On Order Types of Random Point Sets

Olivier Devillers, Philippe Duchon, Marc Glisse, Xavier Goaoc

On Order Types of Random Point Sets*

Olivier Devillers† Philippe Duchon‡ Marc Glisse§ Xavier Goaoc¶

December 20, 2018

Abstract

Let P be a set of n random points chosen uniformly in the unit square. In this paper, we examine the typical resolution of the order type of P. First, we show that with high probability, P can be rounded to the grid of step $\frac{1}{n^{3+\epsilon}}$ without changing its order type. Second, we study algorithms for determining the order type of a point set in terms of the the number of coordinate bits they require to know. We give an algorithm that requires on average $4n \log_2 n + O(n)$ bits to determine the order type of P, and show that any algorithm requires at least $4n \log_2 n - O(n \log \log n)$ bits. Both results extend to more general models of random point sets.

1 Introduction

An order type is a combinatorial abstraction of a finite point configuration. Informally, the order type of a planar point set P records for every triple in P the orientation (clockwise or counterclockwise) of the triangle that they form. This information already determines which subsets of P are in convex position and which pairs of elements of P form intersecting segments; hence, the order type encodes the convex hull, the convex pealing structure, the triangulations of P, or, for instance, which graphs admit straight-line embeddings with vertices mapped to P.

In this paper, we study "how much randomness" is contained in the order type of random point sets, for instance uniform samples of the unit square. We make the meaning of this question clear right after recalling some background.

1.1 Context

Let us briefly recall the notions of order type and the topic of their random generation.

Definitions. The *orientation* of a triple $(a, b, c) \in (\mathbb{R}^2)^3$ is the sign of the determinant

$$\begin{vmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ 1 & 1 & 1 \end{vmatrix},$$

where a_x is the x-coordinate of a, etc. This sign is -1 if the triangle abc is oriented clockwise, 0 if it is flat, and 1 if it is oriented counterclockwise. Two sequence $P = (p_1, p_2, \ldots, p_n) \in (\mathbb{R}^2)^n$ and $Q = (q_1, q_2, \ldots, q_n) \in$

*Funded by grant ANR-17-CE40-0017 of the French National Research Agency (ANR project ASPAG). This work was initiated during the ALEA 2013 conference and the 15th INRIA–McGill–Victoria Workshop on Computational Geometry at the Bellairs Research Institute.

†Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France. Olivier.Devillers@inria.fr

‡LaBRI, Université de Bordeaux, CNRS, Bordeaux INP, F-33504 Talence, France. philippe.duchon@u-bordeaux.fr

§Inria, Centre de recherche Saclay-Île-de-France, France. Marc.Glisse@inria.fr

¶Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France. Partially funded by Institut Universitaire de France. xavier.goaoc@loria.fr

$(\mathbb{R}^2)^n$ *have the same chirotope* if for every indices i, j, k the triples (p_i, p_j, p_k) and (q_i, q_j, q_k) have the same orientation. A related notion is for two finite subsets P and Q of $\mathbb{R}^2$ to *have the same order type*, meaning that there exists a bijection $f : P \to Q$ that preserves orientations. Having the same order type (resp. chirotope) is an equivalence relation, and an *order type* (resp. *chirotope*) is an equivalence class for that relation. An order type or chirotope is *simple* if it can be realized without three collinear points. These definitions extend readily to $\mathbb{R}^d$, but in this paper, we are only interested in planar, simple point sets.

Depending on the context, we will work with chirotopes or with order types. The questions we are interested in are usually oblivious to the labeling of the points, but our methods for addressing them do make explicit use of that labeling. The two notions are related, since an order type of size n corresponds to at most $n!$ chirotopes, possibly fewer if some bijections of the point set into itself preserve orientations. We nevertheless phrase our questions in terms of order types, but state our results in terms of chirotopes for the sake of precision.

Enumerating order types. There are finitely many order types of size n, so, in principle, some properties of planar point sets of small size can be studied by sheer enumeration of order types. Here is an example, coming from geometric Ramsey theory, of such a "constant size" question. Gerken [6] proved that any set of at least 1717 points in the plane without aligned triple contains an *empty hexagon*: six points in convex position with no other point of the set in their convex hull. The largest known point set with no empty hexagon, of size 29, has not been improved for decades [10].

In practice, order types were enumerated (up to possible reflexive symmetry) up to size 11 by Aischolzer *et al.* [1]. They used their database for instance to establish sharp bounds on the minimum and maximum numbers of triangulations on 10 points, a very finite result that they could bootstrap into an asymptotic bound. The number of order types of size n does, however, quickly become overwhelming as n increases: it reaches thousands of billions already for $n = 11$, and grows at least as $n^{3n+o(n)}$ since the number of chirotopes grows as $n^{4n+\Theta\left(\frac{n}{\log n}\right)}$ [2, Theorem 4.1]. It is thus unlikely that the order type database will be extended much beyond size 11, and the geometric Ramsey theory problem above seems out of reach of enumerative methods.

Sampling order types. When a configuration space is too large to be enumerated, it is natural to try and explore it by random sampling. Two desirable properties of a random generator of order types are that it be both efficient (a random order type can be produced quickly, say in time polynomial in n) and reasonably unbiased (it will explore a reasonably large fraction of the space of order types). Satisfying both requirements may be challenging because of two properties of order types. On the one hand, order types enjoy small combinatorial encodings, even of subquadratic size [4], but the set of order types is difficult to describe: already deciding membership is NP-hard [11]. On the other hand, order types can be manipulated through point sets realizing them, so that one needs not worry about remaining in the space of order types, but there are order types of size n for which any realization requires $2^{\Omega(n)}$ bits per coordinate [8].

In fact, little seems known already on the following question. Let m_n be a sequence of positive integers with $m_n \to \infty$, and let μ_n be a probability measure on the set of order types of size m_n. Say that $\{\mu_n\}_{n\in\mathbb{N}}$ exhibits *concentration* if there exists for each n a set S_n of order types of size m_n such that S_n contains a proportion $\epsilon_n \to 0$ of all order types, while $\mu_n(S_n) \to 1$. In other words, μ_n and the uniform measure on order types of size m_n are "asymptotically singular".

Open problem 1. *Does there exist a sequence of measures μ_n on order types of size n such that (i) no subsequence exhibits concentration, and (ii) a random order type of size n according to measure μ_n can be produced in time polynomial in n?*

It is easy to produce a random order type by first generating a random point set, then reading off its order type. In this paper, we study some of the properties of these random generation methods. Let us stress that it is not clear how the probability distribution on point sets translates into a probability distribution

on order types. Naturally, when sampling points independently and from a probability distribution whose support has non-empty interior, every order type appears with positive probability. Indeed, every order type can be realized on an integer grid, and order types are unchanged under rescaling and sufficiently small perturbation. One may nevertheless expect some bias, if only because some order types require exponential precision for their realization [8] and are therefore much more brittle than others. In fact, for order types of small size, bias is unavoidable [7, Prop. 2].

1.2 Questions

We are interested in "how much randomness" there is in the order type of n random points. Formally, we consider two questions:

1. Can we simulate this distribution of order types efficiently, given access to a source of unbiased random bits? (Here we assume a discrete model of computation (e.g., a Turing machine), *not* the real-RAM machine customary in computational geometry.)

2. How biased is the resulting (induced) distribution on order types?

The answers may of course depend on the distribution chosen for the point sets, and we are interested in this dependency. We elaborate on these questions before stating our results.

Algorithmic reformulation. Recall that any real $r \in [0,1]$ has a binary development of the form $0.r_1r_2\ldots$ with $r_i \in \{0,1\}$, so we can identify r with the sequence $r_1r_2\ldots \in \{0,1\}^{\mathbb{N}}$. (In particular, the real 1 is identified with the sequence $1^{\mathbb{N}}$; for dyadic reals, which have two representations, we can choose any.) Our first question has the following algorithmic counterpart. We are given a random point set contained in the unit square (this is not a restriction since order types are invariant under rescaling), in the form of $2n$ infinite binary strings, one per point coordinate. We can read the coordinates, but it has a cost; specifically, accessing the next bit in one of these strings has unit cost, and any other computation is considered free. How efficiently can we (on average) determine the order type of these n points?

Bias. We built some intuition on the second question by running a basic experiment (see Appendix A for details): we picked sets of 10 random points in the square $[1,2]^2$ and read off their order types. A billion repetitions produced only about 10 million different order types (out of the 28.6 million[1] order types of size 10), and a very small fraction of these order types (namely, about 63 thousands of them) represent 99% of the samples. This experiment is very rudimentary (*e.g.* we used floating point coordinates, a standard pseudo-random generator, etc.) so its results should be considered with care; it nevertheless reveals that in practice also, the random generation of diverse order types does require some care.

1.3 Results

We present here some results on the chirotope of certain random point sets. For the sake of clarity, we state and prove our results for a *uniform sample of the unit square*, understood as a *sequence* of random points chosen independently and uniformly in $[0,1]^2$. We comment in Section 5 to what extent our methods generalize. We write log to mean the logarithm of base 2.

First, we give near-tight lower and upper bounds on the expected number of bits needed to determine the chirotope.

Theorem 2. *Let P be a uniform sample of the unit square of size n.*

[1] Aischolzer et al. [1] counted 14 309 547 order types up to reflection; examining all their realizations and their mirror images, we get that the total number of order types is 28 606 030; in other words, 13 064 order types on 10 points are their own mirror image.

(i) Any algorithm that determines the chirotope of P reads on average at least $4n \log n - O(n \log \log n)$ coordinate bits.

(ii) There exists an algorithm that determines the chirotope of P by reading on average $4n \log n + O(n)$ coordinate bits.

As a byproduct of the proof of Theorem 2, we obtain that the typical "resolution" of the chirotope of uniform samples of $[0,1]^2$ is polynomial. We define the *resolution* of chirotope ω to be the smallest integer m such that ω can be realized on a $m \times m$ regular grid. This parameter is related to the *intrinsic spread* studied by Goodman et al. [8] as noted by them. We prove that with high probability, a uniform n-sample of the unit square has resolution no more than $n^{3+\epsilon}$. More precisely:

Theorem 3. *Let $0 < \epsilon < 1$. Let P be a uniform sample of size n of the unit square. With probability at least $1 - O\left(n^{-\epsilon}\right)$, P can be rounded to the regular grid of step $n^{-3-\epsilon}$ without changing its chirotope.*

Do most chirotopes have resolution $O\left(n^{3+\epsilon}\right)$? If not, then Theorem 3 reveals some bias in the probability distribution of the order type of a uniform sample of the unit square. The best bounds that we are aware of, due to Caraballo et al. [3], do not settle this question: they only assert that the number of chirotopes of resolution $n^{-3-\epsilon}$ is at least $n^{3n-O(n \log \log n/ \log n)}$, whereas the number of chirotopes is $n^{4n+\Theta\left(\frac{n}{\log n}\right)}$.

We prove our two theorems in two steps. First, Section 2 answers the questions listed above for an *arbitrary* point set P in terms of two statistics (L and U) of that point set. Sections 3 and 4 then make a probabilistic analysis of these statistics for our random point sets.

2 Setup

In this section, we do not make any probabilistic assumption, and let P be an arbitrary set of n points in the unit square, no three aligned.

Notations. We let G_m denote the partition of $[0,1]^2$ into $m \times m$ square cells of side length $\frac{1}{m}$ where the interior of each cell is of the form $\left(\frac{i}{m}, \frac{i+1}{m}\right) \times \left(\frac{j}{m}, \frac{j+1}{m}\right)$ with $0 \leq i, j < m$. We often set $m = 2^k$, so that two points are in the same cell of G_m if and only if the first k bits of both their $x-$ and $y-$coordinates are equal.

Grids and orientations. We use the following relation between grids and orientations. Consider three points in $[0,1]^2$, and assume that we know only which cells of G_m they lie in (that is, we know the first k bits of each coordinate). Refer to Figure 1. If these three cells cannot be intersected by a line, then the orientation of the three points can be determined solely from these $6k$ bits.

Upper bounds. The algorithm that we propose for Theorem 2 (ii) refines greedily the coordinates of a point involved in a triangle with undetermined orientation, until the chirotope can be determined. We start with no bit read, so we only know that all points are in the unit square. At every step, we select one point and read one more bit for both of its coordinates. So, at every step of the algorithm, we know for each point some grid cell that contains it; the resolution of the grid may of course be different for every point. The selection is done greedily as follows:

> *Find three pairwise distinct indices a, b, c such that the cells known to contain p_a, p_b, p_c can be intersected by a line, and select one among these points known to the coarsest resolution.*

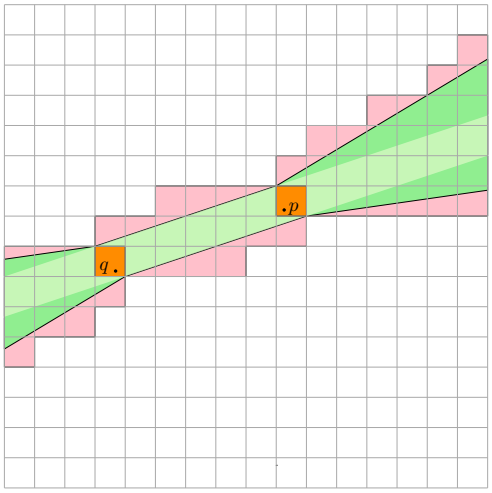

Figure 1: The grid cells containing two of the points (in orange), the union of lines through them (in green), and the cells intersecting such a line (in red).

We break ties arbitrarily, so this is perhaps a method rather than an algorithm. By definition, when the algorithm stops, the chirotope of P can be determined from the precision at which every point is known. The algorithm does *not* stop if P contains three aligned points. The fact that it stops if no three points in P are aligned may be clear; formally, it can be seen via the following statistic.

Definition 4. *Let $U(i)$ be the smallest k such that for any $a, b \in [n] \setminus \{i\}$, there does not exist a line that intersects the cells in G_{2^k} that contain p_a, p_b, and p_i.*

The number $U(i)$ is well-defined if p_i is not aligned with any two other points of P, otherwise we let $U(i) = \infty$. Statistic $U(\cdot)$ bounds from above the complexity of our algorithm on an input P, as measured by the number of bits read. In particular, this implies that our algorithm terminates if P has no aligned triple.

Lemma 5. *In the greedy algorithm above, independently of how ties are resolved, for every $i \in [n]$, at most $U(i)$ bits are read from each coordinate of p_i.*

Proof. Assume that at some point in the algorithm, we read the kth bit of both coordinates of point p_i. To read these bits, our selection method requires that there exist $a, b \in [n] \setminus \{i\}$ such that:

(1) in $\{p_a, p_b, p_i\}$, p_i is one of the points known at coarsest resolution,

(2) there exists a line intersecting the cells known to contain p_a, p_b, and p_i.

Condition (1) ensures that for each of $\{p_a, p_b, p_i\}$, the cell known to contain the point is contained in a cell of $G_{2^{k-1}}$. Condition (2) ensures that these cells in $G_{2^{k-1}}$ can be intersected by a line. Thus, $k - 1 < U(i)$. □

The statistic $U(\cdot)$ also controls the resolution of P.

Corollary 6. *The resolution of P is at most $2^{\max_i U(i)}$.*

Proof. Let $m = 2^{\max_i U(i)}$. When the algorithm terminates, every orientation is determined. Thus, by Lemma 5, knowing, for every $1 \leq i \leq n$, the first $U(i)$ bits of each coordinate of p_i, determines the chirotope of P. Hence, the chirotope of P remains unchanged if we move every point of P to the center of the cell in G_m that contains it. This provides a realization of the order type of P on a grid of size $m \times m$. □

Lower bounds. We also introduce the following statistic.

Definition 7. *Let $L(i)$ denote the smallest k such that* at least one *horizontal or vertical segments of length 2^{-k} starting in p_i is* disjoint *from all lines p_ap_b with $a, b \in [n] \setminus \{i\}$.*

The statistic $L(\cdot)$ bounds from below the complexity of any algorithm determining the chirotope of P.

Lemma 8. *Any algorithm that determines the chirotope of P must read, for every i, at least $L(i) - 1$ bits of each coordinate of p_i.*

Proof. Assume that we know k bits of the x-coordinate of the point p_i. The set of possible positions for p_i then contains a horizontal segment S of length 2^{-k} containing p_i; in fact, it would be exactly such a segment if we knew the y-coordinate of p_i to infinite precision.

By definition of $L(i)$, the two horizontal segments of length $2^{-(L(i)-1)}$ starting in p_i both intersect some line p_ap_b with $a, b \in [n] \setminus \{i\}$ (the lines are different for the two segments). If $2^{-k} \geq 2 \cdot 2^{-(L(i)-1)}$, then the segment S contains at least one of these horizontal segments, and is also intersected by some line p_ap_b with $a, b \in [n] \setminus \{i\}$. Since the possible positions of p_i contain S, this means that the bits read so far from p_i do not suffice to determine the orientation of the triple (p_i, p_a, p_b), even if p_a and p_b were known to infinite precision.

Conversely, if an algorithm that determines the chirotope of P reads k bits from the x-coordinate of p_i, then we must have $2^{-k} < 2 \cdot 2^{-(L(i)-1)}$, that is $k > L(i) - 2$. The same argument applies to the y-coordinate of p_i. □

From here... So the minimal number of bits required to determine the chirotope of P is at least $2\sum_{i=1}^{n}(L(i)-1)$ and at most $2\sum_{i=1}^{n} U(i)$. Note that our lower bound holds for *any* algorithm that determines the chirotope, provided it reads the bits of each coordinate in order, starting from the most significant. It is in particular not assumed that the algorithm always reads as many bits of the two coordinates for a given point, although our proposed algorithm does respect this condition.

We do not know how far apart $2\sum_{i=1}^{n}(L(i) - 1)$ and $2\sum_{i=1}^{n} U(i)$ can be in the worst case but we show, in the next sections, that when P is a uniform sample of the unit square, the expectations of $L(i)$ and of $U(i)$ are equal up to the first order.

3 Analysis of $U(1)$ for a uniform sample of the unit square

Recall that P is a uniform sample of the unit square of size n, and G_m is the partition of $[0, 1]^2$ into $m \times m$ square cells of side length $\frac{1}{m}$.

3.1 Butterflies

Given two points $p, q \in [0, 1]^2$, we first let B be the union of all lines intersecting the cells of G_m containing p and q; we then define $B_m(p, q)$ as the intersection of $[0, 1]^2$ with the Minkowski sum of B with a disk of radius $\frac{\sqrt{2}}{m}$. We call $B_m(p, q)$ the *butterfly* of p and q (at resolution m). Note that the butterfly $B_m(p, q)$ contains all the cells intersecting B. Hence, if there exists a line intersecting the cells of p, q and r, then $r \in B_m(p, q)$. The following lemma bounds the area of $B_m(p, q)$ by $O(\frac{1}{m\delta(p,q)})$.

Lemma 9. *The area of $B_m(p, q)$ is at most $\frac{6}{m} + \frac{4}{m\delta(p,q)}$ where $\delta(p, q)$ is the distance between the centers of the cells of p and q.*

Proof. Note that the bound holds trivially if p and q are in the same cell ($\delta(p, q) = 0$) or in adjacent cells ($\delta(p, q) = 1/m$). Otherwise, the butterfly $B_m(p, q)$ consists of two parts: a strip $S_m(p, q)$ and the union $T_m(p, q)$ of four triangles (shaded in, respectively, green and blue in Figure 2). We have

$$\text{area}\,(S_m(p, q)) \leq \left(3\frac{\sqrt{2}}{m}\right) \cdot \sqrt{2} = \frac{6}{m}.$$

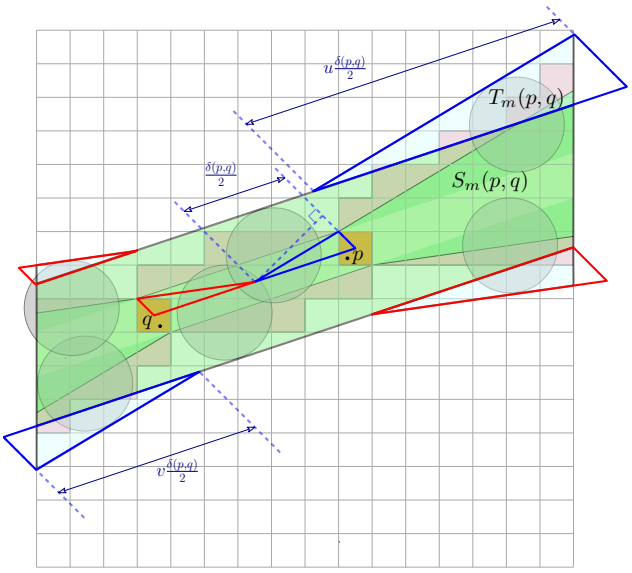

Figure 2: Butterfly of two points.

The four triangles come in two pairs of homothetic triangles, intersected with $[0,1]^2$. Each homothetic pair consists of images under scaling of a triangle whose basis is the half diagonal of a cell of length $\frac{\sqrt{2}}{2m}$ and whose height h is at least $\frac{\delta(p,q)}{2\sqrt{2}}$ (the two kinds of triangles have blue and red boundaries in Figure 2). Letting u and v denote the scaling factors, the areas of the two homothetic triangles sum to $\frac{1}{2}(u^2+v^2)h\frac{\sqrt{2}}{2m}$. Since the scalings turn the height of the reference triangles to two lengths that sum to[2] at most $\sqrt{2}$, we have $(u+v)h \leq \sqrt{2}$. This implies that $u^2+v^2 \leq \left(\frac{\sqrt{2}}{h}\right)^2$ and one pair of homothetic triangles contributes at most $\frac{1}{2}\frac{2}{h^2}h\frac{\sqrt{2}}{2m} = \frac{\sqrt{2}}{2hm} \leq \frac{2}{m\delta(p,q)}$. Altogether, area $(T_m(p,q)) \leq \frac{4}{m\delta(p,q)}$. Finally area $(B_m(p,q)) \leq \frac{6}{m} + \frac{4}{m\delta(p,q)}$. □

3.2 Distribution of $U(1)$

We now analyze the distribution function of the random variable $U(1)$. Recall that the randomness here refers to the choice of the random points $p_1, p_2, \ldots p_n$, which are taken independently and uniformly in $[0,1]^2$.

Lemma 10. $\mathbb{P}[U(1) > k] \leq 57n^2 2^{-k}$.

Proof. We have:

$$\begin{aligned}
\mathbb{P}[U(1) > k] &\leq \mathbb{P}\left[\exists i,j \in \binom{[n]\setminus\{1\}}{2} : p_j \in B_{2^k}(p_1,p_i)\right] \\
&\leq (n-1)\mathbb{P}[\exists j \in [n]\setminus\{1,2\} : p_j \in B_{2^k}(p_1,p_2)] \\
&\leq (n-1)\mathbb{E}\left[1-(1-\text{area}\,(B_{2^k}(p_1,p_2)))^{n-2}\right].
\end{aligned}$$

The geometry of $B_{2^k}(p_1,p_2)$ depends on the distance between the centers of the cells that contain p_1 and p_2. We therefore condition on the cell containing p_1, then sum the contributions of the cell containing p_2 by distance to the cell containing p_1. Accounting for boundary effects, for any $1 \leq t \leq 2^k$ there are at most $8t$

[2]The heights are smaller than the sides and the sides are inside the square $[0,1]^2$ and have disjoint projection on the line (pq).

cells whose center lies at a distance between $t2^{-k}$ and $(t+1)2^{-k}$ from a given cell. We thus have

$$\begin{aligned}
&\mathbb{E}\left[1-(1-\operatorname{area} B_{2^k}(p_1,p_2))^{n-2}\right] \\
&= \sum_{c\in \text{ cells of } G_{2^k}} \mathbb{P}\left[p_2\in c\right]\cdot \mathbb{E}\left[1-(1-\operatorname{area} B_{2^k}(p_1,p_2))^{n-2} \mid p_2\in c\right] \\
&\leq \sum_{t=1}^{2^k} \frac{8t}{(2^k)^2}\left(1-\left(1-\left(\frac{6}{2^k}+\frac{4}{2^k\cdot(t2^{-k})}\right)\right)^{n-2}\right).
\end{aligned}$$

Using $(1-x)^{n-2}\geq 1-(n-2)x$ we get

$$\begin{aligned}
\mathbb{E}\left[1-(1-\operatorname{area} B_{2^k}(p_1,p_2))^{n-2}\right] &\leq (n-2)2^{-2k}\sum_{t=1}^{2^k} 8t\left(6\cdot 2^{-k}+\frac{4}{t}\right) \\
&\leq n2^{-2k}\left(\sum_{t=1}^{2^k} 48t2^{-k}\right)+n2^{-2k}2^k 32 \\
&\leq 24n2^{-k}(1+2^{-k})+32n2^{-k}.
\end{aligned}$$

The statement trivialy bounds a probability by something greater than 1 for $k\leq 5$. For $k\geq 6$, the final term is at most $57n2^{-k}$. □

We can extract an upper bound on the expectation of $U(1)$:

Lemma 11. $\mathbb{E}\left[U(1)\right]\leq 2\log n+8$

Proof. By definition we have

$$\mathbb{E}\left[U(1)\right]=\sum_{k=1}^{\infty} k\mathbb{P}\left[U(1)=k\right]=\sum_{k=0}^{\infty}\mathbb{P}\left[U(1)>k\right].$$

For the first $2\log n+6$ terms, we use the trivial upper bound of 1 and for the remaining terms we use the upper bound of Lemma 10:

$$\mathbb{E}\left[U(1)\right]\leq(2\log n+6)+57n^2\sum_{k\geq 6+2\log n}^{\infty} 2^{-k}=(2\log n+6)+57n^2\cdot 2^{-5-2\log n}.$$

Altogether it comes that $\mathbb{E}\left[U(1)\right]\leq 2\log n+8$. □

3.3 Proofs of Theorems 3 and 2 (ii)

The above analysis of $U(1)$ suffices to prove Theorems 3 and 2 (ii). Again, let P be a uniform sample of size n of the unit square. We first prove that with probability at least $1-O\left(n^{-\epsilon}\right)$, the points of P can be rounded to the regular grid of step $n^{-3-\epsilon}$ without changing the chirotope.

Proof of Theorem 3. By Corollary 6, the resolution of P is at most $2^{\max_i U(i)}$. By union bound, we have

$$\mathbb{P}\left[\max_i U(i)>k\right]\leq n\mathbb{P}\left[U(1)>k\right]\leq 57n^3 2^{-k}$$

so for $k=(3+\epsilon)\log n$ we have

$$\mathbb{P}\left[\max_i U(i)>(3+\epsilon)\log n\right]\leq 57n^{-\epsilon}$$

and the statement follows with Corollary 6. □

We next prove that our greedy algorithm for deciding the chirotope of P reads on average at most $4n \log n + O(n)$ coordinate bits.

Proof of Theorem 2 (ii). By Lemma 5, our greedy algorithm reads at most $U(a)$ bits from each coordinate of point p_a. Thus, using Lemma 11, the average number of bits used by our algorithm is at most:

$$2\mathbb{E}\left[\sum_{i=1}^{n} U(i)\right] = 2\sum_{i=1}^{n} \mathbb{E}\left[U(i)\right] = 2n\mathbb{E}\left[U(1)\right] \leq 4n \log n + 16n.$$

This proves the statement. □

4 Analysis of $L(1)$ for uniform samples of the unit square

Our approach is to look for lines passing close to p_1, as such lines are likely to force $L(1)$ to be large. To do so, we divide the plane into some number of angular sectors around p_1 (Figure 3) and define a blue disk of center p_1 and radius 0.2 and a red annulus with center p_1 and radii 0.3 and 0.4. This discretizes the problem, as if we find two points of P in the blue and red parts of the same or nearby sectors, then they must span a line passing close to p_1. To turn this idea into a lower bound on $L(1)$ we must take care of a few issues; we do this in sections 4.1 to 4.3 and establish:

Lemma 12. *For every $x > 1$, there exists $c > 0$ such that $\mathbb{P}\left[L(1) \geq 2\log n - x \log\log n\right]$ is at least $1 - 2^{-cn}$.*

The proof of Lemma 12 is somewhat technical due to some dependency between the random variables involved. Before we get to that, let us see how it can be used.

Proof of Theorem 2(i). From Lemma 12, we get the following lower bound on the expectation of $L(1)$.

Corollary 13. *For n large enough, $\mathbb{E}\left[L(1)\right]$ is at least $2\log n - 2\log\log n$.*

Proof. As spelled out in the proof of Lemma 11, $\mathbb{E}\left[L(1)\right] = \sum_{k\geq 0} \mathbb{P}\left[L(1) > k\right]$. Note that $\mathbb{P}\left[L(1) > k\right]$ decreases with k. Lemma 12 for $x = \frac{3}{2}$ implies that the first $2\log n - \frac{3}{2}\log\log n$ terms are at least $1 - 2^{-cn}$ for some constant $c > 0$. Keeping only these terms, we get

$$\mathbb{E}\left[L(1)\right] \geq (1 - 2^{-cn})\left(2\log n - \frac{3}{2}\log\log n\right) \geq (1 - 2^{-cn})2\log n - \frac{3}{2}\log\log n.$$

For n large enough, $2^{-cn+1}\log n < \frac{1}{2}\log\log n$ and the statement follows. □

Theorem 2(i) now follows from Lemma 8 and Corollary 13: all n variables $L(i)$ have the same expectation, and any algorithm that determines the chirotope of P must read at least a total of $2(\sum_i L(i) - 1) = 4n\log n - O(n\log\log n)$ coordinate bits.

4.1 Discretization

Let us come back to the proof of Lemma 12. First, if p_1 is close enough to the boundary of $[0,1]^2$, then parts of the red and blue regions will be outside of $[0,1]^2$ and cannot contain any point of P. We handle this by considering the 4 diagonal directions $(\pm 1, \pm 1)$, and picking the one in which the boundary is the furthest away from p_1. Now, in the cone of half-angle $\pi/8$ around that direction, the red and blue parts are contained in the unit square. Letting $8s$ denote the total number of sectors, we therefore focus on the s sectors around that direction. For the rest of this section, we assume that this direction is $(1,1)$ as illustrated in Figure 3; the three other cases are symmetric. We label $B_1, B_2, \ldots B_s$ (resp. $R_1, R_2, \ldots R_s$) the intersection of each of our angular sectors with the blue disk minus p_1 (resp. the red annulus), in counterclockwise order.

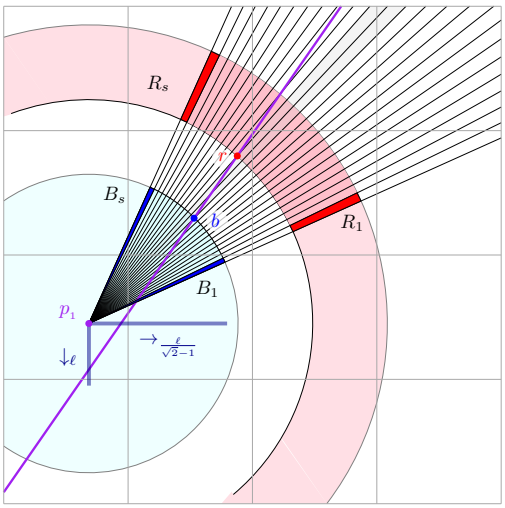

Figure 3: Subdivision in cones around the cell containing p_1.

Next, finding a line close to p_1 is not enough: to ensure that $L(1) > k$, we need to find lines that intersect *all four* horizontal and vertical segments of length 2^{-k} with endpoint p_1. To do that, we look for lines (br) where $b \in B_i$ and $r \in R_{i+1}$. This shift in indices ensures that the line (br) is close to p_1 *and* passes below p_1: indeed, r and b are respectively above and below the ray from p_1 that is a common boundary of B_i and R_{i+1}. Similarly, finding some points $b' \in B_{i'}$ and $r' \in R_{i'-1}$ will provide a line $(b'r')$ passing close to p_1 and above it; together, these two lines will intersect all four horizontal and vertical segments that have p_1 as an endpoint.

From here, two tasks remain: quantify the lower bound on $L(1)$ that comes from finding such indices i and i', and estimate the probability that such indices exist. We do that in the next two subsections.

4.2 Geometric analysis

We now formulate the lower bound on $L(1)$ afforded by the collisions that we want. We focus on a pair $b \in B_i$ and $r \in R_{i+1}$ that yields the line below p_1, as the other pair is symmetric.

Since we consider what happens around the direction $(1, 1)$, the line passing below p_1 will have to intersect both the horizontal segment with p_1 as leftmost point and the vertical segment with p_1 as topmost point. Note, however, that any line (br) that we consider has slope at least $\tan\frac{\pi}{8} = \sqrt{2} - 1$. Let $\rightarrow_\ell$ and $\downarrow_\ell$ denote the segments of length ℓ with p_1 as, respectively, leftmost and topmost point. We thus have that if a line (br) intersects $\downarrow_\ell$, it must also intersect $\rightarrow_{\frac{\ell}{\sqrt{2}-1}}$. We thus focus on finding the smallest ℓ such that $\downarrow_\ell$ is guaranteed to meet (br).

Lemma 14. *If $s \geq 10$, for any point $b \in B_i$ and $r \in R_{i+1}$, the line (br) intersects $\downarrow_{\frac{\pi}{2s}}$.*

Proof. The vertical distance between p_1 and (br) is maximal when b and r are placed in the corners of B_{s-1} and R_s on circles of radii 0.2 and 0.3 as in Figure 4-left. Let us relate this maximal distance h to $\theta = \widehat{bp_1r}$. As spelled out in Figure 4-right, we can express θ as a function of h (for θ sufficiently small):

$$\theta = \arcsin\left(\frac{0.3}{\sqrt{0.09+h^2+0.6h\sin\frac{\pi}{8}}}\sin\frac{\pi}{8}\right) + \arcsin\left(\frac{0.3}{\sqrt{0.09+h^2+0.6h\sin\frac{\pi}{8}}}\frac{h}{0.2}\sin\frac{\pi}{8}\right) - \frac{\pi}{8}.$$

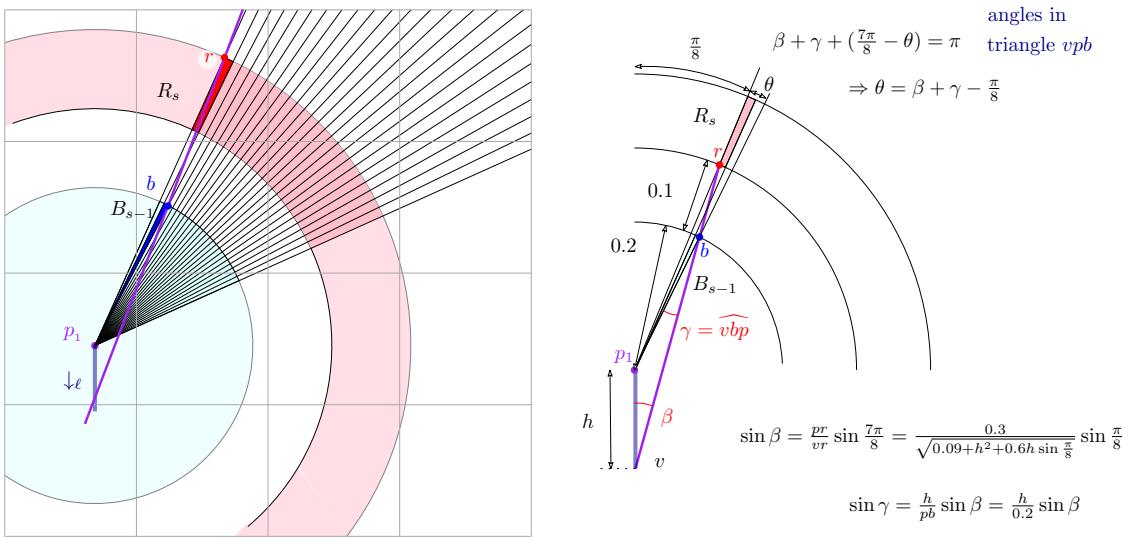

Figure 4: For the proof of Lemma 14.

This function $h \mapsto \theta(h)$ is increasing on $[0, 0.6]$ and $\theta(h) > h$ when $\theta(h) \in [0, 0.17]$. Since θ is the angle of *two* sectors, we have $\theta = 2\frac{\pi}{4s}$. For $s \geq 10$ we have $h < \frac{\pi}{2s}$. □

We can now give our "balls-in-bins" condition that bounds from below the value of $L(1)$.

Corollary 15. *Assume that $k \geq 3$ and that $s = 2^{k+1}$. If there exists i, i' in $[s]$ such that P intersects each of the regions B_i, R_{i+1}, $B_{i'}$, and $R_{i'-1}$, then $L(1) \geq k$.*

Proof. Let $b \in B_i \cap P$ and $r \in R_{i+1} \cap P$. Since $s \geq 16$, Lemma 14 ensures that the line (br) intersects $\downarrow_{\frac{\pi}{2s}}$. As argued before Lemma 14, that line also intersects $\downarrow_\ell$ and $\rightarrow_\ell$ with $\ell = \frac{\pi}{2(\sqrt{2}-1)s}$. A symmetry with respect to the line of slope 1 through p_1 gives the intersection with the two other segments from the points in $B_{i'}$ and $R_{i'-1}$. Since $\frac{\pi}{2(\sqrt{2}-1)} \leq 4$, the existence of i and i' ensures that all four horizontal and vertical segments of length $\frac{4}{s} = 2^{-k+1}$ starting in p_i are intersected by some lines spanned by $P \setminus \{p_1\}$, so $L(1) > k - 1$. □

4.3 A balls-in-bins analysis

We now prove Lemma 12. Since we are interested in the probability that $L(1)$ be at least $2n \log n$ (minus some change), we use Corollary 15 with $s = \frac{n^2}{\log^x n}$ and $x > 1$.

Proof idea. We want to bound from below the probability that there exist $a, b \in [s]$ such that each of B_a, R_{a+1}, B_b, and R_{b-1} is hit by P. For $i \in [s]$ and $j \in [n-1]$ we define the following random variables:

$$\begin{array}{l|l|l} X_{i,j} = \mathbb{1}_{p_{j+1} \in B_i} & X_i = \max_{j \in [n-1]} X_{i,j} & X = \sum_{i \in [s]} X_i \\ Y_{i,j} = \mathbb{1}_{p_{j+1} \in R_i} & Y_i = \max_{j \in [n-1]} Y_{i,j} & Y = \sum_{i \in [s]} Y_i \end{array}$$

(Note that, for a better bookkeeping, we index the events associated with p_j by $j-1$ because p_1 is already chosen.) In plain english, X_i is the indicator variable that B_i is nonempty, and X counts the number of non-empty regions B_i. (The $Y_\bullet$ variables do the same for the regions R_i.) The definition of the regions ensures that each is fully contained in the unit square, that all B_i have the same area, and that all R_i have the same area. So all the $\{X_{i,j}\}_{i,j}$ are identically distributed, and so are the $\{Y_{i,j}\}_{i,j}$, the $\{X_i\}_i$, and the

$\{Y_i\}_i$. Remark, however, that for fixed j, any subset of $\{X_{i,j}\}_i \cup \{Y_{i,j}\}_i$ is dependent as its sum is zero or one.

Conditioned on the fact that a point lands in a red cell, that red cell is chosen uniformly among the s red cells. Thus, conditioning on the values of X and Y, we have

$$\mathbb{P}\left[\exists i\colon B_i \cap P \neq \emptyset \text{ and } R_{i+1} \cap P \neq \emptyset \mid X = \beta, Y = \rho\right] \geq 1 - \left(1 - \frac{\beta - 1}{s}\right)^{\rho}. \tag{1}$$

Indeed, all but at most one of the occupied blue cells are next to a red cell which, if occupied, makes the event true. If the random variables involved were all independent, we could use the Chernoff-Hoeffding concentration bound to bound from below with high probability the values of X and Y. As explained above, however, we have to deal with some dependencies. Also, Inequality (1) takes care of only half of the condition formulated by Corollary 15.

Laws of the random variables. Each B_i has area c_1/s, and each R_i has area c_2/s with $c_1 = \frac{\pi}{200}$ and $c_2 = \frac{7\pi}{800}$. Thus, $X_{i,j}$ and $Y_{i,j}$ are $0-1$ random variables, taking value 1 with probability, respectively, c_1/s and c_2/s. For fixed i, the $\{X_{i,j}\}_{j\in[n-1]}$ are independent, so we have

$$\begin{aligned}\mathbb{E}\left[X_i\right] = \mathbb{P}\left[X_i = 1\right] = 1 - \left(1 - \frac{c_1}{s}\right)^{n-1} &\geq 1 - e^{-c_1\frac{n-1}{s}} \geq c_1\frac{n-1}{s} - \frac{1}{2}\left(c_1\frac{n-1}{s}\right)^2 \\ &= c_1\frac{\log^x n}{n} - O\left(\frac{\log^{2x} n}{n^2}\right).\end{aligned}$$

the first and second inequalities coming, respectively, from the facts that for every $t \geq 0$ we have $1 - t \leq e^{-t}$ and for every $t \in [0,1]$ we have $1 - e^{-t} \geq -t - \frac{t^2}{2}$. Then, we plugged in $s = \frac{n^2}{\log^x n}$. The same computation gives

$$\mathbb{E}\left[Y_i\right] \geq c_2\frac{\log^x n}{n} - O\left(\frac{\log^{2x} n}{n^2}\right).$$

Finally, since the X_i are identically distributed, and so are the Y_i, we have

$$\mathbb{E}\left[X\right] = s\mathbb{E}\left[X_i\right] \geq c_1 n - O\left(\log^x n\right) \quad \text{and} \quad \mathbb{E}\left[Y\right] = s\mathbb{E}\left[Y_i\right] \geq c_2 n - O\left(\log^x n\right).$$

Negatively associated random variables. The variables $\{X_{i,j}\}_i$ are negatively dependent in the sense that when one is 1, the others must be 0. Formally, they can be shown to be *negatively associated*. We do not elaborate on this notion here, but refer to the paper of Dubhashi and Ranjan [5] from which we highlight the following points:

- Any finite set of $0-1$ random variables that sum to 1 is negatively associated [5, Lemma 8]. So, the set $\{X_{i,j}\}_i \cup \{1 - \sum_i X_{i,j}\}$ is negatively associated.

- Any set of increasing functions of pairwise disjoint subsets of negatively associated random variables forms, again, a set of negatively associated random variables [5, Proposition 7]. Thus, each of the sets $\{X_{i,j}\}_i$, $\{Y_{i,j}\}_i$, $\{X_i\}_{i\in[s]}$, and $\{Y_i\}_{i\in[s]}$ consists of negatively associated random variables.

- The Chernoff-Hoeffding bounds apply to sums of any set of negatively associated random variables [5, Proposition 5]. Applying [9, Theorem 4.2] for $\delta = \frac{1}{2}$ for instance yields

$$\mathbb{P}\left[X \leq \frac{\mathbb{E}\left[X\right]}{2}\right] \leq 0.89^{\mathbb{E}[X]} \quad \text{and similarly} \quad \mathbb{P}\left[Y \leq \frac{\mathbb{E}\left[Y\right]}{2}\right] \leq 0.89^{\mathbb{E}[Y]}.$$

Let us return to the event $\mathcal{O}$ that there exist a, b in $[s]$ such that each of B_a, R_{a+1}, B_b, R_{b-1} is hit by P. Let us condition by the event $\mathcal{G} = \{X \geq \mathbb{E}[X]/2 \text{ and } Y \geq \mathbb{E}[Y]/2\}$. A union bound yields

$$\mathbb{P}[\mathcal{G}] \geq 1 - \left(0.89^{\mathbb{E}[X]} + 0.89^{\mathbb{E}[Y]}\right) \geq 1 - \left(0.89^{c_1 n - O(\log^x n)} + 0.89^{c_2 n - O(\log^x n)}\right)$$

which is exponentially close to 1. We thus bound from below $\mathbb{P}[\mathcal{O}] \geq \mathbb{P}[\mathcal{G}]\,\mathbb{P}[\mathcal{O}|\mathcal{G}]$ and concentrate on the conditional probability.

Bichromatic birthday paradox. The probability $\mathbb{P}[\mathcal{O}|\mathcal{G}]$ can be expressed as a convex combination of the conditional probabilities $f(b,r) = \mathbb{P}[\mathcal{O}|\mathcal{G}_{b,r}]$, where for integers $b \geq \mathbb{E}[X]/2$ and $r \geq \mathbb{E}[Y]/2$ we take $\mathcal{G}_{b,r} = \{X = b, Y = r\}$. Conditioned on $\mathcal{G}_{b,r}$, the occupied regions of each type are uniformly random and independent, which will simplify the analysis. Furthermore, the function $f(b,r)$ is increasing in both variables (the more occupied regions there are, the more likely it is that the collisions we desire occur). Thus, we concentrate on finding a lower bound on $f(b,r)$ for $b = \left\lceil \frac{\mathbb{E}[X]}{2} \right\rceil$ and $r = \left\lceil \frac{\mathbb{E}[Y]}{2} \right\rceil$.

Assume the b occupied blue regions have been chosen. Let T_+ (resp. T_-) denote the set of red regions in sectors following counterclockwise (resp. clockwise) the sectors whose blue regions have been chosen. Since the blue regions in the boundary angular sectors may be among those chosen, we have $b-1 \leq |T_+|, |T_-| \leq b$. We now pick the r red regions to be occupied.

Let E_+ (resp. E_-) denote the event that a region of T_+ (resp. T_-) has been chosen among the r red regions. Pretend, for the sake of the analysis, that we choose the red regions one by one. If none of the first i regions chosen is in T_+, then next one has to be picked from the $s-i$ unpicked regions, at least $b-1$ of which are in T_+. Thus,

$$\begin{aligned} 1 - \mathbb{P}[E_+] \leq \prod_{i=0}^{r-1} \left(1 - \frac{b-1}{s-i}\right) &= \frac{(s-b+1)(s-b)\dots(s-b-r+2)}{s(s-1)\dots(s-r+1)} \\ &= \frac{(s-r)!(s-b+1)!}{s!(s-b-r+1)!} \end{aligned}$$

Using a symmetric argument for T_- and applying a union bound, we get

$$1 - f(b,r) \leq 2\,\frac{(s-r)!(s-b+1)!}{s!(s-b-r+1)!}.$$

Note that for $b = \left\lceil \frac{\mathbb{E}[X]}{2} \right\rceil$ and $r = \left\lceil \frac{\mathbb{E}[Y]}{2} \right\rceil$, both b and r are $\Theta(n) = o(s)$. Taking logarithm and using Simpson's approximation formula, which asserts that $\log(N!) = N\log(N) - N + O(\log N)$, we get

$$\begin{aligned} \log(1 - f(r,b)) &= (s-r)\log(s-r) + (s-b+1)\log(s-b+1) \\ &\quad - s\log s - (s-b-r+1)\log(s-b-r+1) + O(\log s) \\ &= s\log\frac{(s-r)(s-b+1)}{s(s-b-r+1)} - r\log\frac{s-r}{s-b-r+1} \\ &\quad - b\log\frac{s-b+1}{s-b-r+1} + O(\log s) \\ &= s\log\left(1 + \frac{r(b-1)}{s(s-b-r+1)}\right) - r\log\left(1 + \frac{b-1}{s-b-r+1}\right) \\ &\quad - b\log\left(1 + \frac{r}{s-b-r+1}\right) + O(\log s). \end{aligned}$$

Now in the regime we are looking at, we have $b = c_1 n/2 - O(\log^x n)$, $r = c_2 n/2 - O(\log^x n)$, and $s = \frac{n^2}{\log^x n}$. Taking first order Taylor expansions, our bound rewrites as

$$\log(1-f(r,b)) = -\frac{rb}{s-b-r+1} + O(\log s) = -\frac{c_1 c_2}{4}\log^x n + O(\log n)\,.$$

provided we have $x > 1$. Hence, $f(r,b) = 1 - \exp(\Theta(\log(n)^x))$. Altogether, we get that $\mathbb{P}[\mathcal{O}|\mathcal{G}]$ is exponentially close to 1. Since $\mathbb{P}[\mathcal{G}]$ is also exponentially close to 1, we finally get that our event $\mathcal{O}$ holds with probability exponentially close to 1. With Corollary 15, this proves Lemma 12.

5 Perspectives

We stated and proved our main results (Theorems 2 and 3) for a uniform sample of the unit square. The careful reader may observe, however, that we have taken care to separate the geometric from the probabilistic arguments. Although the multiplicative constants of the leading terms in the end-results matter (we want both $\mathbb{E}[U(i)]$ and $\mathbb{E}[L(i)]$ to equal $2\log n$ at first order), the multiplicative constants in the geometric arguments do *not* matter:

- Lemma 9 needs only establish an upper bound of $O(\frac{1}{m\delta(p,q)})$ for Lemma 10 to yield that $\mathbb{P}[U(1) > k] \leq O\left(n^2 2^{-k}\right)$, which ensures that $\mathbb{E}[U(1)] \leq 2\log n + O(1)$.

- If the blue disk and red annulus are scaled by a constant factor, Lemma 14 still holds with $\downarrow_{\frac{\pi}{2s}}$ replaced by $\downarrow_{\Theta\left(\frac{1}{s}\right)}$; this changes the choice of s in Section 4.3 to $s = \Theta\left(\frac{n^2}{\log^x n}\right)$, which changes only at *which* exponential speed the probability that $L(1) \geq 2\log n - O(\log\log n)$ converges to 1.

- More generally, the lower bound on $L(1)$ should work for any probability measure for which one can prove a uniform lower bound of $\Omega(1/s)$ for the probabilities of the individual blue and red regions.

It should therefore be clear that the same analysis, with different constants, holds for a variety of more general probability distributions for the points; examples include the uniform distribution on any bounded convex domain with non-empty interior, or even any distribution on such a convex set with a density that is bounded away from 0.

Before examining too closely what other distributions can be handled by our arguments, it would perhaps be interesting to determine tighter upper bounds on the difference between the $U(i)$ and $L(i)$ statistics in the *deterministic* setting. This would quantify the amount by which our algorithm overshoots in the *worst-case*. We currently have no result in this direction.

Another question is whether completely different random point sets would produce order types with typical resolution much higher than that given by Theorem 3. We plan experimentations with determinantal point processes.

References

[1] Oswin Aichholzer, Franz Aurenhammer, and Hannes Krasser. Enumerating order types for small point sets with applications. *Order*, 19(3):265–281, 2002.

[2] Noga Alon. The number of polytopes, configurations and real matroids. *Mathematika*, 33(1):62–71, 1986.

[3] Luis E Caraballo, José-Miguel Díaz-Báñez, Ruy Fabila-Monroy, Carlos Hidalgo-Toscano, Jesús Leaños, and Amanda Montejano. On the number of order types in integer grids of small size. *arXiv preprint arXiv:1811.02455*, 2018.

[4] Jean Cardinal, Timothy M. Chan, John Iacono, Stefan Langerman, and Aurélien Ooms. Subquadratic Encodings for Point Configurations. In Bettina Speckmann and Csaba D. Tóth, editors, *34th International Symposium on Computational Geometry (SoCG 2018)*, volume 99 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 20:1–20:14, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. URL: http://drops.dagstuhl.de/opus/volltexte/2018/8733, doi:10.4230/LIPIcs.SoCG.2018.20.

[5] Devdatt Dubhashi and Desh Ranjan. Balls and bins: A study in negative dependence. *Random Structures and Algorithms*, 13(2):99–124, 1998.

[6] Tobias Gerken. Empty convex hexagons in planar point sets. *Discrete & Computational Geometry*, 39(1-3):239–272, 2008.

[7] Xavier Goaoc, Alfredo Hubard, Rémi de Joannis de Verclos, Jean-Sébastien Sereni, and Jan Volec. Limits of order types. In *LIPIcs-Leibniz International Proceedings in Informatics*, volume 34. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2015.

[8] Jacob E Goodman, Richard Pollack, and Bernd Sturmfels. The intrinsic spread of a configuration in $\mathbb{R}^d$. *Journal of the American Mathematical Society*, pages 639–651, 1990.

[9] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[10] Mark Overmars. Finding sets of points without empty convex 6-gons. *Discrete & Computational Geometry*, 29(1):153–158, 2002.

[11] Peter Shor. Stretchability of pseudolines is NP-hard. *Applied Geometry and Discrete Mathematics-The Victor Klee Festschrift*, 1991.

A Some experimental data

Let us give some details on the experiment mentioned in the introduction.

Disclaimer. Let us stress that this is provided as contextual evidence, and *nothing more*. We do not claim that it meets any experimental standard; for example we did not measure precisely computation times, etc. We hope nevertheless that it gives some idea of the practical limitations one faces when trying to explore order types via random sampling.

Setup. We produced a billion sets of 10 points and recorded the empirical frequencies of their order types. In practice, we recorded the empirical frequencies after every 10 millions point sets. In our experiment, every order type is identified via a signature (see below). We represent our points' coordinates with double precision floating point numbers (`long double` in C++). The mantissa has 52 bits, while 16 suffice to represent any order type of size 10 (*c.f.* Aicholzer et al. [1]).

Order type signature. Given a set P of n points, consider all $n!$ possible labelings of these points by $1, 2, \ldots, n$. For each labeling σ, contruct the word

$$w(\sigma) = 2a_{1,1}a_{1,2}\ldots a_{1,n-2}1a_{2,1}a_{2,2}\ldots a_{2,n-2}1a_{3,1}a_{3,2}\ldots a_{3,n-2}\ldots 1a_{n,1}a_{n,2}\ldots a_{n,n-2}$$

where $1a_{i,1}, a_{i,2}, \ldots$ are the labels of the points in circular CCW order around the ith point, starting from the first point (or from the second point when turning around the 1st point). The lexicographically smallest word $w(\sigma)$ characterizes the order type of P. It can be computed in time $O(n^3)$ by observing that one needs only examine the labellings where 1 and 2 are consecutive on the convex hull, and the other points are labelled in CCW order around 1 following 2. (The geometric computations are done using CGAL's `Exact_predicates_inexact_constructions_kernel`.)

Pseudo-random generation. We generated our point sets by picking the coordinates of each point in $[1.0, 2.0]$, so that the precision is the same everywhere. We used the pseudo-random generators of the standard C++ library:

```
std::random_device rd;

std::mt19937_64 gen(rd());

std::uniform_real_distribution<double> dis(1.0, 2.0);
```

Then we produced each point's coordinate by a call to `dis(gen)`.

Order types of size 10**.** Aicholzer et al. [1] counted 14 309 547 order types of size 10 up to reflection; that is, they identify the order type of a point set with the order type of the reflection of that point set with respect to a line. We examined every realization in their database and checked whether reflecting the points (horizontally) yields the same order type; this happened for 13 064 of the realizations. So, the total number of order types of size 10 is 28 606 030.

Some results. Our one billion point sets produced 10 920 123 distinct order types, of which 2 476 184 were only seen once and the 7 most frequently found were seen respectively 563 409, 375 833, 374 657, 299 907, 277 054, 276 609 and 248 045 times. The following tables give some more general overview of these data.

number of point sets	number of new order types
0-10M	527 885
100-110M	19 679
200-210M	10 882
300-310M	7 558
400-410M	5 886
500-510M	4 926
600-610M	4 091
700-710M	3 522
800-810M	3 044
900-910M	2 731

Table 1: The number of *new* order types discovered per slice of 10 millions point sets, for some slices.

p	s
10	3.331e-06
20	1.3262e-05
30	3.2798e-05
40	6.7136e-05
50	0.000125185
60	0.000223536
70	0.000396444
80	0.000731306
90	0.001544605
92	0.001866255
94	0.002319744
95	0.00262831
96	0.003025812
97	0.003565944
98	0.004370649
99	0.005807947

Table 2: Proportion s of all order types discovered over our billion trials that suffice to make up a percentage p of the trials: half of the trials produce a fraction of only $\simeq \frac{1}{8000}$ of the order types seen, and slightly more than 0.58% of the order types seen account for 99% of the trials.