LARGE CONVEX HOLES IN RANDOM POINT SETS

JÓZSEF BALOGH, HERNÁN GONZÁLEZ-AGUILAR, AND GELASIO SALAZAR

Abstract. A *convex hole* (or *empty convex polygon)* of a point set P in the plane is a convex polygon with vertices in P, containing no points of P in its interior. Let R be a bounded convex region in the plane. We show that the expected number of vertices of the largest convex hole of a set of n random points chosen independently and uniformly over R is $\Theta(\log n/(\log\log n))$, regardless of the shape of R.

1. Introduction

Let P be a set of points in the plane. A *convex hole* (alternatively, *empty convex polygon*) of P is a convex polygon with vertices in P, containing no points of P in its interior.

Questions about (empty or nonempty) convex polygons in point sets are of fundamental importance in discrete and computational geometry. A landmark in this area is the question posed by Erdős and Szekeres in 1935 [10]: "What is the smallest integer $f(k)$ such that any set of $f(k)$ points in the plane contains at least one convex k-gon?".

A variant later proposed by Erdős himself asks for the existence of empty convex polygons [11]: "Determine the smallest positive integer $H(n)$, if it exists, such that any set X of at least $H(n)$ points in general position in the plane contains n points which are the vertices of an empty convex polygon, i.e., a polygon whose interior does not contain any point of X." It is easy to show that $H(3) = 3$ and $H(4) = 5$. Harborth [13] proved that $H(5) = 10$. Much more recently, Nicolás [17] and independently Gerken [12] proved that every sufficiently large point set contains an empty convex hexagon (see also [24]). It is currently known that $30 \leq H(6) \leq 463$ [15, 18]. A celebrated construction of Horton [14] shows that for each $n \geq 7$, $H(n)$ does not exist. For further results and references around Erdős-Szekeres type problems, we refer the reader to the surveys [1, 16] and to the monography [9].

We are interested in the expected size of convex structures in random point sets. This gives rise to a combination of Erdős-Szekeres type problems with variants of Sylvester's seminal question [21]: "What is the probability

Date: November 5, 2012.
2010 *Mathematics Subject Classification.* 52C10, 60D05, 52A22, 52C05, 52A10.
Key words and phrases. Convex hole, Erdős-Szekeres Theorem, random point set.
The first author was supported by NSF CAREER Grant DMS-0745185.
The third author was supported by CONACYT grant 106432.

that four random points chosen independently and uniformly from a convex region form a convex quadrilateral?"

Several fundamental questions have been attacked (and solved) in this direction; see for instance [4, 6, 8]. Particularly relevant to our work are the results of Valtr, who computed exactly the probability that n random points independently and uniformly chosen from a parallelogram [22] or a triangle [23] are in convex position.

Consider a bounded convex region R, and randomly choose n points independently and uniformly over R. We are interested in estimating the *size* (that is, number of vertices) of the largest convex hole of such a randomly generated point set. We will show that the expected size of the largest convex hole is $\Theta(\log n/\log\log n)$. Moreover, we will show that a.a.s. (that is, with probability $1-o(1)$) this size is $\Theta(\log n/\log\log n)$.

The workhorse for our proof for arbitrary regions R is the following statement, which takes care of the particular case in which R is a rectangle.

Theorem 1. *Let R be a rectangle in the plane. Let R_n be a set of n points chosen independently and uniformly at random from R, and let $\textsc{Hol}(R_n)$ denote the random variable that measures the number of vertices of the largest convex hole in R_n. Then*

$$\mathbb{E}(\textsc{Hol}(R_n)) = \Theta\left(\frac{\log n}{\log\log n}\right).$$

Moreover, a.a.s.

$$\textsc{Hol}(R_n) = \Theta\left(\frac{\log n}{\log\log n}\right).$$

Some related questions are heavily dependent on the shape of R. For instance, the expected number of vertices in the convex hull of a random point set, which is $\Theta(\log n)$ if R is the interior of a polygon, and $\Theta(n^{1/3})$ if R is the interior of a convex figure with a smooth boundary (such as a disk) [19, 20]. In the problem under consideration, it turns out that the order of magnitude of the expected number of vertices of the largest convex hole is independent of the shape of R:

Theorem 2. *There exist absolute constants c, c' with the following property. Let R and S be bounded convex regions in the plane. Let R_n (respectively, S_n) be a set of n points chosen independently and uniformly at random from R (respectively, S). Let $\textsc{Hol}(R_n)$ (respectively, $\textsc{Hol}(S_n)$) denote the random variable that measures the number of vertices of the largest convex hole in R_n (respectively, S_n). Then, for all sufficiently large n,*

$$c \le \frac{\mathbb{E}(\textsc{Hol}(R_n))}{\mathbb{E}(\textsc{Hol}(S_n))} \le c'.$$

Moreover, there exist absolute constants d, d' such that a.a.s.

$$d \le \frac{\textsc{Hol}(R_n)}{\textsc{Hol}(S_n)} \le d'.$$

We remark that Theorem 2 is in line with the following result proved by Bárány and Füredi [3]: the expected number of empty simplices in a set of n points chosen uniformly and independently at random from a convex set A with non-empty interior in $\mathbb{R}^d$ is $\Theta(n^d)$, regardless of the shape of A.

Using Theorems 1 and 2, we have determined the expected number of vertices of a largest convex hole, in any bounded convex region R, up to a constant multiplicative factor. Indeed, the following is an immediate consequence of these two results:

Theorem 3. *Let R be a bounded convex region in the plane. Let R_n be a set of n points chosen independently and uniformly at random from R, and let $\textsc{Hol}(R_n)$ denote the random variable that measures the number of vertices of the largest convex hole in R_n. Then*

$$\mathbb{E}(\textsc{Hol}(R_n)) = \Theta\left(\frac{\log n}{\log\log n}\right).$$

Moreover, a.a.s.

$$\textsc{Hol}(R_n) = \Theta\left(\frac{\log n}{\log\log n}\right). \qquad \square$$

For the proof of Theorem 1 (Section 3), in both the lower and upper bounds we use powerful results of Valtr, who computed precisely the probability that n points chosen at random (from a triangle [22] or from a parallelogram [23]) are in convex position. The proof of the lower bound is quite simple: we partition a unit area square R (as we shall note, establishing Theorem 1 for a square implies it for any rectangle) into n/t rectangles such that each of them contains exactly t points, where $t = \frac{\log n}{2\log\log n}$. Using [22], a.a.s. in at least one of the regions the points are in convex position, forming a convex hole. The proof of the upper bound is more involved. We put an n by n lattice in the unit square. The first key idea is that any sufficiently large convex hole H can be well-approximated with *lattice* quadrilaterals Q_0, Q_1 (that is, their vertices are lattice points) such that $Q_0 \subseteq H \subseteq Q_1$. The results about approximation by lattice quadrilaterals are obtained in Section 2. The key advantage of using lattice quadrilaterals is that there are only polynomially many choices (i.e., $O(n^8)$) for each of Q_0 and Q_1. Since H is a hole, then Q_0 contains no point of R_n in its interior. This helps to estimate the area $\mathrm{a}(Q_0)$ of Q_0, and at the same time $\mathrm{a}(H)$ and $\mathrm{a}(Q_1)$ (Claims B,C, and D in Section 3). This upper bound on $\mathrm{a}(Q_1)$ gives that a.a.s. Q_1 contains at most $O(\log n)$ points of R_n. Conditioning that each choice of Q_1 contains at most $O(\log n)$ points, using Valtr [23] (dividing the (≤ 8)-gon $Q_1 \cap R$ into at most eight triangles) we prove that a.a.s. Q_1 does not contain $40\log n/(\log\log n)$ points in convex position (Claim F in Section 3), so a.a.s. there is no hole of that size. A slight complication is that Q_1 may not lie entirely in R; this issue makes the proof somewhat more technical.

Theorem 2 is proved in Section 4. For the proof we use Theorem 1, as well as the approximation results (to convex sets by rectangles) in Section 2. Section 5 contains some concluding remarks.

We make a few final remarks before we move on to the proofs. As we did above, for the rest of the paper we let $\mathrm{a}(U)$ denote the area of a region U in the plane. We also note that, throughout the paper, by $\log x$ we mean the natural logarithm of x. Finally, since we only consider sets of points chosen independently and uniformly at random from a region, for brevity we simply say that such point sets are chosen at random from this region.

2. Approximating convex sets with lattice quadrilaterals

We recall that a rectangle is *isothetic* if each of its sides is parallel to either the x-axis or the y-axis.

In Section 3 we prove Theorem 1 for the case when R is an isothetic unit area square (Theorem 1 then will follow for any rectangle, since whenever Q, Q' are obtained from each other by affine transformations it follows that $\textsc{Hol}(Q_n) = \textsc{Hol}(Q'_n)$). In the proof of the upper bound, we subdivide R into an n by n grid (which defines an $n+1$ by $n+1$ lattice), pick a largest convex hole H, and find lattice quadrilaterals Q_0, Q_1 such that $Q_0 \subseteq H \subseteq Q_1$, whose areas are not too different from the area of H. The caveat is that the circumscribed quadrilateral Q_1 may not fit completely into R; for this reason, we need to extend this grid of area 1 to a grid of area 9 (that is, to extend the $n+1$ by $n+1$ lattice to a $3n+1$ by $3n+1$ lattice).

Proposition 4. *Let R (respectively, S) be the isothetic square of side length* 1 *(respectively,* 3*) centered at the origin. Let $n > 1000$ be a positive integer, and let $\mathcal{L}$ be the* lattice $\{(-3/2 + i/3n, -3/2 + j/3n) \in \mathbb{R}^2 \mid i, j \in \{0, 1, \ldots, 9n\}\}$. *Let $H \subseteq R$ be a closed convex set. Then there exists a* lattice *quadrilateral (that is, a quadrilateral each of whose vertices is a lattice point) Q_1 such that $H \subseteq Q_1$ and $\mathrm{a}(Q_1) \leq 2\,\mathrm{a}(H) + 40/n$. Moreover, if $\mathrm{a}(H) \geq 1000/n$, then there also exists a lattice quadrilateral Q_0 such that $Q_0 \subseteq H$ and $\mathrm{a}(Q_0) \geq \mathrm{a}(H)/32$.*

We remark that some lower bound on the area of H is needed in order to guarantee the existence of a lattice quadrilateral contained in H, as obviously there exist small convex sets that contain no lattice points (let alone lattice quadrilaterals).

Proof. If p, q are points in the plane, we let $\overline{pq}$ denote the closed straight segment that joins them, and by $|\overline{pq}|$ the length of this segment (that is, the distance between p and q). We recall that if C is a convex set, the *diameter* of C is $\sup\{|\overline{xy}| : x, y \in C\}$. We also recall that a *supporting line* of C is a line that intersects the boundary of C and such that all points of C are in the same closed half-plane of the line.

Existence of Q_1

Let a, b be a diametral pair of H, that is, a pair of points in H such that $|\overline{ab}|$ equals the diameter of H (a diametral pair exists because H is closed). Now let ℓ, ℓ' be the supporting lines of H parallel to $\overline{ab}$.

Let ℓ_a, ℓ_b be the lines perpendicular to $\overline{ab}$ that go through a and b, respectively. Since a, b is a diametral pair, it follows that a (respectively, b) is the only point of H that lies on ℓ_a (respectively, ℓ_b). See Figure 1.

Let c, d be points of H that lie on ℓ and ℓ', respectively. Let J be the quadrilateral with vertices a, c, b, d. By interchanging ℓ and ℓ' if necessary, we may assume that a, c, b, d occur in this clockwise cyclic order in the boundary of J.

Let K denote the rectangle bounded by ℓ_a, ℓ, ℓ_b, and ℓ'. Let w, x, y, z be the vertices of K, labelled so that a, w, c, x, b, y, d, z occur in the boundary of K in this clockwise cyclic order. It follows that $\mathrm{a}(K) = 2\mathrm{a}(J)$. Since $\mathrm{a}(H) \geq \mathrm{a}(J)$, we obtain $\mathrm{a}(K) \leq 2\mathrm{a}(H)$. Let T denote the isothetic square of length side 2, also centered at the origin. It is easy to check that since $H \subseteq R$, then $K \subseteq T$.

Let Q_x be the square with side length $2/n$ that has x as one of its vertices, with each side parallel to ℓ or to ℓ_a, and that only intersects K at x. It is easy to see that these conditions define uniquely Q_x. Let x' be the vertex of Q_x opposite to x. Define Q_y, Q_z, Q_w, y', z', and w' analogously.

Since $K \subseteq T$, it follows that Q_x, Q_y, Q_z, and Q_w are all contained in S. Using this, and the fact that there is a circle of diameter $2/n$ contained in Q_x, it follows that there is a lattice point g_x contained in the interior of Q_x. Similarly, there exist lattice points g_y, g_z, and g_w contained in the interior of Q_y, Q_z, and Q_w, respectively. Let Q_1 be the quadrilateral with vertices g_x, g_y, g_z, and g_w.

Let $\mathrm{per}(K)$ denote the perimeter of K. The area of the rectangle K' with vertices w', x', y', z' (see Figure 1) is $\mathrm{a}(K) + \mathrm{per}(K)(2/n) + 4(2/n)^2$. Since the perimeter of any rectangle contained in S is at most 12, then $\mathrm{a}(K') \leq \mathrm{a}(K) + 24/n + 16/n^2 \leq \mathrm{a}(K) + 40/n$. Since $\mathrm{a}(Q_1) \leq \mathrm{a}(K')$, we obtain $\mathrm{a}(Q_1) \leq \mathrm{a}(K) + 40/n \leq 2\,\mathrm{a}(H) + 40/n$.

Existence of Q_0

Suppose without any loss of generality (relabel if needed) that the area of the triangle $\Delta := abd$ is at least the area of the triangle abc. Since $2\mathrm{a}(J) = \mathrm{a}(K) \geq \mathrm{a}(H)$ and $\mathrm{a}(\Delta) \geq \mathrm{a}(J)/2$, we have $\mathrm{a}(\Delta) \geq \mathrm{a}(H)/4$. By hypothesis $\mathrm{a}(H) \geq 1000/n$, and so $\mathrm{a}(\Delta) \geq 1000/(4n)$.

Since a, b is a diametral pair, it follows that the longest side of Δ is $\overline{ab}$. Let e be the intersection point of $\overline{ab}$ with the line perpendicular to $\overline{ab}$ that passes through d. Thus $\mathrm{a}(\Delta) = |\overline{ab}||\overline{de}|/2$. See Figure 1.

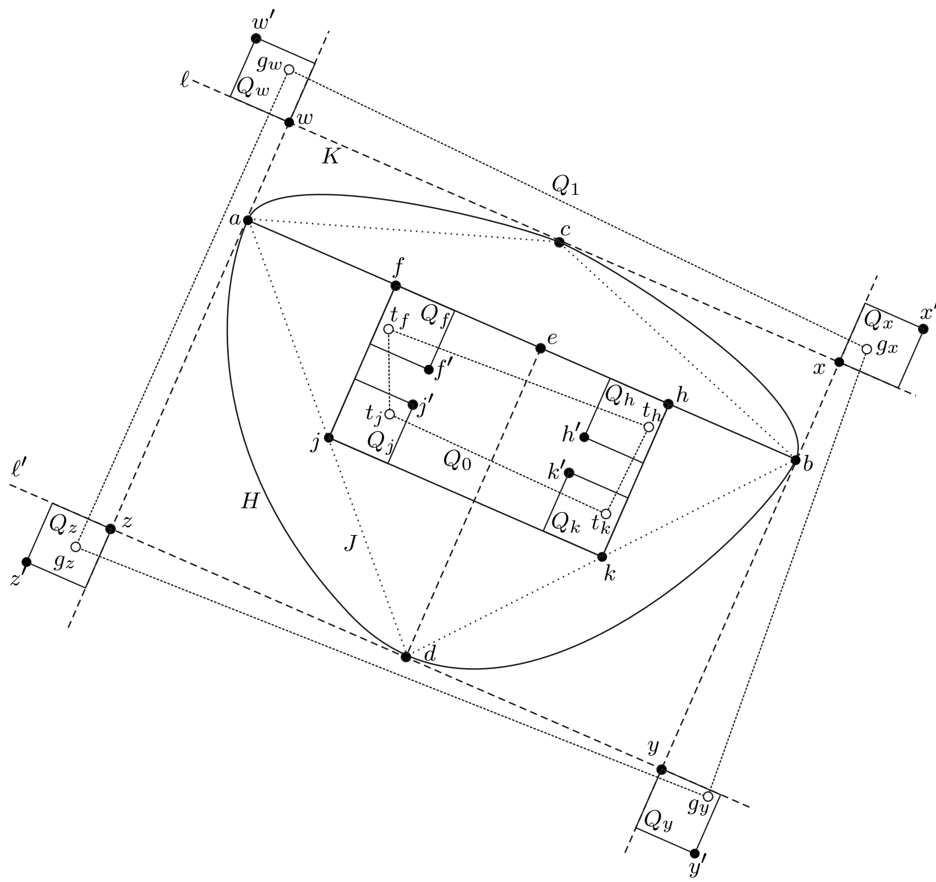

FIGURE 1. Lattice quadrilateral Q_1 has vertices g_w, g_x, g_y, g_z, and lattice quadrilateral Q_0 has vertices t_f, t_h, t_j, t_k.

There exists a rectangle U, with base contained in $\overline{ab}$, whose other side has length $|\overline{de}|/2$, and such that $\mathrm{a}(U) = \mathrm{a}(\Delta)/2$. Let f, h, j, k denote the vertices of this rectangle, labelled so that f and h lie on $\overline{ab}$ (with f closer to a than h), j lies on $\overline{ad}$, and k lies on $\overline{bd}$. Thus $|\overline{fj}| = |\overline{de}|/2$.

Now $|\overline{ab}| < 2$ (indeed, $|\overline{ab}| \leq \sqrt{2}$, since a, b are both in R), and since $|\overline{ab}||\overline{de}|/2 = \mathrm{a}(\Delta) \geq 1000/(4n)$ it follows that $|\overline{de}| \geq 1000/(4n)$. Thus $|\overline{fj}| \geq 1000/(8n)$.

Now since a, b is a diametral pair it follows that $\mathrm{a}(K) \leq |\overline{ab}|^2$. Using $1000/n \leq \mathrm{a}(H) \leq \mathrm{a}(K)$, we obtain $1000/n \leq |\overline{ab}|^2$. Note that $|\overline{fh}| = |\overline{ab}|/2$. Thus $1000/(4n) \leq |\overline{fh}|^2$. Using $|\overline{fh}| = |\overline{ab}|/2$ and $|\overline{ab}| \leq \sqrt{2}$, we obtain $|\overline{fh}| < 1$, and so $|\overline{fh}| > |\overline{fh}|^2 \geq 1000/(4n)$.

Now let Q_f be the square with sides of length $2/n$, contained in U, with sides parallel to the sides of U, and that has f as one of its vertices. Let f' denote the vertex of Q_f that is opposite to f. Define similarly Q_h, Q_j, Q_k, h', j', and k'.

Since $|\overline{fj}|$ and $|\overline{fh}|$ are both at least $1000/(8n)$, it follows that the squares Q_f, Q_h, Q_j, Q_k are pairwise disjoint. Since the sides of these squares are all $2/n$, it follows that (i) each of these squares contains at least one lattice point; and (ii) each square in $\{Q_f, Q_h, Q_j, Q_k\}$ has empty intersection with the convex hull of the other three squares.

By (i), there exist lattice points t_f, t_h, t_j, t_k contained in Q_f, Q_h, Q_j and Q_k, respectively. By (ii), it follows that t_f, t_h, t_j, and t_k define a convex quadrilateral Q_0. Let W denote the rectangle with vertices f', h', j', and k'.

Since $|\overline{fj}|$ and $|\overline{fh}|$ are both at least $1000/(8n)$, and the side lenghts of the squares Q are $2/n$, it follows easily that $|\overline{f'h'}| > (1/2)|\overline{fh}|$ and $|\overline{j'k'}| > (1/2)|\overline{jk}|$. Thus $\mathrm{a}(W) > \mathrm{a}(U)/4$. Now clearly $\mathrm{a}(Q_0) \geq \mathrm{a}(W)$. Recalling that $\mathrm{a}(U) = \mathrm{a}(\Delta)/2$, $\mathrm{a}(\Delta) \geq \mathrm{a}(J)/2$, and $\mathrm{a}(J) = \mathrm{a}(K)/2 \geq \mathrm{a}(H)/2$, we obtain $\mathrm{a}(Q_0) \geq \mathrm{a}(U)/4 \geq \mathrm{a}(\Delta)/8 \geq \mathrm{a}(J)/16 \geq \mathrm{a}(H)/32$. □

We finally note that in the previous proof we have shown the existence of rectangles U, K that approximate the convex body H as follows.

Corollary 5. *Let H be a closed convex set. Then there exist rectangles U, K such that $U \subseteq H \subseteq K$, $\mathrm{a}(U) \geq \mathrm{a}(H)/8$, and $\mathrm{a}(K) \leq 2\mathrm{a}(H)$.* □

3. Proof of Theorem 1

We start by noting that if Q, Q' are two regions such that Q' is obtained from Q by an affine transformation, then $\textsc{Hol}(Q_n) = \textsc{Hol}(Q'_n)$. Thus we may assume without loss of generality that R is the isothetic unit area square centered at the origin.

We prove the lower and upper bounds separately. More specifically, we prove that for all sufficiently large n:

$$\mathbb{P}\left(\textsc{Hol}(R_n) \geq \frac{1}{2}\frac{\log n}{\log\log n}\right) \geq 1 - n^{-2}. \tag{1}$$

$$\mathbb{P}\left(\textsc{Hol}(R_n) \leq 40\,\frac{\log n}{\log\log n}\right) \geq 1 - 2n^{-1.9}. \tag{2}$$

We note that (1) and (2) imply immediately the a.a.s. part of Theorem 1. Now the $\Omega(\log n/\log\log n)$ part of the theorem follows from (1), since $\textsc{Hol}(R_n)$ is a non-negative random variable, whereas the $O(\log n/\log\log n)$ part follows from (2), since $\textsc{Hol}(R_n)$ is bounded by above by n.

Thus we complete the proof by showing (1) and (2).

Proof of (1). Let R_n be a set of n points chosen at random from R. We prove that a.a.s. R_n has a convex hole of size at least t, where $t := \frac{\log n}{2\log\log n}$.

Let $k := n/t$. For simplicity, suppose that both t and k are integers. Let $\{\ell_0, \ell_1, \ell_2, \dots, \ell_k\}$ be a set of vertical lines disjoint from R_n, chosen so that for $i = 0, 1, 2, \dots, k-1$, the set R_n^i of points of R_n contained in the rectangle R^i bounded by R, ℓ_i, and ℓ_{i+1} contains exactly t points. Conditioning that R^i contains exactly t points we have that these t points are chosen at random from R^i.

Valtr [22] proved that the probability that r points chosen at random in a parallelogram are in convex position is $\left(\frac{\binom{2r-2}{r-1}}{r!}\right)^2$. Using the bounds $\binom{2s}{s} \geq 4^s/(s+1)$ and $s! \leq es^{s+1/2}e^{-s}$, we obtain that this is at least r^{-2r} for all $r \geq 3$:

$$\left(\frac{\binom{2r-2}{r-1}}{r!}\right)^2 \geq \left(\frac{\frac{4^{r-1}}{(r-1)+1}}{er^r\sqrt{r}e^{-r}}\right)^2 = \frac{(4e)^{2r}}{16e^2r^3} \cdot r^{-2r}.$$

Since each R^i is a rectangle containing t points chosen at random, it follows that for each fixed $i \in \{0, 1, \dots, k-1\}$, the points of R_n^i are in convex position with probability at least t^{-2t}. Since there are $k = n/t$ sets R_n^i, it follows that none of the sets R_n^i is in convex position with probability at most

$$(1 - t^{-2t})^{n/t} \leq e^{-\frac{n}{t}t^{-2t}} = e^{-nt^{-2t-1}}.$$

If the $t = \log n/(2 \log\log n)$ points of an R_n^i are in convex position, then they form a convex hole of R_n. Thus, the probability that there is a convex hole of R_n of size at least $\log n/(2 \log\log n)$ is at least $1 - e^{-nt^{-2t-1}}$. A routine calculation shows that $e^{-nt^{-2t-1}} < n^{-2}$ for all sufficiently large n, and so (1) follows. □

Proof of (2). Let R_n be a set of n points chosen at random from R. We remark that throughout the proof we always implicitly assume that n is sufficiently large.

We shall use the following easy consequence of Chernoff's bound. This is derived immediately, for instance, from Theorem A.1.11 in [2].

Lemma 6. *Let $Y_1, \dots, Y_m$ be mutually independent random variables with $\mathbb{P}(Y_i = 1) = p$ and $\mathbb{P}(Y_i = 0) = 1 - p$, for $i = 1, \dots, m$. Let $Y := Y_1 + \dots + Y_m$. Then*

$$\mathbb{P}\big(Y \geq (3/2)pm\big) < e^{-pm/16}.$$ □

Let S be the isothetic square of area 9, also (as R) centered at the origin.

Since there are $(9n+1)^2$ lattice points, it follows that there are fewer than $(9n)^8$ lattice quadrilaterals in total, and fewer than n^8 lattice quadrilaterals whose four vertices are in R.

Claim A. *With probability at least $1-n^{-10}$ the random point set R_n has the property that every lattice quadrilateral Q with $\mathrm{a}(Q) < 2000 \log n/n$ satisfies that $|R_n \cap Q| \leq 3000 \log n$.*

Proof. Let Q be a lattice quadrilateral with $\mathrm{a}(Q) < 2000\log n/n$, and let $Z = Z(Q) \subseteq R$ be any lattice quadrilateral containing Q, with $\mathrm{a}(Z) = 2000\log n/n$. Let X_Q (respectively, X_Z) denote the random variable that measures the number of points of R_n in Q (respectively, Z). We apply Lemma 6 with $p = \mathrm{a}(Z)$ and $m = n$, to obtain $\mathbb{P}(X_Z \geq 3000\,\log n) < e^{-125\log n} = n^{-125}$. Since $Q \subseteq Z$, it follows that $\mathbb{P}(X_Q \geq 3000\,\log n) < n^{-125}$. As the number of choices for Q is at most $(9n)^8$, with probability at least $(1 - 9n^8 \cdot n^{-125}) > 1 - n^{-10}$ no such Q contains more than $3000\,\log n$ points of R_n. □

A polygon is *empty* if its interior contains no points of R_n.

Claim B. *With probability at least* $1 - n^{-10}$ *the random point set* R_n *has the property that there is no empty lattice quadrilateral* $Q \subseteq R$ *with* $\mathrm{a}(Q) \geq 20\log n/n$.

Proof. The probability that a fixed lattice quadrilateral $Q \subseteq R$ with $\mathrm{a}(Q) \geq 20\log n/n$ is empty is $(1 - \mathrm{a}(Q))^n < n^{-20}$. Since there are fewer than n^8 lattice quadrilaterals in R, it follows that the probability that at least one of the lattice quadrilaterals with area at least $20\log n/n$ is empty is less than $n^8 \cdot n^{-20} < n^{-10}$. □

Let H be a maximum size convex hole of R_n. We now transcribe the conclusion of Proposition 4 for easy reference within this proof.

Claim C. *There exists a lattice quadrilateral* Q_1 *such that* $H \subseteq Q_1$ *and* $\mathrm{a}(Q_1) \leq 2\,\mathrm{a}(H) + 40/n$. *Moreover, if* $\mathrm{a}(H) \geq 1000/n$, *then there is a lattice quadrilateral* Q_0 *such that* $Q_0 \subseteq H$ *and* $\mathrm{a}(Q_0) \geq \mathrm{a}(H)/32$. □

Claim D. *With probability at least* $1 - 2n^{-10}$ *we have* $\mathrm{a}(Q_1) < 2000\log n/n$ *and* $|R_n \cap Q_1| \leq 3000\log n$.

Proof. By Claim A, it suffices to show that with probability at least $1 - n^{-10}$ we have that $\mathrm{a}(Q_1) < 2000\log n/n$.

Suppose first that $\mathrm{a}(H) < 1000/n$. Then $\mathrm{a}(Q_1) \leq 2\,\mathrm{a}(H) + 40/n < 2040/n$. Since $2040/n < 2000\log n/n$, in this case we are done.

Now suppose that $\mathrm{a}(H) \geq 1000/n$, so that Q_0 (from Claim C) exists. Moreover, $\mathrm{a}(Q_1) \leq 2\mathrm{a}(H) + 40/n < 3\mathrm{a}(H)$. Since $Q_0 \subseteq H$, and H is a convex hole of R_n, it follows that Q_0 is empty. Thus, by Claim B, with probability at least $1 - n^{-10}$ we have that $\mathrm{a}(Q_0) < 20\log n/n$. Now since $\mathrm{a}(Q_1) < 3\,\mathrm{a}(H)$ and $\mathrm{a}(Q_0) \geq \mathrm{a}(H)/32$, it follows that $\mathrm{a}(Q_1) \leq 96\,\mathrm{a}(Q_0)$. Thus with probability at least $1 - n^{-10}$ we have that $\mathrm{a}(Q_1) \leq 96 \cdot 20\log n/n < 2000\log n/n$. □

We now derive a bound from an exact result by Valtr [23].

Claim E. *The probability that r points chosen at random from a triangle are in convex position is at most $(30/r)^{2r}$, for all sufficiently large r.*

Proof. Valtr [23] proved that the probability that r points chosen at random in a triangle are in convex position is $2^r(3r-3)!/\big(((r-1)!)^3(2r)!\big)$. Using the bounds $(s/e)^s < s! \le e\, s^{s+1/2}\, e^{-s}$, we obtain

$$\frac{2^r(3r-3)!}{((r-1)!)^3(2r)!} < \frac{2^r(3r)!}{(r!)^3(2r)!} \le \frac{2^r 3(3r)^{3r}\sqrt{3r}e^{-3r}}{r^{3r}e^{-3r}(2r)^{2r}e^{-2r}} < \sqrt{27r}\left(\frac{27e^2}{2r^2}\right)^r < \left(\frac{30}{r}\right)^{2r},$$

where the last inequality holds for all sufficiently large r. □

For each lattice quadrilateral Q, the polygon $Q \cap R$ has at most eight sides, and so it can be partitioned into at most eight triangles. For each Q, we choose one such decomposition into triangles, which we call the *basic* triangles of Q. Note that there are fewer than $8(9n)^8$ basic triangles in total.

Claim F. *With probability at least $1 - 2n^{-1.9}$ the random point set R_n satisfies that no lattice quadrilateral Q with $\mathrm{a}(Q) < 2000 \log n/n$ contains $40 \log n/(\log\log n)$ points of R_n in convex position.*

Proof. Let $\mathcal{T}$ denote the set of basic triangles obtained from lattice quadrilaterals that have area at most $2000 \log n/n$. By Claim A, with probability at least $1 - n^{-10}$ every $T \in \mathcal{T}$ satisfies $|R_n \cap T| \le 3000 \log n$. Thus it suffices to show that the probability that there exists a $T \in \mathcal{T}$ with $|R_n \cap T| \le 3000 \log n$ and $5 \log n/(\log\log n)$ points of R_n in convex position is at most n^{-2}.

Let $T \in \mathcal{T}$ be such that $|R_n \cap T| \le 3000 \log n$, and let $i := |R_n \cap T|$. Conditioning on i means that the i points in $R_n \cap T$ are randomly distributed in T. By Claim E, the expected number of r-tuples of R_n in T in convex position is at most

$$\binom{i}{r}\left(\frac{30}{r}\right)^{2r} \le \binom{3000 \log n}{r}\left(\frac{30}{r}\right)^{2r} < (10^7 r^{-3} \log n)^r.$$

Since there are at most $8(9n)^8$ choices for the triangle T, it follows that the expected total number of such r-tuples (over all $T \in \mathcal{T}$) with $r = 5 \log n/\log\log n$ is at most $8(9n)^8 \cdot (10^7 r^{-3} \log n)^r < n^{-1.9}$. Hence the probability that one such r-tuple exists (that is, the probability that there exists a $T \in \mathcal{T}$ with $5 \log n/(\log\log n)$ points of R_n in convex position) is at most $n^{-1.9}$. □

To finish the proof of (2), recall that H is a maximum size convex hole of R_n, and that $H \subseteq Q_1$. It follows immediately from Claims D and F that with probability at least $1 - 2n^{-1.9}$ the quadrilateral Q_1 does not contain a set of $40 \log n/(\log \log n)$ points of R_n in convex position. In particular, with probability at least $1-2n^{-1.9}$ the size of H is at most $40 \log n/(\log \log n)$. □

4. Proof of Theorem 2

Claim 7. *For every $\alpha \geq 1$ and every sufficiently large n,*

$$\mathbb{E}(\text{Hol}(R_n)) \geq (1/\alpha)\mathbb{E}(\text{Hol}(R_{\lfloor \alpha \cdot n \rfloor})).$$

Proof. Let $\alpha \geq 1$. We choose a random $\lfloor \alpha \cdot n \rfloor$-point set $R_{\lfloor \alpha \cdot n \rfloor}$ and a random n-point set R_n over R as follows: first we choose $\lfloor \alpha \cdot n \rfloor$ points randomly from R to obtain $R_{\lfloor \alpha \cdot n \rfloor}$, and then from $R_{\lfloor \alpha \cdot n \rfloor}$ we choose randomly n points, to obtain R_n. Now if H is a convex hole of $R_{\lfloor \alpha \cdot n \rfloor}$ with vertex set $V(H)$, then $V(H) \cap R_n$ is the vertex set of a convex hole of R_n. Noting that $\mathbb{E}(|V(H) \cap R_n|) = \frac{n}{\lfloor \alpha n \rfloor}|V(H)| \geq (1/\alpha)|V(H)|$, the claim follows. □

Claim 8. *Suppose that R, U are bounded convex regions in the plane such that $U \subseteq R$. Then, for all sufficiently large n,*

$$\mathbb{E}(\text{Hol}(R_n)) > \frac{\text{a}(U)}{2\text{a}(R)}\mathbb{E}(\text{Hol}(U_n)).$$

Proof. Let $\beta := \text{a}(R)/\text{a}(U)$ (thus $\beta \geq 1$), and let $0 < \epsilon \ll 1$.

Let $R_{\lfloor (1-\epsilon)\beta \cdot n \rfloor}$ be a set of $\lfloor (1-\epsilon)\beta \cdot n \rfloor$ points randomly chosen from R. Let $m := |U \cap R_{\lfloor (1-\epsilon)\beta \cdot n \rfloor}|$, and $\alpha := n/m$. Thus the expected value of α is $(1-\epsilon)^{-1}$, and a standard application of Chernoff's inequality implies that with probability at least $1-e^{\Omega(-n)}$ we have $1 \leq \alpha \leq (1-2\epsilon)^{-1}$. Conditioning on m means that $U_m := U \cap R_{\lfloor (1-\epsilon)\beta \cdot n \rfloor}$ is a randomly chosen m-point set in U.

Since $U \subseteq R$, then every convex hole in U_m is also a convex hole in $R_{\lfloor (1-\epsilon)\beta \cdot n \rfloor}$, and so

$$\text{Hol}(R_{\lfloor (1-\epsilon)\beta \cdot n \rfloor}) \geq \text{Hol}(U_m). \tag{3}$$

From Claim 7 it follows that

$$\mathbb{E}(\text{Hol}(R_n)) \geq ((1-\epsilon)\beta)^{-1}\mathbb{E}(\text{Hol}(R_{\lfloor (1-\epsilon)\beta \cdot n \rfloor})), \tag{4}$$

and that since $\alpha \geq 1$, then $\mathbb{E}(\text{Hol}(U_m)) \geq (1/\alpha)\mathbb{E}(\text{Hol}(U_n))$. Therefore

$$\mathbb{E}(\text{Hol}(U_m)) \geq (1-2\epsilon)\mathbb{E}(\text{Hol}(U_n)), \quad \text{if } 1 \leq \alpha \leq (1-2\epsilon)^{-1}. \tag{5}$$

Since $1 \leq \alpha \leq (1-2\epsilon)^{-1}$ holds with probability at least $1-e^{\Omega(-n)}$, (3), (4), and (5) imply that $\mathbb{E}(\text{Hol}(R_n)) \geq (((1-\epsilon)\beta)^{-1}(1-2\epsilon)\mathbb{E}(\text{Hol}(U_n)) - ne^{\Omega(-n)} > \beta^{-1}(1-2\epsilon)\mathbb{E}(\text{Hol}(U_n)) - ne^{\Omega(-n)}$. Recalling that $\beta = \text{a}(R)/\text{a}(U)$, and making ϵ small enough so that $1-2\epsilon > 1/2$, we conclude that for all sufficiently large n, $\mathbb{E}(\text{Hol}(R_n)) > (\text{a}(U)/2\text{a}(R))\mathbb{E}(\text{Hol}(U_n))$. □

We now note that if Q, Q' are two regions such that Q' is obtained from Q by an affine transformation, then $\mathbb{E}(\text{Hol}(Q_n)) = \mathbb{E}(\text{Hol}(Q'_n))$. It follows that there is a function $f(n)$ such that $\mathbb{E}(\text{Hol}(Q_n)) = f(n)$ for every rectangle Q.

Let U, K be rectangles such that $U \subseteq R \subseteq K$, $\mathrm{a}(U) \geq (1/8)\mathrm{a}(R)$, and $\mathrm{a}(K) \leq 2\mathrm{a}(R)$ (the existence of such rectangles is guaranteed by Corollary 5). It follows from Claim 8 that, for all sufficiently large n, $4\,\mathbb{E}(\textsc{Hol}(K_n)) \geq \mathbb{E}(\textsc{Hol}(R_n)) \geq (1/16)\mathbb{E}(\textsc{Hol}(U_n))$. Since both U and K are rectangles, we conclude that for any bounded convex region R and every sufficiently large n we have $4\,f(n) \geq \mathbb{E}(\textsc{Hol}(R_n)) \geq (1/16)f(n)$. For any bounded convex region S and sufficiently large n we obtain identical inequalities $4\,f(n) \geq \mathbb{E}(\textsc{Hol}(S_n)) \geq (1/16)f(n)$. We conclude that for every sufficiently large n, and every pair R, S of bounded convex regions, $(1/64) \leq \mathbb{E}(\textsc{Hol}(R_n))/\mathbb{E}(\textsc{Hol}(S_n)) \leq 64$, thus proving the expectation part of Theorem 2.

We finally prove the a.a.s. part of Theorem 2. Since $\mathrm{a}(U) \geq (1/8)\mathrm{a}(R)$, it follows by Chernoff's inequality that out of the n points in R_n, a.a.s. at least $n/16$ points are in U. Since U is a rectangle, it follows from (1) that a.a.s.

$$\textsc{Hol}(U_{n/16}) \geq \frac{\log(n/16)}{2\log\log(n/16)} \geq \frac{\log n}{4\log\log n},$$

where the last inequality holds for all sufficiently large n. Since $\textsc{Hol}(R_n) \geq \textsc{Hol}(U_{n/16})$, we obtain that a.a.s. $\textsc{Hol}(R_n) \geq (1/4)(\log n/\log\log n)$.

For the upper bound, we start by noting that we can generate R_n by placing the points one by one on K, and stopping when we have reached exactly n points on R. Since $\mathrm{a}(K) \leq 2\mathrm{a}(R)$, it follows by Chernoff's inequality that a.a.s. when we stop K has s points for some $s \in [n, 4n]$. Since K is a rectangle, it follows from (2) that for each $s \in [n, 4n]$, with probability at least $1 - n^{-1.9}$ we have $\textsc{Hol}(K_s) \leq 40(\log s/\log\log s)$. By the union bound, it then follows that the probability that $\textsc{Hol}(K_s) \leq 40(\log s/\log\log s)$ is at least $1 - (3n)\cdot n^{-1.9} = 1 - 3n^{-0.9}$. Since $\textsc{Hol}(K_s) \geq \textsc{Hol}(R_n)$, it follows that a.a.s. $\textsc{Hol}(R_n) \leq 40(\log s/\log\log s) \leq 40(\log 4n/\log\log 4n) < 50(\log n/\log\log n)$.

Therefore a.a.s. $(1/4)(\log n/\log\log n) \leq \textsc{Hol}(R_n) \leq 50(\log n/\log\log n)$. As we obtain identical bounds for S_n for any bounded convex region S, the a.a.s. part of Theorem 2 follows immediately. □

5. Concluding remarks

The lower and upper bounds we found in the proof of Theorem 1 for the case when R is a square (we proved that a.a.s. $(1/2)\log n/(\log\log n) \leq \textsc{Hol}(R_n) \leq 40\log n/(\log\log n)$) are not outrageously far from each other. It seems plausible to improve the 40 factor. One possible approach would be by decreasing the number of lattice quadrilaterals (from roughly n^8 to n^c for some $c < 8$), by taking care separately of those quadrilaterals with large area. Our belief is that the correct constant is closer to 1/2 than to 40, and we would not be surprised if 1/2 were proved to be the correct constant.

As pointed out by one of the referees, using Bárány's result [6] we could state Claim E for general convex polygons, and consequently in Claim F we would not need to use basic triangles. In this alternative approach, according to our calculations, the bound of 40 would not get improved.

There is great interest not only in the existence, but also on the number of convex holes of a given size (see for instance [7]). Along these lines, let us observe that a slight modification of our proof of Theorem 1 yields the following statement. The details of the proof are omitted.

Proposition 9. *Let R_n be a set of n points chosen independently and uniformly at random from a square. Then, for any positive integer s, the number of convex holes of R_n of size s is a.a.s. at most $9^8 n^9$.*

Sketch of proof. By (2) (in Section 3), we may assume $s \leq 40 \log n / \log\log n$. By Claims C and D, each convex hole of size s is contained in a lattice quadrilateral that has area at most $2000 \log n/n$ and contains at most $3000 \log n$ points of R_n. For each convex hole of size s, find such an enclosing lattice quadrilateral, and let $\mathcal{Q}$ denote the set of all these lattice quadrilaterals. Let $Q \in \mathcal{Q}$. Since Q has at most $3000 \log n$ points of R_n, it obviously follows that the number of convex holes of size s inside Q is at most

$$\binom{3000 \log n}{s} < \binom{3000 \log n}{40 \log n / \log\log n} < n.$$

Since each convex hole is contained in some $Q \in \mathcal{Q}$, and $|\mathcal{Q}| < (9n)^8$, it follows that a.a.s the total number of convex holes of R_n of size s is at most $n \cdot (9n)^8$. □

We made no effort to improve the exponent of n in this statement.

Moreover, for "large" convex holes we can also give lower bounds. Indeed, our calculations can be easily extended to show that for every sufficiently small constant c, there is an $\epsilon(c)$ such that the number of convex holes of size at least $c \cdot \log n/(\log\log n)$ is at most n^8 and at least $n^{1-\epsilon(c)}$.

One of the referees asked if the results in this paper could be generalized to point sets in d-dimensional space, for any $d > 2$. Using Bárány's extension [5] of Valtr's results to higher dimensions, the proofs indeed seem to carry over to all dimensions, albeit with the introduction of a few technical complications.

Acknowledgments

This project initially evolved from conversations of the third author with Ruy Fabila-Monroy. We thank him for these fruitful exchanges, and for several insightful remarks and corrections on this manuscript. We also thank Ferran Hurtado for sharing with us his expertise and providing guidance at the beginning of the project. He pointed us in the right direction by bringing to our attention the importance of settling the (ir)relevance of the shape of the convex sets under consideration. We are also indebted to three anonymous referees for providing substantial feedback on the first version of this manuscript. Their corrections and suggestions helped us improved both the presentation and the results in this paper.

References

[1] O. Aichholzer. [Empty] [colored] k-gons. Recent results on some Erdős-Szekeres type problems. In *Proc. XIII Encuentros de Geometría Computacional*, pp. 43–52, Zaragoza, Spain, 2009.

[2] N. Alon and J. Spencer. The probabilistic method, 3rd. Edition. Wiley, 2008.

[3] I. Bárány and Z. Füredi, Empty simplices in Euclidean space, *Canad. Math. Bull.* **30** (1987) 436–445.

[4] I. Bárány and G. Ambrus, Longest convex chains, *Random Structures and Algorithms* **35** (2009), 137–162.

[5] I. Bárány, A note on Sylvester's four-point problem. *Studia Sci. Math. Hungar.* **38** (2001), 73–77.

[6] I. Bárány, Sylvester's question: the probability that n points are in convex position, *Ann. Probab.* **27** (1999), 2020–2034.

[7] I. Bárány and P. Valtr, Planar point sets with a small number of empty convex polygons. *Studia Sci. Math. Hungar.* **41** (2004), 243–266.

[8] C. Buchta, The exact distribution of the number of vertices of a random convex chain, *Mathematika* **53** (2006), 247–254.

[9] P. Brass, W. Moser, and J. Pach, Research Problems in Discrete Geometry. Springer, 2005.

[10] P. Erdős and G. Szekeres, A combinatorial problem in geometry. *Compositio Math.* **2** (1935), 463–470.

[11] P. Erdős, Some more problems on elementary geometry. *Austral. Math. Soc. Gaz.* **5** (1978), 52–54.

[12] T. Gerken, Empty convex hexagons in planar point sets. *Discrete Comput. Geom.* **39** (2008), 239–272.

[13] H. Harborth, Konvexe Fünfecke in ebenen Punktmengen, *Elem. Math.* **33** (1978), 116–118.

[14] J.D. Horton, Sets with no empty convex 7-gons, *Canad. Math. Bull.* **26** (1983), 482–484.

[15] V.A. Koshelev, The Erdős-Szekeres problem. *Dokl. Math.* **76** (2007), 603–605.

[16] W. Morris and V. Soltan, The Erdős-Szekeres problem on points in convex position — a survey. *Bull. Amer. Math. Soc.* **37** (2000), 437–458.

[17] C. Nicolás, The empty hexagon theorem. *Discrete Comput. Geom.* **38** (2007), 389–397.

[18] M. Overmars, Finding sets of points without empty convex 6-gons. *Discrete Comput. Geom.* **29** (2003), 153–158.

[19] A. Rényi and R. Sulanke, Über die konvexe Hülle von n zufällig gewählten Punkten. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** (1963), 75–84.

[20] A. Rényi and R. Sulanke, Über die konvexe Hülle von n zufällig gewählten Punkten. II. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **3** (1964), 138–147.

[21] J.J. Sylvester, Question 1491. *The Educational Times*, (London). April 1864.

[22] P. Valtr, Probability that n Random Points are in Convex Position. *Discrete and Computational Geometry* **13** (1995), 637–643.

[23] P. Valtr, The Probability that n Random Points in a Triangle Are in Convex Position. *Combinatorica* **16** (1996), 567–573.

[24] P. Valtr, On Empty Hexagons. In: J. E. Goodman, J. Pach, and R. Pollack, *Surveys on Discrete and Computational Geometry, Twenty Years Later.* Contemp. Math. **453**, AMS, 2008, pp. 433–441.

University of Illinois at Urbana-Champaign, USA.
E-mail address: jobal@math.uiuc.edu

Facultad de Ciencias, UASLP. San Luis Potosi, Mexico.
E-mail address: hernan@fc.uaslp.mx

Instituto de Física, UASLP. San Luis Potosi, Mexico.
E-mail address: gsalazar@ifisica.uaslp.mx