

LARGE AREA CONVEX HOLES IN RANDOM POINT SETS

OCTAVIO ARIZMENDI*

Centro de Investigación en Matemáticas. Guanajuato, Mexico.

GELASIO SALAZAR†

Instituto de Física, UASLP. San Luis Potosí, Mexico.

ABSTRACT. Let K, L be convex sets in the plane. For normalization purposes, suppose that the area of K is 1. Suppose that a set K_n of n points are chosen independently and uniformly over K , and call a subset of K a *hole* if it does not contain any point in K_n . It is shown that w.h.p. the largest area of a hole homothetic to L is $(1 + o(1)) \log n/n$. We also consider the problems of estimating the largest area convex hole, and the largest area of a convex polygonal hole with vertices in K_n . For these two problems we show that the answer is $\Theta(\log n/n)$.

1. INTRODUCTION

Let K be a convex set in the plane, and let K_n be a set of n points chosen independently and uniformly at random from K . A K_n -hole (or simply a *hole*) is a subset of K whose interior does not contain any point of K_n .

In [2], Balogh et al. proved that w.h.p. the size of the largest (in the number of vertices) polygonal hole with vertices in K_n is $\Theta(\log n / \log \log n)$.

The present work was motivated by a question raised by Matthew Kahle (private communication to J. Balogh), who asked about the largest hole in terms of area rather than in terms of the number of vertices. Kahle mentioned two possible variants of these problems: either (i) find the area of the largest disk hole; or (ii) find the area of the largest convex hole.

In this note we investigate these questions. For the first question, we give an asymptotically exact answer. Moreover, we show that the answer is the same for any convex K , and it is also independent of the chosen shape of the empty convex region:

*Supported by CONACYT Grant 222668. E-mail: octavio@cimat.mx.

†Supported by CONACYT Grant 222667. E-mail: gsalazar@ifisica.uaslp.mx.

Theorem 1.1. *Let K, L be convex sets in the plane. Suppose for normalization purposes that the area of K is 1. Let K_n be a set of n points chosen independently and uniformly at random from K . Let $\text{MAX}_L(K_n)$ denote the random variable that measures the largest (in terms of area) hole that is homothetic to L . Then w.h.p.*

$$\text{MAX}_L(K_n) = (1 + o(1)) \frac{\log n}{n}.$$

Regarding the area of the largest convex hole, we give an answer that is asymptotically tight within a factor of 4:

Theorem 1.2. *Let K be a convex set in the plane with area 1. Let K_n be a set of n points chosen independently and uniformly at random from K . Let $\text{MAX}(K_n)$ denote the random variable that measures the largest (in terms of area) convex hole. Then w.h.p.*

$$(1 + o(1)) \frac{\log n}{n} \leq \text{MAX}(K_n) \leq (4 + o(1)) \frac{\log n}{n}.$$

A third variant of this problem is to investigate large area polygonal holes (for a related work see [10]). In this direction, our estimate is also asymptotically tight within a factor of 4:

Theorem 1.3. *Let K be a convex set in the plane with area 1. Let K_n be a set of n points chosen independently and uniformly at random from K . Let $\text{POLYMAX}(K_n)$ denote the random variable that measures the largest (in terms of area) convex polygon with vertices in K_n . Then w.h.p.*

$$(1 + o(1)) \frac{\log n}{n} \leq \text{POLYMAX}(K_n) \leq (4 + o(1)) \frac{\log n}{n}.$$

Questions about large convex substructures in point sets are of fundamental importance in discrete and computational geometry; many such variants are typically described as ‘‘Erdős-Szekeres type problems’’, after the seminal paper [8]. Problems that ask for empty (usually convex) substructures are particularly natural and important; see for instance [3–5, 12–14].

From the algorithmical point of view, a well-studied problem in the field is to compute the largest empty rectangle (or, say, the largest axis-parallel d -dimensional empty box). See for instance [1]. A related result by Dumitrescu and Jiang [7] is an efficient $(1 - \epsilon)$ approximation algorithm for computing a maximum-volume empty axis-parallel d -dimensional box contained in an axis-parallel d -dimensional box in \mathbb{R}^d . The problem of investigating empty substructures in random point sets seems to have received far less attention; we could only find two related results in the literature, other than the already mentioned [2]. In [6], Dumitrescu and Jiang found the expected number of maximal empty axis-parallel boxes amidst n random points in the unit hypercube; and, very recently, Fabila-Monroy et al. found the expected number of empty convex (and also non-convex) four-gons with vertices in a finite set randomly chosen from a convex set [9].

Although the proofs of our main results stated above are somewhat technical, at their core they rely on the following well-known fact from probability theory. Suppose that we toss n balls into $c \cdot n / \log n$ bins. If $c < 1$, then w.h.p. none of the bins will be empty, and if $c > 1$, then w.h.p. many bins will be empty. The proofs are also based on some well-known facts from convexity theory that we shall review in Section 2.

Theorems 1.1, 1.2, and 1.3 are proved in Sections 3, 4, and 5, respectively. Section 6 contains some concluding remarks.

2. BASIC FACTS ON PROBABILITY AND CONVEX SETS

In this section we gather some basic facts that will be used in the proofs of the three main theorems.

The following is a straightforward exercise in elementary probability theory.

Proposition 2.1. *Let $k, n \in \mathbb{N}$. Suppose that n balls are thrown independently at random into k bins. Let Y be the random variable counting the number of bins that are empty. Then*

$$\mathbb{E}[Y] = k \left(1 - \frac{1}{k}\right)^n \sim ke^{-n/k}, \quad \text{and}$$

$$\text{Var}[Y] = k \left(1 - \frac{1}{k}\right)^n + k(k-1) \left(1 - \frac{2}{k}\right)^n - k^2 \left(1 - \frac{1}{k}\right)^{2n} \sim ke^{-n/k} - ke^{-2n/k}.$$

In the proofs of Theorems 1.1, 1.2, and 1.3, we rely heavily on the following statement.

Proposition 2.2. *Let K be a region in the plane of area 1. Suppose that n points are chosen independently and uniformly at random from K . Let $\epsilon \in (0, 1)$. Suppose that K is partitioned into $t := n/((1-\epsilon)\log n)$ equal area regions. Then the probability that fewer than $n^\epsilon/(2(1-\epsilon)\log n)$ of these regions are empty is at most $\sim \frac{4(1-\epsilon)\log(n)}{n^\epsilon}$.*

Proof. Let X be the random variable counting the number of empty regions. By Proposition 2.1 we have

$$\mathbb{E}(X) \sim \frac{n}{(1-\epsilon)\log n} e^{-(1-\epsilon)\log(n)} = \frac{n^\epsilon}{(1-\epsilon)\log(n)}.$$

Moreover, also by Proposition 2.1 we have $\text{Var}(X) < \mathbb{E}(X)$, and so it follows from Chebyshev's inequality that

$$P\left(|(X - \mathbb{E}(X))| > \mathbb{E}(X)/2\right) \leq \frac{4\text{Var}(X)}{(\mathbb{E}(X))^2} < \frac{4}{\mathbb{E}(X)} \sim \frac{4(1-\epsilon)\log(n)}{n^\epsilon}.$$

□

The following fact on approximating convex sets by rectangles was proved by Lassak in [11]. In this statement, and in what follows, we let $a(K)$ denote the area of a set K .

Proposition 2.3. *Let L be a convex body in the plane. We can inscribe a rectangle S in L such that a homothetic copy R of S is circumscribed about L . The positive homothety ratio is at most 2 and $\frac{1}{2}a(R) \leq a(L) \leq 2a(S)$.*

The following is a straightforward exercise in convex geometry.

Proposition 2.4. *Let K be a convex set in the plane, and let R be a rectangle. Then there is an integer M such that for all $m \geq M$ there is a partition of K into m equal area regions such that at least $2/3$ of the regions are homothetic to R .*

3. PROOF OF THEOREM 1.1

Let K, L be convex sets in the plane, where K has area 1 and L is convex. Let K_n be a set of n points chosen independently and uniformly at random from K .

Theorem 1.1 is an immediate consequence of the following two statements:

- (1) Fix any $\epsilon > 0$. Then w.h.p. there exists a hole homothetic to L with area at least $(1 - \epsilon) \log n/n$.
- (2) Fix any $\epsilon > 0$. Then w.h.p. there is no hole homothetic to L with area greater than $(1 + 3\epsilon) \log n/n$.

Proof of (1).

We start by using Proposition 2.3 to find a rectangle R that circumscribes L , and such that $a(R) \leq 2a(L)$. Now we invoke Proposition 2.4 (we may assume that n is sufficiently large) to partition K into $2n/((1 - \epsilon) \log n)$ equal area regions, at least $2/3$ of which are homothetic to R . Next, we partition each of the regions homothetic to R into two equal area parts, one of which is homothetic to L ; finally, we partition the part of K not already covered, into regions of area $n/((1 - \epsilon) \log n)$. The result is a partition of K into a collection of $n/((1 - \epsilon) \log n)$ regions of equal area, at least $1/3$ of which are homothetic to L .

By Proposition 2.2, w.h.p. there are at least $\frac{n^\epsilon}{2(1-\epsilon)\log(n)}$ empty regions. Since each empty region is homothetic to L with probability at least $1/3$, (1) follows.

Proof of (2).

We may assume that L has area $(1 + 3\epsilon) \log n/n$, so that the aim is to show that w.h.p. there is no empty translate of L . Let P be a convex set contained in L , whose boundary is a smooth curve, and such that $a(P) = (1 + 2\epsilon) \log n/n$. By Proposition 2.3 there is a rectangle Q of area $2a(P)$ that contains P . By performing an affine transformation on the plane, if necessary, we may assume that Q is a square whose sides are parallel to the Cartesian axes. Let \mathcal{T} denote the set of all translates of P . To prove (2) it suffices to show that w.h.p. no element of \mathcal{T} contained in K is empty.

We let $s(T)$ be the convex set contained in T , defined by the following properties: (i) $s(T)$ has area $(1 + \epsilon) \log n/n$; and (ii) there is an $\omega > 0$ such that $s(T)$ consists of those points whose distance to the boundary of T is at least ω . We let $\mathcal{S} := \{s(T) \mid T \in \mathcal{T}\}$. Note that the map s is invertible: for each $S \in \mathcal{S}$, we let $s^{-1}(S)$ denote the $T \in \mathcal{T}$ such that $s(T) = S$.

For each $T \in \mathcal{T}$, we let $c(T)$ denote the center of mass of T . For each $S \in \mathcal{S}$, we let $c(S)$ denote $c(s^{-1}(S))$. Note that if $S \in \mathcal{S}$ then $c(S)$ is not necessarily the center mass of S .

We claim that $\omega > (\epsilon/8)\sqrt{\log n/n}$. First note that the sides of Q have length $\sqrt{(2 + 4\epsilon) \log n/n}$, and so the perimeter $\text{per}(Q)$ of Q is $4\sqrt{(2 + 4\epsilon) \log n/n}$. Since P is a convex set contained in Q , then the perimeter $\text{per}(P)$ of P is also at most $4\sqrt{(2 + 4\epsilon) \log n/n}$. Now $\epsilon \log n/n = a(P \setminus s(P)) \leq \text{per}(P)\omega$, and so $\omega \geq (\epsilon \log n/n)/(4\sqrt{(2 + 4\epsilon) \log n/n}) > (\epsilon/8)\sqrt{\log n/n}$ (for all sufficiently small ϵ), as claimed.

Let $H := \{(i\omega, j\omega) \mid i, j \in \mathbb{Z}\} \cap K$, and let $\mathcal{S}_H := \{S \in \mathcal{S} \mid c(S) \in H \text{ and } S \subseteq K\}$.

Recall that to prove (2) it suffices to show that w.h.p. no element of \mathcal{T} contained in K is empty. Therefore to finish the proof of (2) it suffices to prove that:

- (A) Every element of \mathcal{T} contained in K contains an element in \mathcal{S}_H ; and
 (B) W.h.p. no element of \mathcal{S}_H is empty.

Consider any $T \in \mathcal{T}$ contained in K . Since every point in K is at distance less than ω from a point in H , it follows that there exists a point p in H that is at distance less than ω from $c(T)$. It follows from the triangle inequality that the element $S \in \mathcal{S}_H$ such that $c(S) = p$ is contained in T . This proves (A).

To prove (B), we start by bounding $|\mathcal{S}_H|$. Let $S \in \mathcal{S}_H$, and let $c(S) = (x, y)$. We define $\square(S)$ to be the square with vertices $(x, y), (x + \omega, y), (x, y + \omega), (x + \omega, y + \omega)$. If $S, S' \in \mathcal{S}_H$, then the interiors of $\square(S)$ and $\square(S')$ are disjoint. Since the area of each such square is ω^2 , it follows that there are at most $a(K)/\omega^2 = 1/\omega^2 \leq (64/\epsilon^2)(n/\log n)$ such squares. Since the map that sends each $S \in \mathcal{S}_H$ to $\square(S)$ is an injection, it follows that $|\mathcal{S}_H| \leq (64/\epsilon^2)(n/\log n)$.

Fix any $S \in \mathcal{S}_H$. The area of S is $(1 + \epsilon) \log n/n$, and so the probability that S is empty is $(1 - (1 + \epsilon) \log n/n)^n \sim n^{-(1+\epsilon)}$. By the union bound, the probability that there is an empty element in \mathcal{S}_H is at most $|\mathcal{S}_H| \cdot n^{-(1+\epsilon)} \leq (64/\epsilon^2)(n/\log n) \cdot n^{-(1+\epsilon)} = o(1)$. Thus (B) follows.

4. PROOF OF THEOREM 1.2

The lower bound of Theorem 1.2 is an immediate consequence of Theorem 1.1. Thus it remains to prove the upper bound.

By Proposition 2.3, every convex set contains a rectangle of half its area. Thus in order to prove Theorem 1.2 it suffices to prove the following:

Theorem 4.1. *Let K be a convex set in the plane with area 1. Let K_n be a set of n points chosen independently and uniformly at random from K . Let $\text{MAXREC}(K_n)$ denote the random variable that measures the largest (in terms of area) empty rectangle contained in K . Then w.h.p.*

$$\text{MAXREC}(K_n) \leq (2 + o(1)) \frac{\log n}{n}.$$

The main tool to prove Theorem 4.1 is the following.

Proposition 4.2. *Let K be a convex set in the plane with area 1. Let K_n be a set of n points chosen independently and uniformly at random from K . Let $\epsilon > 0$ be given. Then there exists a family \mathcal{R} of $O(n^2)$ rectangles, each of area $(2 + \epsilon) \log n/n$, with the following property: every rectangle of area $(2 + 4\epsilon) \log n/n$ contained in K contains a rectangle in \mathcal{R} .*

The proof of Proposition 4.2, although not difficult, is somewhat technical. We defer it for the moment, and show how Theorem 4.1 (and thus Theorem 1.2) follows from it.

Proof of Theorem 4.1. Let \mathcal{R} be as in the statement of Proposition 4.2. The probability that a fixed $R \in \mathcal{R}$ is empty is $(1 - a(R))^n = (1 - (2 + \epsilon) \log n/n)^n \sim n^{-(2+\epsilon)}$. Thus it follows from the union bound that the probability that some $R \in \mathcal{R}$ is empty is at most $O(n^2) \cdot n^{-(2+\epsilon)} = O(n^{-\epsilon})$. Now every rectangle of area $(2 + 4\epsilon) \log n/n$ contained in K contains a rectangle in \mathcal{R} , and so the probability that there is an empty rectangle of area $(2 + 4\epsilon) \log n/n$ is also $O(n^{-\epsilon})$. \square

We devote the rest of the section to the proof of Proposition 4.2.

Proof of Proposition 4.2. Let ρ denote the diameter of K . The *width* $w(R)$ (respectively, *height* $h(R)$) of a rectangle R is the length of its short (respectively, long) sides. (If R is a square, then $w(R) = h(R)$). We recall that the *minor axis* of a rectangle is the line that passes through the center of each long side. We say that the *inclination* of a rectangle is the angle in $[0, \pi)$ that the intersection of its minor axis with the upper halfplane makes with the x -axis. (If the minor axis is parallel to the x -axis, we let its inclination be 0). Finally, let $\theta_0 := \epsilon(2 + 4\epsilon) \log n / (4\rho^2 n)$.

Proposition 4.2 is an immediate consequence of Claims A and B below.

Claim A. *Every rectangle of area $(2+4\epsilon) \log n/n$ contained in K contains a rectangle of area $(2+2\epsilon) \log n/n$ whose inclination is $t \cdot \theta_0$ for some integer $t \in [0, \pi/\theta_0)$.*

Claim B. *There is a family \mathcal{R} of rectangles with the following properties: (i) each $R \in \mathcal{R}$ has area $(2+\epsilon) \log n/n$; (ii) $|\mathcal{R}| = O(n^2)$; and (iii) every rectangle contained in K of area $(2+2\epsilon) \log n/n$, whose inclination is $t\theta_0$ for some integer $t \in [0, \pi/\theta_0)$, contains a rectangle in \mathcal{R} .*

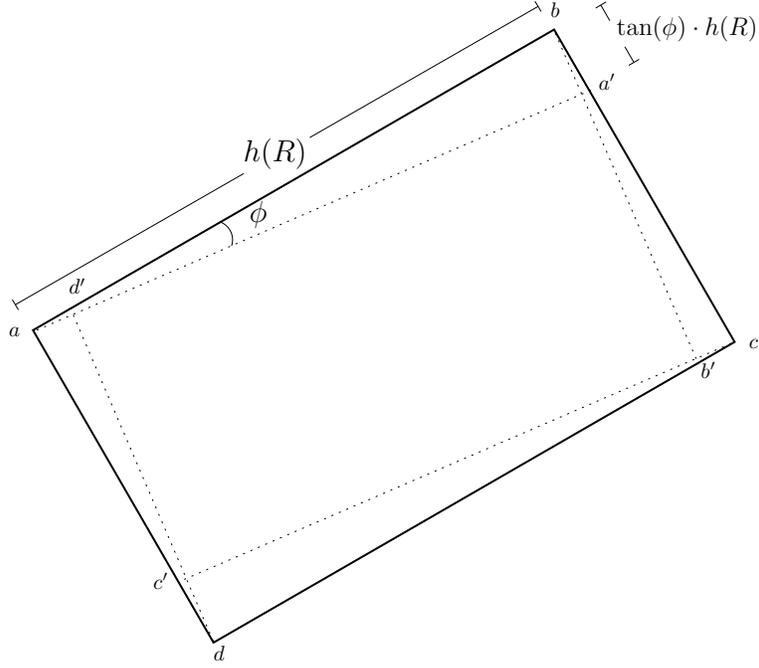
Proof of Claim A. Let R be a rectangle of area $(2+4\epsilon) \log n/n$ contained in K . Let a, b, c, d be the vertices of R , in the clockwise cyclic order in which they appear as we traverse the rectangle, so that ab and cd are the long sides of R . See Figure 1. Let $\ell_{ab}, \ell_{bc}, \ell_{cd}, \ell_{da}$ be the lines that span the sides ab, bc, cd , and da , respectively. We rotate ℓ_{ab} clockwise around a until (for the first time) a line perpendicular to the rotating line reaches an inclination of $t\theta_0$, for some integer $t \in [0, \pi/\theta_0)$. Let $\phi \leq \theta_0$ denote the angle that the line ℓ_{ab} got rotated, and let a' be the point in which the rotated line intersects the side bc . Proceed similarly with the lines ℓ_{bc}, ℓ_{cd} , and ℓ_{da} , to define points b', c' , and d' , respectively. We refer the reader again to Figure 1.

The points in which the segments aa', bb', cc', dd' intersect (each of aa' and cc' intersects each of bb' and dd') define a rectangle R' inscribed in R . Note that the inclination of R' satisfies the condition in Claim A. It remains to show that $a(R') \geq (2+2\epsilon) \log n/n$.

We start by noting that $a(R')$ is clearly at least $a(R)$ minus the sums of the areas of the triangles $aa'b, bb'c, cc'd$, and $dd'a$. Since $a(aa'b) = a(cc'd) \geq a(bb'c) = a(dd'a)$, it follows that $a(R') \geq a(R) - 4a(aa'b)$. Let $|a'b|$ denote the length of the segment $a'b$. Then $|a'b| = \tan(\phi) \cdot h(R) < \tan(\theta_0) \cdot h(R) < \theta_0 \cdot h(R) = (\epsilon(2+4\epsilon) \log n / (4\rho^2 n)) h(R)$. Thus $a(aa'b) = |a'b|h(R)/2 \leq (\epsilon(2+4\epsilon) \log n / (8\rho^2 n)) \cdot (h(R))^2$.

Note that since $R \subseteq K$ and $a(R) = (2+4\epsilon) \log n/n$, then the aspect ratio $w(R)/h(R)$ attains its minimum value when $h(R) = \rho$ (and so $w(R) = (2+4\epsilon) \log n / (\rho n)$). Thus $(2+4\epsilon) \log n / (\rho^2 n) \leq w(R)/h(R)$. Therefore $a(aa'b) \leq \epsilon \cdot w(R)h(R)/8 = (\epsilon/8)a(R)$. Since $a(R') \geq a(R) - 4a(aa'b)$, then $a(R') \geq a(R)(1 - \epsilon/2) = (2+4\epsilon)(1 - \epsilon/2) \log n/n$. Since this last expression is greater than $(2+2\epsilon) \log n/n$ for all sufficiently small ϵ , the claim follows. \square

Proof of Claim B. Let $\gamma := \left(\frac{2+2\epsilon}{2+\epsilon}\right)^{1/3}$. Let $w_0 := (2+2\epsilon) \log n / (\rho \cdot n)$. Note that, since ρ is the diameter of K , then w_0 is the smallest possible width of a rectangle of area $(2+2\epsilon) \log n/n$ contained in K . On the other hand, the largest possible width of a rectangle of area $(2+2\epsilon) \log n/n$ contained in K is $\sqrt{(2+2\epsilon) \log n/n}$. Let M be the smallest integer m such that $\gamma^m w_0 \geq \sqrt{(2+2\epsilon) \log n/n}$. We note for future

FIGURE 1. Finding the rectangle R' inscribed in R (proof of Claim A).

reference that since $\gamma^{M-1}w_0 < \sqrt{(2+2\epsilon)\log n/n}$, a simple calculation shows that $M = O(\log n)$.

For the rest of the proof we regard points in the plane as column vectors.

Let $\Delta_x^m := \gamma^m(\gamma-1)w_0/2$ and $\Delta_y^m := \frac{\rho(\gamma-1)}{2\gamma^{m+3}}$. For each integer $t \in [0, \pi/\theta_0]$ we let A^t be the rotation matrix $\begin{pmatrix} \cos(t\theta_0) & -\sin(t\theta_0) \\ \sin(t\theta_0) & \cos(t\theta_0) \end{pmatrix}$.

Now for each $-1 \leq m \leq M-2$, and each integer $t \in [0, \pi/\theta_0]$, we define the set (or *grid*)

$$G^{m,t} := \left\{ A^t \begin{pmatrix} i\Delta_x^m \\ j\Delta_y^m \end{pmatrix} \mid i, j \in \mathbb{Z} \right\} \cap K,$$

and let $\mathcal{R}^{m,t}$ be the set of those rectangles R contained in K such that: (i) the center of R is in $G^{m,t}$; (ii) the inclination of R is $t\theta_0$; (iii) $w(R) = \gamma^m w_0$; (iv) $h(R) = \rho/\gamma^{m+3}$.

Now define

$$\mathcal{R} := \bigcup_{m,t} \mathcal{R}^{m,t},$$

where the union is over all $m \in \{-1, 0, \dots, M-2\}$ and all integers $t \in [0, \pi/\theta_0]$.

We claim that \mathcal{R} satisfies the properties in Claim B.

Property (i) is trivial: every $R \in \mathcal{R}$ has area $w_0\rho/\gamma^3 = (2+\epsilon)\log n/n$.

To prove (ii), in order to bound $|\mathcal{R}|$ we first estimate $|\mathcal{R}^{m,t}|$ for any two fixed integers $m \in \{-1, 0, \dots, M-2\}$, $t \in [0, \pi/\theta_0]$. Let $H^{m,t}$ be the set of points in $G^{m,t}$ that are centers of rectangles in $\mathcal{R}^{m,t}$. Note that $|H^{m,t}| = |\mathcal{R}^{m,t}|$.

Now consider the map g that sends each point $A^t \begin{pmatrix} x \\ y \end{pmatrix}$ in $H^{m,t}$ to the *grid rectangle* with vertices $A^t \begin{pmatrix} x \\ y \end{pmatrix}$, $A^t \begin{pmatrix} x + \Delta_x^m \\ y \end{pmatrix}$, $A^t \begin{pmatrix} x \\ y + \Delta_y^m \end{pmatrix}$, and $A^t \begin{pmatrix} x + \Delta_x^m \\ y + \Delta_y^m \end{pmatrix}$. Since every $A^t \begin{pmatrix} x \\ y \end{pmatrix} \in H^{m,t}$ is the center of a (much larger) rectangle (in \mathcal{R}) contained in K , it follows that the grid rectangle $g(A^t \begin{pmatrix} x \\ y \end{pmatrix})$ is also contained in K .

Moreover, the interiors of any two distinct such rectangles that are images of g are disjoint. Each grid rectangle has area $\Delta_x^m \Delta_y^m$. It follows that the number of grid rectangles that are images under g (and thus also $|H^{m,t}|$ and $|R^{m,t}|$) is at most $a(K)/(\Delta_x^m \Delta_y^m) = 1/(\Delta_x^m \Delta_y^m) = (4\gamma^3)/((\gamma-1)^2 \rho w_0) = 4\gamma^3 n / ((2+2\epsilon)(\gamma-1)^2 \log n)$.

Thus for every $m \in \{-1, 0, \dots, M-2\}$ and every integer $t \in [0, \pi/\theta_0)$ we have $|R^{m,t}| = O(n/\log n)$. Since $M = O(\log n)$ and $\pi/\theta_0 = (4\pi\rho^2 n)/(\epsilon(2+4\epsilon)\log n) = O(n/\log n)$, then $|\mathcal{R}| = (M \cdot \lceil \pi/\theta_0 \rceil) |\mathcal{R}_{m,t}| = O(n^2)$. This proves (ii).

To prove (iii), let Q be a rectangle contained in K , of area $(2+2\epsilon)\log n/n$, whose inclination is $t\theta_0$ for some integer $t \in [0, \pi/\theta_0)$. Since w_0 is the smallest possible width of Q , and $\sqrt{(2+2\epsilon)\log n/n} \leq \gamma^M w_0$ is the largest possible width of Q , it follows that there is an integer m in $\{-1, 0, \dots, M-2\}$ such that $w(Q) \in [\gamma^{m+1} w_0, \gamma^{m+2} w_0]$. Since $a(Q) = w(Q)h(Q) = \rho w_0$, it follows that $h(Q) \in [\frac{\rho}{\gamma^{m+2}}, \frac{\rho}{\gamma^{m+1}}]$.

Now let (x, y) be the point in $G^{m,t}$ that is closest to the center of Q , and let R be the rectangle in $\mathcal{R}^{m,t}$ centered at (x, y) . Since $w(Q) \geq \gamma^{m+1} w_0$, it follows that the distance from (x, y) to each long side of Q is at least $\gamma^{m+1} w_0/2 - \Delta_x^m = \gamma^m w_0/2$. Similarly, since $h(Q) \geq \rho/\gamma^{m+2}$, it follows that the distance from (x, y) to each short side of Q is at least $\rho/2\gamma^{m+2} - \Delta_y^m = \rho/2\gamma^{m+3}$. Since $w(R) = \gamma^m w_0$ and $h(R) = \rho/\gamma^{m+3}$, it follows that $R \subseteq Q$. \square

\square

5. PROOF OF THEOREM 1.3

Proof. It suffices to prove the lower bound, since the upper bound for $\text{POLYMAX}(K_n)$ follows at once from Theorem 1.2.

We prove the lower bound for the case in which K is the unit square. As we explain at the end of this proof, the ideas in the proof carry over in a straightforward way to the general case. We prefer to focus on the case in which K is a square since the main ideas will not be hidden behind the necessarily more technical details required in the general case.

We will prove the following. Let $\delta, \epsilon > 0$. Then w.h.p. there is an empty convex quadrilateral of area at least $(1-2\delta)(1-\epsilon)\log n/n$. This clearly implies the required lower bound.

Let $t := 1/((1-\epsilon)\log n/n)$. For simplicity we assume that t is an integer. We partition K into t consecutive rectangles (or *strips*) s_1, s_2, \dots, s_t , each of width n/t and height 1. It follows from Proposition 2.2 that the probability that there are fewer than $n^\epsilon/(2(1-\epsilon)\log n)$ empty strips is at most $\frac{4(1-\epsilon)\log(n)}{n^\epsilon}$.

We now estimate the probability that there exist two consecutive strips that are empty. The combined area of any two consecutive strips is $2(1-\epsilon)\log n/n$, and so the probability that it is empty is $\sim n^{-2-2\epsilon}$. There are $t-1$ pairs of consecutive

strips, and so it follows from the union bound that the probability that one such consecutive pair is empty is smaller than $t \cdot n^{-2-2\epsilon} = n^{-1-\epsilon}/((1-\epsilon)\log n)$.

Let e_1, e_2, \dots, e_p denote the empty strips, labeled so that e_{i+1} is to the right of e_i for $i = 1, \dots, p-1$. Suppose for simplicity that $p = 4q + 2$ for some integer q . Now for $j = 1, 2, \dots, q$, let l_1^j, l_2^j (respectively, r_1^j, r_2^j) be the points with the largest (respectively, smallest) x -coordinates that are to the left (respectively, to the right) of the strip e_{4j} .

Let us call A the event that for all $j \neq k$, the sets $\{l_1^j, l_2^j, r_1^j, r_2^j\}$ and $\{l_1^k, l_2^k, r_1^k, r_2^k\}$ are disjoint (the probability that A does *not* occur is at most the probability that one consecutive pair of strips is empty, that is, smaller than $n^{-1-\epsilon}/((1-\epsilon)\log n)$). Since the y -coordinates of the points in K_n are independent of their x -coordinates, it follows that for each $j = 1, \dots, q$, if A occurs then the following occurs with probability δ^4 : the points r_1^j and l_1^j have y -coordinates larger than $1 - \delta$, and the points r_2^j and l_2^j have y -coordinates smaller than δ . Therefore for each such j , with probability δ^4 the points l_1^j, l_2^j, r_1^j , and r_2^j form an empty convex quadrilateral of area at least $(1 - 2\delta) \cdot (1 - \epsilon) \log n/n$.

Thus (i) w.h.p. the number p of empty strips is at least $n^\epsilon/(2(1-\epsilon)\log n)$, and so w.h.p. $q = (p-2)/4$ is $\Omega(n^\epsilon/\log n)$; (ii) w.h.p. the event A occurs; and (iii) if A occurs, then for each $j = 1, 2, \dots, q$, with probability δ^4 the points l_1^j, l_2^j, r_1^j , and r_2^j form an empty convex quadrilateral of area at least $(1 - 2\delta) \cdot (1 - \epsilon) \log n/n$. Clearly (i), (ii), and (iii) combine to prove that w.h.p. there is an empty convex quadrilateral of area at least $(1 - 2\delta) \cdot (1 - \epsilon) \log n/n$, as claimed.

For the general case, very few adaptations of substance are needed. For instance, one may start by approximating K with a convex set T contained in K whose boundary is smooth, and whose area is arbitrarily close to the area of K (the smoothness of the boundary of T is not necessary, but it simplifies somewhat the ensuing discussion). Then, as in the unit square case, one partitions T into t vertical strips of equal area (thus each vertical strip is bounded by two vertical segments and by two pieces of the boundary of T). All the arguments from the unit square carry over so far to this case: the estimates for the number of empty strips and the probability that two of them are consecutive are the same. The only technical complication arises when one needs to do the equivalent step of choosing points r_1^j and l_1^j with large y -coordinates (and points r_2^j and l_2^j with small y -coordinates): in this general case, one needs to define, for each empty strip, small regions at the top and at the bottom of its neighboring strips (each of these small regions must have area $\delta \cdot (1 - \epsilon) \log n/n$, as in the unit square case). Taking care rigorously of the details is of course not a deep difficulty (here is where the smoothness of the boundary of T comes handy), but the simplicity of the basic ideas seems much more apparent by focusing on the unit square case. \square

6. CONCLUDING REMARKS

We conjecture that in both Theorems 1.2 and 1.3 the upper bounds should match the lower bounds $(1 + o(1)) \log n/n$, as in Theorem 1.1.

It is worth noting that in none of the main theorems we make full use of the convexity assumption for K and L . Actually, all these results hold as long as K is (for instance) a finite union of convex sets. The convexity of L is only required so that we can invoke Proposition 2.3, but for any fixed (not necessarily convex)

bounded L one can obviously find inscribed and circumscribed rectangles that approximate its area within constant factors. Thus the proofs of the main theorems can be adapted to prove the following:

Theorem 6.1. *Let K, L be sets in the plane, where K is the finite union of convex sets and L is bounded. Suppose for normalization purposes that the area of K is 1. Let K_n be a set of n points chosen independently and uniformly at random from K . Let $\text{MAX}_L(K_n)$ denote the random variable that measures the largest (in terms of area) hole that is homothetic to L ; let $\text{MAX}(K_n)$ denote the random variable that measures the largest (in terms of area) convex hole; and let $\text{POLYMAX}(K_n)$ denote the random variable that measures the largest (in terms of area) convex polygon with vertices in K_n . Then w.h.p.*

$$\begin{aligned}\text{MAX}_L(K_n) &= \Theta\left(\frac{\log n}{n}\right), \\ \text{MAX}(K_n) &= \Theta\left(\frac{\log n}{n}\right), \quad \text{and} \\ \text{POLYMAX}(K_n) &= \Theta\left(\frac{\log n}{n}\right).\end{aligned}$$

We also note that all these results and their proofs carry over easily to the d -dimensional case, although the multiplicative constants one obtains depend on d .

ACKNOWLEDGMENTS

We thank Gerardo Arizmendi and Edgardo Ugalde for stimulating discussions.

REFERENCES

- [1] A. Aggarwal and S. Suri, Fast algorithms for computing the largest empty rectangle. In: *Proceedings of the 3rd Annual Symposium on Computational Geometry* (1987), pp. 278–290.
- [2] J. Balogh, H. González-Aguilar, and G. Salazar, Large convex holes in random point sets. *Comput. Geom.* **46** (2013), 725–733.
- [3] I. Bárány, Z. Füredi, Empty simplices in Euclidean space, *Canad. Math. Bull.* **30** (1987), 436–445.
- [4] I. Bárány, P. Valtr, Planar point sets with a small number of empty convex polygons, *Studia. Sci. Math. Hungar.* **41** (2004), 243–266.
- [5] A. Dumitrescu, Planar sets with few empty convex polygons, *Studia. Sci. Math. Hungar.* **36** (2000), 93–109.
- [6] A. Dumitrescu and M. Jiang, Maximal Empty Boxes Amidst Random Points. *Lecture Notes in Computer Science* **7408** (2012), pp. 529–540.
- [7] A. Dumitrescu and M. Jiang, On the Largest Empty Axis-Parallel Box Amidst n Points. *Algorithmica* **66** (2013), 225–248.
- [8] P. Erdős and G. Szekeres, A combinatorial problem in geometry. *Compositio Math.* **2** (1935), 463–470.
- [9] R. Fabila-Monroy, C. Huemer, and D. Mitsche, Empty non-convex and convex four-gons. *Studia Scientiarum Mathematicarum Hungarica* **52** (2015), 52–64.
- [10] C. Hervías, N. Hitschfeld-Kahler, L.E. Campusano, and G. Font, On Finding Large Polygonal Voids Using Delaunay Triangulation: The Case of Planar Point Sets. *Proceedings of the 22nd International Meshing Roundtable 2014*, pp. 275–292. Springer.
- [11] M. Lassak, Approximation of convex bodies by rectangles. *Geometriae Dedicata* **47**(1993), 111–117.
- [12] P. Valtr, Probability that n Random Points are in Convex Position. *Discrete and Computational Geometry* **13** (1995), 637–643.

- [13] P. Valtr, The Probability that n Random Points in a Triangle Are in Convex Position. *Combinatorica* **16** (1996), 567–573.
- [14] P. Valtr, On the minimum number of empty polygons in planar point sets, *Studia. Sci. Math. Hungar.* **30** (1995) 155–163.