Empty convex polytopes in random point sets

József Balogh*[1], Hernán González-Aguilar†[2], and Gelasio Salazar‡[3]

[1]Department of Mathematics, University of Illinois at Urbana-Champaign. Urbana, IL, United States.
[2]Facultad de Ciencias, Universidad Autónoma de San Luis Potosí. San Luis Potosí, SLP, México.
[3]Instituto de Física, Universidad Autónoma de San Luis Potosí. San Luis Potosí, SLP, México.

Abstract

Given a set P of points in $\mathbb{R}^d$, a *convex hole* (alternatively, *empty convex polytope*) of P is a convex polytope with vertices in P, containing no points of P in its interior. Let R be a bounded convex region in $\mathbb{R}^d$. We show that if P is a set of n random points chosen independently and uniformly over R, then the expected number of vertices of the largest hole of P is $\Theta(\log n/(\log\log n))$, regardless of the shape of R. This generalizes the analogous result proved for the case $d = 2$ by Balogh, González-Aguilar, and Salazar.

Introduction

Given a set P of points in $\mathbb{R}^d$, a *convex hole* (alternatively, *empty convex polytope*) of P is a convex polytope with vertices in P, containing no points of P in its interior.

Recently, we showed that the expected size of the largest convex hole in a random n-point set in the plane is $\Theta(\log n/\log\log n)$ [3]. One anonymous referee of this paper asked if this could be generalized to $d > 2$ dimensions. Joe O'Rourke asked the same question in MathOverflow, and Douglas Zare replied that the $\Omega(\log n/\log\log n)$ lower bound carries over easily to the d-dimensional case [10]. At the end of his reply, Zare wrote: "I don't know whether their harder upper bound of the same form also extends to higher dimensions, but I suspect that it does."

Our aim in this note is to show that, indeed, the upper bound also holds for higher dimensions. Thus, our main result is:

Theorem 1 *Let $d \geq 2$ be an integer, and let R be a bounded convex region in $\mathbb{R}^d$. 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*

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

Moreover, a.a.s.

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

The proof, which is an immediate consequence of Theorems 2 and 3 below, follows very closely the main ideas of the proof of [3, Theorem 3]. Indeed, the strategy and the main ideas are so close that it seems best to follow as closely as possible the structure of [3]. As we shall see below, some of the results proved in [3] follow without any modification to arbitrary dimensions. The main adaptations needed are:

1. a generalization of the results in [3, Section 2] to $d > 2$ dimensions, to approximate convex sets in $\mathbb{R}^d$ with lattice polytopes; and
2. an adaptation to $d > 2$ dimensions of the results on the probability that a random n-point set is in convex position, from the exact results of Valtr [11, 12] in $\mathbb{R}^2$ to the asymptotic results of Bárány [6] in $\mathbb{R}^d$, for any $d \geq 2$.

The workhorse for the proof of Theorem 1 for arbitrary regions R is the following statement, which takes care of the particular case in which R is a parallelotope.

Theorem 2 *Let R be a parallelotope in $\mathbb{R}^d$. 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*

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

Moreover, a.a.s.

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

*Email: jobal@math.uiuc.edu. Research supported by NSF CAREER Grant DMS-0745185.

†Email: hernan@fc.uaslp.mx. Research supported by PROMEP.

‡Email: gsalazar@ifisica.uaslp.mx. Research supported by CONACYT Grant 106432.

The other essential fact is that the order of magnitude of the expected number of vertices of the largest convex hole is independent of the shape of R:

Theorem 3 *There exist absolute constants b, b' with the following property. Let R and S be bounded convex regions in $\mathbb{R}^d$. 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,*

$$b \leq \frac{\mathbf{E}(\textsc{Hol}(R_n))}{\mathbf{E}(\textsc{Hol}(S_n))} \leq b'.$$

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

$$c \leq \frac{\textsc{Hol}(R_n)}{\textsc{Hol}(S_n)} \leq c'.$$

We remark that Theorem 3 is in line with the following result proved by Bárány and Füredi [5]: 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.

Remark (Proof of Theorem 1). Theorem 1 is an immediate consequence of Theorems 2 and 3.

The proof of Theorem 2 is in Section 1. As we explain in Section 2, the proof of Theorem 3 is totally analogous to the proof of [3, Theorem 2].

We make a few final remarks before we move on to the proofs. For the rest of the paper we let $\mathrm{Vol}(U)$ denote the volume of a region U in $\mathbb{R}^d$. 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.

1 Proof of Theorem 2

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:

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

$$\Pr\left(\textsc{Hol}(R_n) \leq d(2+2d^2)\,\frac{\log n}{\log\log n}\right) \geq 1 - n^{-1}. \quad (2)$$

We note that (1) and (2) imply immediately the a.a.s. part of Theorem 2. 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 an empty convex polytope of size at least $\frac{\log n}{2\log\log n}$.

Consider the 2-dimensional projection $\pi : \mathbb{R}^d \to \mathbb{R}^2$ defined by $(x_1, x_2, x_3, x_4, \ldots, x_d) \to (x_1, x_2, 0, 0, \ldots, 0)$. Note that $\pi(R_n)$ is a set of n points chosen (independently and uniformly) at random from the unit square. Thus it follows from Eq. (1) in [3] that a.a.s. $\pi(R_n)$ has a convex hole H of size at least $\frac{\log n}{2\log\log n}$. Clearly, $\pi^{-1}(H)$ is an empty convex polytope of R_n of size at least $\frac{\log n}{2\log\log n}$. □

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 [1].

Lemma 4 *Let $Y_1, \ldots, Y_m$ be mutually independent random variables with $\Pr(Y_i = 1) = p$ and $\Pr(Y_i = 0) = 1 - p$, for $i = 1, \ldots, m$. Let $Y := Y_1 + \ldots + Y_m$. Then*

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

Let S be the isothetic d-cube of volume 3^d, also (as R) centered at the origin.

We need the following result on approximating convex sets by lattice parallelotopes.

Claim A. *For each positive integer $d > 0$ there exist integers $f_1(d)$ and $f_2(d)$ with the following property. Let H be a convex set in $\mathbb{R}^d$. Then there exists a lattice parallelotope Q_1 such that $H \subseteq Q_1$ and $\mathrm{Vol}(Q_1) \leq (f_1(d)+1)\,\mathrm{Vol}(H)$. Moreover, if $\mathrm{Vol}(H) \geq 2^{d-1} \cdot 1000/n$, then there is a lattice parallelotope Q_0 such that $Q_0 \subseteq H$ and $\mathrm{Vol}(Q_0) \geq (f_2(d) - 1)\mathrm{Vol}(H)$.*

Sketch of Proof. By a the theorem of M. Balla [2], for every convex compact set $H \subset \mathbb{R}^d$ there exists a parallelotope P such that $P \subset H \subset dP = \widehat{P}$ where dP is the image of P under a homothety with ratio d. This implies that $d^{-d}\,\mathrm{Vol}(P) \leq \mathrm{Vol}(H) \leq d^d\,\mathrm{Vol}(P)$.

For each vertex v_i, $i = 0, \ldots, 2^d$, of $\widehat{P}$, let us denote by Q_{v_i} the parallelotope with side length $2/n$ with facets parallel to the facets of $\widehat{P}$ that has v_i as one of its vertices and $\widehat{P} \cap Q_{v_i} = \{v_i\}$. Observe that each Q_{v_i} contains a d-ball of diameter $2/n$ and for that, there is a lattice point v'_i in the interior of each Q_{v_i}. Let Q_1 be the convex hull of the points $v'_1, \ldots, v'_{2^d}$. Note that $d(v_i, v'_i) \leq \frac{2}{n}\sqrt{d}$ for each $i = 1, \ldots, 2^d$, this implies that $\varrho(\widehat{P}, Q_1) \leq \frac{2}{n}\sqrt{d}$ where $\varrho(\cdot,\cdot)$ is the Hausdorff metric. Then

$$\begin{aligned} \mathrm{Vol}(Q_1) &\leq \mathrm{Vol}(\widehat{P}) + \frac{2}{n}\sqrt{d} \cdot \mathrm{Surf}(\widehat{P}) + \mathrm{Vol}(B) \\ &\leq d^d\, \mathrm{Vol}(H) + \frac{2}{n}\sqrt{d} \cdot \mathrm{Surf}(\widehat{P}) + \mathrm{Vol}(B) \\ &\leq (f_1(d) + 1)\, \mathrm{Vol}(H) \end{aligned}$$

where B is the d-ball of diameter $\frac{2}{n}\sqrt{d}$ and $\mathrm{Surf}(\cdot)$ is the volume $(d-1)$-dimensional.

Now, for each vertex w_i, $i = 0, \ldots, 2^d$, of P consider the parallelotope Q_{w_i} with side length $2/n$ with facets parallel to the facets of P that has w_i as one of its vertices and $Q_{w_i} \subset P$. Because each Q_{w_i} contains a d-ball of diameter $\frac{2}{n}$, then there exists a lattice point in each Q_{w_i}, let w'_i this point. The existence of these points is guaranteed provided that $\mathrm{Vol}(H) \geq 2^{d-1} \cdot 1000/n$ Let Q_0 be the convex hull of the points $w'_1, \ldots, w'_{2^d}$. Note that $d(w_i, w'_i) \leq \frac{2}{n}\sqrt{d}$ for each $i = 1, \ldots, 2^d$, this implies that $\varrho(P, Q_0) \leq \frac{2}{n}\sqrt{d}$. Then $\mathrm{Vol}(P) - \frac{2}{n}\sqrt{d} \cdot \mathrm{Surf}(Q_0) - \mathrm{Vol}(B) \leq \mathrm{Vol}(Q_0)$ This implies

$$\begin{aligned} \mathrm{Vol}(Q_0) &\geq \mathrm{Vol}(P) - \frac{2}{n}\sqrt{d} \cdot \mathrm{Surf}(Q_0) - \mathrm{Vol}(B) \\ &\geq \mathrm{Vol}(P) - \frac{2}{n}\sqrt{d} \cdot \mathrm{Surf}(P) - \mathrm{Vol}(B) \\ &\geq d^{-d}\, \mathrm{Vol}(H) - \frac{2}{n}\sqrt{d} \cdot \mathrm{Surf}(H) - \mathrm{Vol}(B) \\ &\geq (f_2(d) - 1)\mathrm{Vol}(H). \end{aligned}$$ □

For the rest of the proof, for simplicity we define $f_3(d) := f_3(d)$, where $f_1(d)$ and $f_2(d)$ are as in Claim A.

Since there are $(9n+1)^d < (10n)^d$ lattice points, out of which $(n+1)^d$ are in R, it follows that there are fewer than $\binom{(10n)^d}{2^d} < (10n)^{d2^d}$ lattice d-parallelotopes in total, and fewer than $\binom{n^d}{2^d} < n^{d2^d}$ lattice d-parallelotopes all of whose vertices are in R.

Claim B. *With probability at least $1 - n^{-10}$ every lattice parallelotope Q with $\mathrm{Vol}(Q) < 20 f_3(d) \log n/n$ satisfies that $|R_n \cap Q| \leq (3/2) \cdot 20 f_3(d) \log n$.*

Proof. We note that, since $(f_1(d)+1)/f_2(d) > 1$, it follows that $20 f_3(d) > d2^d + 10$. Let Q be a lattice parallelotope with $\mathrm{Vol}(Q) < 20\, f_3(d) \log n/n$, and let $Z = Z(Q) \subseteq R$ be any lattice parallelotope containing Q, with $\mathrm{Vol}(Z) = 20\, f_3(d) \log n/n$. 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 4 with $p = \mathrm{Vol}(Z)$ and $m = n$, to obtain $\Pr(X_Z \geq (3/2) \cdot 20\, f_3(d) \log n) < e^{-(3/2)20\, f_3(d)/24 \log n} = n^{-(5/4) f_3(d)}$. Since $Q \subseteq Z$, it follows that $\Pr(X_Q \geq (3/2) \cdot 20\, f_3(d) \log n) < n^{-(5/4) f_3(d)}$. As the number of choices for Q is at most $(10n)^{d2^d}$, with probability at least $(1 - (10n)^{d2^d} \cdot n^{-(5/4) f_3(d)}) > 1 - n^{-10}$, no such Q contains more than $(3/2) \cdot 20\, f_3(d) \log n$ points of R_n. □

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

Claim C. *With probability at least $1 - n^{-10}$, there is no empty lattice parallelotope $Q \subseteq R$ with $\mathrm{Vol}(Q) \geq 20\,(d2^d + 10) \log n/n$.*

Proof. The probability that a fixed lattice parallelotope $Q \subseteq R$ with $\mathrm{Vol}(Q) \geq d2^d + 10 \log n/n$ is empty is $(1 - \mathrm{Vol}(Q))^n < n^{-d2^d+10}$. Since there are fewer than n^{d2^d} lattice parallelotopes in R, it follows that the probability that at least one of the lattice parallelotope with area at least $(d2^d + 10) \log n/n$ (and hence with area at least $20\,(d2^d + 10) \log n/n$) is empty is less than $n^{d2^d} \cdot n^{-(d2^d+10)} \leq n^{-10}$. □

For the rest of the proof, we let H be a maximum size convex hole of R_n.

Claim D. *With probability at least $1 - n^{-10}$ we have $\mathrm{Vol}(Q_1) < f_3(d) \log n/n$.*

Proof. Suppose first that $\mathrm{Vol}(H) < 2^{d-1} 1000/n$. Then $\mathrm{Vol}(Q_1) \leq 2^{d-1} \cdot 1000(f_1(d)+1)/n$. Since this is obviously smaller than $f_3(d) \log n/n$, in this case we are done. Now suppose that $\mathrm{Vol}(H) \geq 2^{d-1} \cdot 1000/n$, so that Q_0 (from Claim A) exists. Moreover, $\mathrm{Vol}(Q_1) \leq (f_1(d)+1)\mathrm{Vol}(H)$. Since $Q_0 \subseteq H$, and H is a hole of R_n, it follows that Q_0 is empty. Thus, by Claim C, with probability at least $1 - n^{-10}$ we have that $\mathrm{Vol}(Q_0) < (d2^d + 10) \log n/n$. Now since $\mathrm{Vol}(Q_1) < (f_1(d) + 1)\,\mathrm{Vol}(H)$ and $\mathrm{Vol}(Q_0) \geq f_2(d) \cdot \mathrm{Vol}(H)$, it follows that $\mathrm{Vol}(Q_1) \leq (f_1(d)+1)\,\mathrm{Vol}(Q_0)/f_2(d)$. Thus with probability at least $1 - n^{-10}$ we have that $\mathrm{Vol}(Q_1) \leq \big((f_1(d)+1)(d2^d+10)/f_2(d)\big) \log n/n = f_3(d) \log n/n$. □

Claim E. *For each fixed integer $d > 0$, there exist a universal positive constant $c_2 := c_2(d)$ with the following property. Let K be any convex polytope in $\mathbb{R}^d$. Then the probability that r points chosen at random from K are in convex position is at most $(c_2\, n^{\frac{2}{d-1}})^{-n}$.*

Proof. This is an immediate consequence of [6] (see for instance [4, Theorem 2.1]). □

Claim F. *With probability at least* $1 - 2n^{-2}$ *the random point set* R_n *satisfies that no lattice parallelotope* Q *with* $\mathrm{Vol}(Q) < 20\, f_3(d) \log n/n$ *contains* $d(2+2d^2)\log n/(\log\log n)$ *points of* R_n *in convex position.*

Proof. Let Q be a lattice parallelotope with $\mathrm{Vol}(Q) < 20\, f_3(d)\log n/n$. By Claim B, with probability at least $1 - n^{-10}$ we have $|R_n \cap Q| \le (3/2) \cdot 20\, f_3(d) \log n$. Thus it suffices to show that the probability that there exists a lattice parallelotope Q with with $|R_n \cap Q| \le (3/2) \cdot 20\, f_3(d) \log n$ and $d(2+2d^2)\log n/(\log\log n)$ points of R_n in convex position is at most n^{-2}.

Let $c_2 := c_2(d)$ be as in Claim E. Thus the expected number of r-tuples of R_n in Q in convex position is at most

$$\begin{aligned}\binom{|R_n \cap Q|}{r} & \left(c_2 r^{\frac{2}{d-1}}\right)^{-r} \\ & \le \binom{(3/2)\cdot f_3(d)\log n}{r}\left(c_2 r^{\frac{2}{d-1}}\right)^{-r} \\ & \le \left(\frac{e\cdot(3/2)\cdot F_2 \log n}{r}\right)^{r}\left(c_2 r^{\frac{2}{d-1}}\right)^{-r} \\ & < \left(c_3 \log n \cdot r^{-1-\frac{2}{d-1}}\right)^{r},\end{aligned}$$

where $c_3 := 3ef_3(d)/2c_2$.

Since there are at most n^{d2^d} choices for Q, it follows that the expected total number of such r-tuples with $r = d(2+2d^2)\log n/\log\log n$ is at most $n^{d2^d} \cdot \left(c_3 \log n \cdot r^{-1-\frac{2}{d-1}}\right)^{r} < n^{-2}$ (this last inequality follows from an elemenary but long manipulation). This completes the proof, since it follows that the probability that such an r-tuple exists is at most n^2. □

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

2 Proof of Theorem 3

The proof of Theorem 3 is totally analogous to the proof of [3, Theorem 2]. Indeed, in that proof, essentially all the arguments are independent of the dimension. The only adaptation that needs to be done is that we need a version of [3, Corollary 6] for $d > 2$ dimensions. We recall that [3, Corollary 6] claims that if H is a closed convex set in $\mathbb{R}^2$, then there exist rectangles U, K such that $U \subseteq H \subseteq K$, $\mathrm{Vol}(U) \ge \mathrm{Vol}(H)/8$, and $\mathrm{Vol}(K) \le 2\mathrm{Vol}(H)$.

A d-dimensional analogue of this statement follows from the following result in [9]: if H is a convex body in $\mathbb{R}^d$, then H contains a parallelotope P such that some translate of dP contains K. Indeed, this implies at once that if H is a closed convex set in $\mathbb{R}^d$, then there exist parallelotopes U, K such that $U \subseteq H \subseteq K$, $\mathrm{Vol}(U) \ge \mathrm{Vol}(H)/d$, and $\mathrm{Vol}(K) \le d \cdot \mathrm{Vol}(H)$.

References

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

[2] M. Y. Balla, Approximation of convex bodies by parallelotopes, *Internal Report IC/87/310, International Centre for Theoretical Physics, ICTP*, Trieste, 1988

[3] J. Balogh, H. González-Aguilar, and G. Salazar, Large convex holes in random point sets. *Comp. Geom.* **46** (2013), 725–733.

[4] I. Bárány, Random points, convex bodies, lattices. *Proceedings of the International Congress of Mathematicians, Vol. III* (Beijing, 2002), 527–535. Higher Ed. Press, Beijing, 2002.

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

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

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

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

[9] G.D. Chakerian and S. K. Stein, Some intersection properties of convex bodies. *Proc. Amer. Math. Soc.* **18** (1967), 109–112.

[10] J. O'Rourke, Empty convex polytopes for random point sets. `http://mathoverflow.net/questions/107875` (2012).

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

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