

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

ERY ARIAS-CASTRO* AND

DAVID L. DONOHO,* *Stanford University*

XIAOMING HUO** *** AND

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

Abstract

Given a class Γ of curves in $[0, 1]^2$, we ask: in a cloud of n uniform random points, how many points can lie on some curve $\gamma \in \Gamma$? Classes studied here include curves of length less than or equal to L , Lipschitz graphs, monotone graphs, twice-differentiable curves, and graphs of smooth functions with m -bounded derivatives. We find, for example, that there are twice-differentiable curves containing as many as $O_P(n^{1/3})$ uniform random points, but not essentially more than this. More generally, we consider point clouds in higher-dimensional cubes $[0, 1]^d$ and regular hypersurfaces of specified codimension, finding, for example, that twice-differentiable k -dimensional hypersurfaces in \mathbb{R}^d may contain as many as $O_P(n^{k/(2d-k)})$ uniform random points. We also consider other notions of ‘incidence’, such as curves passing through given location/direction pairs, and find, for example, that twice-differentiable curves in \mathbb{R}^2 may pass through at most $O_P(n^{1/4})$ uniform random location/direction pairs. Idealized applications in image processing and perceptual psychophysics are described and several open mathematical questions are identified for the attention of the probability community.

Keywords: Curve detection; filament detection; ε -entropy; configuration function; concentration of measure; longest increasing subsequence; traveling salesman problem; pattern recognition

2000 Mathematics Subject Classification: Primary 60D05

Secondary 62M40

1. Introduction

Several well-known games in popular culture require a player to travel through an arena scattered with ‘goodies’ and to gather as many of them as possible in a given amount of time; examples include the 1960s television show *Supermarket Sweep* (see <http://www.supermarketsweep.com/>) and the 1980s video game *PacMan* (see, for example, <http://www.miniclip.com/pacman.htm>).

Consider a geometric probability problem. Suppose that we have n points, X_1, \dots, 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 less than or equal to λ . This problem is similar

Received 16 September 2004; revision received 2 May 2005.

* Postal address: Department of Statistics, Stanford University, Stanford, CA 94305, USA.

** Postal address: School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA.

*** Email address: xiaoming@isye.gatech.edu

to the above-mentioned popular games, if we think of the goodies as being located at the points X_i , and imagine ‘participants’ traveling at unit speed for a length of time less than or equal to λ . The probability problem is to estimate the maximal number of points we can visit, i.e. to determine how many goodies we could typically gather by following the best possible path among (some subsets of) the random points.

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

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

$$P\{\frac{1}{5}\lambda\sqrt{n} \leq N_n(\mathcal{C}_\lambda) \leq 17\lambda\sqrt{n}\} \rightarrow 1, \quad n \rightarrow \infty.$$

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 result 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 and a class, Γ , of curves, γ , 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 λ , and we get order-of-magnitude estimates $O_P(\lambda\sqrt{n})$.

Before continuing with a 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 *nonrandom* 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 [30], [10], [24], where CTD refers to the identification of a small subset of facts among many random conflicting ones, thereby detecting a subtle pattern. Thus, journalists writing in [30], [10], and [24] 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 sources of confusion that one does not connect. The modern journalistic usage of the term CTD is broadly consistent with our own usage, and with potential applications in signal detection and pattern recognition (see below).

In our study of CTD problems, we will consider several choices for Γ : twice-differentiable curves, graphs of functions of bounded variation, graphs of increasing functions, and 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 size of the class Γ . 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 less than or equal to λ and curvature less than or equal to κ , pointwise. Then $\rho(\mathcal{C}_\lambda(2, \kappa)) = \frac{1}{3}$.
- *Graphs of bounded variation.* Let BVGr_τ denote the class of graphs $(x, f(x))$ taking values in $[0, 1]^2$, where the total variation is $\|f\|_{\text{TV}} \leq \tau$. Then $\rho(\text{BVGr}_\tau) = \frac{1}{2}$.

- *Lipschitz graphs.* Let LipGr_τ denote the class of graphs $(x, f(x))$ in which f is a function with range $[0, 1]$ and slope controlled by $|f(x) - f(x')| \leq \tau|x - x'|$. Then $\rho(\text{LipGr}_\tau) = \frac{1}{2}$.
- *m-fold-differentiable graphs.* Let $\text{DiffGr}_{m,\tau}$ denote the class of graphs $(x, f(x))$ in which f is an m -times-differentiable function with range $[0, 1]$ and $\|f^{(m)}\|_\infty \leq \tau$, where $\|\cdot\|_\infty$ is the maximum value function. Then $\rho(\text{DiffGr}_{m,\tau}) = 1/(m + 1)$.

In all these results, the size of the class Γ enters through the growth exponent of the ε -entropy of the class.

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

We also generalize the problem from the connection of points to where we have a curve passing through points with tangents having prescribed orientations; we call this the *connect-the-darts* problem.

1.2. Motivation

The CTD problem is interesting from several viewpoints.

1. Probability theory. The CTD problem generalizes two known problems of considerable interest among probabilists.

1. *Length of the longest increasing subsequence.* Suppose that we let IncrGr denote the class of increasing curves, i.e. of sets $\{(x, f(x)) : x \in [0, 1]\}$ in which $f : [0, 1] \rightarrow [0, 1]$ is monotone increasing. Then $N_n(\text{IncrGr})$ measures the result of *last-passage percolation* [4]. Also, if we write $X_i = (x_i, y_i)$, let π denote the sorting permutation (such that $x_{\pi(1)} \leq x_{\pi(2)} \leq \dots$), and define $w_i = y_{\pi(i)}, i = 1, \dots, n$, then $N_n(\text{IncrGr})$ is the length of the *longest increasing subsequence* among the numbers w_1, w_2, \dots, 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 1990s, with concentration of measure estimates [17], massive computational studies [28], and, finally, results on asymptotic distributions [5]. In the 1970s, Vershik and Kerov [35] and Logan and Shepp [25] showed that, asymptotically, $N_n(\text{IncrGr}) \sim 2\sqrt{n}$. (Groeneboom [18] gave a particularly simple proof of this.) The more delicate fluctuation distributional properties have been determined by Baik *et al.* [5], who showed that the asymptotic distribution follows the Tracy–Widom distribution [34]. It turns out that similar asymptotic results hold for a different CTD problem, in which the class Γ consists of Lipschitz graphs. The fact that study of the CTD problem extends the range of such limit phenomena seems interesting per se, especially in view of other universality results regarding the Tracy–Widom distribution [32].
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 $0.7124\sqrt{n}$, where 0.7124 is an approximation to the Beardwood–Halton–Hammersley constant [20]. The CTD problem instead considers the maximum number of points on a curve of fixed length, λ , independent of n . While the two problems would thus be closely connected if λ were

variable, with $\lambda = \lambda_n \approx 0.7124\sqrt{n}$, we do not consider this case in our approach. This leads to differences in application and interpretation – differences that might be stimulating to traveling salesman problem researchers. In addition, an interesting connection between these two problems is described at the end of Section 2.1.2.

II. *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, etc.) [6], [26]. As a referee has pointed out, CTD could be considered a variant of this approach, maximizing discrepancy over classes of curve. Since classes of curve involve objects of zero volume, measuring the discrepancy reduces 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 Appendix A.8; results on CTD imply bounds on the geometric discrepancy of random point-sets.

III. *Filament detection.* The CTD problem is relevant to inference problems in image analysis. Suppose that we observe n points scattered about the unit square. Consider the hypothesis testing problem, in which

- under hypothesis H_0 , the points are independent, random, and uniform on $[0, 1]^2$, and
- under hypothesis $H_{1,\Gamma,n}$, the vast majority of the points are again independent, random, and uniform on $[0, 1]^2$, but a small fraction ε_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 [1] or, more recently, sensor networks forming a smart dust [21]. 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 these false detections.

It is clear that, if the fraction ε_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, which shows three examples of scattered points in $[0, 1]$ and the corresponding CTD solution. In each case the class, Γ , of curves is the set of Lipschitz graphs. Panels (a), (c), and (e) show the point-sets only, while panels (b), (d), and (f) show the maximal Lipschitz curves, with 31, 35, and 54 points, respectively, out of a total of $n = 500$ points. Panel (a) shows a uniformly distributed random point-set. 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 nonuniform structure in (c) (compare (a)); however, a statistical test based on CTD counts can reliably establish its presence. 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 [2] and [19].

IV. *Vision research.* An interesting stream of vision research started with the experiments described in [16] and [23]. Both experiments presented specially prepared images to human

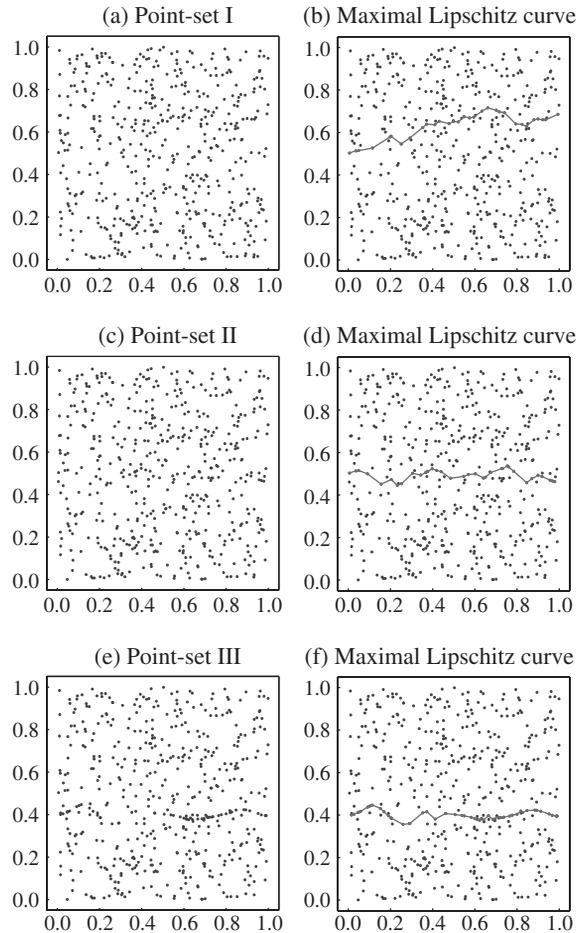


FIGURE 1: Examples of connect-the-dots.

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 that, 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 of this experiment was that the graphical elements were not points, but instead oriented patches. Panel (a) shows a collection of $n = 200$ darts taking uniform random positions and directions, panel (b) shows the maximal Hölder-2 function with first-order contact, and panel (c) extracts this function and its points of contact. Panels (d)–(f) are similar to panels (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 is $\beta = 15$. The power of the test is nearly 1.

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.

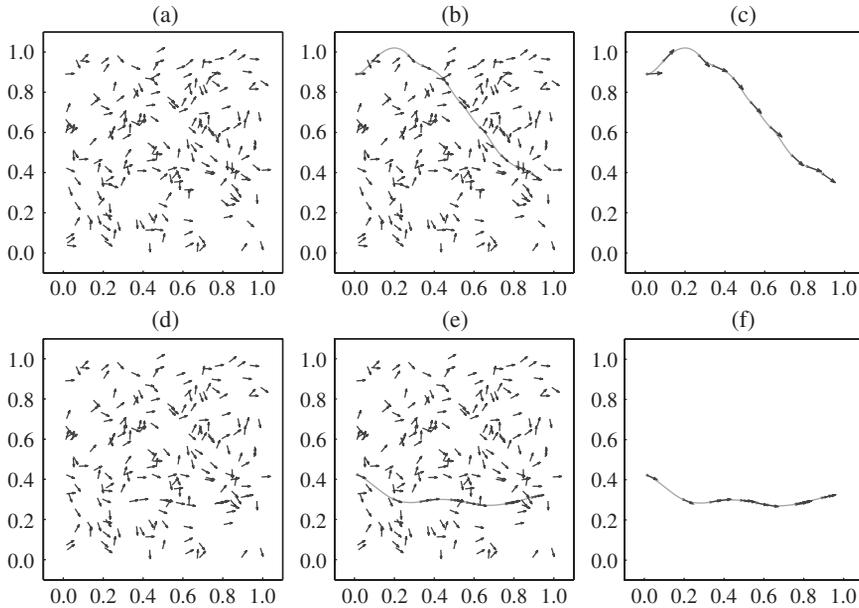


FIGURE 2: Examples of connect-the-darts.

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 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 ρ 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, in Section 2 we provide 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 found useful in building the algorithms in [2]. Then, in Section 3, we develop an abstract approach based on ε -entropy that allows us to efficiently derive growth exponents in a wide range of classes; examples such as graphs of smooth functions and graphs of bounded variation functions are also given in that section. In Section 4, we apply 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 controls the standard deviation of $N_n(\Gamma)$ and gives exponential bounds on fluctuations of $N_n(\Gamma) - \text{median}\{N_n(\Gamma)\}$. Perhaps some probabilists will be inspired by this article to complete the picture and derive more detailed distributional properties of $N_n(\Gamma)$.

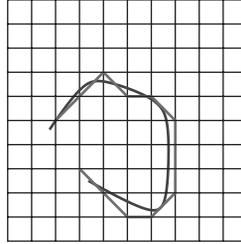


FIGURE 3: A rectifiable curve γ and its associated piecewise-linear path $\pi^n(\gamma)$.

2. Concrete approaches

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

2.1. CTD with bounded length

First consider the case $\Gamma = \mathcal{C}_\lambda$ mentioned in Theorem 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 numbers 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 integer k_i such that $0 \leq k_1, k_2 \leq \sqrt{n} = 1/\varepsilon_1$. 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 π is a sequence of vertices in \mathcal{V}_n connected by edges in \mathcal{E}_n . Such a path π has an *image* $\text{Im}(\pi)$ in the unit square, defined as the ε -neighborhood (in Hausdorff distance) of the union of the line segments that π traverses, where $\varepsilon = \frac{5}{4}\varepsilon_1$. The choice of ε comes from the proof of Lemma 1, as follows.

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

Proof. 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 a $\frac{1}{2}\varepsilon_1$ -covering of $[0, \ell]$, denoted by $\{s_j : j = 1, \dots, J\}$, chosen so that $J \leq \ell/\varepsilon_1 + 1 \leq \lambda/\varepsilon_1 + 1$. For each $j = 1, \dots, J$, let b_j be a closest grid point to $\gamma(s_j)$. Note that $|\gamma(s_j) - b_j| \leq \varepsilon_1/\sqrt{2}$. Since

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

the line segment $[b_j, b_{j+1}]$ is in \mathcal{E}_n . Therefore, we may define π to be the path b_1, \dots, b_J . Moreover, since

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

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

We extend our definition of $X^n(S)$ so that, whenever S is a subset $S \subset [0, 1]^2$, not necessarily a curve, we have $X^n(S) = \text{card}\{i : X_i \in S\}$ (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, we have $N_n(\mathcal{S}) = \max\{X^n(S) : S \in \mathcal{S}\}$. Also define Y^n so that if π is a path in \mathcal{G}_n , then $Y^n(\pi) = X^n(\text{Im}(\pi))$; and for Π a family of paths, let $M_n(\Pi) = \sup\{Y^n(\pi) : \pi \in \Pi\}$.

Now let

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

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

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

Since Π_λ^n is finite, Boole’s inequality gives, for all $B > 0$,

$$P\{M_n(\Pi_\lambda^n) > B\lambda\sqrt{n}\} \leq \text{card}(\Pi_\lambda^n) \max_{\pi \in \Pi_\lambda^n} P\{Y^n(\pi) > B\lambda\sqrt{n}\}.$$

Lemma 2. *For $n > n_0(\lambda)$, Y^n and Π_λ^n have the following properties.*

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

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

Using this lemma, we obtain the bound

$$\max_{\pi \in \Pi_\lambda^n} P\{Y^n(\pi) > B\lambda\sqrt{n}\} \leq P\{\text{bin}(n, 6\lambda/\sqrt{n}) > B\lambda\sqrt{n}\},$$

which is valid for all $B > 0$. Hoeffding’s inequality [31] gives us control over the tail of the binomial distribution.

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

$$P\{\text{bin}(n, p) > Cnp\} \leq \varphi(C)^{np},$$

where

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

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

$$P\{\text{bin}(n, 6\lambda/\sqrt{n}) > B\lambda\sqrt{n}\} \leq \varphi\left(\frac{1}{6}B\right)^{6\lambda\sqrt{n}}.$$

Combining all the above, we have

$$\begin{aligned} P\{N_n(\mathcal{C}_\lambda) > B\lambda\sqrt{n}\} &\leq 1.1n \cdot 9^{\lambda\sqrt{n}} \varphi\left(\frac{1}{6}B\right)^{6\lambda\sqrt{n}} \\ &= 1.1n \exp(-\lambda\sqrt{n}(-6 \log(\varphi(\frac{1}{6}B)) - \log(9))). \end{aligned}$$

We now choose $B = 17$, so that $-6 \log(\varphi(\frac{1}{6}B)) > \log(9)$, making the right-hand side tend to 0 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 sufficiently small set of regions (property 1), each of which is of sufficiently small area (property 2) that every curve in the class is contained in a region. The abstraction in Section 3 treats the regions as a set of ε -balls in the associated Hausdorff metric, such that they cover the class of curves. Property 1 corresponds to the volume of each ε -ball being sufficiently small, and property 2 corresponds to the ε -entropy (the logarithm of the least number of balls in an ε -covering) being sufficiently small.

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, \dots . We connect successive points by linear interpolation, argue that the result has length less than or equal to λ , and then show that, with probability tending to 1, $J_{n,\lambda} \geq A\lambda\sqrt{n}$.

To extract the points, subdivide $[0, 1]^2$ into $\varepsilon_1 \times \varepsilon_1$ squares as in Section 2.1.1. We define a zigzag ordering of the squares, starting in the upper-left corner as shown in the following array:

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 nonempty 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 λ . Denote the constructed curve by $\gamma_{n,\lambda}$. It is random, as its construction depends on X^n , and certainly belongs to \mathcal{C}_λ .

Let $J_{n,\lambda}$ denote the (random) number of nonempty 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 nonempty, connecting points associated with such squares could not lead to a length exceeding λ . 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, given in Appendix A.1, show that the following lemma holds.

Lemma 4. $P\{J_{n,\lambda} \geq \frac{1}{5}\lambda\sqrt{n}\} \rightarrow 1$ as $n \rightarrow \infty$.

The lower bound in Theorem 1 follows.

The above probabilistic analysis used a ‘selection plus interpolation’ (select–interpolate) approach that can serve as a common framework in later results. In this particular setting, a nonprobabilistic analysis is also possible, invoking previous research about the traveling salesman problem. In 1955, L. Few proved that, given n points on a unit square, there is a curve of length not exceeding $\sqrt{2n} + \frac{7}{4}$ that traverses *all* the n points [15]. Divide such a path into $m = \lceil (\sqrt{2n} + \frac{7}{4})/\lambda \rceil$ consecutive pieces $\gamma_1, \dots, \gamma_m$, with lengths less than or equal to λ . By definition, $N_n(\mathcal{C}_\lambda)$ is larger than each $X^n(\gamma_i)$ and, so, larger than their average. Hence,

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

2.2. CTD with bounded curvature

We now increase the assumed regularity of our curves, controlling both the curvature and length. Let $\mathcal{C}_\lambda(2, \kappa)$ denote the class of C^2 curves with lengths bounded by λ and curvatures bounded pointwise by κ . We are interested in $N_n(\mathcal{C}_\lambda(2, \kappa))$.

Note that a constraint on κ 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 less than or equal to κ and its image lies in $[0, 1]^2$, then, for small enough κ , the curve will not be able to ‘wind around’ within the unit square; this constrains the achievable lengths. For a given bound κ on curvature, let $\lambda_0(\kappa)$ give the maximum length subject to that bound. There is a threshold κ_0 such that $\lambda_0(\kappa) < \infty$ for $\kappa < \kappa_0$. In the extreme case $\kappa \rightarrow 0$, curves in $\mathcal{C}_\lambda(2, \kappa)$ are almost straight; hence, $\lambda_0(\kappa) \rightarrow \sqrt{2}$ as $\kappa \rightarrow 0$ and, if $\kappa < \kappa_0$, we assume that $\lambda < \lambda_0(\kappa)$, ensuring that at least one curve in $\mathcal{C}_\lambda(2, \kappa)$ has length greater than or equal to λ .

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

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

$$P\{A\lambda\kappa^{1/3}n^{1/3} \leq N_n(\mathcal{C}_\lambda(2, \kappa)) \leq B\lambda\kappa^{1/3}n^{1/3}\} \rightarrow 1, \quad n \rightarrow \infty.$$

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

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

2.2.1. Upper bound. As in the case of unconstrained curvature, 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 with regions in the plane will reflect the underlying reason for the $\rho = \frac{1}{3}$ 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 and horizontal grid points $\{(k_1\varepsilon_1, k_2\varepsilon_2)\}$ and $\{(k_2\varepsilon_2, k_1\varepsilon_1)\}$, respectively. Here $0 \leq k_1 \leq 1/\varepsilon_1$ and $0 \leq k_2 \leq 1/\varepsilon_2$, where $\varepsilon_i = 2^{-m_i}$ with

- $m_2 = \frac{1}{3}(2\lceil \log_2(n) \rceil - \lfloor \log_2(\kappa) \rfloor)$, so that $\varepsilon_2 \approx \kappa^{1/3}n^{-2/3}$, and
- $m_1 = \frac{1}{2}(m_2 + k + \lceil \log_2(\kappa) \rceil)$, where k is a universal (integer) constant defined in Appendix A.2.

Then consider line segments that join two vertical or horizontal grid points such that their angle with the horizontal or, respectively, vertical direction does not significantly exceed 45° ; formally, such a line segment is defined by its endpoints $(k_1\varepsilon_1, k_2\varepsilon_2)$ and $(k'_1\varepsilon_1, k'_2\varepsilon_2)$, or, respectively, $(k_2\varepsilon_2, k_1\varepsilon_1)$ and $(k'_2\varepsilon_2, k'_1\varepsilon_1)$, with the restriction that

$$|k'_1 - k_1| = 1 \quad \text{and} \quad |k'_2 - k_2| < \frac{\varepsilon_1}{\varepsilon_2} + 1.$$

Such line segments will be respectively called *horizontal or vertical beamlets*; see Figure 4. Roughly speaking, the beamlets have lengths comparable to $n^{-1/3}$ and slopes chosen from a grid of spacing $\asymp n^{-2/3}$, i.e. the spacing has an asymptotic order of $n^{-2/3}$.

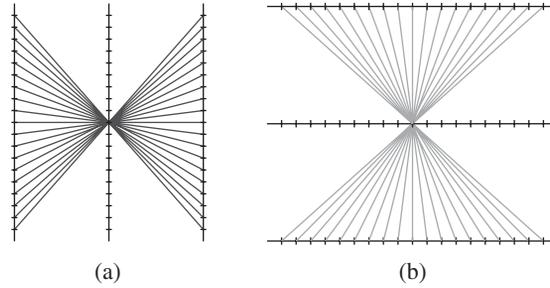


FIGURE 4: Panel (a) shows all the horizontal beamlets passing through a vertical grid point, while panel (b) shows all the vertical beamlets passing through a horizontal grid point.

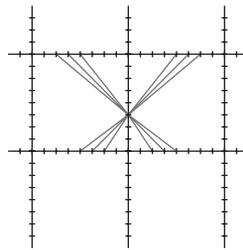


FIGURE 5: Diagonal beamlets passing through a vertical grid point.

We also consider line segments joining vertical and horizontal grid points belonging to the same square and making an angle close to 45° 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 $k'_1 = k_{21} + 1$, 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 up of all beamlets defined above. They provide an efficient, dyadically organized, multiscale, multiorientation organizational structure. The graphical structure we build using them is a variant of the beamlet graph defined in [12], [3], and [11].

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 that their directions are sufficiently similar [2]. Formally, two horizontal beamlets connected in \mathcal{E}_n are of the form

$$[(k_1\varepsilon_1, k_2\varepsilon_2), ((k_1 + 1)\varepsilon_1, k'_2\varepsilon_2)] \quad \text{and} \quad [((k_1 + 1)\varepsilon_1, k'_2\varepsilon_2), ((k_1 + 2)\varepsilon_1, k''_2\varepsilon_2)],$$

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 the same sense); see Figure 7.

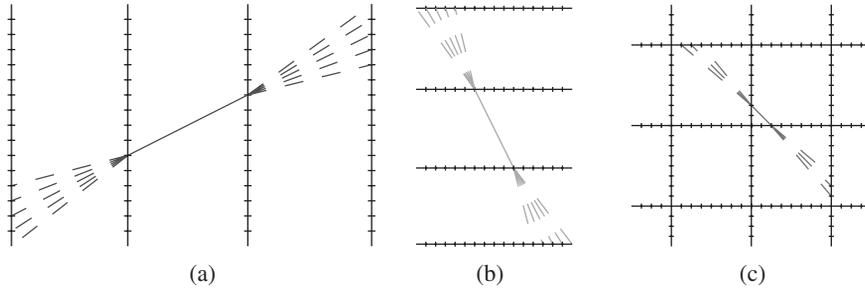


FIGURE 6: Panel (a) shows a horizontal beamlet and its horizontal neighbors, panel (b) shows a vertical beamlet and its vertical neighbors, while panel (c) shows a diagonal beamlet and its diagonal neighbors.

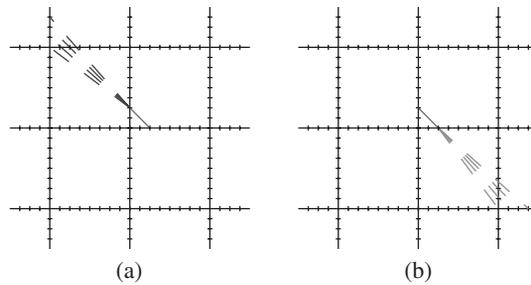


FIGURE 7: Panel (a) shows a diagonal beamlet and its neighboring horizontal beamlets, while panel (b) shows a diagonal beamlet and its neighboring vertical beamlets.

In the graph $\mathcal{G}_n = (\mathcal{V}_n, \mathcal{E}_n)$, a path π is a sequence of vertices in \mathcal{V}_n connected by edges in \mathcal{E}_n . Such a path π has an image $\text{Im}(\pi)$ in the unit square: the tubular region defined as the ε -neighborhood of the union $\bigcup_{p \in \pi} p$ of beamlets visited by π , 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 $\text{Im}(\pi)$ covers γ . Moreover, π 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.

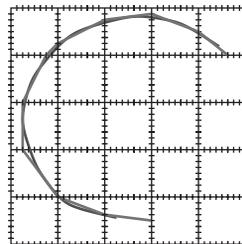


FIGURE 8: A smooth curve γ and its associated piecewise-linear path $\pi^n(\gamma)$.

Now define $\Pi_{\lambda,\kappa}^n = \{\pi^n(\gamma) : \gamma \in \mathcal{C}_\lambda(2, \kappa)\}$, $Y^n(\pi) = X^n(\text{Im}(\pi))$, and $|\pi| = \text{area}(\text{Im}(\pi))$.

Lemma 6. Y^n and $\Pi_{\lambda,\kappa}^n$ satisfy the following properties.

1. $\text{card } \Pi_{\lambda,\kappa}^n \leq cn^2 \cdot 14^{c\lambda\kappa^{1/3}n^{1/3}}$.
2. For any path $\pi \in \Pi_{\lambda,\kappa}^n$, $Y^n(\pi) \sim \text{bin}(n, |\pi|)$ with $|\pi| \leq c\lambda\kappa^{1/3}n^{-2/3}$.

The first property comes from the facts that vertices in \mathcal{G}_n have degree less than or equal to 14 and there are fewer than $(2\varepsilon_1^{-1}\varepsilon_2^{-1})^2 \leq cn^2$ vertices in total. The second property comes from the fact that $\text{Im}(\pi)$ is contained in the union of the $\kappa^{1/3}n^{-2/3}$ -neighborhoods of the beamlets that π connects; there are at most $c\lambda\kappa^{1/3}n^{1/3}$ beamlets in π , and the $\kappa^{1/3}n^{-2/3}$ -neighborhood of each one 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 α ?

Our answer requires a more formal statement of the question.

- For α a positive integer, let $\mathcal{H}^{k,1}(\alpha, \beta)$ denote the functions $\{g : [0, 1]^k \rightarrow [0, 1]\}$ 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 $\alpha_1 + \dots + \alpha_k = \alpha - 1$. Here $g^{(\alpha_1, \dots, \alpha_k)}$ is a partial derivative, with α_i being the order of differentiation with respect to the i th coordinate.

- For $\alpha > 1$ not an integer, let $\mathcal{H}^{k,1}(\alpha, \beta)$ denote the functions $\{g : [0, 1]^k \rightarrow [0, 1]\}$ 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 $\alpha_1 + \dots + \alpha_k = \lfloor \alpha \rfloor$.

Here $\lfloor \cdot \rfloor$ denotes the integer part and $\{\cdot\}$ denotes the fractional part (not to be confused with the single-member set – this will be clear from context).

Speaking very loosely, β measures the size of the α 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 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 $\tau \equiv \tau(k, d, \alpha, \beta)$,

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

We 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 μ . Let $X_1, \dots, X_n \in \mathbb{X}$ be independent with common distribution μ . We consider \mathcal{S} , a class of subsets of \mathbb{X} , and let $N_n(\mathcal{S}) = \max_{S \in \mathcal{S}} X^n(S)$, where, as before, if $S \subset \mathbb{X}$ then $X^n(S) = \text{card}\{i : X_i \in S\}$. Let δ be a semimetric on \mathbb{X} and let Δ be the corresponding Hausdorff semimetric on the subsets of \mathbb{X} . For a subset $S \subset \mathbb{X}$, and $\varepsilon > 0$, let $(S)_\varepsilon$ denote the ε -neighborhood of S in the Δ semimetric.

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

We measure the volume of ε -neighborhoods of sets in \mathcal{S} as follows:

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

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

$$P\{N_n(\mathcal{S}) > Bn^{b/(a+b)}\} \rightarrow 0, \quad n \rightarrow \infty. \tag{1}$$

Proof. We just saw that

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

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

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

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

$$X^n((S)_\varepsilon) \sim \text{bin}(n, \mu\{(S)_\varepsilon\}).$$

Hence,

$$\max_i P\{(S_i)_\varepsilon > u\} \leq P\{\text{bin}(n, c_2\varepsilon^a) > u\},$$

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

Again using Hoeffding’s inequality, as in Lemma 3, and with φ as defined there, we have

$$P\{\text{bin}(n, c_2\varepsilon^a) > Cc_2n\varepsilon^a\} \leq \varphi(C)c_2n\varepsilon^a.$$

Hence,

$$P\{N_n(\mathcal{S}) > Cc_2n\varepsilon^a\} \leq \exp(c_1\varepsilon^{-b} + c_2n\varepsilon^a \log \varphi(C)). \tag{2}$$

We choose $\varepsilon = \varepsilon_n$, solving $\varepsilon^{-b} = n\varepsilon^a$. Then, (2) becomes

$$P\{N_n(\mathcal{S}) > Cc_2n\varepsilon^a\} \leq \exp((c_1 + c_2 \log \varphi(C))n^{b/(a+b)}). \tag{3}$$

For C such that $c_1 + c_2 \log \varphi(C) < 0$, the right-hand side of (3) tends to 0 as n increases. Rewriting the left-hand side of (3), we obtain

$$P\{N_n(\mathcal{F}) > Cc_2n^{b/(a+b)}\} \rightarrow 0, \quad n \rightarrow \infty,$$

for C sufficiently large. Equation (1) follows upon defining $B = Cc_2$. From the fact that $\varphi(C) \leq \exp(-\frac{1}{6}C)$ for $C > 2$, we see that (1) holds for $B \geq 6 \max\{c_2, c_1\}$.

3.2. Application to Hölder immersions

We now use the abstract approach to prove the upper bound in Theorem 3. In that setting $\mathbb{X} = [0, 1]^d$, μ is the uniform density on \mathbb{X} , δ is the usual Euclidean distance, and Δ is the usual Hausdorff distance. The classes $\mathcal{H}^{k,d}(\alpha, \beta)$ and $\mathcal{G}^{k,d}(\alpha, \beta)$ are as 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 lemma is basically self-evident.

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

$$\Delta(f([0, 1]^k), g([0, 1]^k)) \leq \|f - g\|_\infty. \tag{4}$$

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

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

We now recall a very well-known result on the ε -entropy of Hölder classes.

Theorem 4. (Kolmogorov and Tikhomirov [22].) *For a constant, $c_{k,d,\alpha}$,*

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

Finally, we estimate the key volumetric quantity $M(\varepsilon)$. We note the connection with the notion of Minkowski content (see [14]). Indeed, fix $S \in \mathcal{G}^{k,d}(\alpha, \beta)$; then

$$\frac{\mu((S)_\varepsilon)}{v_{d-k}\varepsilon^{d-k}} \rightarrow \text{vol}(S), \quad \varepsilon \rightarrow 0,$$

where v_{d-k} denotes the volume of the $(d - k)$ -dimensional unit ball and $\text{vol}(S)$ is the k -dimensional Hausdorff measure of S (see [14, Equation (3.2.39)]). We have the following lemma.

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

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

Then, for sufficiently small ε and a constant, $c_{k,d}$,

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

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

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

Therefore, without loss of generality, we assume that $\alpha = 1$.

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

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

For d, k , and β fixed and ε small, this implies that

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

We now combine these results with Lemma 7 to obtain the upper bound of Theorem 3.

The authors do not know an abstract lower bound technique. Using the select–interpolate method, we give a lower bound for Hölder objects in Appendix A.4. This matches the order of magnitude of the 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 now consider the class $BVGr_\tau$ mentioned in the introduction. We make use of the following theorem.

Theorem 5. (Clements [8].) *Let $H_\varepsilon(BVGr_\tau)$ denote the ε -entropy of the class of graphs of functions f with $|f|_{TV} \leq \tau$ and $|f|_\infty \leq 1$, with respect to the usual Hausdorff distance. Then*

$$H_\varepsilon(BVGr_\tau) \leq C\tau/\varepsilon, \quad \varepsilon \in (0, 1).$$

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

Lemma 10. *Fix $\tau > 0$ and let*

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

Then

$$M(\varepsilon) = O(\varepsilon), \quad \varepsilon \rightarrow 0.$$

Theorem 5 and Lemma 10 give us the raw ingredients to apply Lemma 7 with exponents $a = 1$ and $b = 1$, yielding $\rho(BVGr_\tau) \geq \frac{1}{2}$. A matching lower bound follows immediately, from the inclusion

$$\mathcal{H}^{1,1}(1, \tau) \subset BVGr_\tau, \quad \tau > 0,$$

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

3.4. Increasing graphs

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

$$IncrGr \subset BVGr_1,$$

from which it follows that $N_n(IncrGr) = O_P(n^{1/2})$. Of course, we have already pointed out that $N_n(IncrGr) = 2n^{1/2}(1 + o_P(1))$ [35], [25], 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 now consider the class, $ConvGr_d$, of graphs $\{(x, f(x))\}$ in \mathbb{R}^d of convex functions $f: [0, 1]^{d-1} \rightarrow [0, 1]$.

Theorem 6. (Bronšteĭn [7].) *Let $H_\varepsilon(ConvGr_d)$ denote the ε -entropy for the class $ConvGr_d$ with respect to the usual Hausdorff distance. Then*

$$H_\varepsilon(ConvGr_d) \leq C_d \varepsilon^{-(d-1)/2}, \quad \varepsilon \in (0, 1).$$

We state the next lemma without proof.

Lemma 11. *Fix d and let*

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

Then

$$M(\varepsilon) = O(\varepsilon), \quad \varepsilon \rightarrow 0.$$

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

3.6. Lipschitz graphs

We now consider the class $LipGr_\tau$, also mentioned in the introduction. This may be viewed as the special, particularly elementary case $\alpha = 1$ in the scale of Hölder classes.

Now let Lip_τ be the set of functions f with Lipschitz constant less than or equal to τ and $0 \leq f \leq 1$, and let $LipGr_\tau$ be the collection of graphs of such functions.

Theorem 7. (Kolmogorov and Tikhomirov [22].) *Let $H_\varepsilon(Lip_\tau; |\cdot|_\infty)$ denote the ε -entropy of the class of functions f with Lipschitz constant less than or equal to τ and $|f|_\infty \leq 1$, with respect to the usual supremum norm. Then*

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

Let $H_\varepsilon(LipGr_\tau; \Delta)$ denote the ε -entropy of this class of graphs with respect to the Hausdorff distance Δ . Then, from (4), we have

$$H_\varepsilon(LipGr_\tau; \Delta) \leq H_\varepsilon(Lip_\tau; |\cdot|_\infty), \quad \varepsilon \in (0, 1).$$

By arguments similar to those supporting Lemma 9, we obtain our next result.

Lemma 12. Fix $\tau > 0$ and let

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

Then

$$M(\varepsilon) = O(\varepsilon), \quad \varepsilon \rightarrow 0.$$

By combining Theorem 7 and Lemma 12, we obtain the exponents $a = 1$ and $b = 1$. Lemma 7 then gives $\rho(\text{LipGr}_\tau) \geq \frac{1}{2}$. A matching lower bound follows immediately, from the inclusion

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

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

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

3.7. Besov and 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 [27], [9], [13, p. 105]. It turns out that results very similar to those in the Hölder case hold over these more general scales, even though the notion of regularity can be quite different.

We use $\mathcal{B}_{p,q}^{k,d}(\alpha, \beta)$ to denote classes of immersions built from Besov function classes $B_{p,q}^\alpha[0, 1]^d$ 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 $\|f_j\|_{H^\alpha} \leq \beta$. In a parallel fashion, we now constrain the components of such an f to each have Besov norm $\|f_j\|_{B_{p,q}^\alpha} \leq \beta$.

In the Besov scale, $\alpha > 0$ measures smoothness, while p and q are second-order parameters measuring the ‘uniformity’ of that smoothness. Thus, $B_{\infty,\infty}^\alpha[0, 1]^d$ with α nonintegral is equivalent to the usual Hölder class, and $B_{2,2}^\alpha[0, 1]^d$ is equivalent to the usual L^2 -Sobolev class of smoothness α . The bump algebra $B_{1,1}^1[0, 1]$ is an interesting approximation to bounded variation [27], and several other interesting spaces can be obtained by a proper choice of $\alpha, p,$ and q .

We use $\mathcal{F}_{p,q}^{k,d}(\alpha, \beta)$ to denote classes of immersions built from Triebel function classes $F_{p,q}^\alpha[0, 1]^d$. Again, $\alpha > 0$ measures smoothness, p and q are second-order parameters measuring the uniformity of that smoothness, and $F_{2,2}^\alpha[0, 1]^d$ is the usual L^2 -Sobolev space of smoothness α , while several other interesting spaces (e.g. bounded mean oscillation and H^1 [27]) can be obtained by a proper choice of $\alpha, p,$ and q .

We extend the notion of growth exponent slightly.

Definition 1. A sequence (N_n) has *critical growth exponent* $\hat{\rho}$ if and only if $N_n = O_P(n^{\hat{\rho}+\varepsilon})$ for each $\varepsilon > 0$ but for no $\varepsilon < 0$.

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

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

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 that we are given a sequence of *points and directions* (x_i, θ_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 as tangents the directions given at those points.

We say that a C^2 unit-speed curve γ makes *first-order contact* with this point-set at $N \equiv N(\gamma, (x_i, \theta_i)_{i=1}^n)$ points if there are N indices $1 \leq i_1 < i_2 < \dots < i_N \leq n$ and corresponding arc length parameters s_j such that $\gamma(s_j) = x_{i_j}$ and, at such intersections, the tangent $\dot{\gamma}(s_j) = \exp\{i\theta_{i_j}\}$. 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 [16], [23], in which human subjects are shown pictures containing many randomly oriented objects. It may or may not be the case, in a given picture, 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 that we have n observations

$$X_i = (x_i, e^{i\theta_i}) \stackrel{\text{i.i.d.}}{\sim} \mu = \text{uniform}(0, 1)^2 \times \text{uniform}(\mathbb{S}_1),$$

where ‘i.i.d.’ means that the observations are independently and identically distributed. We are interested in C^2 unit-speed curves of length less than or equal to λ and curvature less than or equal to κ . (This is the class $\mathcal{C}_\lambda(2, \kappa)$ that was defined earlier.) Each such curve γ 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 point-sets. As before, for a subset S of \mathbb{X} , write $X^n(S) = \text{card}\{i : X_i \in S\}$ and, for the class $\Sigma_\lambda(2, \kappa)$, write

$$N_n(\Sigma_\lambda(2, \kappa)) = \sup_{S \in \Sigma_\lambda(2, \kappa)} X^n(S).$$

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

$$P\{A\lambda\kappa^{1/2}n^{1/4} \leq N_n(\Sigma_\lambda(2, \kappa)) \leq B\lambda\kappa^{1/2}n^{1/4}\} \rightarrow 1, \quad n \rightarrow \infty.$$

Comparing this result with the CTD problem for C^2 curves, we see that incorporating the direction-matching constraint reduces the exponent ρ from $\frac{1}{3}$ to $\frac{1}{4}$. 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 the connect-the-darts problem were shown in Figure 2. There are noticeably more points on the curves in Figure 9 than there are in Figure 2.

In Figure 10, the maximum number of points and darts on C^2 curves in some finite simulations are compared. The histograms demonstrate that a detector decision based on unusually many points in a connect-the-darts problem can be dramatically more sensitive than a decision based on unusually many points in a connect-the-dots problem. These ‘ideal observer’ results correlate strongly with the performance of humans in curve detection experiments [16]. The results are also consistent with our theoretical results on growth exponents.

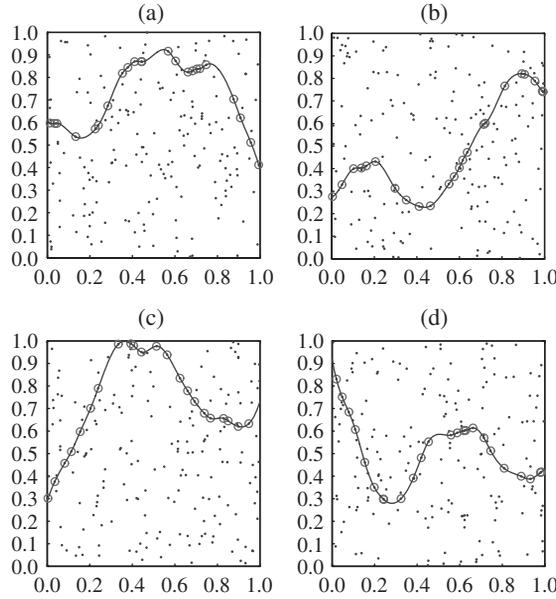


FIGURE 9: Examples of maximal Hölder-2 curves, with $n = 200$. The Hölder constant is 30.

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

$$\Delta(S_1, S_2) = \inf \|\gamma_1 - \gamma_2\|_\infty, \tag{5}$$

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 at first be surprising that we use Δ , 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 transpires (see Appendix A.6) that, on $\Sigma_\lambda(2, \kappa)$, Δ 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 ε -entropy of $\Sigma_\lambda(2, \kappa)$ for the metric Δ .

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

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

Proof. To prove the lemma, recall that the metric Δ introduced in (5) ignores the orientation component. Hence, we have the identity

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

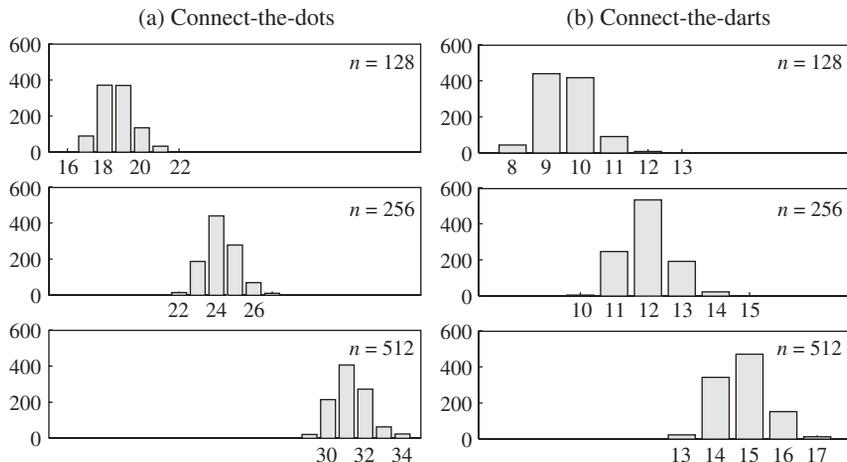


FIGURE 10: Histograms for the maximal number of points on a Hölder-2 curve (panel (a)) and the maximal number of darts tangent to a Hölder-2 curve (panel (b)). The sample size, n , is shown in each chart. The Hölder constant is chosen to be 30 in all cases.

where, on the right-hand side, Δ denotes Hausdorff distance with respect to the Euclidean distance.

From here, we use the graph \mathcal{G}_n constructed in Section 2.2.1, but with

$$m_2 = \lfloor \log_2(1/\varepsilon) \rfloor$$

so that $\varepsilon_2 \leq \varepsilon$. In the proof of Lemma 5 (in Appendix A.2), we see that a curve $\gamma \in \mathcal{C}_\lambda(2, \kappa)$ is contained in the ε_2 -neighborhood of its associated path $\pi^n(\gamma)$. Since there are at most

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

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

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

We next estimate the order of magnitude of

$$M(\varepsilon) = \sup\{\mu((S)_\varepsilon) : S \in \Sigma_\lambda(2, \kappa)\},$$

with $(S)_\varepsilon$ the ε -neighborhood of S with respect to Δ . For the proof, see Appendix A.6.

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

$$M(\varepsilon) \leq c_2 \lambda \kappa^{1/2} \varepsilon^{3/2}, \quad \varepsilon \rightarrow 0.$$

Now, by applying Lemma 7 with exponents $a = \frac{3}{2}$ and $b = \frac{1}{2}$, we obtain the upper bound presented 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. $N_n(\Gamma) = O_P(n^{\rho(\Gamma)})$. The following refinements seem plausible.

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

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

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

2. *Negligible fluctuations.* The fluctuations might be relatively small, i.e.

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

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, in the 1990s there was intense interest in understanding the longest increasing subsequence problem. This is essentially the CTD problem with $\Gamma = \text{IncrGr}$, the class of increasing graphs. In this case, we know that for $N_n \equiv N_n(\text{IncrGr})$, the longest increasing subsequence,

- $\text{median}\{N_n\} \sim 2\sqrt{n}$ (see [35] and [25]);
- $|N_n - \text{median}\{N_n\}| = O_P(n^{1/3})$ (see [17]); and
- $(N_n - E\{N_n\})/\text{SD}\{N_n\}$ has the Tracy–Widom distribution [5].

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

While the longest increasing subsequence 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 longest increasing subsequence problem appear to hold in at least one other CTD problem.

5.2. Lipschitz graphs

Consider the collection of Lipschitz graphs LipGr_1 , mentioned both in the introduction and in Section 3.6. We have observed the following properties for $N_n \equiv N_n(\text{LipGr}_1)$:

- $\text{median}\{N_n\} \sim 2\sqrt{n}$, both empirically and in theory;
- $|N_n - \text{median}\{N_n\}| = O_P(n^{1/3})$, empirically; and
- $(N_n - E\{N_n\})/\text{SD}\{N_n\}$ has the Tracy–Widom distribution, empirically.

These properties are entirely analogous to those in the longest increasing subsequence 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 [19]. Furthermore, 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.

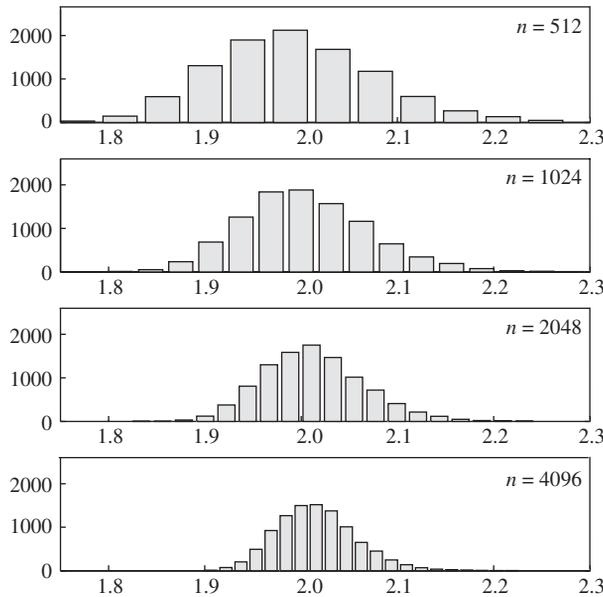


FIGURE 11: Histograms of the maximal number of points on Lipschitz graphs for uniform random samples of different sample sizes n . The values of $N_n(\text{LipGr}_1)$ are divided by \sqrt{n} . Clearly, the median is converging to $2\sqrt{n}$. Statistical analysis shows that the histogram shape is nonnormal and matches the Tracy–Widom distribution.

5.3. Simulations

In [19], the authors described a software library, CTDLab, that 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 $\text{median}\{N_n(\Gamma)\}$ for the classes studied. As a simple example, we display in Figure 11 a sequence of histograms of $N_n \equiv N_n(\text{LipGr}_1)$ for various values of 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 $\text{median}\{N_n(\mathcal{S})\} \sim An^\rho$ for general \mathcal{S} , there is a general proof of the asymptotic negligibility of fluctuations, in a very strong sense.

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

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

Deviations of size $t\sqrt{M_n}$ have a probability that decays exponentially in $|t|$, uniformly 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 [33]. Let us introduce Talagrand’s key abstract notion.

Definition 2. (Talagrand [33, Definition 7.1.7].) $L: \mathbb{X}^n \rightarrow \mathbb{R}$ is a *configuration function* if, given any $x^n = (x_i) \in \mathbb{X}^n$, there exists a subset J of $\{1, \dots, n\}$, with $\text{card}(J) = L(x^n)$, such that, for each $y^n = (y_i) \in \mathbb{X}^n$, we have $L(y^n) \geq \text{card}\{i \in J : y_i = x_i\}$.

We now 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. If there is an $S_0 \in \mathcal{S}$ 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 in S_0 and, hence, the maximal number of entries of Y^n in any $S \in \mathcal{S}$ is at least k .

Theorem 10 is now immediate from Talagrand’s concentration-of-measure result for configuration functions, as follows.

Theorem 11. (Talagrand [33, Theorem 7.1.3].) Let $L = L(X^n)$ be a configuration function applied to a random uniform point cloud X^n with $\text{median}\{M\} = \text{median}\{L\}$. Then, for $u > 0$,

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

6. Conclusions

We have introduced the notion of connect-the-dot problems and derived asymptotic growth properties for a range of such problems. We have 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 have pointed out that concentration-of-measure results hold for such problems, and have 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 [19]. These showed clearly that the asymptotic theory of growth exponents developed in this paper accurately describes the behavior in moderately sized random point clouds. We hope our work inspires 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:

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

Now,

$$P\{J_{n,\lambda} \geq \ell\} = P\{J_{n,\lambda} \geq \ell \mid K_{n,m} \geq \frac{1}{2}m\} P\{K_{n,m} \geq \frac{1}{2}m\}.$$

By the law of large numbers,

$$P\{K_{n,m} \geq \frac{1}{2}m\} \rightarrow 1, \quad n \rightarrow \infty.$$

Clearly,

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

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

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

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

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

Now, for $u = 0.29$,

$$\log(u^u(1-u)^{1-u}) - \frac{1}{2} \log(u) > 0,$$

and, so, $um \geq \frac{1}{3}\lambda\sqrt{n}$ for $n \geq n_1(\lambda)$. This completes the proof of Lemma 4.

A.2. Proof of Lemma 5

Here, by vertical and horizontal grid lines we mean the lines that traverse the vertical and horizontal grid points, respectively.

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

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

$$|\gamma(t) - \gamma(s) - (t-s)\gamma'(s)| \leq \frac{1}{2}\kappa(t-s)^2.$$

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

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

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 $|\cos(\angle(\gamma'(0), u))|$, where $\gamma'(0)$ is the unit-norm tangent vector at the beginning of the curve γ , is greater than or equal to $1/\sqrt{2}$, or less than $1/\sqrt{2}$, then let a_0 be the first intersection of the ray $\{\gamma(0) - t\gamma'(0) : t \geq 0\}$ with a vertical or horizontal grid line, respectively. We then extend γ by adding the line segment $[a_0, \gamma(0)]$. Hence, we may assume that γ starts on a vertical or horizontal grid line with $|\cos(\angle(\gamma'(0), u))|$ greater than or equal to $1/\sqrt{2}$ or less than $1/\sqrt{2}$, respectively.

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

Lemma 18. *There exists a constant $c_1 > 0$ such that, for a sufficiently small $\varepsilon_1 > 0$, we have*

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

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

Lemma 19. *There exists a constant $c_2 > 0$ such that, for a sufficiently small $\varepsilon_1 > 0$, we have $|a_{i+1} - a'_i| \leq c_2\kappa\varepsilon_1^2$.*

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

Definition 3. (*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 $i = 0, \dots, I + 1$, let b_i be one of the grid points closest to a_i . Lemma 19 implies that, for all $i = 0, \dots, I + 1$, $[b_i, b_{i+1}]$ is a beamlet. 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 that they are both horizontal. Then b_{i+2} is a vertical grid point. Let V be the vertical grid line it belongs to, i.e. such that $b_{i+2} \in V$. Then, if b'_{i+2} is the grid point defined by the intersection of the line segment $[b_i, b_{i+1}]$ with V , we need to show that $|b'_{i+2} - b_{i+2}| < 3\varepsilon_2$. For that, it suffices to show that $|a_{i+2} - b'_{i+2}| < \frac{5}{2}\varepsilon_2$. Now, because $|a_i - b_i| \leq \frac{1}{2}\varepsilon_2$ and $|a_{i+1} - b_{i+1}| \leq \frac{1}{2}\varepsilon_2$, we find that $|a'_{i+2} - b'_{i+2}| \leq \frac{3}{2}\varepsilon_2$. Hence, by Lemma 19,

$$|a_{i+2} - b'_{i+2}| \leq \frac{3}{2}\varepsilon_2 + c\kappa\varepsilon_1^2 \leq 2\varepsilon_2.$$

Moreover, with $\xi_b = \bigcup_i [b_i, b_{i+1}]$, the Hausdorff distance between ξ_b and γ is bounded by ε_2 . This comes from the fact that the Hausdorff distance between ξ_b and ξ_a is bounded by $\frac{1}{2}\varepsilon_2$, since this is the case for $[b_i, b_{i+1}]$ and $[a_i, a_{i+1}]$ for all $i = 0, \dots, I$.

Finally, since $\gamma([s_i, s_{i+2}])$ has a length of at least ε_1 , we have $I \leq 2(1 + \ell/\varepsilon_1) \leq 2\lambda/\varepsilon_1$ (for $\varepsilon_1 < \varepsilon_0(\lambda)$).

A.3. Proof of a 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 $\psi : \mathbb{R} \rightarrow [0, 1]$ that has support in $[-\frac{1}{2}, \frac{1}{2}]$ and satisfies $\psi(0) = 1$. Introduce η_n and v , chosen so that

$$\begin{aligned} v &= \kappa/\|\psi''\|_\infty, \\ v\eta_n^3 &= 1/n. \end{aligned} \tag{6}$$

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

$$J \geq 0.3161m \approx 0.3161\lambda/2\eta > 0.1508\|\psi''\|_\infty^{-1/3}\lambda\kappa^{1/3}n^{1/3}$$

rectangles R_j . We will next show that at least the J points $X_{i_j} = (t_j, z_j) \in R_j$ lie on a curve in $\mathcal{S}_{\lambda, \kappa}$. Hence, we can set $A = 0.1508 \|\psi''\|_\infty^{-1/3}$ in the statement of Theorem 2.

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 regions of 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 a finite sum of smooth functions. Moreover, since, for fixed t , there is at most one nonzero 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 κ . We also need to control the total length of $\text{graph}(f)$. On $[0, \frac{1}{2}\lambda] \setminus \bigcup_j [t_j - \frac{1}{2}\eta, t_j + \frac{1}{2}\eta]$, f takes the value 0 and, thus, its graph is flat in this region. On $[t_j - \frac{1}{2}\eta, t_j + \frac{1}{2}\eta]$, the length of $\text{graph}(f)$ is bounded by $\eta\sqrt{1 + v\eta\|\psi''\|_\infty}$. Therefore, the length of $\text{graph}(f)$ tends to $\frac{1}{2}\lambda$ as $\eta \rightarrow 0$. Hence, for a sufficiently small η , the total length of $\text{graph}(f)$ does not exceed λ . By (6), we have $\eta \rightarrow 0$ as $n \rightarrow \infty$ and $\text{graph}(f) \in \mathcal{C}_\lambda(2, \kappa)$ for sufficiently large n .

A.4. Proof of a lower bound for Hölder objects

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

$$v\|\psi^\alpha\|_\infty \leq (\alpha - 1)! \beta, \tag{7}$$

while if α is not an integer, then

$$v\|\psi^{\lfloor \alpha \rfloor}\|_\infty \leq (\lfloor \alpha \rfloor)! \beta. \tag{8}$$

Also introduce an $\eta \equiv \eta_n$ chosen so that

$$v\eta^{k+\alpha(d-k)} = 1/n. \tag{9}$$

We partition $[0, 1]^k$ into hypercubes with sides of length η and denote these hypercubes by I_i , where $i = (i_1, \dots, i_k)$. To each such hypercube, we associate the hyperrectangle $Q_i = I_i \times [0, v\eta^\alpha]^{d-k}$ in $[0, 1]^d$. According to (9), each Q_i has volume $1/n$ and, so, we expect to find one point of X^n in each Q_i . Now consider only those even-numbered Q_i (i.e. with i_1, \dots, i_k all even) that are not empty (i.e. $X^n(Q_i) > 0$). Rename these rectangles $R_j, j = 1, \dots, J$; again, notice that both J and the $\{R_j\}$ are random. In each R_j , select one point X_{i_j} and group the ‘independent’ and ‘dependent’ components together as $(t_j, z_j) \in [0, 1]^k \times [0, 1]^{d-k}$.

We now define the ‘independent’ and ‘dependent’ components of g :

- for $i = 1, \dots, k$, set $g_i(s) = s_i$;
- for $i = 1, \dots, d - k$, set $g_{k+i}(s) = \sum_j f_{ij}(s)$, where $f_{ij}(s) = z_{ji} \psi((s - t_j)/\eta)$ and z_{ji} is the i th coordinate of the vector z_j .

This construction guarantees that the image $g([0, 1]^k)$ includes $\{(t_j, z_j) : j = 1, \dots, J\}$. Again, J is a random variable and we are able to give a concrete estimate for an $A > 0$ such that $P\{J \geq Ac(\alpha, \beta, k)n^{k/(k+\alpha(d-k))}\} \rightarrow 1, n \rightarrow \infty$, where the constant $c(\alpha, \beta, k)$ only depends on α, β , and k . 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.

Since 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 that $\alpha > 1$, 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 this also holds for the dependent components. We can write

$$\|g_{k+i}^{[\alpha]}\|_\infty = \max_j \|f_{ij}^{[\alpha]}\|_\infty \leq v \|\psi^{[\alpha]}\|_\infty \eta^{[\alpha]}, \tag{10}$$

$$\|g_{k+i}^{[\alpha]}\|_\infty = \max_j \|f_{ij}^{[\alpha]}\|_\infty \leq v \|\psi^{[\alpha]}\|_\infty \eta^{-r(\alpha)}, \tag{11}$$

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

Consider the following two cases.

- α is an integer. Using (11) and a Taylor expansion gives

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

- α is not an integer. Using (10) and the fact that g_{k+i} takes the value 0, we can prove that

$$|g_{k+i}^{[\alpha]}(t) - g_{k+i}^{[\alpha]}(s)| \leq v \|\psi^{[\alpha]}\|_\infty \eta^{[\alpha]}.$$

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

$$|g_{k+i}^{[\alpha]}(t) - g_{k+i}^{[\alpha]}(s)| \leq v \|\psi^{[\alpha]}\|_\infty \|t - s\|^{[\alpha]}.$$

Using (11) and a Taylor expansion gives

$$|g_{k+i}^{[\alpha]}(t) - g_{k+i}^{[\alpha]}(s)| \leq v \|\psi^{[\alpha]}\|_\infty \eta^{-(1+\{\alpha\})} \|t - s\|.$$

Hence, for $\|t - s\| \leq \eta$, we also have

$$|g_{k+i}^{[\alpha]}(t) - g_{k+i}^{[\alpha]}(s)| \leq v \|\psi^{[\alpha]}\|_\infty \|t - s\|^{[\alpha]}.$$

In both cases, (7) and (8) guarantee that $g_{k+i} \in \mathcal{H}^{k,d}(\alpha, \beta)$, as required.

A.5. Besov and Triebel objects

The following results are classical (see [13, p. 105]).

1. We have the inclusions

$$B_{p,p \wedge q}^\alpha \subset \mathcal{F}_{p,q}^\alpha \subset B_{p,p \vee q}^\alpha, \tag{12}$$

where $F \subset B$, say, means that $\|f\|_B \leq C \|f\|_F$ 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

$$B_{p_1,q_1}^{\alpha_1} \subset B_{p_2,q_2}^{\alpha_2}, \tag{13}$$

where ‘ \subset ’ again means continuous inclusion.

- For $\alpha > k/p$, the ε -entropy of the unit ball in $B_{p,q}^\alpha$, for the supremum norm, 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 that (using (12)) $N_n(\mathcal{B}_{p,q}^{k,d}(\alpha, \beta))$ and $N_n(\mathcal{F}_{p,q}^{k,d}(\alpha, \beta))$ are of the same order of magnitude in most cases.

From now on, we assume that $\alpha > k/p$. Using (13), for all $\alpha' > \alpha$ and a certain constant $C(\alpha, \alpha')$, we have $\mathcal{H}^{k,d}(\alpha', C(\alpha, \alpha')\beta) \subset \mathcal{B}_{p,q}^{k,d}(\alpha, \beta)$. Hence, $N_n(\mathcal{B}_{p,q}^{k,d}(\alpha, \beta))$ 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 (1)) of order $n^{1/(1+\alpha(d/k-1))}$. Therefore, the order of $N_n(\mathcal{B}_{p,q}^{k,d}(\alpha, \beta))$ is somewhere between these bounds, and $N_n(\mathcal{F}_{p,q}^{k,d}(\alpha, \beta))$ is of the same order because of (12).

A.6. Proof of Lemma 14

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

$$(S)_\varepsilon \subset \bigcup_{(x,v) \in W} \text{ball}(x, \varepsilon) \times \text{cone}(v, \theta_{\varepsilon,\ell}), \tag{14}$$

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

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

Assume that $\ell \geq 6\varepsilon$ and that we have proved (14). Choose a unit-speed parametrization of γ (also denoted $\gamma(\cdot)$) and notice that $S = \{(\gamma(s), \gamma'(s)) : s \in [0, \ell]\}$. Now consider an ε -covering of $[0, \ell]$, which we denote by $\{s_j : j = 1, \dots, J\}$ 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 γ being Lipschitz with constant 1, while the second inclusion comes from γ' being Lipschitz with constant κ . 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 $\mu(\text{ball}(x, r) \times \text{cone}(v, \theta)) = r^2 \min\{\theta, \pi\}$, we have

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

and

- if $\ell \leq \varepsilon^{1/2}/\kappa^{1/2}$ then $\theta_{\varepsilon,\ell} = 16\varepsilon/\ell$ and, so, $\mu((S)_\varepsilon) \leq 64\varepsilon^2 + 4\pi\varepsilon^2$; while
- if $\ell > \varepsilon^{1/2}/\kappa^{1/2}$ then $\theta_{\varepsilon,\ell} = 16\kappa^{1/2}\varepsilon^{1/2}$ and, so,

$$\mu((S)_\varepsilon) \leq 64\ell\kappa^{1/2}\varepsilon^{3/2} + 4\pi\varepsilon^2 \leq 64\lambda\kappa^{1/2}\varepsilon^{3/2} + 4\pi\varepsilon^2.$$

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

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

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

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

We still must prove (14). To do so, let $S_1 \in (S)_\varepsilon$. By definition, there exist $\gamma, \gamma_1 \in \mathcal{C}_\lambda(2, \kappa)$ such that $S = \{(\gamma(s), \gamma'(s)) : s \geq 0\}$, $S_1 = \{(\gamma_1(s), \gamma'_1(s)) : s \geq 0\}$, and $\|\gamma - \gamma_1\|_\infty \leq \varepsilon$. Let $\ell = \text{length}(\gamma)$ and $\ell_1 = \text{length}(\gamma_1)$. Also, let $\theta_0 = \max_{s \leq 0} |\angle(\gamma'(s), \gamma'_1(s))|$ and let s_0 be a maximizer (we need only consider $s_0 \leq \max\{\ell, \ell_1\}$).

If $s_0 > \min\{\ell, \ell_1\}$ then, because

$$|\gamma'(s_0) - \gamma'(\ell_1)| \leq \kappa|s_0 - \ell_1| \quad \text{and} \quad |\gamma'_1(s_0) - \gamma'_1(\ell)| \leq \kappa|s_0 - \ell|,$$

we have $|\gamma'(s_0) - \gamma'_1(s_0)| \leq \kappa|\ell - \ell_1|$. We now use the following result.

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

$$|\text{length}(\gamma) - \text{length}(\gamma_1)| \leq 3\varepsilon.$$

Proof. Assume that $\ell \leq \ell_1$. For $\ell \leq s \leq \ell_1$, we have $|\gamma_1(s) - \gamma(\ell)| \leq \varepsilon$ and $|\gamma_1(\ell) - \gamma(\ell)| \leq \varepsilon$; hence, $|\gamma_1(s) - \gamma_1(\ell)| \leq 2\varepsilon$. Now, using Lemma 16 with $\ell \leq s$, we obtain $(s - \ell) - \frac{1}{2}\kappa(s - \ell)^2 \leq 2\varepsilon$. From this, we see that $s - \ell$ cannot take the value 3ε when $\varepsilon < 2/9\kappa$. Hence, $\ell_1 < \ell + 3\varepsilon$.

In light of this result, we assume that $s_0 \leq \min\{\ell, \ell_1\}$. Since $\ell \geq 6\varepsilon$, we have $\frac{1}{2}\ell \leq \ell_1 \leq \frac{3}{2}\ell$. This implies that either $s_0 + \frac{1}{4}\ell \leq \min\{\ell, \ell_1\}$ or $s_0 - \frac{1}{4}\ell \geq 0$. Both cases are treated similarly, so we will assume that the former is true. Then, for $s_0 \leq s \leq s_0 + \frac{1}{4}\ell$, we have

$$\begin{aligned} \gamma(s) - \gamma_1(s) &= \gamma(s) - \gamma(s_0) - \gamma'(s_0)(s - s_0) + \gamma(s_0) - \gamma_1(s_0) \\ &\quad + (\gamma'(s_0) - \gamma'_1(s_0))(s - s_0) - (\gamma_1(s) - \gamma_1(s_0) - \gamma'_1(s_0)(s - s_0)). \end{aligned}$$

By Lemma 16,

$$|\gamma(s) - \gamma(s_0) - \gamma'(s_0)(s - s_0)| \leq \frac{1}{2}\kappa(s - s_0)^2$$

and

$$|\gamma_1(s) - \gamma_1(s_0) - \gamma'_1(s_0)(s - s_0)| \leq \frac{1}{2}\kappa(s - s_0)^2.$$

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

$$\|\gamma'(s_0) - \gamma'_1(s_0)\|_\infty(s - s_0) - \kappa(s - s_0)^2 \leq 2\varepsilon.$$

Let $D_0 = \|\gamma'(s_0) - \gamma'_1(s_0)\|_\infty$ and $\eta = s - s_0$. The inequality then becomes

$$p(\eta) = \kappa\eta^2 - D_0\eta + 2\varepsilon \geq 0$$

for all $\eta \in [0, \frac{1}{4}\ell]$. There are two possibilities: either

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

In both cases, $D_0 \leq 16 \max\{\varepsilon/\ell, \kappa^{1/2}\varepsilon^{1/2}\}$ and (14) follows.

A.7. Proof of a lower bound for connect-the-darts

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

- $0 \leq \psi \leq 1$ with $\psi(0) = 1$, and
- $-1 \leq \phi \leq 1$ with $\phi'(0) = 1$.

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

$$v = \kappa/2 \|\psi''\|_\infty, \tag{15}$$

$$w = \kappa/2 \|(\psi\phi)''\|_\infty, \tag{16}$$

$$2v\eta_n^3 \arctan(w\eta_n) = 1/n. \tag{17}$$

Partition $[0, \frac{1}{2}\lambda]$ into intervals of length η , and denote these intervals by I_i , where $i = 1, \dots, \lfloor \lambda/2\eta \rfloor$. To each interval we associate a $Q_i = I_i \times [0, v\eta^2] \times \{e^{i\theta} : |\theta| \leq w\eta\}$. By (17), the volume of each Q_i is $1/n$, and so we expect to find one point in each Q_i . We now consider only those even-numbered Q_i that are not empty (i.e. $X''(Q_i) > 0$). Rename these sets $R_j, j = 1, \dots, J$; once again, both J and the R_j are random. In each R_j , select one point $X_{i_j} = (t_j, z_j, e^{i\theta_j})$.

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

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

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 sufficiently small A_0 , and with overwhelming probability as n increases, there are at least $J \geq A_0\lambda/\eta = A_0\kappa^{-1}\lambda\kappa^{1/2}n^{1/4}$ points X_{i_j} lying on f and with directions tangent to f ; for the purposes of Theorem 9, we may take $A = A_0\kappa^{-1}$.

We now verify that f has sufficient regularity. By construction, f is nonnegative and twice differentiable. Moreover, since for fixed t there is at most one nonzero term in the sum, we have

$$\|f\|_\infty \leq \max_j \|f_j\|_\infty \quad \text{and} \quad \|f''\|_\infty \leq \max_j \|f_j''\|_\infty.$$

First, $\|f_j\|_\infty \leq z_j(1 + \eta|\tan \theta_j|/z_j)$, meaning that $\|f\|_\infty \leq v\eta^2 + \eta \tan(w\eta) \leq 1$ for sufficiently small η , i.e. sufficiently large n . Next, since

$$f_j''(t) = \frac{z_j}{\eta^2} \psi''\left(\frac{t-t_j}{\eta}\right) + \frac{\tan \theta_j}{\eta} (\psi\phi)''\left(\frac{t-t_j}{\eta}\right),$$

we have $\|f''\|_\infty \leq v\|\psi''\|_\infty + w\|(\psi\phi)''\|_\infty$. By (15) and (16), the graph of f has maximum curvature bounded by κ . The length is controlled exactly as in Section A.3, ensuring that, for sufficiently small η , the total length of $\text{graph}(f)$ does not exceed λ . However, since $\eta_n \rightarrow 0$ as $n \rightarrow \infty$, we conclude that $\text{graph}(f) \in \mathcal{C}_\lambda(2, \kappa)$ for sufficiently large n .

A.8. CTD and geometric discrepancy theory

We mentioned in the introduction that there is a quantitative connection between our CTD problem and geometric discrepancy theory. Here 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, from work of Schmidt [29] and [6, 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 and random. Consider the class 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 γ that 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 γ . For $\varepsilon > 0$, consider also the slightly shifted curve γ_ε based on the convex function $f_\varepsilon = f + \varepsilon$, and define the set C_ε to be the convex hull of the range of γ_ε . Now set $\varepsilon = 1/n^2$; we can be practically certain that C_ε contains exactly the same points as C , and that their areas differ by at most $c\varepsilon$. Define the discrepancy for a specific set by

$$D_n(C) = |N_n(C) - n \text{area}(C)|;$$

clearly

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

We conclude that

$$\max\{D_n(C), D_n(C_\varepsilon)\} \geq \frac{1}{2}N_n(\gamma) - c/2n.$$

Hence, with overwhelming probability for large n ,

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

Thus, CTD theory gives lower bounds on the discrepancy over the class of convex sets, among random point-sets.

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 that can lie on the boundary of a convex set, and that uniform random point-sets 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] 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.
- [2] ARIAS-CASTRO, E., DONOHO, D. L. AND HUO, X. (2005). Adaptive multiscale detection of filamentary structures embedded in a background of uniform random points. To appear in *Ann. Statist.*

- [3] ARIAS-CASTRO, E., DONOHO, D. L. AND HUO, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans. Inf. Theory* **51**, 2402–2425.
- [4] BAIK, J. (2003). Limiting distribution of last passage percolation models. In *Proc. 14th Internat. Congr. Math. Phys.* Available at <http://www.math.lsa.umich.edu/~baik/>.
- [5] 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.
- [6] BECK, J. AND CHEN, W. W. L. (1987). *Irregularities of Distribution* (Cambridge Tracts Math. **89**). Cambridge University Press.
- [7] BRONŠTEĪN, E. M. (1976). ε -entropy of convex sets and functions. *Sibirsk. Mat. Ž.* **17**, 508–514, 715.
- [8] CLEMENTS, G. F. (1963). Entropies of sets of functions of bounded variation. *Canad. J. Math.* **15**, 422–432.
- [9] DEVORE, R. A. AND LORENTZ, G. G. (1993). *Constructive Approximation*. Springer, New York.
- [10] DIAMOND, J. AND KIELY, K. (2002). Administration, agencies failed to connect the dots. *USA Today*, 17 May 2002. Available at <http://www.usatoday.com/news/washington/2002/05/17/failure-usatcov.htm>.
- [11] DONOHO, D. L. (1999). Wedgelets: nearly minimax estimation of edges. *Ann. Statist.* **27**, 859–897.
- [12] DONOHO, D. L. AND HUO, X. (2002). Beamlets and multiscale image analysis. In *Multiscale and Multiresolution Methods* (Lecture Notes Comput. Sci. Eng. **20**), eds T. Barth, T. Chan and R. Haimes, Springer, Berlin, pp. 149–196.
- [13] EDMUNDS, D. E. AND TRIEBEL, H. (1996). *Function Spaces, Entropy Numbers, and Differential Operators*. Cambridge University Press.
- [14] FEDERER, H. (1969). *Geometric Measure Theory*. Springer, New York.
- [15] FEW, L. (1955). The shortest path and the shortest road through n points. *Mathematika* **2**, 141–144.
- [16] FIELD, D. J., HAYES, A. AND HESS, R. F. (1993). Contour integration by the human visual system: evidence for a local ‘association field’. *Vision Res.* **33**, 173–193.
- [17] FRIEZE, A. M. (1991). On the length of the longest monotone subsequence in a random permutation. *Ann. Appl. Prob.* **1**, 301–305.
- [18] GROENEBOOM, P. (2001). Ulam’s problem and Hammersley’s process. *Ann. Prob.* **29**, 683–690.
- [19] HUO, X., DONOHO, D. L., TOVEY, C. A. AND ARIAS-CASTRO, E. (????). Dynamic programming methods for ‘connecting the dots’. Submitted.
- [20] 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 ACM/SIAM Symp. Discrete Algorithms* (Atlanta, GA), ACM, New York, pp. 341–350.
- [21] KAHN, J. M., KATZ, R. H. AND PISTER, K. S. J. (2000). Emerging challenges: mobile networking for smart dust. *J. Commun. Networks* **2**, 188–196.
- [22] KOLMOGOROV, A. N. AND TIKHOMIROV, V. M. (1959). ε -entropy and ε -capacity of sets in function spaces. *Uspekhi Mat. Nauk* **14**, 3–86 (in Russian). English translation: *Trans. Amer. Math. Soc.* **17**, 277–364.
- [23] KOVACS, I. AND JULESZ, B. (1993). A closed curve is much more than an incomplete one – effect of closure in figure–ground segmentation. *Proc. Nat. Acad. Sci. USA* **90**, 7495–7497.
- [24] LANE, E. (2003). Board not ready to connect the dots. Probe into shuttle tragedy still working out the details. *NewYork Newsday*.
- [25] LOGAN, B. F. AND SHEPP, L. A. (1977). A variational problem for random Young tableaux. *Adv. Math.* **26**, 206–222.
- [26] MATOUŠEK, J. (1999). *Geometric Discrepancy. An Illustrated Guide* (Algorithms and Combinatorics **18**). Springer, Berlin.
- [27] MEYER, Y. (1992). *Wavelets and Operators*. Cambridge University Press.
- [28] ODLYZKO, A. M. AND RAINS, E. M. (2000). On longest increasing subsequences in random permutations. In *Analysis, Geometry, Number Theory: The Mathematics of Leon Ehrenpreis* (Philadelphia, PA, 1998; Contemp. Math. **251**), American Mathematical Society, Providence, RI, pp. 439–451.
- [29] SCHMIDT, W. M. (1969). Irregularities of distribution. III. *Pacific J. Math.* **29**, 225–234.
- [30] SCHNEIDER, B. (2002). Connecting the dots. *CNN.com, Inside Politics*, 17 May 2002. Available at <http://www.cnn.com/2002/ALLPOLITICS/05/17/pol.play.dots/>.
- [31] SHORACK, G. R. AND WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley, New York.
- [32] SOSHNIKOV, A. (1999). Universality at the edge of the spectrum in Wigner random matrices. *Commun. Math. Phys.* **207**, 697–733.
- [33] TALAGRAND, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Études Sci. Pub. Math.* **81**, 73–205.
- [34] TRACY, C. A. AND WIDOM, H. (2001). On the distributions of the lengths of the longest monotone subsequences in random words. *Prob. Theory Relat. Fields* **119**, 350–380.
- [35] 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 (in Russian).