*Applied Probability Trust* (1 May 2005)

# CONNECT-THE-DOTS: HOW MANY RANDOM POINTS CAN A REGULAR CURVE PASS THROUGH?

ERY ARIAS-CASTRO,* *Stanford University*

DAVID L. DONOHO,* *Stanford University*

XIAOMING HUO,* *Georgia Institute of Technology*

CRAIG A. TOVEY,* *Georgia Institute of Technology*

* Postal address: Department of Statistics, Stanford University, Stanford, CA 94309
* Postal address: Department of Statistics, Stanford University, Stanford, CA 94309
* Postal address: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205
* Postal address: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205

**Abstract**

Suppose $n$ points are scattered uniformly at random in the unit square $[0,1]^2$. Question: How many of these points can possibly lie on some curves of length $\lambda$? Answer, proved here: $O_P(\lambda \cdot \sqrt{n})$.

We consider a general class of such questions; in each case, we are given a class $\Gamma$ of curves in the square, and we ask: in a cloud of $n$ uniform random points, how many can lie on some curve $\gamma \in \Gamma$? Classes of interest include (in addition to the rectifiable curves mentioned above): Lipschitz graphs, monotone graphs, twice-differentiable curves, graphs of smooth functions with $m$-bounded derivatives. In each case we get order-of-magnitude estimates; for example, there are twice-differentiable curves containing as many as $O_P(n^{1/3})$ uniform random points, but not essentially more than this.

We also consider generalizations to higher dimensions and to hypersurfaces of various co-dimensions. Thus, twice-differentiable $k$-dimensional hypersurfaces in $R^d$ may contain as many as $O_P(n^{k/(2d-k)})$ uniform random points. We also consider other notions of 'passing through' such as passing through given space/direction pairs. Thus, twice-differentiable curves in $R^2$ may pass through at most $O_P(n^{1/4})$ uniform random location/direction pairs.

We give both concrete approaches to our results, based on geometric multiscale analysis, and abstract approaches, based on $\varepsilon$-entropy. Several open mathematical questions are identified here for the attention of the probability community.

Stylized applications in image processing and perceptual psychophysics are described.

*Keywords:* Curve/filament detection; $\varepsilon$-entropy; configuration functions; concentration of measure; longest increasing subsequence; travelling salesman problem; pattern recognition.

2000 Mathematics Subject Classification: Primary 60D05

Secondary 62M40

## 1. Introduction

Several well-known games in pop culture require a player to travel through an arena scattered with 'goodies' and to gather as many 'goodies' as possible in a given amount of time; examples include the 60's TV show *Supermarket Sweep* [2] and the 80's video game *PacMan* [1].

Consider now a geometric probability problem. Suppose we have $n$ points $X_1, \ldots, X_n$ scattered uniformly at random in the unit square $[0,1]^2$. We try to visit as many points as possible while traveling a path with total length $\leq \lambda$. This problem is similar to the above-mentioned popular games, if we think of the 'goodies' as located at the points $X_i$, and imagine 'participants' traveling at unit speed for a length of time $\leq \lambda$. The probability problem is to estimate the maximal number of points we can visit, i.e., to determine how many 'goodies' we could typically get by following the best possible path among (some subsets of) the random points.

To answer this question formally, we introduce some notation: let $X^n = \{X_1, \ldots, X_n\}$ and let $\mathcal{C}_\lambda$ be the class of rectifiable curves of length $\leq \lambda$. For a curve $\gamma \in \mathcal{C}_\lambda$, let $X^n(\gamma)$ be the number of points $X_i$ found along $\gamma$. Let $N_n(\mathcal{C}_\lambda) = \max\{X^n(\gamma) : \gamma \in \mathcal{C}_\lambda\}$. In this paper we will establish:

**Theorem 1.** *For each $\lambda > 0$,*

$$\mathbf{P}\left\{1/5\ \lambda\sqrt{n} \leq N_n(\mathcal{C}_\lambda) \leq 17\ \lambda\sqrt{n}\right\} \to 1, \quad n \to \infty.$$

We mention this result not because it is our principal aim in this paper – it is actually an easy warm-up exercise – but to make concrete the kind of results we pursue in this paper.

### 1.1. Generalization

This 'Supermarket Sweep/PacMan' problem is a particular instance of a class which we call *connect-the-dots* (CTD) problems. In each such problem, we have a uniformly-scattered set of points, we have a class $\Gamma$ of curves $\gamma$, and we ask for the maximum number of points on any curve $\gamma \in \Gamma$. In the case just discussed, $\Gamma = \mathcal{C}_\lambda$, the collection of all finite-length curves with length at most $\lambda$, and we get order-of-magnitude estimates $O_P(\lambda\sqrt{n})$.

Before continuing with technical discussion, we comment on our terminology. To avoid confusion, we mention that CTD terminology is currently used in popular culture in two different ways, evoking different responses. On one hand, the term is used in discussing a classic children's coloring-book game—where a cloud of *non-random* points is presented to the budding artist whose task is to connect *every* dot in a particular sequence and see a picture emerge. We are *not* thinking of this usage. We think instead of recent usage in political discourse [32, 13, 26] where CTD refers to identifying a small subset of facts among many random conflicting ones, thereby detecting a subtle pattern. Thus, journalists writing in [32, 13, 26] all used the CTD phraseology to convey the failure of policymakers to winnow from many apparently random pieces of information at their disposal a few specific precursors to important events. The lengthier terminology "connect the dots amid heavy clutter" would also be appropriate, where the term "clutter" evokes the many irrelevant confusers that one does not connect. The modern journalistic usage of the CTD phrase is broadly consistent with our own usage; and with potential applications in signal detection and pattern recognition (see below).

Our study of CTD problems will consider several choices for $\Gamma$: twice-differentiable curves, graphs of functions of bounded variation, graphs of increasing functions, graphs of $m$-times-differentiable functions. We will see that in each case the maximum number of points on a curve in the given class will grow as $N_n(\Gamma) = O_P(n^\rho)$ where the *growth exponent* $\rho = \rho(\Gamma)$ depends on the massivity of the class $\Gamma$. Thus, we obtain the following results.

- *2-smooth Curves.* Let $\mathcal{C}_\lambda(2,\kappa)$ denote the class of twice differentiable curves taking values in $[0,1]^2$ with length $\leq \lambda$ and curvature $\leq \kappa$ pointwise. Then $\rho(\mathcal{C}_\lambda(2,\kappa)) = 1/3$.
- *Graphs of Bounded Variation.* Let $\mathrm{BVGr}_\tau$ denote the class of graphs $(x, f(x))$ taking values in $[0,1]^2$ where the total variation $\|f\|_{TV} \leq \tau$. Then $\rho(\mathrm{BVGr}_\tau) = 1/2$.
- *Lipschitz Graphs.* Let $\mathrm{LipGr}_\tau$ denote the class of graphs $(x, f(x))$ where $f$ is a function with range $[0,1]$ and slope controlled by $|f(x) - f(x')| \leq \tau|x - x'|$. Then $\rho(\mathrm{LipGr}_\tau) = 1/2$.
- *m-fold Differentiable Graphs.* Let $\mathrm{DiffGr}_{m,\tau}$ denote the class of graphs $(x, f(x))$ where $f$ is an $m$-times differentiable function with range $[0,1]$ and $\|f^{(m)}\|_\infty \leq \tau$. Then $\rho(\mathrm{DiffGr}_{m,\tau}) = 1/(m+1)$.

In all these results, the massivity of the class $\Gamma$ enters through the growth exponent of the $\varepsilon$-entropy of the class.

We generalize beyond point clouds in dimension 2, considering the case where $X_i$, $i = 1, \ldots, n$, are random points in the $d$-dimensional hypercube, with $d > 2$. In that setting, we also generalize the CTD problem from connecting dots using curves to connecting dots using hypersurfaces, and more generally using $k$-dimensional immersions with $1 \leq k \leq d-1$.

We also generalize the problem from connecting points to passing through points with tangents having prescribed orientations at those points; we call this the *connect-the-darts* problem.

### 1.2. Motivation

The CTD problem is interesting from several viewpoints.

- *Probability Theory.* CTD generalizes two known problems of considerable interest among probabilists.

  1. *Length of the Longest Increasing Subsequence (LIS).* Suppose we let IncrGr denote the class of increasing curves, i.e., of sets $\{(x, f(x)) : x \in [0,1]\}$ where $f : [0,1] \mapsto [0,1]$ is monotone increasing. Then $N_n(\text{IncrGr})$ measures the result of *last-passage percolation* [6]. Also, if we write $X_i = (x_i, y_i)$, let $\pi$ denote the sorting permutation $x_{\pi(1)} \leq x_{\pi(2)} \leq \ldots$, and define $w_i = y_{\pi(i)}$, $i = 1, \ldots, n$, then $N_n(\text{IncrGr})$ is the length of the *longest increasing subsequence* among the numbers $w_1, w_2, \ldots, w_n$. The problem of determining the asymptotic behavior of the length of the longest increasing subsequence of $n$ such random numbers (sometimes called *Ulam's Problem*) attracted considerable attention in the 1990's, with concentration of measure estimates [20], massive computational studies [30], and finally results on asymptotic distributions [7]. In the 1970's, Vershik, Logan, and Shepp [37, 27] showed that asymptotic behavior of $N_n(\text{IncrGr}) \sim 2\sqrt{n}$. (Groeneboom [21] gives a particular simple proof of this). The more delicate fluctuation distributional properties have been determined by Baik, Deift, and Johanssen [7] who showed that the asymptotic distribution follows the Tracy-Widom distribution [36]. It turns out that similar asymptotic results hold for a different CTD problem where the class $\Gamma$ is made of

Lipschitz graphs. The fact that study of CTD extends the range of such limit phenomena seems *per se* interesting, especially in view of other universality results regarding the Tracy-Widom distribution [34].

2. *Traveling Salesman Problem.* Probabilists and operations researchers have been interested for decades in the problem of determining the *shortest path* through *every* point in a cloud of $n$ uniform random points. This path length grows like $.7124\sqrt{n}$ , where .7124 is an approximation to the Beardswood-Halton-Hammersley constant [23]. The CTD problem considers instead the maximum number of points on a curve of fixed length $\lambda$ independent of $n$. While the two problems would thus be closely connected if $\lambda$ were variable, $\lambda = \lambda_n \approx .7124\sqrt{n}$, we do not consider this case in our approach, which leads to differences in application and interpretation — differences which might be stimulating to TSP researchers. In addition, an interesting connection between these two problems is described at the end of Section 2.1.2.

- *Geometric Discrepancy Theory.* Number theorists, harmonic analysts, and numerical analysts have long been interested in the problem of determining whether a set of points is nearly uniformly distributed. It is standard to measure the discrepancy from uniform by comparing the fraction of points in a set with the fraction of volume in that set, and one maximizes the discrepancy over a class of sets (rectangles, disks, convex sets, ...) [8, 28]. As a referee has pointed out, CTD could be considered a variant of this approach, maximizing discrepancy over classes of curves. Since classes of curves involve objects of zero volume, discrepancy boils down to measuring the maximal number of points on a curve. Despite the apparent differences – studying 'vanishingly thin sets' rather than 'thick' geometric objects – a quantitative connection is sketched in the Appendix; results on CTD imply bounds on the geometric discrepancy of random point sets.
- *Filament Detection.* CTD is relevant to inference problems in image analysis. Suppose we observe $n$ points scattered about the unit square. Consider the hypothesis testing problem:
  - Under $H_0$, the points are independent and random uniform on $[0,1]^2$.
  - Under $H_{1,\Gamma,n}$, the vast majority of the points are again independent, random

uniform on $[0,1]^2$, but a small fraction $\varepsilon_n$ of points are actually uniformly sampled at random along an unknown curve $\gamma \in \Gamma$.

This inference problem models data given as the output of a spatially distributed array of detectors, such as particle detectors in high energy physics [3] or, more recently, sensor networks forming a Smart Dust [24]. Sensor alarms caused by 'background' are 'false detections' uniformly scattered in space; sensor alarms caused by something interesting (a particle or intruder) are scattered along the path of the interesting object, but immersed in the irrelevant 'clutter' of 'false alarms'.

It is clear that if the fraction $\varepsilon_n$ defining the alternative hypothesis exceeds the typical behavior for $N_n(\Gamma)/n$ under $H_0$ then reliable detection is possible. A sketch of the idea is illustrated in Figure 1. Hence it is of some interest to determine the asymptotic behavior of $N_n(\Gamma)$, as we do in this paper. For more on such problems, see [4, 22].

- *Vision Research.* An interesting stream of vision research started with the two papers [19, 25]. Both experiments presented specially prepared images to human subjects who were asked to (quickly and reflexively) judge whether the images were 'purely random' or 'contained a curve buried in clutter.' In detail, the images showed a collection of graphical elements which, like in the signal detection problem mentioned above, were either purely randomly scattered or else contained, in addition to randomly scattered points, a small fraction scattered along a curve. Compare Figure 2. A special feature in this experiment was that the graphical elements were not points, but instead oriented patches.

  In one experiment, the patches were used as if they were simply points; their orientation was meaningless and chosen at random. In another experiment, the patches were used as tangents, i.e., when points were sampled from a curve, the patches were chosen tangent to the curve.

  A question of particular interest to vision researchers is: how does the detection performance of the human visual system compare to that of an *ideal observer*—a mathematically optimal detector? Our results provide such an ideal observer with which to compare human performance. They also explain why the second

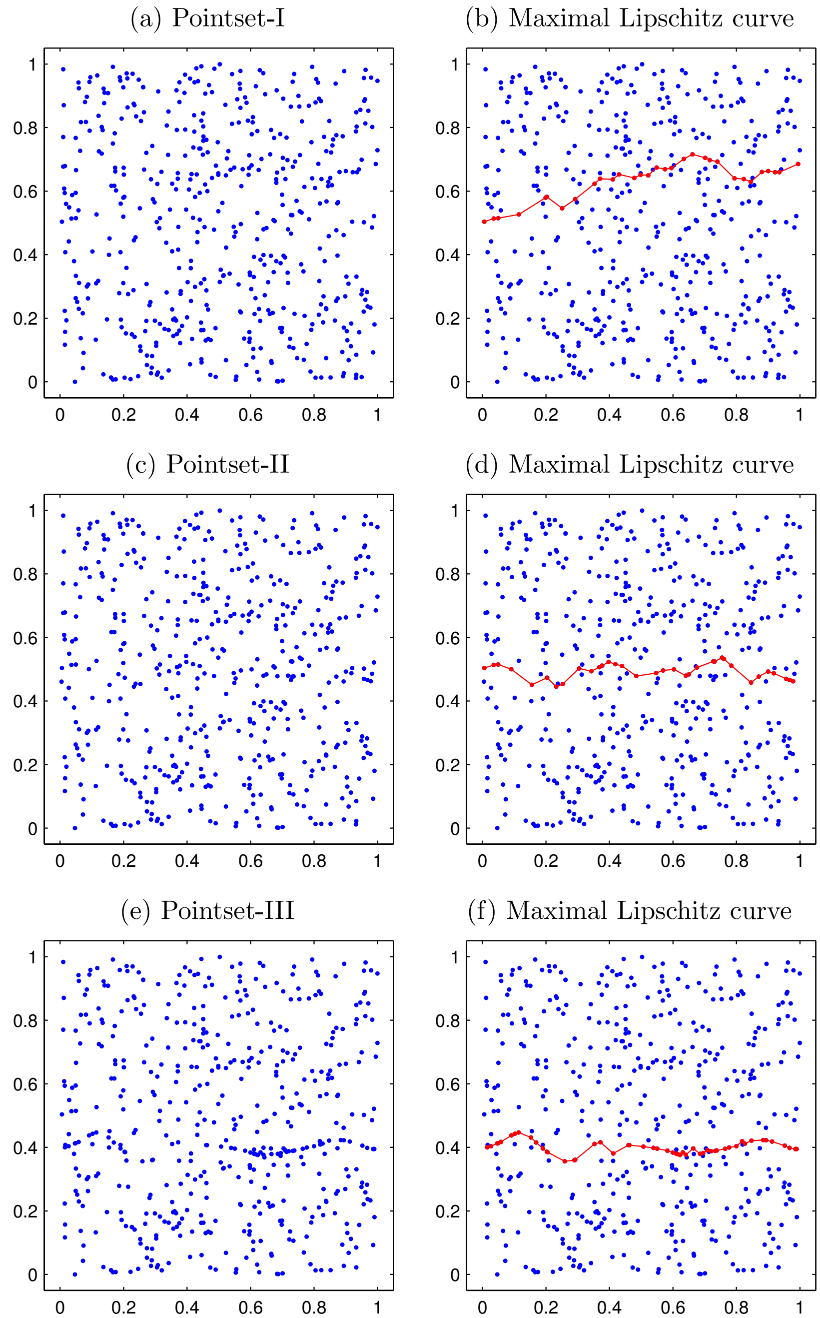



FIGURE 1: Examples for Connect-The-Dots. Three examples of scattered points in $[0, 1]$, and the corresponding CTD solution. In each case the class $\Gamma$ of curves is the set of Lipschitz graphs. Panels (a),(c), and (e) show the pointsets only, while panels (b), (d), and (f) show the maximal Lipschitz curves, with $31, 35, 54$ points respectively – out of $n = 500$ points total. Panel (a) shows a uniformly distributed random pointset. Panels (c) and (e) show clouds with a small number of points on a Lipschitz curve, in addition to uniform random points. It seems unlikely that visual inspection would detect non-uniform structure in (c) (compare (a)); however, a statistical test based on CTD counts can reliably establish its presence.

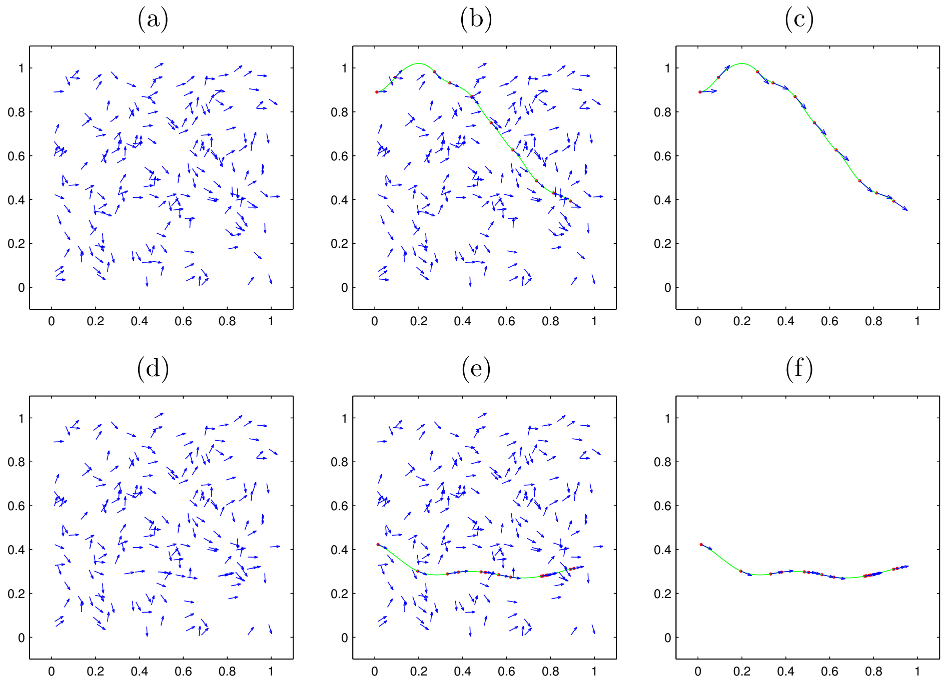



FIGURE 2: Examples for Connect-The-Darts. Panel (a) shows a collection $n = 200$ of 'darts' taking uniform random positions and directions; Panel (b) shows the maximal Hölder-2 function with first-order contact; Panel (c) extracts this function and its points of contact. Panels(d)-(f) are similar to (a)-(c), except that 15 of the 200 points are now nonuniform; they are chosen to be in first-order contact with a specific Hölder-2 function. In both cases, the Hölder constant $\beta = 15$. The power of the test is nearly 1.

experiment is substantially easier for humans than the first. Indeed, the second experiment involves not just passing through points, but passing through points at given angles. The growth exponents $\rho$ are smaller in such cases; this provides a rigorous and quantitative sense in which one can say that detection exploiting orientations can be more sensitive than detection based on points alone.

### 1.3. Contents

Our aim in this paper is to formalize a class of problems in stochastic geometry and give some initial results and methods. Accordingly, we provide in Section 2 a concrete approach to determining growth exponents for the class of curves of bounded length and curves of bounded curvature. Our concrete arguments use discrete structures we found useful in building the algorithms in [4]. We then develop in Section 3 an abstract approach based on $\varepsilon$-entropy which allows us to efficiently derive growth exponents in a wide range of classes; examples are given in that section, such as graphs of smooth functions and graphs of bounded variation functions. Section 4 applies this abstract machinery to the connect-the-darts case. Our main question about this class of problems concerns the set of questions beyond growth exponents. In Section 5 we consider the question of whether $N_n(\Gamma)/n^\rho$ tends to a limit in probability, and the behavior of the fluctuations of $N_n(\Gamma)$. Such results are known in the case of longest increasing subsequences mentioned above, and we review evidence indicating that they hold more generally. We point out that a simple application of a concentration of measure result of Talagrand's controls the standard deviation of $N_n(\Gamma)$ and gives exponential bounds on fluctuation of $N_n(\Gamma) - \text{median}\{N_n(\Gamma)\}$. Perhaps some probabilists will be inspired by this article to complete the picture and derive finer distributional properties of $N_n(\Gamma)$.

## 2. Concrete Approaches

We begin by discussing concrete methods for estimating the order of growth of $N_n(\Gamma)$, for two specific choices of $\Gamma$.

### 2.1. CTD with bounded length

Consider first the case $\Gamma = \mathcal{C}_\lambda$ mentioned in Theorem 1.1. We consider upper and lower bounds for $N_n(\mathcal{C}_\lambda)$ separately.

2.1.1. *Upper Bound.* We construct a graph $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$, associate paths in the graph with tubular sets in the plane, and estimate the number of points in such tubes. The vertex set $\mathcal{V}_n$ corresponds to the grid points $\{(k_1\varepsilon_1, k_2\varepsilon_1)\}$ with $k_i$ integer, $0 \leq k_1, k_2 \leq \sqrt{n} = 1/\varepsilon_1$. Consider the line segments joining grid points $(k_1\varepsilon_1, k_2\varepsilon_1)$ and $(k_1'\varepsilon_1, k_2'\varepsilon_1)$, with the restriction that $|k_1' - k_1| \leq 1$ and $|k_2' - k_2| \leq 1$. These line segments form the edges $\mathcal{E}_n$ of our graph $\mathcal{G}_n$. We permit $k_i' = k_i, i = 1, 2$, in which case the line segment consists of the single point $(k_1, k_2)$.

In this graph, a path $\pi$ is a sequence of vertices in $\mathcal{V}_n$ connected by edges in $\mathcal{E}_n$. Such a path $\pi$ has an *image* $\text{Im}(\pi)$ in the unit square, defined as the $\varepsilon$-neighborhood (in

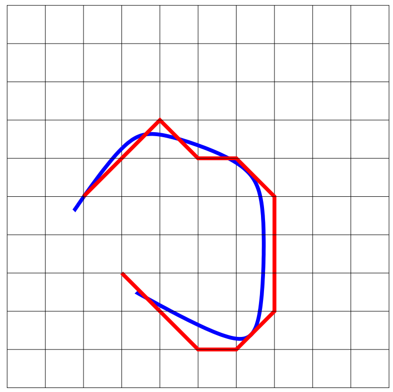


FIGURE 3: A rectifiable curve $\gamma$ and its associated path $\pi^n(\gamma)$ (piecewise linear).

Hausdorff distance) of the union of the line segments that $\pi$ traverses, where $\varepsilon = \frac{5}{4}\varepsilon_1$. The choice of $\varepsilon$ comes from the proof of Lemma 1 following.

**Lemma 1.** *Fix $\lambda > 0$ and assume $n > n_0$ large enough. For each curve $\gamma \in \mathcal{C}_\lambda$, there is a path $\pi$ through $\mathcal{G}_n$ whose image in the unit square covers $\gamma$, i.e., $\gamma \subset \mathrm{Im}(\pi)$. Moreover, $\pi$ may be chosen ($\pi = \pi^n(\gamma)$) so that it traverses at most $\lambda\sqrt{n} + 1$ vertices.*

See Figure 3. Fix $\gamma \in \mathcal{C}_\lambda$, and choose a unit-speed parametrization, denoted by $\gamma([0, \ell])$ where $\ell = \text{length}(\gamma)$. Consider an $\varepsilon_1/2$-covering of $[0, \ell]$, denoted by $\{s_j : j = 1, \ldots, J\}$, chosen so that $J \le \ell/\varepsilon_1 + 1 \le \lambda/\varepsilon_1 + 1$. For each $j = 1, \ldots, J$, let $b_j$ be a closest grid point to $\gamma(s_j)$. Note that $|\gamma(s_j) - b_j| \le \varepsilon_1/\sqrt{2}$. Since

$$\begin{aligned} |b_{j+1} - b_j| &\le |\gamma(s_{j+1}) - b_{j+1}| + |\gamma(s_{j+1}) - \gamma(s_j)| + |\gamma(s_j) - b_j| \\ &\le \varepsilon_1/\sqrt{2} + \varepsilon_1/2 + \varepsilon_1/\sqrt{2} \\ &< 2\varepsilon_1, \end{aligned}$$

the line segment $[b_j, b_{j+1}]$ is in $\mathcal{E}_n$. Therefore, we may define $\pi$ to be the path $b_1, \ldots, b_J$. Moreover, since for $s_j \le s \le s_{j+1}$,

$$\max\{|\gamma(s) - \gamma(s_{j+1})|, |\gamma(s) - \gamma(s_j)|\} \le \varepsilon_1/2,$$

the piece of curve $\gamma([s_j, s_{j+1}])$ is within Hausdorff distance $\varepsilon_1/2 + \varepsilon_1/\sqrt{2} < \frac{5}{4}\varepsilon_1$ from the edge $[b_j, b_{j+1}]$. Thus $\pi$ passes through at most $\lambda\sqrt{n}$ vertices and its image contains $\gamma$. □

Extend our definition of $X^n(S)$ so that, whenever $S$ is a subset $S \subset [0,1]^2$, not necessarily a curve, then $X^n(S) = \#\{i : X_i \in S\}$ (i.e., the size of $S$) and so that, whenever $\mathcal{S}$ is a collection of subsets $S \in [0,1]^2$, not necessarily a collection of curves, then $N_n(\mathcal{S}) = \max\{X^n(S) : S \in \mathcal{S}\}$. Define also $Y^n$ so that, if $\pi$ is a path in $\mathcal{G}_n$, then $Y^n(\pi) = X^n(\mathrm{Im}(\pi))$; and for $\Pi$ a family of paths, put $M_n(\Pi) = \sup\{Y^n(\pi) : \pi \in \Pi\}$.

Let now

$$\Pi_\lambda^n = \{\pi^n(\gamma) : \gamma \in \mathcal{C}_\lambda\}.$$

It follows from our definition of $Y^n$ and the previous lemma that $N_n(\mathcal{C}_\lambda) \le M_n(\Pi_\lambda^n)$, which in turn implies

$$\mathbf{P}\left\{N_n(\mathcal{C}_\lambda) > B\ \lambda\sqrt{n}\right\} \le \mathbf{P}\left\{M_n(\Pi_\lambda^n) > B\ \lambda\sqrt{n}\right\}.$$

Since $\Pi_\lambda^n$ is finite, Boole's inequality gives, for all $B > 0$,

$$\mathbf{P}\left\{M_n(\Pi_\lambda^n) > B\ \lambda\sqrt{n}\right\} \le \#(\Pi_\lambda^n)\ \cdot\ \max_{\pi \in \Pi_\lambda^n} \mathbf{P}\left\{Y^n(\pi) > B\ \lambda\sqrt{n}\right\}.$$

**Lemma 2.** *For $n > n_0(\lambda)$, $Y^n$ and $\Pi_\lambda^n$ have these properties:*

1. $\#(\Pi_\lambda^n) \le 1.1n \cdot 9^{\lambda\sqrt{n}}$;
2. *For any path $\pi \in \Pi_\lambda^n$, $Y^n(\pi) \backsim \mathrm{Bin}(n, |\pi|)$, where $\mathrm{Bin}(n,p)$ denotes the usual binomial distribution and $|\pi|$ is the area of $\mathrm{Im}(\pi)$; moreover, $|\pi| \le 6\lambda/\sqrt{n}$.*

Property 1 simply combines the fact that every element in $\Pi_\lambda^n$ is a chain of at most $\lambda\sqrt{n}+1$ vertices, with the observations that, from each vertex, there are (at most) 9 possibilities for the next vertex on the path, and that there are $(\sqrt{n}+1)^2 \le 1.1n$ (for sufficiently large $n$) possible starting vertices. For property 2, $\mathrm{Im}(\pi)$ is contained in the union of the $5/4\ n^{-1/2}$-neighborhoods of the line segments that $\pi$ connects; there are no more than $\lambda\sqrt{n}+1$ such regions and each one of them has area not exceeding $(\sqrt{2}\varepsilon_1 + 2\varepsilon)\varepsilon \le 11/(2n)$.

Using this lemma, we obtain the bound, valid for all $B > 0$,

$$\max_{\pi \in \Pi_\lambda^n} \mathbf{P}\left\{Y^n(\pi) > B\ \lambda\sqrt{n}\right\} \le \mathbf{P}\left\{\mathrm{Bin}(n, 6\lambda/\sqrt{n}) > B\ \lambda\sqrt{n}\right\}.$$

Hoeffding's inequality [33] gives us control over the tail of the Binomial distribution:

**Lemma 3.** *For $C > 2$,*

$$\mathbf{P}\left\{\mathrm{Bin}(n,p) > C\ np\right\} \leq \varphi(C)^{np},$$

*where*

$$\varphi(C) = \exp\left(-\frac{3(C-1)^2}{2(C+2)}\right). \tag{2.1}$$

We get immediately that for $B > 1$ and $n > n_0(\lambda)$,

$$\mathbf{P}\left\{\mathrm{Bin}(n, 6\lambda/\sqrt{n}) > B\ \lambda\sqrt{n}\right\} \leq \varphi(B/6)^{6\lambda\sqrt{n}}.$$

Combining all the above,

$$\begin{aligned}\mathbf{P}\left\{N_n(\mathcal{C}_\lambda) > B\ \lambda\sqrt{n}\right\} &\leq 1.1n9^{\lambda\sqrt{n}}\ \varphi(B/6)^{6\lambda\sqrt{n}} \\ &= 1.1n\exp\left(-\lambda\sqrt{n}(-6\log(\varphi(B/6)) - \log(9))\right).\end{aligned}$$

Now, choose $B = 17$ so that $-6\log(\varphi(B/6)) > \log(9)$, making the right-hand side tend to zero as $n$ tends to infinity. This proves the upper bound in Theorem 1.

We remark here that the general idea of the upper bound proof is to find a small enough (Property 1) set of regions, each of which has small enough area (Property 2), such that every curve in the class is contained in a region. The abstraction in Section 3 treats the regions as a set of $\varepsilon$-balls in the associated Haussdorff metric, such that they cover the class of curves. Property 1 corresponds to the volume of each $\varepsilon$-ball being small enough, and Property 2 corresponds to the $\varepsilon$-entropy (log of the least number of balls in an $\varepsilon$-covering) being small enough.

2.1.2. *Lower Bound.* To control $N_n(\mathcal{C}_\lambda)$ from below, we start from the $n$ random points in $X^n$. We analyze these points and extract a random subset of $J_{n,\lambda}$ ordered points $p_1, p_2, \ldots$. We connect successive points by linear interpolation. We argue that the result has length $\leq \lambda$ and then show that, with probability tending to one, $J_{n,\lambda} \geq A\ \lambda\sqrt{n}$.

To extract the points, subdivide $[0,1]^2$ as in Section 2.1.1 into $\varepsilon_1 \times \varepsilon_1$ squares. We define a zig-zag ordering of the squares starting in the upper left corner. See below.

| | | | |
|---|---|---|---|
| $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ |
| $Q_8$ | $Q_7$ | $Q_6$ | $Q_5$ |
| $Q_9$ | $Q_{10}$ | $Q_{11}$ | $Q_{12}$ |
| $Q_{16}$ | $Q_{15}$ | $Q_{14}$ | $Q_{13}$ |

Let $I = \{i : X^n(Q_i) \geq 1\}$, indexing the non-empty squares. For each $i \in I$, pick one distinguished point $p_i \in Q_i$. We arrive in this way at a well-defined sequence of points $(p_i)$. Connect successive points by linear interpolation, stopping just when the total length of the curve reaches $\lambda$. Denote the constructed curve by $\gamma_{n,\lambda}$ (it is random, as its construction depends on $X^n$). Certainly it belongs to $\mathcal{C}_\lambda$.

Let $J_{n,\lambda}$ denote the (random) number of non-empty squares in the first $\lfloor \lambda\sqrt{n}/\sqrt{2} \rfloor$ squares of our ordering. Even if all the squares in this initial set were non-empty, connecting points associated with such squares cannot lead to a length exceeding $\lambda$. Hence $Y^n(\gamma_{n,\lambda}) \geq J_{n,\lambda}$; and so, $N_n(\mathcal{C}_\lambda) \geq J_{n,\lambda}$. Now $J_{n,\lambda}$ is essentially a binomial random variable. Elementary calculations in Appendix A.1 show:

**Lemma 4.**

$$\mathbf{P}\left\{J_{n,\lambda} \geq 1/5\ \lambda\sqrt{n}\right\} \to 1, n \to \infty.$$

The lower bound in Theorem 1 follows.

The above probabilistic analysis used a “Selection + Interpolation” approach which can serve as a common framework in later results. In this particular setting, a non-probabilistic analysis is also possible, invoking previous research about the Travelling Salesman Problem (TSP). Nearly half century ago, L. Few proved that given $n$ points on a unit square, there is a curve of length not exceeding $\sqrt{2n} + 7/4$ that traverses *all* the $n$ points [18]. Divide such a path into $m = \lceil \frac{\sqrt{2n}+7/4}{\lambda} \rceil$ consecutive pieces $\gamma_1, \ldots, \gamma_m$, with length $\leq \lambda$. By definition, $N_n(\mathcal{C}_\lambda)$ is larger than each $X^n(\gamma_i)$, and so larger than their average. Hence,

$$N_n(\mathcal{C}_\lambda) \geq \frac{X^n(\gamma_1) + \cdots + X^n(\gamma_m)}{m} = \frac{n}{m} \approx \frac{1}{\sqrt{2}}\lambda\sqrt{n}.$$

### 2.2. CTD with bounded curvature

We now increase the assumed regularity on our curves, controlling both the curvature and length. Let then $\mathcal{C}_\lambda(2, \kappa)$ denote the class of $C^2$ curves with length bounded by $\lambda$

and curvature bounded pointwise by $\kappa$. We are interested in $N_n(\mathcal{C}_\lambda(2,\kappa))$.

Note that a constraint on $\kappa$ may imply a constraint on the length. The 'turning radius' for a curve in the plane is of course $1/\kappa$, so, if a curve has curvature $\leq \kappa$ and its image lies in $[0,1]^2$, then, for small enough $\kappa$, the curve will not be able to 'wind around' within the unit square; this constrains achievable lengths. For a given bound $\kappa$ on curvature, let $\lambda_0(\kappa)$ give the maximum length subject to that bound. There is a threshold $\kappa_0$ so that $\lambda_0(\kappa) < \infty$ for $\kappa < \kappa_0$. In the extreme where $\kappa \to 0$, curves in $\mathcal{C}_\lambda(2,\kappa)$ are almost straight, which forces their maximal length $\lambda_0(\kappa) \to \sqrt{2}$ as $\kappa \to 0$. Hence, if $\kappa < \kappa_0$, we assume $\lambda < \lambda_0(\kappa)$, ensuring that at least one curve in $\mathcal{C}_\lambda(2,\kappa)$ has length $\geq \lambda$.

By imposing the curvature constraint we dramatically reduce the maximum number of points which can lie on a curve.

**Theorem 2.** *There exist $A, B > 0$, so that, for each pair $(\lambda,\kappa)$ with $0 < \lambda < \lambda_0(\kappa)$,*

$$\mathbf{P}\left\{A \cdot \lambda\kappa^{1/3}n^{1/3} \leq N_n(\mathcal{C}_\lambda(2,\kappa)) \leq B \cdot \lambda\kappa^{1/3}n^{1/3}\right\} \to 1, n \to \infty.$$

In particular, the growth exponent $\rho(\mathcal{C}_\lambda(2,\kappa)) = 1/3$ is considerably smaller than the value $\rho(\mathcal{C}_\lambda) = 1/2$, when curvature was unconstrained. This is reflected in the difference between Figure 3 and Figure 8.

We again divide the proof into separate arguments for upper and lower bounds.

2.2.1. *Upper Bound.* As in the curvature-unconstrained case, we construct a graph, associate paths in the graph with tubular sets in the plane, and estimate the number of points in such tubes. The graph will have a more complex structure than before, and the association of paths to regions in the plane will reflect the underlying reason for the $\rho = 1/3$ growth exponent.

Our graph $\mathcal{G}_n$ has a vertex set $\mathcal{V}_n$ with vertices corresponding to special planar line segments called *beamlets*, defined as follows. We first define special collections of grid points, then join pairs of endpoints to form our special line segments. Consider the vertical (resp. horizontal) grid points $\{(k_1\varepsilon_1, k_2\varepsilon_2)\}$ (resp. $\{(k_2\varepsilon_2, k_1\varepsilon_1)\}$), with $0 \leq k_1 \leq 1/\varepsilon_1$ and $0 \leq k_2 \leq 1/\varepsilon_2$, where $\varepsilon_i = 2^{-m_i}$ with

- $m_2 = 1/3\ (2\lceil\log_2(n)\rceil - \lfloor\log_2(\kappa)\rfloor)$, so that $\varepsilon_2 \approx \kappa^{1/3}n^{-2/3}$, and

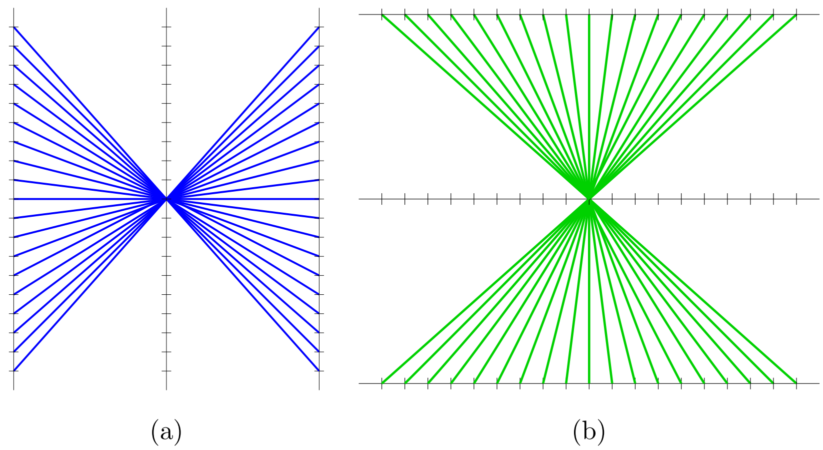



FIGURE 4: Panel (a): all the horizontal beamlets passing through a vertical grid point; Panel (b): all vertical beamlets passing through a horizontal grid point.

- $m_1 = 1/2 \ (m_2 + k + \lceil \log_2(\kappa) \rceil)$, where $k$ is a universal (integer) constant defined in Appendix A.2.

We then consider line segments that join two vertical (resp. horizontal) grid points such that their angle with the horizontal (resp. vertical) direction does not significantly exceed $45^o$; formally, such a line segment is defined by its endpoints $(k_1\varepsilon_1, k_2\varepsilon_2)$ and $(k_1'\varepsilon_1, k_2'\varepsilon_2)$ (resp. $(k_2\varepsilon_2, k_1\varepsilon_1)$ and $(k_2'\varepsilon_2, k_1'\varepsilon_1)$), with the restriction that $|k_1' - k_1| = 1$ and $|k_2' - k_2| < \varepsilon_1/\varepsilon_2 + 1$. Such line segments will be called *horizontal (resp. vertical) beamlets*. See Figure 4. Roughly speaking, the beamlets have length comparable to $n^{-1/3}$ and slope chosen from a grid of spacing $\asymp n^{-2/3}$.

We also consider line segments joining vertical and horizontal grid points belonging to the same square and making an angle close to $45^o$ with the horizontal direction; formally, such a line segment is defined by its endpoints $(k_1\varepsilon_1, k_2\varepsilon_2)$ and $(k_2'\varepsilon_2, k_1'\varepsilon_1)$, with the restriction that, if $k_2 = k_{21}\varepsilon_1 + k_{22}\varepsilon_2$ with $0 < k_{22} < \varepsilon_1/\varepsilon_2$,

- $k_1' = k_{21}$ or $k_{21} + 1$, and
- $|k_2' - (k_1\varepsilon_1/\varepsilon_2 + k_{22})| \leq 1$ or $|k_2' - (k_1\varepsilon_1/\varepsilon_2 - k_{22})| \leq 1$.

Such line segments will be called diagonal beamlets. See Figure 5.

The set of vertices $\mathcal{V}_n$ is made of all beamlets defined above. They provide an efficient, dyadically organized, multiscale, multi-orientation organizational structure. The graphical structure we build using them is a variant of the beamlet graph defined

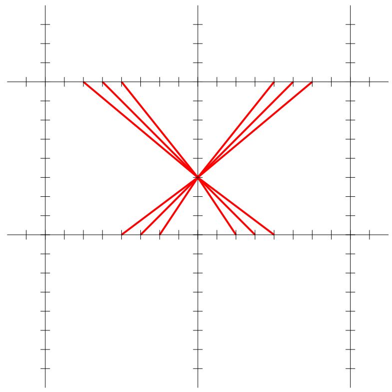
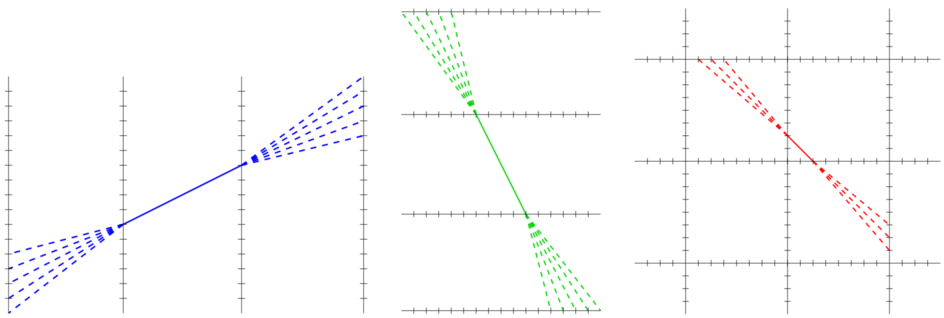


FIGURE 5: Diagonal beamlets passing through a vertical grid point.



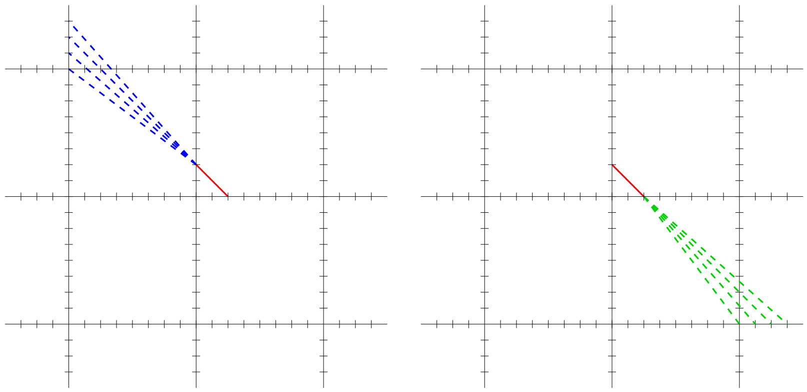
(a) (b) (c)

FIGURE 6: Panel (a): A horizontal beamlet and its horizontal neighbors; Panel (b): a vertical beamlet and its vertical neighbors; Panel (c): a diagonal beamlet and its diagonal neighbors.

in [15, 5, 14].

The set of edges in $\mathcal{G}_n$, which is denoted by $\mathcal{E}_n$, links any two line segments in $\mathcal{V}_n$ that are in "good continuation", which here means their directions are close enough [4]. Formally, two horizontal beamlets connected in $\mathcal{E}_n$ are of the form $[(k_1\varepsilon_1, k_2\varepsilon_2), ((k_1+1)\varepsilon_1, k_2'\varepsilon_2)]$ and $[((k_1+1)\varepsilon_1, k_2'\varepsilon_2), ((k_1+2)\varepsilon_1, k_2''\varepsilon_2)]$, with the restriction that $|k_2'' - 2k_2' + k_2| \leq 2$. Similar statements hold for two vertical beamlets or two diagonal beamlets. See Figure 6. Moreover, a diagonal beamlet and a horizontal (or vertical) beamlet are connected if they share an endpoint and are in good continuation in a similar way. See Figure 7.

In graph $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$, a path $\pi$ is a sequence of vertices in $\mathcal{V}_n$ connected by edges in $\mathcal{E}_n$. Such a path $\pi$ has an image $\mathrm{Im}(\pi)$ in the unit square: the tubular region defined as



(a) (b)

FIGURE 7: A diagonal beamlet and its neighboring horizontal (resp. vertical) beamlets [Panel (a)] (resp. [Panel (b)]).

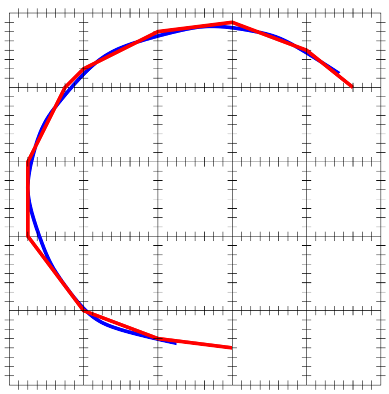


FIGURE 8: A smooth curve $\gamma$ and its associated path $\pi^n(\gamma)$ (piecewise linear).

the $\varepsilon$-neighborhood of the union $\cup_{p \in \pi} p$ of beamlets visited by $\pi$, where $\varepsilon = \kappa^{1/3} n^{-2/3}$. The following analog of Lemma 1 is proved in Appendix A.2.

**Lemma 5.** *For each curve $\gamma \in \mathcal{C}_\lambda(2, \kappa)$, there is a path $\pi \in \mathcal{G}_n$ such that $\mathrm{Im}(\pi)$ covers $\gamma$. Moreover, $\pi$ may be chosen ($\pi = \pi^n(\gamma)$) so that it visits at most $c\lambda\kappa^{1/3} n^{1/3}$ vertices of $\mathcal{G}_n$.*

See Figure 8.

Define now

$$\Pi^n_{\lambda,\kappa} = \{\pi^n(\gamma) : \gamma \in \mathcal{C}_\lambda(2, \kappa)\}.$$

Define also $Y^n(\pi) = X^n(\mathrm{Im}(\pi))$, and $|\pi| = \mathrm{Area}(\mathrm{Im}(\pi))$.

**Lemma 6.** *$Y^n$ and $\Pi^n_{\lambda,\kappa}$ satisfy the following properties:*

1. *$\#\Pi^n_{\lambda,\kappa} \leq cn^2\ 14^{c\lambda\kappa^{1/3}n^{1/3}}$;*
2. *For any path $\pi \in \Pi^n_{\lambda,\kappa}$, $Y^n(\pi) \backsim \mathrm{Bin}(n, |\pi|)$ with $|\pi| \leq c\lambda\kappa^{1/3}n^{-2/3}$.*

The first property comes from the fact that vertices in $\mathcal{G}_n$ have degree $\leq 14$, and there are fewer than $(2\varepsilon_1^{-1}\varepsilon_2^{-1})^2 \leq cn^2$ vertices total. The second comes from the fact that $\mathrm{Im}(\pi)$ is contained in the union of the $\kappa^{1/3}n^{-2/3}$-neighborhoods of the beamlets that $\pi$ connects; there are at most $c\lambda\kappa^{1/3}n^{1/3}$ beamlets in $\pi$ and for each of them, its $\kappa^{1/3}n^{-2/3}$-neighborhood has area bounded by $cn^{-1}$.

With these lemmas established, the upper bound follows in the same way as in the corresponding proof of Theorem 1.

2.2.2. *Lower Bound.* Here again, as in Subsection 2.1.2, we subdivide into boxes, select one point per box, and interpolate. Details are given in Appendix A.3.

## 3. Abstraction and Generalization

As implied in the Introduction, the two results proved so far are merely examples of a much wider class of possible results. Here is a far-reaching example. Suppose now that $X_1, \dots, X_n$ are points in $[0,1]^d$; we ask: how many of these points can possibly lie on a $k$-dimensional surface with smoothness index $\alpha$?

Our answer requires a more formal statement of the question.

- For $\alpha$ a positive integer, let $\mathcal{H}^{k,1}(\alpha,\beta)$ denote the functions $\{g : [0,1)^k \to [0,1)\}$ obeying
$$|g^{(\alpha_1,\dots,\alpha_k)}(s) - g^{(\alpha_1,\dots,\alpha_k)}(t)| \leq (\alpha-1)!\beta\|s-t\|,$$
for all $s,t \in [0,1)^k$ and all $(\alpha_1,\dots,\alpha_k) \in \mathbb{N}^k$ such that $\alpha_1 + \cdots + \alpha_k = \alpha - 1$.
- For $\alpha > 1$ not an integer, let $\mathcal{H}^{k,1}(\alpha,\beta)$ denote the functions $\{g : [0,1)^k \to [0,1)\}$ obeying
$$|g^{(\alpha_1,\dots,\alpha_k)}(s) - g^{(\alpha_1,\dots,\alpha_k)}(t)| \leq (\lfloor\alpha\rfloor!)\beta\|s-t\|^{\{\alpha\}},$$
for all $s,t \in [0,1)^k$ and all $(\alpha_1,\dots,\alpha_k) \in \mathbb{N}^k$ such that $\alpha_1 + \cdots + \alpha_k = \lfloor\alpha\rfloor$.

Here $\lfloor\cdot\rfloor$ denotes the integer part and $\{\cdot\}$ denotes fractional part.

Speaking very loosely, $\beta$ measures the size of the $\alpha$-th derivative. Define, for $k < d$, a class of $k$-dimensional immersions in $\mathbb{R}^d$:

$$\mathcal{H}^{k,d}(\alpha,\beta) = \{g = (g_1,\dots,g_d) : g_j \in \mathcal{H}^{k,1}(\alpha,\beta), j = 1,2,\dots,d\},$$

and finally, consider the corresponding class of 'surfaces' viewed merely as point sets:

$$\mathcal{S}^{k,d}(\alpha,\beta) = \{g([0,1)^k) : g \in \mathcal{H}^{k,d}(\alpha,\beta)\}.$$

Let $N_n(\mathcal{S}^{k,d}(\alpha,\beta))$ be the maximum number of uniform random points lying on any such surface.

**Theorem 3.** *For some $A, B > 0$ and a function $\tau(k,d,\alpha,\beta)$,*

$$P\{A\ \tau\ n^{1/(1+\alpha(d/k-1))} \le N_n(\mathcal{S}^{k,d}(\alpha,\beta)) \le B\ \tau\ n^{1/(1+\alpha(d/k-1))}\} \to 1, \quad n \to \infty.$$

One could approach this by adapting our previous methods; however, we prefer to develop an abstract upper bound, which we then apply to Theorem 3 as well as several other examples.

### 3.1. Upper Bound, Abstract Setting

Let $\mathbb{X}$ be a probability space, with probability measure denoted by $\mu$. Let $X_1,\dots,X_n \in \mathbb{X}$ be independent with common distribution $\mu$. We consider $\mathcal{S}$, a class of subsets of $\mathbb{X}$, and $N_n(\mathcal{S}) = \max_{S\in\mathcal{S}} X^n(S)$ – where, as before, if $S \subset \mathbb{X}$, $X^n(S) = \#\{i : X_i \in S\}$. Let $\delta$ be a semi-metric on $\mathbb{X}$ and let $\Delta$ be the corresponding Hausdorff semi-metric on the subsets of $\mathbb{X}$. For a subset $S \subset \mathbb{X}$, and $\varepsilon > 0$, let $(S)_\varepsilon$ denote the $\varepsilon$-neighborhood of $S$ in the $\Delta$ semi-metric.

We denote by $H(\varepsilon)$ the $\varepsilon$-entropy of $\mathcal{S}$ with respect to $\Delta$. Let $\{S_i : i = 1,\dots,\exp(H(\varepsilon))\}$ be a minimal $\varepsilon$-covering with respect to $\Delta$. Let $S \in \mathcal{S}$ and $i$ such that $\Delta(S,S_i) \le \varepsilon$. By definition, $S \subset (S_i)_\varepsilon$. Therefore, $N_n(\mathcal{S}) \le \max_i X^n((S_i)_\varepsilon)$.

Measure the volume of $\varepsilon$-neighborhoods of sets in $\mathcal{S}$ with:

$$M(\varepsilon) = \sup_{S\in\mathcal{S}} \mu\{(S)_\varepsilon\}.$$

**Lemma 7.** (Upper Bound.) *Suppose there are $a, b > 0$ and $c_1, c_2 > 0$ such that $H(\varepsilon) \le c_1\ \varepsilon^{-b}$ and $M(\varepsilon) \le c_2\ \varepsilon^a$. Then, for $B$ large enough,*

$$\mathbf{P}\left\{N_n(\mathcal{S}) > B\ n^{b/(a+b)}\right\} \to 0, \quad n \to \infty. \tag{3.1}$$

*Proof.* We just saw that

$$\mathbf{P}\left\{N_n(\mathcal{S}) > u\right\} \leq \mathbf{P}\left\{\max_i X^n((S_i)_\varepsilon) > u\right\}.$$

Since there are a finite number of $S_i$, we are able to use Boole's inequality:

$$\mathbf{P}\left\{\max_i X^n((S_i)_\varepsilon) > u\right\} \leq \exp\{H(\varepsilon)\} \ \max_i \ \mathbf{P}\left\{X^n((S_i)_\varepsilon) > u\right\}.$$

Now, for a given $S$, $X^n((S)_\varepsilon)$ follows a binomial distribution

$$X^n((S)_\varepsilon) \backsim \mathrm{Bin}(n, \mu\{(S)_\varepsilon\}).$$

Hence,

$$\max_i \mathbf{P}\left\{(S_i)_\varepsilon > u\right\} \leq \mathbf{P}\left\{\mathrm{Bin}(n, c_2\varepsilon^a) > u\right\},$$

since $\mu\{(S)_\varepsilon\} \leq c_2\varepsilon^a$ for all $S \in \mathcal{S}$.

Using again Hoeffding as in Lemma 3, and with $\varphi$ as defined there,

$$\mathbf{P}\left\{\mathrm{Bin}(n, c_2\varepsilon^a) > C \cdot c_2 n\varepsilon^a\right\} \leq \varphi(C)^{c_2 n\varepsilon^a}.$$

Hence,

$$\mathbf{P}\left\{N_n(\mathcal{S}) > C \cdot c_2 n\varepsilon^a\right\} \leq \exp\{c_1\varepsilon^{-b} + c_2 n\varepsilon^a \log\varphi(C)\}. \tag{3.2}$$

We choose $\varepsilon = \varepsilon_n$ solving $\varepsilon^{-b} = n\varepsilon^a$. Then, (3.2) becomes

$$\mathbf{P}\left\{N_n(\mathcal{S}) > C \cdot c_2 n\varepsilon^a\right\} \leq \exp\{(c_1 + c_2\log\varphi(C))n^{b/(a+b)}\}. \tag{3.3}$$

For $C$ such that $c_1 + c_2\log\varphi(C) < 0$, the right-hand side in (3.3) tends to zero as $n$ increases. Rewriting the left-hand side of (3.3), we get

$$\mathbf{P}\left\{N_n(\mathcal{S}) > C \cdot c_2 n^{b/(a+b)}\right\} \to 0, \qquad n \to \infty,$$

for $C$ large enough. (3.1) follows upon defining $B \equiv C \cdot c_2$. From $\varphi(C) \leq \exp\{-C/6\}$ for $C > 2$, we see that (3.1) holds for $B \geq 6\max\{c_2, c_1\}$. $\square$

### 3.2. Application to Hölder Immersions

We now use the abstract approach to prove the upper bound in Theorem 3. In that setting, $\mathbb{X} = [0,1]^d$, $\mu$ is the uniform density on $\mathbb{X}$, $\delta$ is the usual Euclidean distance and $\Delta$ is the usual Hausdorff distance. The classes $\mathcal{H}^{k,d}(\alpha,\beta)$ and $\mathcal{S}^{k,d}(\alpha,\beta)$ were defined at the beginning of Section 3.

Our first step is to transfer the calculation of entropies from the class of sets to the function class. The following is basically self-evident:

**Lemma 8.** *Let $f, g : [0,1]^k \mapsto [0,1]^d$. Then with $f([0,1]^k)$ denoting the image of $[0,1]^k$ under $f$, and similarly for $g([0,1]^k)$,*

$$\Delta(f([0,1]^k), g([0,1]^k)) \leq \|f - g\|_\infty. \tag{3.4}$$

Let now $H_\varepsilon(\mathcal{S}^{k,d}(\alpha,\beta);\Delta)$ denote the $\varepsilon$-entropy of the collection of sets $\mathcal{S}^{k,d}(\alpha,\beta)$ with respect to Hausdorff distance, and let $H_\varepsilon(\mathcal{H}^{k,d}(\alpha,\beta), |\cdot|_\infty)$ be the $\varepsilon$-entropy of the function class $\mathcal{H}^{k,d}(\alpha,\beta)$ with respect to the supnorm. From Lemma 8, we get

$$H_\varepsilon(\mathcal{S}^{k,d}(\alpha,\beta);\Delta) \leq H_\varepsilon(\mathcal{H}^{k,d}(\alpha,\beta), |\cdot|_\infty), \qquad \varepsilon \in (0,1).$$

We now recall the very well-known results on $\varepsilon$-entropy of Hölder classes:

**Theorem 4.** (Kolmogorov and Tikhomirov, 1958.)

$$H_\varepsilon(\mathcal{H}^{k,d}(\alpha,\beta), |\cdot|_\infty) \leq c_{k,d,\alpha}\ (\beta/\varepsilon)^{k/\alpha}, \qquad \varepsilon \in (0,1).$$

Finally we estimate the key volumic quantity $M(\varepsilon)$. We note the connection with the notion of Minkowski content (see [17]). Indeed, fix $S \in \mathcal{S}^{k,d}(\alpha,\beta)$; then,

$$\mu((S)_\varepsilon)/(v_{d-k}\ \varepsilon^{d-k}) \to \mathrm{vol}(S), \qquad \varepsilon \to 0,$$

where $v_{d-k}$ denotes the volume of the $(d-k)$-dimensional unit ball, and $\mathrm{vol}(S)$ is the $k$-dimensional Hausdorff measure of $S$ (see [17], 3.2.39). We have the following lemma:

**Lemma 9.** *Fix $\alpha \geq 1$, $\beta > 0$, $k$, $d$. Set*

$$M(\varepsilon) = \sup_{S \in \mathcal{S}^{k,d}(\alpha,\beta)} \mu((S)_\varepsilon).$$

*Then, for $\varepsilon$ small enough,*

$$M(\varepsilon) \leq c_{k,d}\ \beta^k\ \varepsilon^{d-k}.$$

First, notice that, if $\alpha \geq 1$, we have (by repeated integration by parts),

$$\mathcal{S}^{k,d}(\alpha,\beta) \subset \mathcal{S}^{k,d}(1,\beta).$$

Therefore, WLOG, we assume $\alpha = 1$.

Let $S \in \mathcal{S}^{k,d}(1,\beta)$ and $g \in \mathcal{H}^{k,d}(1,\beta)$ such that $g([0,1]^k) = S$. Fix $\varepsilon > 0$ and take a $\varepsilon/\beta$-covering of $[0,1]^k$ – call it $\{s_j : j = 1,\dots,J\}$. Note that $J \le \min\{v_k^{-1}(\beta/\varepsilon)^k, 1\}$, where $v_k$ is the unit ball in $k$ dimensions. Let $x \in (S)_\varepsilon$; by definition, there is $s \in [0,1]^k$ such that $|g(s) - x| \le \varepsilon$. Also by definition, there is $j$ such that $|s - s_j| \le \varepsilon/\beta$, which implies $|g(s) - g(s_j)| \le \varepsilon$. Therefore,

$$(S)_\varepsilon \subset \cup_{j=1}^J B(g(s_j), 2\varepsilon).$$

This implies, for $d, k, \beta$ fixed and $\varepsilon$ small,

$$\mu((S)_\varepsilon) \le \sum_{j=1}^J \mu(B(g(s_j), 2\varepsilon)) = J v_d (2\varepsilon)^d \le c\ 2^d v_d / v_k\ \beta^k\ \varepsilon^{d-k}.$$

□

We now combine these results with the Upper Bound Lemma 7 and obtain the upper bound in Theorem 3.

We don't know an abstract lower bound technique. Using the select/interpolate method, we give a lower bound for Hölder objects in Appendix A.4. It matches the order of magnitude of upper bound and completes the proof of Theorem 3.

### 3.3. Graphs of Bounded Variation

To illustrate the generality of the abstract upper bound, we consider now the class $\mathrm{BVGr}_\tau$ from the Introduction. We invoke the following:

**Theorem 5.** (Clements, 1963.) *Let $H_\varepsilon(\mathrm{BVGr}_\tau)$ denote the $\epsilon$-entropy for the class of graphs of functions $f$ with $|f|_{TV} \le \tau$ and $|f|_\infty \le 1$, with respect to the usual Hausdorff distance. Then*

$$H_\varepsilon(\mathrm{BVGr}_\tau) \le C\ (\tau/\varepsilon), \qquad \varepsilon \in (0,1).$$

In a fashion analogous to Lemma 9, we can prove the following;

**Lemma 10.** *Fix $\tau > 0$. Set*

$$M(\varepsilon) = \sup_{S \in \mathrm{BVGr}_\tau} \mu((S)_\varepsilon).$$

*Then,*

$$M(\varepsilon) = O(\varepsilon), \qquad \varepsilon \to 0.$$

These two lemmas give us the raw ingredients for applying our Upper Bound Lemma 7, with exponents $a = 1$ and $b = 1$, yielding $\rho(\mathrm{BVGr}_\tau) \geq 1/2$. A matching lower bound follows immediately by the inclusion

$$\mathcal{H}^{1,1}(1,\tau) \subset BV_\tau, \qquad \tau > 0,$$

and the monotonicity of $N_n(\cdot)$ under inclusion.

### 3.4. Increasing Graphs

For later reference we point out a special case of $\mathrm{BVGr}_\tau$. As in the Introduction, let IncrGr denote the class of graphs of increasing functions $f$ with values in $[0,1]$. We have the inclusion

$$\mathrm{IncrGr} \subset \mathrm{BVGr}_1,$$

from which follows $N_n(\mathrm{IncrGr}) = O_P(n^{1/2})$. Of course, we pointed out already that $N_n(\mathrm{IncrGr}) = 2 \cdot n^{1/2} \cdot (1 + o_P(1))$ [37, 27], which illustrates that our general approach does not give the sharpest known results in specific cases. We return to this point in Section 5 below.

### 3.5. Convex Graphs

Continuing to illustrate the generality of the abstract upper bound, we consider now the class $\mathrm{ConvGr}_d$ of graphs $\{(x, f(x))\}$ in $R^d$ of convex functions $f : [0,1]^{d-1} \mapsto [0,1]$. We invoke the following:

**Theorem 6.** (Bronstein, 1976.) *Let $H_\varepsilon(\mathrm{ConvGr}_d)$ denote the $\epsilon$-entropy for the class $\mathrm{ConvGr}_d$ with respect to the usual Hausdorff distance. Then*

$$H_\varepsilon(\mathrm{ConvGr}_d) \leq C_d\, \varepsilon^{-(d-1)/2}, \qquad \varepsilon \in (0,1).$$

We state, without proof:

**Lemma 11.** *Fix $d$. Set*

$$M(\varepsilon) = \sup_{S \in \mathrm{ConvGr}_d} \mu((S)_\varepsilon).$$

*Then,*

$$M(\varepsilon) = O(\varepsilon), \qquad \varepsilon \to 0.$$

Combining Theorem 6 and Lemma 11 with the Upper Bound Lemma 7, we get the exponents $a = 1$ and $b = (d-1)/2$, yielding $\rho(\text{ConvGr}_d) \geq (d-1)/(d+1)$. A matching lower bound follows by a separate argument, which we omit. Notice that this exponent matches the result for twice-differentiable surfaces $\rho(\mathcal{S}^{d-1,d}(2, \beta)) = (d-1)/(d+1)$.

### 3.6. Lipschitz Graphs

We consider now the class $\text{LipGr}_\tau$ from the Introduction. This may be viewed as the special case $\alpha = 1$ in the scale of Hölder classes, but a particularly elementary case.

**Theorem 7.** (Kolmogorov and Tikhomirov, 1958.) *Let $H_\varepsilon(Lip_\tau, |\cdot|_\infty)$ denote the $\epsilon$-entropy for the class of functions $f$ with Lipschitz constant $\leq \tau$ and $|f|_\infty \leq 1$, with respect to the usual supremum norm. Then*

$$H_\varepsilon(Lip_\tau, |\cdot|_\infty) \leq C\ (\tau/\varepsilon), \qquad \varepsilon \in (0, 1).$$

Let now $Lip_\tau$ be the set of functions $f$ with Lipschitz constant $\leq \tau$ and $0 \leq f \leq 1$, and $\text{LipGr}_\tau$ be the collection of graphs of such functions. Let $H_\varepsilon(\text{LipGr}_\tau)$ denote the $\varepsilon$-entropy of this class of graphs with respect to Hausdorff distance $\Delta$. Then from (3.4) we have

$$H_\varepsilon(\text{LipGr}_\tau, \Delta) \leq H_\varepsilon(Lip_\tau, |\cdot|_\infty), \qquad \varepsilon \in (0, 1).$$

By arguments similar to those supporting Lemma 9, we obtain

**Lemma 12.** *Fix $\tau > 0$. Set*

$$M(\varepsilon) = \sup_{S \in \text{LipGr}_\tau} \mu((S)_\varepsilon).$$

*Then,*

$$M(\varepsilon) = O(\varepsilon), \qquad \varepsilon \to 0.$$

Combining these two, we get the exponents $a = 1$ and $b = 1$. The Upper Bound Lemma 7, gives $\rho(\text{LipGr}_\tau) \geq 1/2$. A matching lower bound follows immediately by the inclusion

$$\mathcal{H}^{1,1}(1, \tau) \subset \text{LipGr}_\tau, \qquad \tau > 0,$$

and the monotonicity of $N_n(\cdot)$ under inclusion. Our empirical data and theoretical analysis show that

$$N_n(\text{LipGr}_\tau) = 2\sqrt{n}(1 + o_p(1)).$$

### 3.7. Besov/Triebel Graphs

The modern trend in studying functional classes views the traditional regularity classes based on Hölder, Lipschitz and Sobolev regularity simply as special cases of the more general Besov and Triebel scales [29, 12, 16, p. 105]. It turns out that results very similar to the Hölder case hold over these more general scales, even though the notion of regularity can be quite different.

We use $\mathcal{B}^{k,d}_{p,q}(\alpha,\beta)$ for classes of immersions built from Besov function classes $B^{\alpha}_{p,q}[0,1]^d$, in a fashion analogous to our earlier construction of Hölder graphs. Recall that $\mathcal{H}^{k,d}(\alpha,\beta)$ consists of graphs of vector functions $f(x) = (f_1(x),\dots,f_{d-k}(x))$ with components $f_j$ obeying the Hölder condition $\|f_j\|_{H^\alpha} \le \beta$. Operating in parallel fashion we now constrain the components of such $f$ to each have Besov norm $\|f_j\|_{B^{\alpha}_{pq}} \le \beta$.

In the Besov scale, $\alpha > 0$ measures smoothness, while $p$ and $q$ are second-order parameters measuring the 'uniformity' of that smoothness. Thus $B^{\alpha}_{\infty,\infty}[0,1]^d$ with $\alpha$ non-integral is equivalent to the usual Hölder class, and $B^{\alpha}_{2,2}[0,1]^d$ is equivalent to the usual $L^2$-Sobolev class of smoothness $\alpha$. The bump algebra $B^{1}_{1,1}[0,1]$ is an interesting approximation to BV [29], and several other interesting spaces can be obtained by proper choice of $\alpha, p, q$.

We use $\mathcal{F}^{k,d}_{p,q}(\alpha,\beta)$ for classes of immersions built from Triebel function classes $F^{\alpha}_{p,q}[0,1]^d$. Again $\alpha > 0$ measures smoothness, $p$ and $q$ are second-order parameters measuring the 'uniformity' of that smoothness. Again $F^{\alpha}_{2,2}[0,1]^d$ is the usual $L^2$-Sobolev space of smoothness $\alpha$ and several other interesting spaces (e.g., BMO and $H^1$ [29]) can be obtained by proper choice of $\alpha, p, q$.

We extend the notion of growth exponent slightly:

**Definition 1.** A sequence $(N_n)$ has *critical growth exponent* $\hat{\rho}$ iff $N_n = O_P(n^{\hat{\rho}+\varepsilon})$ for each $\varepsilon > 0$ but not for any $\varepsilon < 0$.

In Appendix A.5, we sketch the proof of the following:

**Theorem 8.** *Let $\alpha > k/p$ and $1 \le p,q \le \infty$. Then with $\Gamma$ being either $\mathcal{B}^{k,d}_{p,q}(\alpha,\beta)$ or $\mathcal{F}^{k,d}_{p,q}(\alpha,\beta)$, $N_n(\Gamma)$ has critical growth exponent $\hat{\rho} = \frac{1}{1+\alpha(d/k-1)}$.*

In fact, the most points any Besov or Triebel graph can possibly carry will be $O_P(n^{\hat{\rho}})$ for $\hat{\rho} = 1/(1+\alpha(d/k-1))$; also, for each $\varepsilon > 0$, with overwhelming probability for large $n$ there are graphs carrying at least $n^{\hat{\rho}-\varepsilon}$ points.

## 4. Extension: Connect-the-Darts

We now generalize the CTD problem to encompass 'higher-order' interpolation. We suppose we are given a sequence of *points and directions* $(x_i, \theta_i)$, where $x_i \in [0,1]^2$ and $\theta_i \in [0, 2\pi)$, and we are interested in whether a curve of given smoothness can pass through a large collection of such points, while taking the indicated tangent directions at those points.

We say that a $C^2$ unit-speed curve $\gamma$ makes *first-order contact* with this pointset in $N = N(\gamma, (x_i, \theta_i)_{i=1}^n)$ points if there are $N$ indices $1 \le i_1 < i_2 < \cdots < i_N \le n$ and corresponding arc-length parameters $s_j$ so that $\gamma(s_j) = x_{i_j}$ and, at such intersections, the tangent $\dot{\gamma}(s_j) = \exp\{\sqrt{-1}\cdot\theta_{i_j}\}$. This was illustrated in Figure 2 above. Thinking of the point and direction together as a 'dart', we call this a *Connect-the-Darts* problem.

As mentioned in the Introduction, this problem is motivated by perceptual psychophysics [19, 25], where human subjects are shown pictures containing many randomly-oriented objects. It may or may not be the case in a given showing that a small fraction of objects lie distributed along a smooth curve, with each object oriented parallel to the curve.

In fact this problem can be treated as an ordinary CTD problem in a more abstract setting. Let $\mathbb{S}_1$ denote the unit circle, and let $\mathbb{X} = (0,1)^2 \times \mathbb{S}_1$ . Suppose we have $n$ observations

$$X_i = (x_i, e^{\sqrt{-1}\cdot\theta_i}) \overset{\text{i.i.d.}}{\sim} \mu = \text{Uniform}(0,1)^2 \times \text{Uniform}(\mathbb{S}_1).$$

We are interested in $C^2$ unit-speed curves of length $\le \lambda$ and curvature $\le \kappa$. (This is the class $\mathcal{C}_\lambda(2,\kappa)$ that was defined earlier.) Each such curve $\gamma$ in $[0,1]^2$ induces a curve $(\gamma(s), \dot{\gamma}(s))$ in $\mathbb{X}$. In this way, the class $\mathcal{C}_\lambda(2,\kappa)$ of curves induces the collection $\Sigma_\lambda(2,\kappa)$ of corresponding pointsets. Write, as before, for $S$ a subset of $\mathbb{X}$, $X^n(S) = \#\{i : X_i \in S\}$, and for the class $\Sigma_\lambda(2,\kappa)$, $N_n(\Sigma_\lambda(2,\kappa)) = \sup_{S \in \Sigma_\lambda(2,\kappa)} X^n(S)$.

**Theorem 9.** *There are constants $A, B > 0$ so that, for each pair $(\lambda, \kappa)$ with $0 < \lambda < \lambda_0(\kappa)$,*

$$\mathbf{P}\left\{A\ \lambda\kappa^{1/2}\ n^{1/4} \le N_n(\Sigma_\lambda(2,\kappa)) \le B\ \lambda\kappa^{1/2}\ n^{1/4}\right\} \to 1, \qquad n \to \infty.$$

Comparing this result with the CTD problem for $C^2$ curves, we see that incorporating the direction-matching constraint reduces the exponent $\rho$ from $1/3$ to $1/4$. We will see below that this difference in growth exponents is reflected in finite-sample behavior. The behavior matches results of psychophysical experiments; for each given number of objects on a curve embedded in clutter, the curve is far more detectable when the objects are aligned with the curves' tangent field than when the objects are randomly-aligned.

Figure 9 displays maximal Hölder-2 curves for four random point clouds. Comparable examples for connect-the-darts were found in Figure 2. There are noticeably more points on the curves in Figure 9 than in Figure 2.

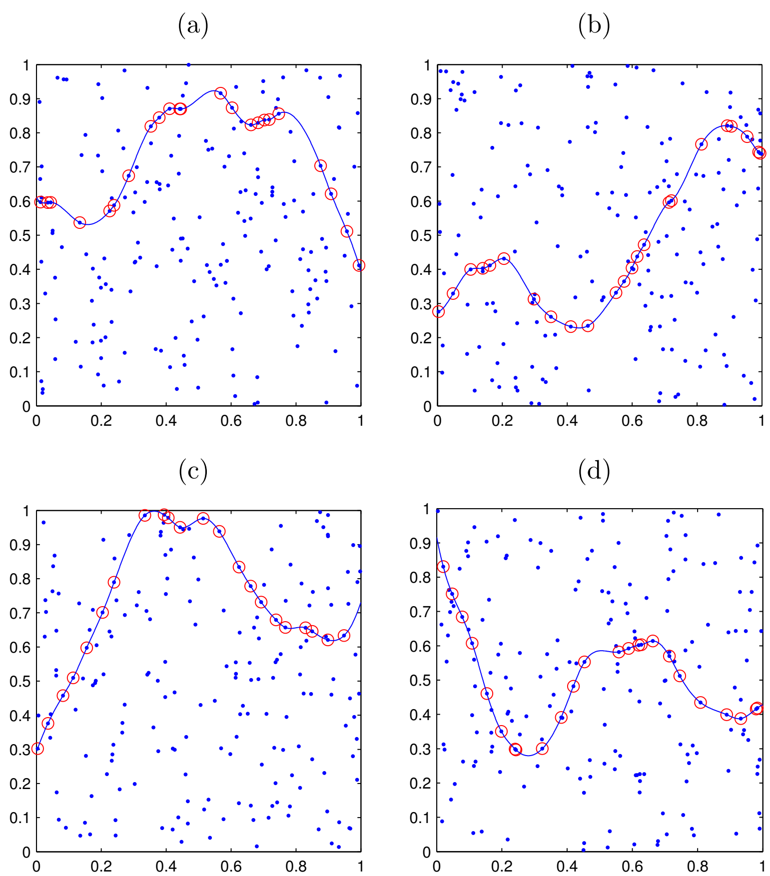



FIGURE 9: Examples of maximal Hölder-2 curves, with $n = 200$. The Hölder constant is 30.

In Figure 10, the maximum number of 'dots' and 'darts' on $C^2$ curves in some finite simulations are compared. The results are consistent with our theoretical results on growth exponents.

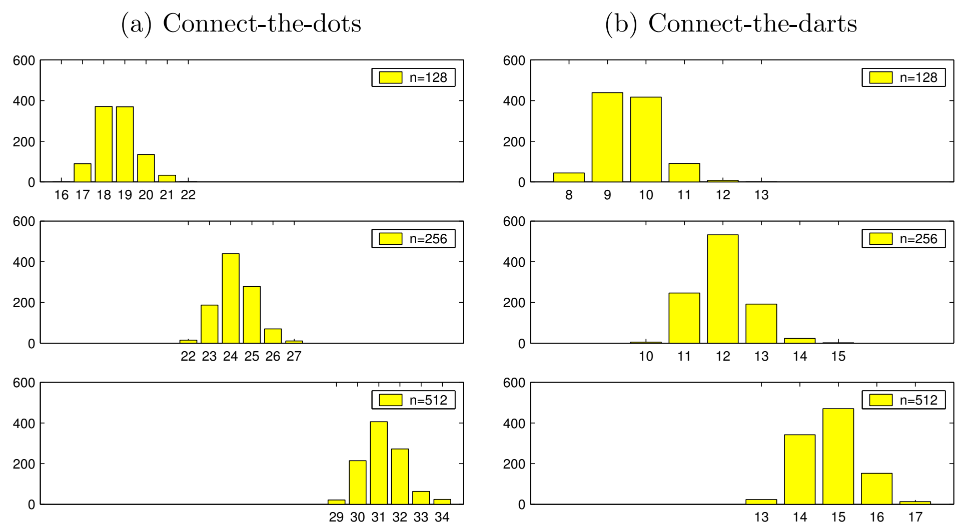



FIGURE 10: Histograms for the maximal number of 'dots' tangent to a Hölder-2 curve, Panel (a), and the maximal number of 'darts' on a Hölder-2 curve, Panel (b). The sample sizes are $n = 128, 256$, and 512 respectively. The Hölder constant is chosen to be 30 in all cases. These figures demonstrate that a detector decision based on unusually many points in connect-the-darts can be dramatically more sensitive than a decision based on unusually many points in connect-the-dots. These 'ideal observer' results correlate strongly with performance of humans in curve detection experiments [19].

We will use the abstract upper-bound machinery. However, this time we directly define a metric on $\Sigma_\lambda(2, \kappa)$. For $S_1, S_2 \in \Sigma_\lambda(2, \kappa)$, define

$$\Delta(S_1, S_2) = \inf \|\gamma_1 - \gamma_2\|_\infty, \tag{4.1}$$

where the infimum is over $\gamma_1, \gamma_2 \in \mathcal{C}_\lambda(2, \kappa)$ such that $S_i = \{(\gamma_i(s), \gamma_i'(s)) : s \geq 0\}$.

It may be surprising at first that we use $\Delta$ which is not sensitive to orientation; a metric like

$$\Delta^*(S_1, S_2) = \inf \ \max\{\|\gamma_1 - \gamma_2\|_\infty, \|\gamma_1' - \gamma_2'\|_\infty\},$$

seems intuitively more appropriate. However, it turns out (see Appendix A.6) that on $\Sigma_\lambda(2, \kappa)$, $\Delta$ is essentially equivalent to the less controversial discrepancy

$$\Delta^{**}(S_1, S_2) = \inf \ \max\{\|\gamma_1 - \gamma_2\|_\infty, \|\gamma_1' - \gamma_2'\|_\infty^2\}.$$

We first derive a bound on $H_\varepsilon(\Sigma_\lambda(2,\kappa);\Delta)$, the $\varepsilon$-entropy of $\Sigma_\lambda(2,\kappa)$ for the metric $\Delta$.

**Lemma 13.** *There is a constant $c_1 > 0$ such that, for each pair $(\lambda,\kappa)$ with $0 < \lambda < \lambda_0(\kappa)$ and $0 < \varepsilon < \varepsilon_0(\lambda,\kappa)$,*

$$H_\varepsilon(\Sigma_\lambda(2,\kappa);\Delta) \le c_1 \lambda \kappa^{1/2} \varepsilon^{-1/2}.$$

To prove the lemma, recall that the metric $\Delta$ introduced in (4.1) ignores the orientation component. Hence, we have the following identity

$$H_\varepsilon(\Sigma_\lambda(2,\kappa);\Delta) = H_\varepsilon(\mathcal{C}_\lambda(2,\kappa);\Delta),$$

where the $\Delta$ on the RHS denotes Hausdorff distance with respect to the Euclidean distance.

From there, we use our graph $\mathcal{G}_n$ constructed in Section 2.2.1, but with $m_2 = \lfloor \log_2(1/\varepsilon) \rfloor$, so that $\varepsilon_2 \le \varepsilon$. In the proof of Lemma 5, we saw that a curve $\gamma \in \mathcal{C}_\lambda(2,\kappa)$ is contained in the $\varepsilon_2$-neighborhood of its associated path $\pi^n(\gamma)$. Since there are at most

$$c(\varepsilon_1\varepsilon_2)^{-1}\ 14^{c\lambda/\varepsilon_1} \le c\kappa^{1/2}\varepsilon^{-3/2}\ 14^{c\lambda\kappa^{1/2}\varepsilon^{-1/2}}$$

such paths, we have, for $0 < \varepsilon < \varepsilon_0(\lambda,\kappa)$,

$$H_\varepsilon(\mathcal{C}_\lambda(2,\kappa);\Delta) \le c_1 \lambda \kappa^{1/2} \varepsilon^{-1/2}.$$

□

We then estimate the order of magnitude of $M(\varepsilon) = \sup\{\mu((S)_\varepsilon) : S \in \Sigma_\lambda(2,\kappa)\}$, with $(S)_\varepsilon$ the $\varepsilon$-neighborhood of $S$ with respect to $\Delta$. For the proof, see Appendix A.6.

**Lemma 14.** *There is a constant $c_2$ such that, for each pair $(\lambda,\kappa)$ with $0 < \lambda < \lambda_0(\kappa)$ and $0 < \varepsilon < \varepsilon_0(\lambda,\kappa)$,*

$$M(\varepsilon) \le c_2 \lambda \kappa^{1/2} \varepsilon^{3/2}, \qquad \varepsilon \to 0.$$

Applying now the Upper Bound Lemma 7 with exponents $a = 3/2$ and $b = 1/2$ we get the upper bound announced in Theorem 9. To establish a matching lower bound with the same growth exponent, we use the (by now familiar) select/interpolate method. For details, see Appendix A.7.

## 5. More Precise Asymptotics

So far we have only mentioned general 'rate' results, e.g., $N_n(\Gamma) = O_P(n^{\rho(\Gamma)})$. Refinements seem plausible:

1. *Scaling Laws for Centering.* $N_n(\Gamma)$ might fluctuate around a center obeying the power law scaling principle

$$\text{Median}(N_n(\Gamma)) = An^{\rho} \cdot (1 + o(1))$$

for some $A > 0$, with growth exponent $\rho = \rho(\Gamma)$; and

2. *Negligible Fluctuations.* The fluctuations might be relatively small:

$$|N_n(\Gamma) - \text{Median}(N_n(\Gamma))| = o_P(n^{\rho}).$$

In this section we summarize evidence supporting such refinements. In particular, the relative negligibility of fluctuations can be strengthened and proven in considerable generality.

### 5.1. Longest Increasing Subsequence

As mentioned in the Introduction, there has been intense interest in the last decade to understand the longest increasing subsequence (LIS) problem. This is essentially the CTD problem with $\Gamma$ = IncrGr, the class of increasing graphs. In this case we know that for $N_n = N_n(\text{IncrGr})$, the longest increasing subsequence:

- $\text{Median}(N_n) \sim 2\sqrt{n}$ [37, 27];
- $|N_n - \text{Median}(N_n)| = O_p(n^{1/3})$ [20]; and
- $(N_n - E(N_n))/SD(N_n)$ has the Tracy-Widom Distribution [7].

Thus for example, the mean, median, standard deviation and even the (non-Gaussian) shape of the limit distribution are known.

While the LIS problem is undoubtedly very special, it has connections to problems in Quantum Gravity, Random Matrix Theory, and Growth in Random Media, so there appears to be some universality to the conclusions. As it turns out, properties like those seen in the LIS problem appear to hold in at least one other CTD problem.

### 5.2. Lipschitz Graphs

Consider the collection of Lipschitz graphs $\mathrm{LipGr}_1$, mentioned both in the Introduction and in Section 3.6. We have observed the following properties for $N_n = N_n(\mathrm{LipGr}_1)$:

- $\mathrm{Median}(N_n) \sim 2\sqrt{n}$, both empirically and in theory;
- $|N_n - \mathrm{Median}(N_n)| = O_p(n^{1/3})$, empirically; and
- $(N_n - E(N_n))/SD(N_n)$ has the Tracy-Widom Distribution (empirically).

These properties are entirely analogous to those in the LIS problem. The similarity has a simple explanation. There is a simple transformation that maps one problem onto a variant of the other, and the variation seems not to affect these characteristics; see [22]. On the other hand, the fact that we have nice centering and fluctuation properties for at least two CTD problems encourages us to suppose that such nice properties might hold in many other CTD problems.

### 5.3. Simulations

In [22] we describe a software library, CTDLab, which can solve a range of CTD problems. We have applied this library to (pseudo-) random uniform point clouds and observed numerous patterns consistent with power-law scaling of $\mathrm{Median}(N_n(\Gamma))$ for the classes studied. As a simple example, we display in Figure 11 a sequence of histograms of $N_n = N_n(\mathrm{LipGr}_1)$ at various $n$. The histograms show increasing concentration around $2\sqrt{n}$ as $n$ increases. They also show convergence towards the Tracy-Widom law.

### 5.4. Concentration of Measure

While at the moment, we have no proof of the scaling law $\sim An^\rho$ for the centering of $N_n(\mathcal{S})$ for general $\mathcal{S}$, there is a general proof for the asymptotic negligibility of fluctuations, in a very strong sense.

**Theorem 10.** *Let $\mathcal{S}$ be a nonempty class of sets; and let $M_n = \mathrm{Median}(N_n(\mathcal{S}))$. Then*

$$|N_n(\mathcal{S}) - M_n| = O_p(\sqrt{M_n}), \quad n \to \infty.$$

*Deviations of size $t\sqrt{M_n}$ have a probability that decays exponentially in $|t|$, uniformly*

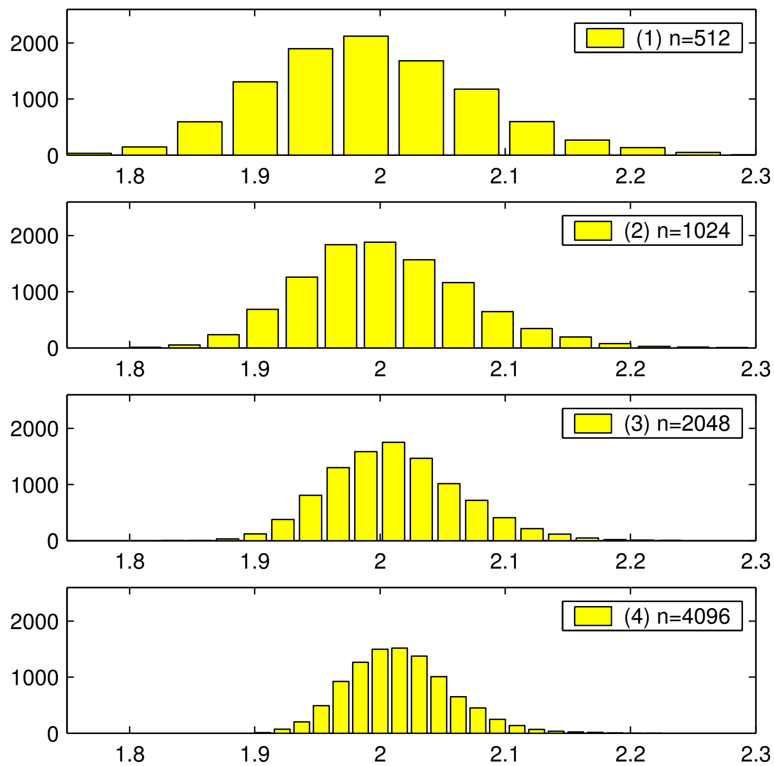



FIGURE 11: Histograms of the maximal number of points on **Lipschitz** graphs for uniform random samples of different sample sizes $n$. The values of $N_n(\mathrm{LipGr}_1)$'s are divided by $\sqrt{n}$. The sample sizes are $n = 512, 1024, 2048$, and 4096 respectively. Clearly the median is converging to $2\sqrt{n}$. Statistical analysis shows that the histogram shape is definitely non-normal and matches the Tracy-Widom distribution.

*in* $\{n : M_n \geq 1\}$.

Note that this applies to all kinds of CTD problems: curves, hypersurfaces, etc. The proof of this apparently powerful result is actually just a simple application of a framework of Talagrand's [35]. First, we introduce Talagrand's key abstract notion:

**Definition 2.** (*Talagrand 1995, Def. 7.1.7.*) $L : \mathbb{X}^n \mapsto \mathbb{R}$ is a **configuration function** if given any $x^n = (x_i) \in \mathbb{X}^n$, there exists a subset $J$ of $\{1, \ldots, n\}$, with $\#(J) = L(x^n)$ and such that, for each $y^n = (y_i) \in \mathbb{X}^n$, we have

$$L(y^n) \geq \#\{i \in J : y_i = x_i\}.$$

Next we check that this notion covers our setting.

**Lemma 15.** *For every CTD problem in our sense, $N_n(\mathcal{S})$ is the value $L(X^n)$ of a configuration function in Talagrand's sense.*

**Proof.** Indeed, if there is a $S_0 \in \mathcal{S}$ interpolating $X_{i_j}$, $j = 1, \dots, N_n(\mathcal{S})$, then take $J = \{i_1, \dots i_{N_n(\mathcal{S})}\}$. Then if $(Y_j)_{j \in J}$ has $k$ entries in common with $(X_j)_{j \in J}$, there are of course at least $k$ entries of $Y^n$ on $S_0$, and hence the maximal number of entries of $Y^n$ on any $S \in \mathcal{S}$ is at least $k$. □

Theorem 10 is now immediate from Talagrand's concentration of measure result for configuration functions:

**Theorem 11.** (Talagrand 1995, Thm. 7.1.3.) *Let $L = L(X^n)$ be a configuration function applied to a random uniform point cloud $X^n$ with median $M$ = median $(L)$. Then for $u > 0$,*

$$P\{L > M + u\} \leq 2 \exp\left(\frac{-u^2}{4M + u}\right) \qquad \text{and} \quad P\{L < M - u\} \leq 2 \exp\left(\frac{-u^2}{4M}\right).$$

## 6. Conclusion

We introduced the notion of connect-the-dot problems and derived asymptotic growth properties for a range of such problems. We developed an abstract approach for upper bounds, which is easy to apply in all the CTD problems we considered, and for which there is a lower bound with matching order.

We pointed out that concentration of measure results hold for such problems, and reviewed evidence for the conjecture that $N_n(\Gamma)/n^\rho$ tends in probability towards a limit, while a suitably standardized version of $N_n(\Gamma)$ converges in distribution – in a few cases to the Tracy-Widom limit.

We remind the reader of the computational efforts reported in [22]. These show clearly that the asymptotic theory of growth exponents developed in this paper accurately describes the behavior in moderate-sized random point clouds. We hope to inspire further research into this class of problems.

## Appendix A. Proofs

### A.1. Proof of Lemma 4

The total number of points falling in the first $m = \lfloor \lambda\sqrt{n}/\sqrt{2} \rfloor$ squares, denoted by $K_{n,m}$, has distribution $\text{Bin}(n, m/n)$. Also, given $K_{n,m}$, the distribution of $J_{n,\lambda}$ is simplified,

$$\mathbf{P}\left\{J_{n,\lambda} \geq \ell | K_{n,m}\right\} \geq 1 - \binom{m}{\ell} (\ell/m)^{K_{n,m}}.$$

Now

$$\mathbf{P}\left\{J_{n,\lambda} \geq \ell\right\} = \mathbf{P}\left\{J_{n,\lambda} \geq \ell | K_{n,m} \geq m/2\right\} \cdot \mathbf{P}\left\{K_{n,m} \geq m/2\right\}.$$

By the Law of Large Numbers,

$$\mathbf{P}\left\{K_{n,m} \geq m/2\right\} \to 1, \qquad n \to \infty.$$

Clearly

$$\mathbf{P}\left\{J_{n,\lambda} \geq \ell | K_{n,m} \geq m/2\right\} \geq 1 - \binom{m}{\ell} (\ell/m)^{m/2}.$$

Now use the fact that, for $u \in (0,1)$ fixed and $m \geq m_0(u)$,

$$\binom{m}{\lceil um \rceil} \leq \left(u^u(1-u)^{1-u}\right)^{-m}.$$

Hence for $n \geq n_0(\lambda)$,

$$\log\left[\binom{m}{\lceil um \rceil} \cdot u^{m/2}\right] \leq -m \cdot (\log[u^u(1-u)^{1-u}] - 1/2\ \log(u)).$$

Now, for $u = 0.29$,

$$\log[u^u(1-u)^{1-u}] - 1/2\ \log(u) > 0.$$

Then, $um \geq 1/5\ \lambda\sqrt{n}$, for $n \geq n_1(\lambda)$. This completes the proof of Lemma 4. □

### A.2. Proof of Lemma 5

Here by vertical (resp. horizontal) grid lines we mean the lines that traverse the vertical (resp. horizontal) grid points.

Fix $\gamma \in \mathcal{C}_\lambda(2, \kappa)$. We choose a smooth parametrization by arc-length, also denoted $\gamma(\ )$. Let $\ell \equiv \text{length}(\gamma)$. We need two inequalities.

**Lemma 16.** *Let $\gamma \in \mathcal{C}_\lambda(2,\kappa)$. Then, for arclengths $s, t$,*

$$|\gamma(t) - \gamma(s) - (t-s)\gamma'(s)| \le \kappa/2 \ (t-s)^2.$$

**Lemma 17.** *Let $\gamma \in \mathcal{C}_\lambda(2,\kappa)$. Then, for all arclengths $r < s < t$,*

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

Lemma 16 is a simple Taylor expansion, and yields Lemma 17 after some simple algebra.

Let $u = (1,0)$ be the horizontal unit vector pointing to the right. If $|\cos(\angle(\gamma'(0),u))| \ge 1/\sqrt{2}$ (resp. $< 1/\sqrt{2}$), then let $a_0$ be the first intersection of the ray $\{\gamma(0) - t\gamma'(0) : t \ge 0\}$ with a vertical (resp. horizontal) grid line. We then extend $\gamma$ by adding the line segment $[a_0, \gamma(0)]$. Hence, we may assume $\gamma$ starts on a vertical (resp. horizontal) grid line with $|\cos(\angle(\gamma'(0),u))| \ge 1/\sqrt{2}$ (resp. $< 1/\sqrt{2}$).

Put $a_0 = \gamma(0)$, $v_0 = \gamma'(0)$ and $s_0 = 0$. Starting at $s = s_0 = 0$, we recursively define $a_1, \dots, a_I$, in the following way: from $a_i, v_i$ and $s_i$, if $|\cos(\angle(v_i,u))| \ge 1/\sqrt{2}$ (resp. $< 1/\sqrt{2}$), let $a_{i+1} = \gamma(s_{i+1})$ be the first intersection of $\gamma$ with a vertical (resp. horizontal) grid line, if $\gamma((s_i,\ell])$ intersects such a line; otherwise call $I = i$ and stop. If $a_{i+1}$ has been defined, then $v_{i+1} = \gamma'(s_{i+1})$. If $s_I < \ell$ and $|\cos(\angle(v_I,u))| \ge 1/\sqrt{2}$ (resp. $< 1/\sqrt{2}$), define $a_{I+1}$ as the first intersection of the semi-line $\{\gamma(\ell) + t\gamma'(\ell) : t \ge 0\}$ with a vertical (resp. horizontal) grid line. Lemma 18 below implies that this $a_{I+1}$ is well defined. It is a direct consequence of Lemma 16.

**Lemma 18.** *There is a constant $c_1 > 0$ so that, for $\varepsilon_1 > 0$ that is small enough,*

$$s_{i+1} - s_i \le \sqrt{2}\varepsilon_1 + c_1\kappa\varepsilon_1^2.$$

If $|\cos(\angle(v_i,u))| \ge 1/\sqrt{2}$ (resp. $< 1/\sqrt{2}$), let $a_i'$ be the first intersection of the ray $\{a_i + tv_i : t > 0\}$ with a vertical (resp. horizontal) grid line. The following lemma is a direct consequence of Lemma 17.

**Lemma 19.** *There is a constant $c_2 > 0$ so that, for $\varepsilon_1 > 0$ that is small enough, $|a_{i+1} - a_i'| \le c_2\kappa\varepsilon_1^2$.*

Let $\xi_a = \cup_i [a_i, a_{i+1}]$. Then, Lemma 17 implies that the Hausdorff distance between $\xi_a$ and $\gamma$ does not exceed $c_3\kappa\varepsilon_1^2$, for a constant $c_3 > 0$.

*Definition of $k$:* We choose $k$ large enough that $2^k \geq 2\max\{c_1, c_2, c_3\}$ – so that $\varepsilon_2 \geq 2c_i\kappa\varepsilon_1^2$, for $i = 1, 2, 3$.

We now define a sequence of beamlets. For $i = 0, \ldots, I + 1$ let $b_i$ be one of the closest grid point to $a_i$. Lemma 19 implies that, for all $i = 0, \ldots, I + 1$, $[b_i, b_{i+1}]$ is a beam.

Also, two successive beamlets, e.g., $[b_i, b_{i+1}]$ and $[b_{i+1}, b_{i+2}]$ are in good continuation. Indeed, by construction two successive beamlets are either both vertical, both horizontal or one of them is diagonal. For example, assume they are both horizontal beamlets. Then, $b_{i+2}$ is a vertical grid point; let $V$ be the vertical grid line it belongs to, i.e., $b_{i+2} \in V$. Then, if $b'_{i+2}$ is the grid point defined by the intersection of line $(b_i, b_{i+1})$ with $V$, we need to show that $|b'_{i+2} - b_{i+2}| < 3\varepsilon_2$. For that, it suffices to show that $|a_{i+2} - b'_{i+2}| < 5/2\ \varepsilon_2$. Now, since $|a_i - b_i| \leq \varepsilon_2/2$ and $|a_{i+1} - b_{i+1}| \leq \varepsilon_2/2$, we get $|a'_{i+2} - b'_{i+2}| \leq 3/2\ \varepsilon_2$. Hence, by Lemma 19, $|a_{i+2} - b'_{i+2}| \leq 3/2\ \varepsilon_2 + c\kappa\varepsilon_1^2 \leq 2\varepsilon_2$.

Moreover, with $\xi_b = \cup_i[b_i, b_{i+1}]$, the Hausdorff distance between $\xi_b$ and $\gamma$ is bounded by $\varepsilon_2$. This comes from the fact that the Hausdorff distance between $\xi_b$ and $\xi_a$ is bounded by $\varepsilon_2/2$ – since this is the case for $[b_i, b_{i+1}]$ and $[a_i, a_{i+1}]$ for all $i = 0, \ldots, I$.

Finally, since $\gamma([s_i, s_{i+2}])$ has a length of at least $\varepsilon_1$, $I \leq 2(1 + \ell/\varepsilon_1) \leq 2\lambda/\varepsilon_1$ (for $\varepsilon_1 < \varepsilon_0(\lambda)$). □

### A.3. Proof of Lower Bound for $\mathcal{C}_\lambda(2, \kappa)$

We extend the interpolation methods used earlier. We assume that $\lambda \leq 1$ for simplicity. Pick a $C^2$ "bump" function $\psi : \mathbb{R} \mapsto [0, 1]$ which is supported in $[-1/2, 1/2]$ and satisfies $\psi(0) = 1$. Introduce $\eta_n, v$ chosen so that

$$v = \kappa/\|\psi''\|_\infty, \tag{A.1}$$

$$v\eta_n^3 = 1/n. \tag{A.2}$$

Partition $[0, \lambda/2]$ into $m = \lfloor \lambda/(2\eta) \rfloor$ intervals of length $\eta$ – denoted by $I_i, i = 1, \ldots, m$. Thicken these intervals so they become a row of $\eta \times v\eta^2$ rectangles $Q_i = I_i \times [0, v\eta^2]$. Note that the area $|Q_i| = 1/n$, so we expect $Q_i$ to contain one point of $X^n$. Now consider only the even numbered $Q_i$ which are not empty (i.e., $X^n(Q_i) > 0$). Rename these rectangles $R_j, j = 1, \ldots, J$ – notice that both $J$ and the $\{R_j\}$ are random. In each $R_j$ select one point – say $X_{i_j}$; label its components $(t_j, z_j)$. With overwhelming probability, $J/m$ is roughly $(1-e^{-1})/2$. Hence, with overwhelming probability for large $n$, there are

$$J \geq .3161 \cdot m \approx .3161 \cdot \lambda/(2\eta) > .1508 \cdot \|\psi''\|_\infty^{-1/3}\ \lambda\kappa^{1/3}n^{1/3}.$$

We will next show that at least the $J$ points $X_{i_j} = (t_j, z_j) \in R_j$ live on a curve in $\mathcal{S}_{\lambda,\kappa}$. Hence for the statement of the lemma we can set $A = .1508 \cdot \|\psi''\|_\infty^{-1/3}$.

We now construct a smooth function whose graph passes through the $X_{i_j}$. Let $f_j(t) = z_j\psi((t-t_j)/\eta)$, for $j = 1,\dots,J$. By construction, $f_j(t_j) = z_j$, and the support of the different $f_j$ are disjoint. Hence $f(t) = \sum_{j=1}^J f_j(t)$ obeys $f(t_j) = z_j$, $j = 1,\dots,J$. Also, $f$ is twice continuously differentiable, being the finite sum of smooth functions. Moreover, since for fixed $t$ there is at most one non-zero term in the sum, we have

$$\|f\|_\infty \leq \max_j \|f_j\|_\infty = \max_j z_j \leq v\eta^2 \leq 1$$

and

$$\|f''\|_\infty \leq \max_j \|f_j''\|_\infty \leq v\|\psi''\|_\infty = \kappa.$$

Hence the graph of $f$ is contained in the unit square and has maximum curvature bounded by $\kappa$. We also need to control the total length of graph$(f)$. On $[0,\lambda/2] \setminus \cup_j[t_j - \eta/2, t_j + \eta/2]$, $f$ takes the value zero – thus its graph is flat in this region. On $[t_j - \eta/2, t_j + \eta/2]$, the length of graph$(f)$ is bounded by $\eta \cdot \sqrt{1 + v\eta\|\psi'\|_\infty}$. Therefore, the length of graph$(f)$ tends to $\lambda/2$ as $\eta \to 0$. Hence, for $\eta$ small enough, the total length of graph$(f)$ does not exceed $\lambda$. By (A.2), we have $\eta \to 0$ as $n \to \infty$, and graph$(f) \in \mathcal{C}_\lambda(2,\kappa)$ for $n$ large enough. □

### A.4. Lower Bound for Hölder objects

We extend the interpolation methods used earlier. We give a proof valid for $\alpha > 1$. Take $\psi : \mathbb{R}^k \to [0,1]$ which is $\lceil\alpha\rceil$ times continuously differentiable, supported in $[-1/2,1/2]^k$, and satisfying $\psi(0) = 1$. Introduce $v > 0$ chosen so that, if $\alpha$ is an integer, then

$$v\|\psi^{\alpha}\|_\infty \leq (\alpha-1)!\beta \quad \text{(A.3)}$$

while if $\alpha$ is not an integer

$$v\|\psi^{\lfloor\alpha\rfloor}\|_\infty \leq (\lfloor\alpha\rfloor)!\beta; \quad \text{(A.4)}$$

and introduce $\eta = \eta_n$ chosen so that

$$v\eta^{k+\alpha(d-k)} = 1/n. \tag{A.5}$$

We partition $[0,1]^k$ into hypercubes of sidelength $\eta$ – denoted by $I_i$, where $i = (i_1, \ldots, i_k)$. To each such hypercube, we associate the hyperrectangle $Q_i = I_i \times [0, v\eta^\alpha]^{d-k}$ in $[0,1]^d$. Because of (A.5), each $Q_i$ has volume $1/n$, and so we expect one point of $X^n$ in $Q_i$. Now consider only the even numbered $Q_i$ (i.e., $i_1, \ldots, i_k$ all even) which are not empty (i.e., $X^n(Q_i) > 0$). Rename these rectangles $R_j, j = 1, \ldots, J$ – notice that both $J$ and the $\{R_j\}$ are random. In each $R_j$ select one point $X_{i_j}$, group the 'independent' and 'dependent' components as $(t_j, z_j) \in [0,1]^k \times [0,1]^{d-k}$.

Now, we define both the 'independent' and 'dependent' components of $g$.

- For $i = 1, \ldots, k$, set $g_i(s) = s_i$.
- For $i = 1, \ldots, d-k$, $g_{k+i}(s) = \sum_j f_{ij}(s)$, where $f_{ij}(s) = z_{ji}\psi((s - t_j)/\eta)$.

This construction guarantees that the image $g([0,1]^k)$ includes $\{(t_j, z_j) : j = 1, \ldots, J\}$. Again, $J$ is a random variable and we can (if we like) give a concrete estimate for an $A > 0$ so that

$$\mathbf{P}\left\{ J \geq A\ c(\alpha, \beta, k)\ n^{k/(k+\alpha(d-k))} \right\} \to 1 \quad n \to \infty.$$

Hence we have at least $n^{k/(k+\alpha(d-k))}$ points on a $k$-surface in dimension $d$. We next show that this surface has the required regularity.

Because at most one term in the sum defining $g$ is positive, $g$ is $\lceil \alpha \rceil$ times differentiable. However, the regularity condition we seek is quantitative and componentwise. Since we assume $\alpha > 1$, and the 'independent' components $g_i$ are all linear for $i \leq k$, they all belong to $\mathcal{H}^{k,d}(\alpha, \beta)$.

It remains to show that the 'dependent' components are as well.

$$(H1) \quad \|g_{k+i}^{\lfloor \alpha \rfloor}\|_\infty = \max_j \|f_{ij}^{\lfloor \alpha \rfloor}\|_\infty \leq v\|\psi^{\lfloor \alpha \rfloor}\|_\infty \cdot \eta^{\{\alpha\}}.$$

$$(H2) \quad \|g_{k+i}^{\lceil \alpha \rceil}\|_\infty = \max_j \|f_{ij}^{\lceil \alpha \rceil}\|_\infty \leq v\|\psi^{\lceil \alpha \rceil}\|_\infty \cdot \eta^{-r(\alpha)},$$

with $r(\alpha) = 0$ if $\alpha$ is an integer and $r(\alpha) = 1 - \{\alpha\}$ otherwise.

Two cases:

- $\alpha$ is an integer. Using $(H2)$ and a Taylor expansion gives

$$\left|g_{k+i}^{\alpha-1}(s) - g_{k+i}^{\alpha-1}(t)\right| \leq v\|\psi^\alpha\|_\infty \cdot \|s-t\|.$$

- $\alpha$ is not an integer. Using $(H1)$ and the fact that $g_{k+i}$ takes the value 0, we prove that

$$\left|g_{k+i}^{\lfloor\alpha\rfloor}(t) - g_{k+i}^{\lfloor\alpha\rfloor}(s)\right| \leq v\|\psi^{\lfloor\alpha\rfloor}\|_\infty \cdot \eta^{\{\alpha\}}.$$

Hence, for $\|t-s\| > \eta$,

$$\left|g_{k+i}^{\lfloor\alpha\rfloor}(t) - g_{k+i}^{\lfloor\alpha\rfloor}(s)\right| \leq v\|\psi^{\lfloor\alpha\rfloor}\|_\infty \cdot \|t-s\|^{\{\alpha\}}.$$

Using $(H2)$ and a Taylor expansion gives

$$\left|g_{k+i}^{\lfloor\alpha\rfloor}(t) - g_{k+i}^{\lfloor\alpha\rfloor}(s)\right| \leq v\|\psi^{\lceil\alpha\rceil}\|_\infty \cdot \eta^{-(1+\{\alpha\})} \cdot \|t-s\|.$$

Hence, for $\|t-s\| \leq \eta$,

$$\left|g_{k+i}^{\lfloor\alpha\rfloor}(t) - g_{k+i}^{\lfloor\alpha\rfloor}(s)\right| \leq v\|\psi^{\lfloor\alpha\rfloor}\|_\infty \cdot \|t-s\|^{\{\alpha\}}.$$

In all cases, (A.3)-(A.4) guarantee $g_{k+i} \in \mathcal{H}^{k,d}(\alpha,\beta)$. □

### A.5. Besov and Triebel Objects

The following results are classical (see [16], p. 105).

1. We have the inclusions:

$$B^\alpha_{p,p\wedge q} \subset \mathcal{F}^\alpha_{p,q} \subset B^\alpha_{p,p\vee q}, \tag{A.6}$$

where $F \subset B$, say, means $\|f\|_B \leq C\|f\|_F$ for some $C > 0$.

2. If $\alpha_1 - \alpha_2 - k(p_1^{-1} - p_2^{-1})_+ > 0$, where $p_1, p_2, q_1, q_2 \in (0,\infty]$ and $-\infty < \alpha_2 < \alpha_1 < \infty$, then

$$B^{\alpha_1}_{p_1,q_1} \subset B^{\alpha_2}_{p_2,q_2}, \tag{A.7}$$

where $\subset$ means (again) continuous inclusion.

3. For $\alpha > k/p$, the $\varepsilon$-entropy of the unit ball in $B^\alpha_{p,q}$ for the supnorm, is of order $\varepsilon^{-k/\alpha}$.

Notice that the parameter $q$ does not play a critical role in these results. This will translate into the fact (using (A.6)) that $N_n(\mathcal{B}^{k,d}_{p,q}(\alpha,\beta))$ and $N_n(\mathcal{F}^{k,d}_{p,q}(\alpha,\beta))$ are of the same order of magnitude in most cases.

From now on, we assume $\alpha > k/p$. Using (A.7), we have that for all $\alpha' > \alpha$ and a certain constant $C(\alpha,\alpha')$,

$$\mathcal{H}^{k,d}(\alpha', C(\alpha,\alpha')\beta) \subset \mathcal{B}^{k,d}_{p,q}(\alpha,\beta).$$

Hence, $N_n(\mathcal{B}^{k,d}_{p,q}(\alpha,\beta))$ is of order at least $n^{1/(1+\alpha'(d/k-1))}$ for all $\alpha' > \alpha$.

On the other hand, the usual entropy approach yields an upper bound (using (3.1)) of order $n^{1/(1+\alpha(d/k-1))}$.

So the order of $N_n(\mathcal{B}^{k,d}_{p,q}(\alpha,\beta))$ is somewhere in between. And $N_n(\mathcal{F}^{k,d}_{p,q}(\alpha,\beta))$ is of the same order because of (A.6).

### A.6. Proof of Lemma 14

Fix $S \in \Sigma_\lambda(2,\kappa)$. The first component of $S$ traces out a curve $\gamma$ in the unit square. If $\ell \equiv \text{length}(\gamma) \geq 6\varepsilon$, we show below that

$$(S)_\varepsilon \subset \bigcup_{(x,v)\in W} \text{Ball}(x,\varepsilon) \times \text{Cone}(v,\theta_{\varepsilon,\ell}), \tag{A.8}$$

where $\theta_{\varepsilon,\ell} = 16\max\{\varepsilon/\ell, \kappa^{1/2}\varepsilon^{1/2}\}$. If $\ell < 6\varepsilon$, we use the obvious fact

$$(S)_\varepsilon \subset \bigcup_{x\in\gamma} \text{Ball}(x,\varepsilon) \times \mathbb{S}_1.$$

Assume $\ell \geq 6\varepsilon$ and that we proved (A.8). Choose a unit-speed parametrization of $\gamma$ (also denoted $\gamma()$). Notice that $S = \{(\gamma(s),\gamma'(s)) : s \in [0,\ell]\}$. Now, consider an $\varepsilon$-covering of $[0,\ell]$, that we denote by $\{s_j : j = 1,\ldots,J\}$, with $J \leq \ell/\varepsilon + 1$. For $|s - s_j| \leq \varepsilon$,

$$\text{Ball}(\gamma(s),\varepsilon) \subset \text{Ball}(\gamma(s_j), 2\varepsilon)$$

and

$$\text{Cone}(\gamma'(s),\theta_{\varepsilon,\ell}) \subset \text{Cone}(\gamma'(s_j),\theta_{\varepsilon,\ell} + \kappa\varepsilon).$$

The first inclusion is due to $\gamma$ being Lipschitz with constant 1, while the second inclusion comes from $\gamma'$ being Lipschitz with constant $\kappa$. Therefore, for $\varepsilon < \varepsilon_0(\kappa)$ small,

$$(S)_\varepsilon \subset \bigcup_{j=1}^{J} \text{Ball}(\gamma(s_j), 2\varepsilon) \times \text{Cone}(\gamma'(s_j), 2\theta_{\varepsilon,\ell}).$$

Since $\mu(\mathrm{Ball}(x,r)\times\mathrm{Cone}(v,\theta)) = r^2\min\{\theta,\pi\}$, we have

$$\mu((S)_\varepsilon) \le J\cdot(2\varepsilon)^2\min\{\theta_{\varepsilon,\ell},\pi\} \le 4\ell\varepsilon\min\{\theta_{\varepsilon,\ell},\pi\} + 4\pi\varepsilon^2.$$

1. If $\ell \le \varepsilon^{1/2}/\kappa^{1/2}$, then $\theta_{\varepsilon,\ell} = 16\varepsilon/\ell$ and so $\mu((S)_\varepsilon) \le 64\varepsilon^2 + 4\pi\varepsilon^2$.

2. If $\ell > \varepsilon^{1/2}/\kappa^{1/2}$ then $\theta_{\varepsilon,\ell} = 16\kappa^{1/2}\varepsilon^{1/2}$ and so $\mu((S)_\varepsilon) \le 64\ell\kappa^{1/2}\varepsilon^{3/2} + 4\pi\varepsilon^2 \le 64\lambda\kappa^{1/2}\varepsilon^{3/2} + 4\pi\varepsilon^2$.

In any case, when $\varepsilon < \varepsilon_0(\lambda,\kappa)$, we have $\mu((S)_\varepsilon) \le 65\lambda\kappa^{1/2}\varepsilon^{3/2}$.

In a similar way, if $\ell < 6\varepsilon$,

$$(S)_\varepsilon \subset \bigcup_{j=1}^{J}\mathrm{Ball}(\gamma(s_j),2\varepsilon),$$

and so $\mu((S)_\varepsilon) \le 28\pi\varepsilon^2 \le 65\lambda\kappa^{1/2}\varepsilon^{3/2}$ for $\varepsilon < \varepsilon_0(\lambda,\kappa)$.

We still need to prove (A.8). For that, take $S_1 \in (S)_\varepsilon$. By definition, there is $\gamma,\gamma_1 \in \mathcal{C}_\lambda(2,\kappa)$ such that $S = \{(\gamma(s),\gamma'(s)) : s \ge 0\}$, $S_1 = \{(\gamma_1(s),\gamma_1'(s)) : s \ge 0\}$ and $\|\gamma-\gamma_1\|_\infty \le \varepsilon$. Let $\ell_1 = \mathrm{length}(\gamma)$ and $\ell_1 = \mathrm{length}(\gamma_1)$. Also, let $\theta_0 = \max_{s\le 0}|\angle(\gamma'(s),\gamma_1'(s))|$ and let $s_0$ be a maximizer (we need only consider $s_0 \le \max\{\ell,\ell_1\}$).

If $s_0 > \min\{\ell,\ell_1\}$, then, because $|\gamma'(s_0)-\gamma'(\ell_1)| \le \kappa|s_0-\ell_1|$ and $|\gamma_1'(s_0)-\gamma_1'(\ell)| \le \kappa|s_0-\ell|$, $|\gamma'(s_0)-\gamma_1'(s_0)| \le \kappa|\ell-\ell_1|$. We then use the following result.

**Lemma 20.** *Let $\gamma,\gamma_1 \in \mathcal{C}_\lambda(2,\kappa)$ with $\|\gamma-\gamma_1\|_\infty \le \varepsilon$. Then, for $\varepsilon < \varepsilon_0(\kappa)$, $|\mathrm{length}(\gamma)-\mathrm{length}(\gamma_1)| \le 3\varepsilon$.*

*Proof.* Assume $\ell \le \ell_1$. For $\ell \le s \le \ell_1$, $|\gamma_1(s)-\gamma(\ell)| \le \varepsilon$ and $|\gamma_1(\ell)-\gamma(\ell)| \le \varepsilon$; hence, $|\gamma_1(s)-\gamma_1(\ell)| \le 2\varepsilon$. Now, using Lemma 16 for $\ell \le s$, we get $(s-\ell)-\kappa/2\,(s-\ell)^2 \le 2\varepsilon$. From this, we see that, when $\varepsilon < 2/(9\kappa)$, $s-\ell$ cannot take the value $3\varepsilon$. Hence, $\ell_1 < \ell + 3\varepsilon$. ∎

Therefore, assume $s_0 \le \min\{\ell,\ell_1\}$. Because $\ell \ge 6\varepsilon$, we have $\ell/2 \le \ell_1 \le 3\ell/2$. This implies that either $s_0 + \ell/4 \le \min\{\ell,\ell_1\}$ or $s_0 - \ell/4 \ge 0$. Both cases are treated similarly, so we assume the former is true. Then, for $s_0 \le s \le s_0 + \ell/4$, we have

$$\begin{aligned}\gamma(s) - \gamma_1(s) \quad = \quad & \gamma(s) - \gamma(s_0) - \gamma'(s_0)(s - s_0) \\ & + \gamma(s_0) - \gamma_1(s_0) \\ & + (\gamma'(s_0) - \gamma_1'(s_0))(s - s_0) \\ & - \left(\gamma_1(s) - \gamma_1(s_0) - \gamma_1'(s_0)(s - s_0)\right).\end{aligned}$$

By Lemma 16,

$$|\gamma(s) - \gamma(s_0) - \gamma'(s_0)(s - s_0)| \le \kappa/2 \ (s - s_0)^2$$

and

$$|\gamma_1(s) - \gamma_1(s_0) - \gamma_1'(s_0)(s - s_0)| \le \kappa/2 \ (s - s_0)^2.$$

We also have $|\gamma(s) - \gamma_1(s)| \le \varepsilon$ and $|\gamma(s_0) - \gamma_1(s_0)| \le \varepsilon$. Hence, we get

$$\|\gamma'(s_0) - \gamma_1'(s_0)\|_\infty (s - s_0) - \kappa(s - s_0)^2 \le 2\varepsilon.$$

Call $D_0 = \|\gamma'(s_0) - \gamma_1'(s_0)\|_\infty$ and $\eta = s - s_0$. The inequality becomes $p(\eta) = \kappa\eta^2 - D_0\eta + 2\varepsilon \ge 0$, for all $\eta \in [0, \ell/4]$. There are two possibilities:

1. either $D_0 \le 2^{3/2}\kappa^{1/2}\varepsilon^{1/2}$, in which case $\theta_0 = \mathrm{acos}(1 - D_0^2/2) \le 6\kappa^{1/2}\varepsilon^{1/2}$, for $\varepsilon < \varepsilon_0(\lambda, \kappa)$;
2. or $D_0 > 2^{3/2}\kappa^{1/2}\varepsilon^{1/2}$, in which case $p(4\varepsilon/D_0) < 0$, implying $\ell/4 < 4\varepsilon/D_0$.

In each case, $D_0 \le 16 \max\{\varepsilon/\ell, \kappa^{1/2}\varepsilon^{1/2}\}$. (A.8) follows. □

### A.7. Lower Bound for Connect-The-Darts

We again use the method of Section A.3. We assume that $\lambda \le 1$ for simplicity. We take $\psi, \phi : \mathbb{R} \to \mathbb{R}$ twice continuously differentiable, supported in $[-1/2, 1/2]$, and satisfying

- $0 \le \psi \le 1$ with $\psi(0) = 1$;
- $-1 \le \phi \le 1$ with $\phi'(0) = 1$.

Introduce $v, w, \eta_n > 0$ chosen as follows:

$$v = \kappa/(2\|\psi''\|_\infty), \tag{A.9}$$

$$w = \kappa/(2\|(\psi\phi)''\|_\infty), \tag{A.10}$$

$$v\eta_n^3 \cdot 2\arctan(w\eta_n) = 1/n. \tag{A.11}$$

Partition $[0, \lambda/2]$ into intervals of length $\eta$ – denoted by $I_i, i = 1, \dots, \lfloor \lambda/(2\eta) \rfloor$. To each interval we associate $Q_i = I_i \times [0, v\eta^2] \times \{e^{\sqrt{-1}\theta} : |\theta| \le w\eta\}$. By (A.11) the volume of each $Q_i$ is $1/n$, and so we expect one point in each such rectangle. We now consider only the even-numbered $Q_i$ which are not empty (i.e., $X^n(Q_i) > 0$). Rename these sets $R_j, j = 1, \dots, J$ – both $J$ and the $R_j$ are random. In each $R_j$ select one point $X_{i_j} = (t_j, z_j, e^{\sqrt{-1}\theta_j})$.

Now, consider $f(t) = \sum_j \ f_j(t)$, where

$$f_j(t) = z_j \psi((t - t_j)/\eta) \left(1 + \frac{\eta \tan\theta_j}{z_j} \ \phi((t - t_j)/\eta)\right).$$

By construction, $f$ interpolates the points $X_{i_j}$, i.e., $f(t_j) = z_j$ and $f'(t_j) = \tan(\theta_j)$. Again, for $A_0$ small enough, and with overwhelming probability as $n$ increases, there are at least $J \ge A_0 \ \lambda/\eta = A_0 c^{-1} \ \lambda\kappa^{1/2} n^{1/4}$ points $X_{i_j}$ with locations and directions agreeing with the position and tangent of $f$; for the purposes of the Theorem, we may take $A = A_0 c^{-1}$.

Now we verify that $f$ has sufficient regularity. By construction, $f$ is non-negative and twice differentiable. Moreover, since for fixed $t$ there is at most one non-zero term in the sum, we have

$$\|f\|_\infty \le \max_j \|f_j\|_\infty \qquad \text{and} \qquad \|f''\|_\infty \le \max_j \|f_j''\|_\infty.$$

First, $\|f_j\|_\infty \le z_j \left(1 + \frac{\eta|\tan\theta_j|}{z_j}\right)$, so that $\|f\|_\infty \le v\eta^2 + \eta\tan(w\eta) \le 1$ for $\eta$ small enough, i.e., $n$ large enough. Next, since

$$f_j''(t) = z_j/\eta^2 \ \psi''((t - t_j)/\eta) + (\tan\theta_j)/\eta \cdot (\psi\phi)''((t - t_j)/\eta),$$

we have $\|f''\|_\infty \le v\|\psi''\|_\infty + w\|(\psi\phi)''\|_\infty$. By (A.9)-(A.10) the graph of $f$ has maximum curvature bounded by $\kappa$. The length is controlled exactly as in Section A.3, ensuring that for $\eta$ small enough, the total length of graph$(f)$ does not exceed $\lambda$. But since $\eta_n \to 0$ as $n \to \infty$, we conclude that graph$(f) \in \mathcal{C}_\lambda(2, \kappa)$ for $n$ large enough. $\square$

### A.8. CTD and Geometric Discrepancy Theory

We mentioned in the introduction that there is a quantitative connection between our CTD problem and GDT. We give a simple example. Suppose that $\mathcal{C}$ is the class of all convex sets in $[0,1]^2$. We are given a set $X_i$ of points in $[0,1]^2$ and are interested in the discrepancy

$$\Delta(\mathcal{C}) = \sup_{C \in \mathcal{C}} |N_n(C) - n\text{Area}(C)|.$$

It is known, by work of Schmidt [31], [8, Theorem 15], that for *any* collection of $n$ points,

$$\Delta(\mathcal{C}) \geq cn^{1/3}.$$

CTD leads to the same conclusion for random point sets. Suppose that the point set $(X_i)$ is uniform random. Consider the class $\text{ConvGr}_2$ of convex graphs in the unit square. From our analysis in Section 3.5, we know that with overwhelming probability, there is a curve $\gamma$ which is the graph of a convex function $f$, say, and which passes through $cn^{1/3}$ points.

Consider the convex set $C$ formed as the convex hull of the range of $\gamma$.

For $\varepsilon > 0$, consider also the slightly shifted curve $\gamma_\varepsilon$ based on the convex function $f_\varepsilon = f + \varepsilon$, and define the set $C_\varepsilon$ as convex hull of the range of $\gamma_\varepsilon$. Now set $\varepsilon = 1/n^2$; we can be practically certain that $C_\varepsilon$ contains exactly the same points as $C$; and its area differs by at most $c\varepsilon$. Define the discrepancy at a specific set by

$$D_n(C) = |N_n(C) - n\text{Area}(C)|,$$

so clearly,

$$|D_n(C) - D_n(C_\varepsilon)| > N_n(\gamma) - c/n.$$

We conclude that

$$max(D_n(C), D_n(C_\varepsilon)) \geq N_n(\gamma)/2 - c/2n.$$

Hence, with overwhelming probability for large $n$,

$$\Delta(\mathcal{C}) = \sup_{C \in \mathcal{C}} |D_n(C)| \geq max(D_n(C), D_n(C_\varepsilon)) \geq cn^{1/3}.$$

So CTD theory gives lower bounds on discrepancy over the class of convex sets - among random pointsets.

Since Schmidt's bound for arbitrary point sets is of the same order, one gets the crude impression that the 'cause' of discrepancy is the number of points which can lie on the boundary of a convex set, and that uniform random pointsets are near-optimal for near-uniform behavior near boundaries. (They are badly suboptimal for behavior over classes of 'large' sets such as squares and rectangles). This set of connections seems worth pursuing. (We thank a referee for asking us whether there was a connection between the two types of problems).

## Acknowledgements

This work was partially supported by NSF grants DMS-0077261, DMS 01-40587 (FRG) and DMS-0346307. We would like to thank Hagit Hel-Or and Brian Wandell for helpful discussions and references. Thanks also to the anonymous referees for insightful questions and suggestions.

## References

[1] Pacman-miniclip.com. http://www.miniclip.com/pacman.htm.

[2] Supermarket sweep. http://www.supermarketsweep.com/.

[3] Abramowicz, H., Horn, D., Naftali, U. and Sahar-Pikielny, C. (1996). An orientation-selective neural network and its application to cosmic muon identification. *Nucl. Instr. Meth. Phys. Res. A* **378,** 305–311.

[4] Arias-Castro, E., Donoho, D. L. and Huo, X. (2003). Adaptive multiscale detection of filamentary structures embedded in a background of uniform random points. *Annals of Statistics*. Accepted and in press.

[5] Arias-Castro, E., Donoho, D. L. and Huo, X. (2003). Asymptotically optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Information Theory*. Accepted and in press.

[6] Baik, J. (2003). Limiting distribution of last passage percolation models. In *Proceedings, 14th International Congress of Mathematical Physics.* http://www.math.lsa.umich.edu/~baik/research/ICMP.pdf.

[7] Baik, J., Deift, P. and Johansson, K. (1999). On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.* **12,** 1119–1178.

[8] Beck, J. and Chen, W. W. L. (1987). *Irregularities of distribution* vol. 89 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, UK.

[9] Bronstein, E. M. (1976). $\varepsilon$-entropy of convex sets and functions. *Sibirsk. Mat. Ž.* **17,** 508–514, 715.

[10] Chui, C. K. and Smith, P. W. (1975). A note on Landau's problem for bounded intervals. *Amer. Math. Monthly* **82,** 927–929.

[11] Clements, G. F. (1963). Entropies of sets of functions of bounded variation. *Canadian J. Math.* **15,** 422–432.

[12] DeVore, R. and Lorentz, G. (1993). *Constructive Approximation*. Springer -Verlag, New York, USA.

[13] Diamond, J. and Kiely, K. (2002). Administration, agencies failed to connect the dots. *USA Today*. http://www.usatoday.com/news/washington/2002/05/17/failure-usatcov.htm.

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

[15] Donoho, D. L. and Huo, X. (2002). Beamlets and multiscale image analysis. In *Multiscale and Multiresolution Methods*. ed. T. Barth, T. Chan, and R. Haimes. vol. 20 of *Lecture Notes in Computational Science and Engineering*. Springer pp. 149–196.

[16] Edmunds, D. E. and Triebel, H. (1996). *Function spaces, entropy numbers, and differential operators*. Cambridge Univ. Press.

[17] Federer, H. (1969). *Geometric measure theory* vol. 153 of *Die Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, New York, USA.

[18] Few, L. (1955). The shortest path and the shortest road through $n$ points. *Mathematika* **2,** 141–144.

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

[20] Frieze, A. M. (1991). On the length of the longest monotone subsequence in a random permutation. *Ann. Appl. Probab.* **1,** 301–305.

[21] Groeneboom, P. (2001). Ulam's problem and Hammersley's process. *Ann. Probab.* **29,** 683–690.

[22] Huo, X., Donoho, D. L., Tovey, C. A. and Arias-Castro, E. Dynamic programming methods for "connecting the dots". *Technical report*.

[23] Johnson, D. S., McGeoch, L. A. and Rothberg, E. E. (1996). Asymptotic experimental analysis for the Held-Karp traveling salesman bound. In *Proc. 7th Annual Symposium on Discrete Algorithms (SODA), ACM-SIAM.*

[24] Kahn, J. M., Katz, R. H. and Pister, K. S. J. (2000). Emerging challenges: mobile networking for smart dust. *Journal of Communications and Networks* **2,** 188–196.

[25] Kovacs, I. and Julesz, B. (1993). A closed curve is much more than an incomplete one - effect of closure in figure-ground segmentation. *P. Natl. Acad. Sci. USA* **90,** 7495–7497.

[26] Lane, E. (2003). Board not ready to connect the dots. Probe into shuttle tragedy still working out the details. *NewYork Newsday*. http://www.newsday.com/news/ny-hsshut253253376apr25,0,2593631.story.

[27] Logan, B. F. and Shepp, L. A. (1977). A variational problem for random Young Tableaux. *Advances in Math.* **26,** 206–222.

[28] Matousek, J. (1999). *Geometric Discrepancy. An Illustrated Guide* vol. 18 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, Germany.

[29] Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press, Cambridge, UK.

[30] Odlyzko, A. M. and Rains, E. M. (2000). On longest increasing subsequences in random permutations. *Contemp. Math.* **251,** 439–451. Also published in *Analysis, geometry, number theory: the mathematics of Leon Ehrenpreis* (Philadelphia, PA, 1998).

[31] Schmidt, W. M. (1969). Irregularities of distribution iii. *Pacific J. Math.* **29,** 225–234.

[32] Schneider, B. Connecting the dots. CNN.com: Inside Politics 2002. http://www.cnn.com/2002/ ALLPOLITICS/05/17/pol.play.dots/.

[33] Shorack, G. and Wellner, J. (1986). *Empirical processes with applications to statistics*. Series in Probability and Mathematical Statistics. Wiley, New York, USA.

[34] Soshnikov, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Commun. Math. Phys.* **207,** 697–733. arXiv:math-ph/9907013.

[35] Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Pub. Math. IHES* **81,** 74–205.

[36] Tracy, C. A. and Widom, H. (2001). On the distributions of the lengths of the longest monotone subsequences in random words. *Probab. Theory Related Fields* **119,** 350–380.

[37] Vershik, A. M. and Kerov, S. V. (1977). Asymptotic behavior of the Plancherel measure of the symmetric group and the limit form of young tableaux. *Dokl. Akad. Nauk SSSR* **233,** 1024–1027. Russian.

**Authors's affiliations and addresses.**


- **Ery Arias-Castro**

  `acery@stanfordalumni.org; http://www-stat.stanford.edu/∼acery`

  Statistics Department,

  Stanford University,

  Stanford, CA 94305, USA

- **David Donoho**

  `donoho@stat.stanford.edu; http://www-stat.stanford.edu/∼donoho`

  Statistics Department,

  Stanford University,

  Stanford, CA 94305, USA

- **Xiaoming Huo**

  `xiaoming@isye.gatech.edu; http://www.isye.gatech.edu/∼xiaoming/`

  School of Industrial and System Engineering,

  Georgia Institute of Technology,

  765 Ferst Dr. Atlanta, GA 30332-0205

- **Craig Tovey**

  `craig.tovey@isye.gatech.edu; http://www.isye.gatech.edu/∼ctovey/`

  School of Industrial and System Engineering,

  Georgia Institute of Technology,

  765 Ferst Dr. Atlanta, GA 30332-0205