \cdot (6^{(d-k)c_3})^{\Delta^{-k}}), \tag{17}$$

since there are $O(\delta^{-(d-k)c_3} \Delta^{(d-k)c_4})$ choices for the starting point and $6^{(d-k)c_3}$ at each step after that, along a path of length Δ^{-k} (see Lemma 2.6). Hence, $\log(|\mathcal{P}|) \leq a_1 \delta^{-k/\alpha}$ for some constant a_1 not depending on δ.

Under H_0, $N(R_\delta(P))$ is binomial with parameters n and $|R_\delta(P)|_d$, so that

$$\max_{P \in \mathcal{P}} \mathbf{P}\left\{ N(R_\delta(P)) > C\delta^{d-k} n | H_0 \right\} = \mathbf{P}\left\{ \mathrm{Bin}(n, \max_{P \in \mathcal{P}} |R_\delta(P)|_d) > C\delta^{d-k} n \right\}.$$

By integrating with respect to $\mathbf{z}$ (the last $d - k$ coordinates) first, we have:

$$|R_\delta(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k})|_d = (c_0 + 1)\Delta^k\delta^{d-k}, \quad \forall(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}). \tag{18}$$

Hence, for all $P \in \mathcal{P}$, $|R_\delta(P)|_d = (c_0 + 1)\delta^{d-k}$, and therefore, since $\delta = n^{-\alpha/(k+\alpha(d-k))}$,

$$\max_{P \in \mathcal{P}} \mathbf{P}\left\{N(R_\delta(P)) > C\delta^{d-k}n | H_0\right\} = \mathbf{P}\left\{\mathrm{Bin}(n, (c_0 + 1)\delta^{d-k}) > C\delta^{d-k}n\right\}.$$

By standard large deviation bounds like Bernstein's Inequality [52], the logarithm of the right hand side is bounded from above by $-a_2C\delta^{d-k}n$ for $C \geq 2(c_0 + 1)$, where a_2 does not depend on δ or n.

Collecting terms, we get the following bound:

$$\log \mathbf{P}\left\{M_{\mathcal{P}} > C\delta^{d-k}n | H_0\right\} \leq a_1\delta^{-k/\alpha} - Ca_2\delta^{d-k}n, \quad \forall C \geq 2(c_0 + 1).$$

By choosing $C = 2((c_0 + 1) \vee (a_1/a_2))$, the right hand side is negative if $\delta \geq n^\rho$. $\qquad\square$

Proof of Theorem 3.3

We use simplified notation for clarity. By (18), for all $(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) \in \mathbb{G}^{k,d-k}(\alpha, \beta)$, we have:

$$\mathbf{P}\left\{S(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) = 1 | H_0\right\} = \mathbf{P}\left\{\mathrm{Bin}(n, a_1\delta^{k/\alpha+d-k}) > \tau n\delta^{k/\alpha+d-k}\right\},$$

where $a_1 = (c_0 + 1)(c_1\beta)^{-k/\alpha}$. Let $q_0(\tau) = \max \mathbf{P}\left\{\mathrm{Bin}(n, a_1p) > \tau np\right\}$ over $n \in \mathbb{N}$ and $p \in (0, 1)$ such that $np \geq 1$; note that $q_0(\tau) \to 0$ as $\tau \to \infty$. In the same way that we obtained (17), we find that the number of paths of the form

$$\{(\mathbf{m}_{\mathrm{zz}}(t), \mathbf{h}_1(t), \ldots, \mathbf{h}_{d-k}(t)) : t = t_0, \ldots, t_0 + \ell - 1\}$$

is bounded above by $a_2\delta^{-(d-k)c_3}\Delta^{(d-k)c_4}\Delta^{-k}6^{(d-k)c_3\ell}$, for a_2 not depending on μ or η. With this fact and Boole's Inequality, we get:

$$\mathbf{P}\left\{L > \ell | H_0\right\} \leq a_2\delta^{-(d-k)c_3}\Delta^{(d-k)c_4-k}6^{(d-k)c_3\ell} \cdot q_0^\ell.$$

Hence, with q_0 small enough (i.e. τ large enough), we have

$$\mathbf{P}\left\{L > \log(1/\delta) | H_0\right\} \to 0, \quad n \to \infty,$$

where we used (6). We then conclude with the fact that $\delta = n^{-\alpha/(k+\alpha(d-k))} \vee \eta$.

Assume for concreteness that $\eta > 0$. Under H_1, let $P^* \in \mathcal{P}$ such that $\mathrm{graph}_\eta(f^*) \subset R_\delta(P^*)$. Note that $|\mathrm{graph}_\eta(f^*)|_d = \eta^{d-k}$ and $|\mathrm{graph}_\eta(f^*) \cap R_\delta(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k})|_d = \Delta^k\eta^{d-k}$ for all $(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) \in P^*$. Therefore, together with the behavior under the null, for all $(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) \in P^*$, we have:

$$\mathbf{P}\left\{S(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) = 1 | H_1\right\} \geq \mathbf{P}\left\{\mathrm{Bin}(n, \varepsilon_n\Delta^k) > \tau n\delta^{k/\alpha+d-k}\right\}.$$

Hence, with $\varepsilon_n > Cn\delta^{d-k}$ we have, for all $(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) \in P^*$,

$$\mathbf{P}\left\{S(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) = 1|H_1\right\} \geq \mathbf{P}\left\{\mathrm{Bin}(n, C(c_1\beta)^{-k/\alpha}n\delta^{k/\alpha+d-k}) > \tau n\delta^{k/\alpha+d-k}\right\}.$$

With τ fixed as above, let $q_1(b) = \min \mathbf{P}\left\{\mathrm{Bin}(n, bp) > \tau np\right\}$ over $n \in \mathbb{N}$ and $p \in (0,1)$ such that $np \geq 1$; note that $q_1(b) \to 1$ as $b \to \infty$. Assume that n is large enough that $n^\rho > 1$, so that $n\delta^{k/\alpha+d-k} > 1$. Then, for $C = b(c_1\beta)^{k/\alpha}$,

$$\mathbf{P}\left\{S(\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) = 1|H_1\right\} \geq q_1(b), \quad \forall (\mathbf{m}, \mathbf{h}_1, \ldots, \mathbf{h}_{d-k}) \in P^*.$$

Now, by the Erdös-Rényi Law [9] (in fact a modified version allowing for weak dependencies, found in Appendix A.3 in [5]), the longest significant run along P^* has length at least $\log(|P^*|)/\log(1/q_1)(1+o(1))$ with high probability. Hence, for q_1 large enough (i.e. C large enough),

$$\mathbf{P}\left\{L > \log(1/\delta)|H_1\right\} \to 1, \quad n \to \infty,$$

where we used $|P^*| = \Delta^{-k}$ together with (6). $\qquad\square$

Proof of Lemma 4.3

There are $O(\Delta^{-d})$ Δ-hypercubes and each one of them has $O(\Delta\delta_0^{-1})^{d-1}$ gridpoints on anyone of its $2d$ faces. Therefore, there are $O(\Delta^{-d}d^2(\Delta\delta_0^{-1})^{2(d-1)}) = O(d^2 2^{2(d-1)J-(d-2)j})$ beamlets at scale j.

We now look at the degree of a beamlet $B = [b_1, b_2] \in \mathbb{B}_{j,J}^d$. Let r be such that $|b_{r,2} - b_{r,1}| = \|b_2 - b_1\|$. Consider another beamlet of the form $[b_2, b_3]$, and define

$$b_3^* = b_2 + \frac{b_{r,3} - b_{r,2}}{b_{r,2} - b_{r,1}}(b_2 - b_1).$$

If $[b_1, b_2]$ and $[b_2, b_3]$ are neighbors,

$$\begin{aligned}
\|b_3 - b_3^*\| &\leq 2^{j-J}\frac{\|b_3 - b_2\| + \|b_2 - b_1\|}{|b_{r_2,2} - b_{r_2,1}|} \\
&\leq 2^{j-J}(\Delta/\delta_0 + 1) \\
&= O(2^{-J}|B|^{-1})
\end{aligned}$$

where the first inequality comes from (15) and the second from the fact that for any beamlet $[b, b']$, $\delta \leq \|b' - b\| \leq \Delta$. Now, in a ball of radius A there are at most $O(A/\delta_0)^{d-1}$ r-gridpoints, for any $r = 1, \ldots, d$. Therefore, B has $O(d|B|^{-(d-1)})$ neighbors. $\qquad\square$

Proof of Lemma 4.5

The number of r-gridpoints is of order $O(\Delta^{-1}\delta^{-(d-1)})$. For a fixed r-gridpoint b_1, there are order $O(\Delta\delta^{-1})^{d-1}$ r-gridpoints b_2 such that $[b_1, b_2]$ forms an r-beam. Therefore, the number of r-beams is of order $O(\Delta^{d-2}\delta^{-2d+2})$.

The number of r_1r_2-beams is of smaller order of magnitude, because they are more constrained. Indeed, $b_{r_1,2}$ and $b_{r_2,2}$ are determined by $b_{r_1,1}$ and $b_{r_2,1}$ up to $O(\delta)$, which corresponds to order $O(1)$ choices; while for each $b_{r_3,2}, r_3 \neq r_1, r_2$ there are order $O(\Delta\delta^{-1})$ choices. Hence, there are order $O(\Delta^{d-3}\delta^{-2d+3})$ r_1r_2-beams.

All together, the number of beams is of order $O(d\Delta^{d-2}\delta^{-2d+2}(1 + d\Delta^{-1}\delta))$.

We now bound the degree of a vertex in $\overline{\mathbb{B}}^d_{j,J}$. By construction, an r_1-beam is connected only with r_1-beams, and (possibly) r_1r_2-beams; similarly, an r_1r_2-beam is connected only with r_1- and r_2-beams, and (possibly) r_3r_1- and r_2r_3-beams.

Fix an r_1r_2-beam $[b_1, b_2]$. First, take an r_2-beam $[b_2, b_3]$ and define

$$b_3^* = b_2 + \frac{b_{r_2,3} - b_{r_2,2}}{b_{r_2,2} - b_{r_2,1}}(b_2 - b_1),$$

which is the intersection of the line (b_1, b_2) with the r_2-hyperplane b_3 belongs to. If $[b_1, b_2]$ and $[b_2, b_3]$ are neighbors,

$$\begin{aligned}
\|b_3 - b_3^*\| &\leq (11\delta/20)\frac{\|b_3 - b_2\| + \|b_2 - b_1\|}{|b_{r_2,2} - b_{r_2,1}|} \\
&\leq (11\delta/20)(4 + O(\delta)) \\
&< (5/2)\delta, \quad \text{for } \delta \text{ small enough,}
\end{aligned}$$

where the first inequality comes from (15) and the second from (11)-(14). Therefore, there are at most 5 choices per coordinate of b_3 (except for $b_{r_2,3}$). Hence, $[b_1, b_2]$ has at most 5^{d-1} neighbors that are r_2-beams. Similarly, $[b_1, b_2]$ has at most 5^{d-1} neighbors that are r_1-beams.

Next, take an r_2r_3-beam $[b_2, b_3]$ and define

$$b_1^* = b_2 + \frac{b_{r_1,3} - b_{r_1,2}}{b_{r_1,2} - b_{r_1,1}}(b_2 - b_3),$$

which is the intersection of the line (b_2, b_3) with the r_1-hyperplane b_1 belongs to. If $[b_1, b_2]$ and $[b_2, b_3]$ are neighbors, performing the corresponding computations we arrive at

$$\|b_1 - b_1^*\| < (11\delta/20)(6 + O(\delta)) < (7/2)\delta, \quad \text{for } \delta \text{ small enough.}$$

Note that $[b_2, b_1^*]$ is proportional to $[b_2, b_3]$, with a constant of proportionality between 1 and 2, and $[b_2, b_3]$ satisfies (12)-(14) with r_1 replaced by r_3; this together with the bound above (applied with $r_4 = r_3$) and the triangle inequality implies:

$$|b_{r_3,1} - b_{r_3,2}| > |b_{r_2,1} - b_{r_2,2}| - 6\delta.$$

Turning things around, define

$$b_3^* = b_2 + \frac{b_{r_3,3} - b_{r_3,2}}{b_{r_3,2} - b_{r_3,1}}(b_2 - b_1),$$

which is the intersection of the line (b_1, b_2) with the r_3-hyperplane b_3 belongs to. Again using (15) together with properties (11)-(14), and the above inequality, we get

$$\|b_3 - b_3^*\| < (11\delta/20)(6 + O(\delta)) < (7/2)\delta, \quad \text{for } \delta \text{ small enough.}$$

Therefore, there are at most 7 choices per coordinate of b_3 (except for $b_{r_3,3}$). Hence, $[b_1, b_2]$ has at most 7^{d-1} neighbors that are $r_2 r_3$-beams. Similarly, $[b_1, b_2]$ has at most 7^{d-1} neighbors that are $r_3 r_1$-beams.

The reasoning is similar when $[b_1, b_2]$ is an r_1-beam. In particular, an r-beam that does not make an angle close a 45° with the r-hyperplanes it connects only has r-beams as neighbors, at most 5^{d-1} on each side. $\square$

Proof of Theorem 4.6

Assume the conditions Theorem 4.6 are satisfied, with the constant K chosen as to make all the forthcoming appearances of $O(\kappa \Delta^\alpha)$ sufficiently small compared to δ. Note that under these conditions, both δ, Δ and δ/Δ are all decreasing functions of J. We choose a smooth parametrization by arc-length of γ. Let $\ell = \text{length}(\gamma)$.

We say that $(u_1, \ldots, u_d) \in \mathbb{R}^d$ is an r-vector if $|u_r| \geq |u_{r_1}|, \forall r_1$. Without loss of generality, assume $\gamma'(0)$ is an r-vector with $\gamma'_r(0) > 0$. We may also assume that $\gamma(0)$ belongs to an r-hyperplane, for otherwise we work with an extension of γ that reaches an r-hyperplane in the direction $-\gamma'(0)$. We recursively define an increasing sequence of arclengths $\{s_i : 0 \leq i \leq I\}$. First, let $s_0 = 0$. Suppose s_{i-1} has been defined and assume, without loss of generality, that $\gamma'(s_{i-1})$ is an r-vector; then, let

$$s_i = \inf\{s > s_{i-1} : \gamma(s) \text{ belongs to an } r\text{-hyperplane and } |\gamma_r(s) - \gamma_r(s_{i-1})| \geq \Delta\},$$

with the usual convention $\inf\{\emptyset\} = \infty$. We may also assume that $s_i < \infty$, for otherwise we work with an extension of γ that reaches an r-hyperplane (in the case above) in the direction $\gamma'(\ell)$. If $s_i = \ell$, then let $I = i$ and stop the recursion.

Define b_i to be the gridpoint closest to $\gamma(s_i)$, and $B_i = [b_i, b_{i+1}]$ the line-segment joining b_i and b_{i+1}.

- **Claim 1.** $I \leq \lambda \Delta^{-1} + 2$.

- **Claim 2.** For $i = 0, \ldots, I - 1$, B_i is a vertex in $\overline{\mathbb{B}}_{j,J}^d$.

- **Claim 3.** For $i = 0, \ldots, I - 2$, B_i and B_{i+1} are neighbors in $\overline{\mathbb{B}}_{j,J}^d$.

- **Claim 4.** For $i = 0, \ldots, I - 1$, $\gamma([s_i, s_{i+1}]) \subset R(B_i)$.

Preliminaries.

- For Δ small enough,

$$\Delta \leq s_{i+1} - s_i < 3d^{1/2}\Delta, \quad \forall i = 0, \ldots, I - 1. \tag{19}$$

  Proof. We start with the lower bound. A Taylor expansion and the fact that γ is parametrized by arc gives

$$\|\gamma(s_{i+1}) - \gamma(s_i)\| \leq s_{i+1} - s_i,$$

and by construction, $\|\gamma(s_{i+1}) - \gamma(s_i)\| \geq \Delta$.

We now turn to proving the upper bound. Suppose without loss of generality that $\gamma'(s_i)$ is an r-vector. Then, by construction $|\gamma_r(s) - \gamma_r(s_i)| \leq 2\Delta, \forall s \in [s_i, s_{i+1}]$, and $|\gamma'_r(s_i)| \geq d^{-1/2}$. This, (9) and the triangle inequality imply

$$2\Delta + \kappa(s - s_i)^\alpha - d^{-1/2}(s - s_i) \geq 0, \quad \forall s \in [s_i, s_{i+1}].$$

And for $\Delta < (3^\alpha d^{\alpha/2}\kappa)^{1/(\alpha-1)}$, the left handside is negative for $s - s_i$ replaced by $3d^{1/2}\Delta$, so that $s_{i+1} - s_i < 3d^{1/2}\Delta$. □

- For $i = 1, \ldots, I$, if $\gamma'(s_i)$ is an r_1-vector (say) and $\gamma'(s_{i-1})$ is a r_2-vector (say), then

$$|\gamma'_{r_1}(s_i)| - |\gamma'_{r_2}(s_i)| \leq 4\kappa(s_i - s_{i-1})^{\alpha-1}. \tag{20}$$

Proof. Because $|\gamma'_{r_2}(s_{i-1})| \geq |\gamma'_{r_1}(s_{i-1})|$, we have

$$|\gamma'_{r_1}(s_i)| - |\gamma'_{r_2}(s_i)| \leq |\gamma'_{r_1}(s_i)| - |\gamma'_{r_1}(s_{i-1})| - (|\gamma'_{r_2}(s_i)| - |\gamma'_{r_2}(s_{i-1})|).$$

Using Lemma 4.1, this implies

$$|\gamma'_{r_1}(s_i)| - |\gamma'_{r_2}(s_i)| \leq 4\kappa(s_i - s_{i-1})^{\alpha-1}.$$

□

- For $i = 1, \ldots, I$,
$$\|b_i - \gamma(s_i)\| \leq \delta/2. \tag{21}$$

Proof. By construction. □

Proof of Claim 1. A straightforward consequence of the fact that $s_{i+1} - s_i \geq \Delta$ for all $i = 0, \ldots, I - 1$. The '+2' comes from possibly extending the curve as described above.

□

Proof of Claim 2. We first prove that B_0 is a beam. Assume without loss of generality that $\gamma(s_0)$ is on an r_1-hyperplane and $\gamma'(s_0)$ is a r_1-vector; we now show that B_0 is an r_1-beam. Applying (9) and the fact that $|\gamma'_{r_2}(s_0)| \leq |\gamma'_{r_1}(s_0)|$ we get

$$|\gamma_{r_2}(s_1) - \gamma_{r_2}(s_0)| \leq |\gamma'_{r_1}(s_0)|(s_1 - s_0) + \kappa(s_1 - s_0)^\alpha.$$

Using (9) again, together with $|\gamma_{r_1}(s_1) - \gamma_{r_1}(s_0)| = \Delta$, we have

$$|\gamma'_{r_1}(s_0)|(s_1 - s_0) \leq \Delta + \kappa(s_1 - s_0)^\alpha.$$

Therefore,

$$|\gamma_{r_2}(s_1) - \gamma_{r_2}(s_0)| \leq \Delta + 2\kappa(s_1 - s_0)^\alpha.$$

On the right handside, we use (19) to get $\kappa(s_1 - s_0)^\alpha = O(\kappa\Delta^\alpha)$. All together,

$$|\gamma_{r_2}(s_1) - \gamma_{r_2}(s_0)| \le \Delta + O(\kappa\Delta^\alpha);$$

similarly

$$|\gamma_{r_3}(s_1) - \gamma_{r_3}(s_0)| \le \Delta + O(\kappa\Delta^\alpha).$$

This, together with the triangle inequality and (21), shows that

$$\|b_1 - b_0\| \le \delta + \Delta + O(\kappa\Delta^\alpha) < 2\delta + \Delta,$$

which implies that $\|b_1 - b_0\| \le \delta + \Delta$ since $\|b_1 - b_0\|$ is an integer multiple of δ. This proves that B_0 is an r_1-beam.

We next consider $i = 1, \ldots, I - 1$ and prove that B_i is a beam. Assume without loss of generality that $\gamma'(s_i)$ is an r_1-vector. If $\gamma'(s_{i-1})$ is an r_1-vector, then $\gamma(s_i)$ is on an r_1-hyperplane and the situation is as above. Therefore, assume without loss of generality that $\gamma'(s_{i-1})$ is a r_2-vector; we now show that B_i is an $r_1 r_2$-beam. With (9), we get

$$\big|\, |\gamma_{r_1}(s_{i+1}) - \gamma_{r_1}(s_i)| - |\gamma_{r_2}(s_{i+1}) - \gamma_{r_2}(s_i)| \,\big| \le (|\gamma'_{r_1}(s_i)| - |\gamma'_{r_2}(s_i)|)(s_{i+1} - s_i) + 2\kappa(s_{i+1} - s_i)^\alpha;$$

together with (20), this implies

$$\big|\, |\gamma_{r_1}(s_{i+1}) - \gamma_{r_1}(s_i)| - |\gamma_{r_2}(s_{i+1}) - \gamma_{r_2}(s_i)| \,\big| \le 6\kappa(s_{i+1} - s_i)^\alpha,$$

and with (19),

$$\big|\, |\gamma_{r_1}(s_{i+1}) - \gamma_{r_1}(s_i)| - |\gamma_{r_2}(s_{i+1}) - \gamma_{r_2}(s_i)| \,\big| = O(\kappa\Delta^\alpha).$$

On the other hand, using (9) and the fact that $|\gamma'_{r_1}(s_i)| \ge |\gamma'_{r_3}(s_i)|$, we also get

$$|\gamma_{r_1}(s_{i+1}) - \gamma_{r_1}(s_i)| - |\gamma_{r_3}(s_{i+1}) - \gamma_{r_3}(s_i)| \ge -2\kappa(s_{i+1} - s_i)^\alpha,$$

which by (19) implies

$$|\gamma_{r_1}(s_{i+1}) - \gamma_{r_1}(s_i)| - |\gamma_{r_3}(s_{i+1}) - \gamma_{r_3}(s_i)| \ge -O(\kappa\Delta^\alpha).$$

Using the equations above together with the triangle inequality and (21),

$$\big|\, |b_{r_1,i+1} - b_{r_1,i}| - |b_{r_2,i+1} - b_{r_2,i}| \,\big| \le \delta + O(\kappa\Delta^\alpha) < 2\delta,$$

$$|b_{r_1,i+1} - b_{r_1,i}| - |b_{r_3,i+1} - b_{r_3,i}| \ge -\delta - O(\kappa\Delta^\alpha) > -2\delta.$$

Since the differences above are multiple integers of δ, the first inequality may be replaced by $\le \delta$ and the second by $\ge -\delta$. Therefore B_i is an $r_1 r_2$-beam. $\square$

Proof of Claim 3. Fix coordinate r_1; we want to show that:

$$\|(b_{r_1,i+1} - b_{r_1,i})(b_{i+2} - b_{i+1}) - (b_{r_1,i+2} - b_{r_1,i+1})(b_{i+1} - b_i)\|$$
$$< (11\delta/20)(\|b_{i+2} - b_{i+2}\| + \|b_{i+1} - b_i\|).$$

We first prove a similar inequality involving γ:

$$\|(\gamma_{r_1}(s_{i+1}) - \gamma_{r_1}(s_i))(\gamma(s_{i+2}) - \gamma(s_{i+1})) - (\gamma_{r_1}(s_{i+2}) - \gamma_{r_1}(s_{i+1}))(\gamma(s_{i+1}) - \gamma(s_i))\|$$
$$\leq O(\kappa\Delta^\alpha)(\|\gamma(s_{i+2}) - \gamma(s_{i+1})\| + \|\gamma(s_{i+1}) - \gamma(s_i)\|).$$

This simply comes from applying (9) to get

$$\gamma_{r_1}(s_{i+2}) - \gamma_{r_1}(s_{i+1}) = \gamma'_{r_1}(s_{i+1})(s_{i+2} - s_{i+1}) + O(\kappa(s_{i+2} - s_{i+1})^\alpha),$$
$$\gamma_{r_1}(s_{i+1}) - \gamma_{r_1}(s_i) = \gamma'_{r_1}(s_{i+1})(s_{i+1} - s_i) + O(\kappa(s_{i+1} - s_i)^\alpha),$$
$$\gamma(s_{i+2}) - \gamma(s_{i+1}) = \gamma'(s_{i+1})(s_{i+2} - s_{i+1}) + O(\kappa(s_{i+2} - s_{i+1})^\alpha),$$
$$\gamma(s_{i+1}) - \gamma(s_i) = \gamma'(s_{i+1})(s_{i+1} - s_i) + O(\kappa(s_{i+1} - s_i)^\alpha),$$

and then using the triangle inequality and (19).

Using this inequality, the triangle inequality and (21), we then get

$$\|(b_{r_1,i+1} - b_{r_1,i})(b_{i+2} - b_{i+1}) - (b_{r_1,i+2} - b_{r_1,i+1})(b_{i+1} - b_i)\|$$
$$< (\delta/2 + O(\kappa\Delta^\alpha))(\|b_{i+2} - b_{i+2}\| + \|b_{i+1} - b_i\| + 2\delta)$$
$$< (\delta/2 + O(\kappa\Delta^\alpha))(1 + \Delta^{-1}\delta)(\|b_{i+2} - b_{i+2}\| + \|b_{i+1} - b_i\|),$$

where we used the fact that any beam has (as a vector) supnorm at least Δ. We conclude by making $O(\kappa\Delta^\alpha)$ sufficiently small compared with δ and δ itself sufficiently small compared with Δ.

Therefore, $B_i = [b_i, b_{i+1}]$ and $B_{i+1} = [b_{i+1}, b_{i+2}]$ are neighbors in $\overline{\mathbb{B}}^d_{j,J}$. $\qquad\square$

Proof of Claim 4. Applying Lemma 4.2 with $r = s_i$ and $t = s_{i+1}$, it follows that $\gamma([s_i, s_{i+1}])$ belongs to the $2\kappa(s_{i+1} - s_i)^\alpha$-neighborhood of $[\gamma(s_i), \gamma(s_{i+1})]$. Because of (19), (21) and the triangle inequality, this implies that $\gamma([s_i, s_{i+1}]) \subset R(B_i)$. $\qquad\square$

References

[1] Sloan Digital Sky Survey (`http://www.sdss.org`).

[2] Library of Efficient Data types and Algorithms (`http://www.mpi-inf.mpg.de/LEDA`).

[3] Software used in this paper (`http://www.cbl.uh.edu/~efros/research.html`).

[4] N. Alon, R. Yuster, and U. Zwick. Color-coding. *J. Assoc. Comput. Mach.*, 42(4):844–856, 1995.

[5] E. Arias-Castro. *Graphical structures for geometric detection.* PhD thesis, Stanford University, 2004.

[6] E. Arias-Castro, D. Donoho, and X. Huo. Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inform. Theory*, 51(7):2402–2425, 2005.

[7] E. Arias-Castro, D. Donoho, and X. Huo. Adaptive multiscale detection of filamentary structures in a background of uniform random points. *Ann. Statist.*, 34(1):326–349, 2006.

[8] E. Arias-Castro, D. Donoho, X. Huo, and C. Tovey. Connect the dots: how many random points can a regular curve pass through? *Adv. in Appl. Probab.*, 37(3):571–603, 2005.

[9] R. Arratia, L. Goldstein, and L. Gordon. Poisson approximation and the Chen-Stein method. *Statist. Sci.*, 5(4):403–434, 1990. With comments and a rejoinder by the authors.

[10] A. Averbuch, R. Coifman, D. Donoho, and M. Israeli. Fast slant stack: A notion of radon transform for data in a cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible. Tech. report, Stanford University, 2001.

[11] B. Awerbuch, Y. Azar, A. Blum, and S. Vempala. New approximation guarantees for minimum-weight k-trees and prize-collecting salesmen. *SIAM J. Comput.*, 28(1):254–262 (electronic), 1999.

[12] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 26–33 vol.1, Oct. 2003.

[13] E. J. Candes. Ridgelets and the representation of mutilated Sobolev functions. *SIAM J. Math. Anal.*, 33(2):347–368 (electronic), 2001.

[14] E. J. Candès. Multiscale chirplets and near-optimal recovery of chirps. Technical report, Stanford University, 2002.

[15] E. J. Candès, P. R. Charlton, and H. Helgason. Detecting highly oscillatory signals by chirplet path pursuit. *Appl. Comput. Harmon. Anal.*, 24(1):14–40, 2008.

[16] E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.

[17] B. DasGupta, J. Hespanha, and E. Sontag. Computational complexities of honey-pot searching with local sensory information. In *2004 American Control Conference (ACC 2004)*, pages 2134–2138, 2004.

[18] A. Desolneux, L. Moisan, and J.-M. Morel. A grouping principle and four applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(4):508–513, 2003.

[19] A. Desolneux, L. Moisan, and J.-M. Morel. Maximal meaningful events and applications to image analysis. *Ann. Statist.*, 31(6):1822–1851, 2003.

[20] D. Donoho, O. Levi, J.-L. Starck, and V. Martinez. Multiscale geometric analysis for 3-d catalogues. In J.-L. Starck and F. Murtagh, editors, *SPIE Conference on Astronomical Data Analysis*, volume 4847, 2002.

[21] D. L. Donoho. Unconditional bases are optimal bases for data compression and for statistical estimation. *Appl. Comput. Harmon. Anal.*, 1(1):100–115, 1993.

[22] D. L. Donoho. Wedgelets: nearly minimax estimation of edges. *Ann. Statist.*, 27(3):859–897, 1999.

[23] D. L. Donoho and X. Huo. Beamlets and multiscale image analysis. In *Multiscale and multiresolution methods*, volume 20 of *Lect. Notes Comput. Sci. Eng.*, pages 149–196. Springer, Berlin, 2002.

[24] D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26(3):879–921, 1998.

[25] D. L. Donoho and O. Levi. Fast X-ray and beamlet transforms for three-dimensional data. In *Modern signal processing*, volume 46 of *Math. Sci. Res. Inst. Publ.*, pages 79–116. Cambridge Univ. Press, Cambridge, 2004.

[26] D. Geiger, A. Gupta, L. Costa, and J. Vlontzos. Dynamic programming for detecting, tracking and matching deformable contours. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(3):294–302, 1995.

[27] J. Glaz and N. Balakrishnan, editors. *Scan statistics and applications*. Statistics for Industry and Technology. Birkhäuser Boston Inc., Boston, MA, 1999.

[28] J. Glaz, J. Naus, and S. Wallenstein. *Scan statistics*. Springer Series in Statistics. Springer-Verlag, New York, 2001.

[29] T. C. Hu, A. B. K. T, and G. Robins. Optimal minimum-surface computations using network flow. To appear in Mathematical Programming.

[30] X. Huo and J. Chen. Jbeam: multiscale curve coding via beamlets. *Image Processing, IEEE Transactions on*, 14(11):1665–1677, Nov. 2005.

[31] X. Huo, D. Donoho, C. Tovey, and E. Arias-Castro. Dynamic programming methods for 'connecting-the-dots' in scattered point sets. In preparation.

[32] I. Johnstone. Function estimation in gaussian noise sequence models. Draft of a monograph, available at www-stat.stanford.edu/ imj/baseb.pdf.

[33] D. Karger, R. Motwani, and G. D. S. Ramkumar. On approximating the longest path in a graph. *Algorithmica*, 18(1):82–98, 1997.

[34] D. Kirsanov and S. Gortler. A discrete global minimization algorithm for continuous variational problems. Technical report, Harvard University, Computer Science, 2004.

[35] A. N. Kolmogorov. *Selected works of A. N. Kolmogorov. Vol. III*, volume 27 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1993. Information theory and the theory of algorithms. Edited by A. N. Shiryayev. Translated from the 1987 Russian original by A. B. Sossinsky.

[36] A. N. Kolmogorov and V. M. Tikhomirov. ε-entropy and ε-capacity of sets in functional space. *Amer. Math. Soc. Transl. (2)*, 17:277–364, 1961.

[37] V. Kolmogorov and Y. Boykov. What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. *iccv*, 1:564–571, 2005.

[38] A. P. Korostelëv and A. B. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993.

[39] E. Le Pennec and S. Mallat. Sparse geometric image representations with bandelets. *IEEE Trans. Image Process.*, 14(4):423–438, 2005.

[40] O. Levi and B. A. Efros. A new fast algorithm for exact calculation of the discrete 2-d and 3-d x-ray transform. In *Lecture Series on Computer and Computational Sciences*, volume 4, pages 319–322. ICCMSE, 2005.

[41] O. Levi, S. Rotman, and B. Efros. Multiframe dim target detection using 3d multiscale geometric analysis. IMA Annual Program Year Workshop New Mathematics and Algorithms for 3-D Image Analysis, January 2006.

[42] M. Lindenbaum and A. Berengolts. A probabilistic interpretation of the saliency network. In *ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II*, pages 257–272, London, UK, 2000. Springer-Verlag.

[43] A. Lyazghi, C. Decaestecker, I. Camby, R. Kiss, and V. Ham. Characterization of actin filament in cancer cells by the hough transform. In *Proceedings of the IASTED International Conference : Signal Processing, Pattern Recognition and Applications*, pages 183–142, Rhodes, Greece, July 2001.

[44] S. Mallat. *A wavelet tour of signal processing*. Academic Press Inc., San Diego, CA, 1998.

[45] V. Martínez and E. Saar. *Statistics of the Galaxy Distribution*. Chapman and Hall/CRC press, Boca Raton, 2002.

[46] Y. Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1992. Translated from the 1990 French original by D. H. Salinger.

[47] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[48] G. P. Patil, J. Balbus, G. Biging, J. Jaja, W. L. Myers, and C. Taillie. Multiscale advanced raster map analysis system: definition, design and development. *Environ. Ecol. Stat.*, 11(2):113–138, 2004.

[49] G. P. Patil and C. Taillie. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environ. Ecol. Stat.*, 11(2):183–197, 2004.

[50] V. Prinet, O. Monga, and S. Ma. Extraction of vascular network in 3d images. In *Proceedings: The IEEE International Conference on Image Processing*, pages 307–310, September 1996.

[51] A. Sha'asua and S. Ullman. Structural saliency: The detection of globally salient structures using a locally connected network. *Computer Vision., Second International Conference on*, pages 321–327, Dec 1988.

[52] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[53] G. Wahba. *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.

[54] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK, 1994.

[55] R. Willett and R. Nowak. Platelets: a multiscale approach for recovering edges and surfaces in photon-limited medical imaging. *Medical Imaging, IEEE Transactions on*, 22(3):332–350, March 2003.