Holes and Islands in Random Point Sets

Martin Balko
Department of Applied Mathematics, Faculty of Mathematics and Physics,
Charles University, Prague, Czech Republic
balko@kam.mff.cuni.cz

Manfred Scheucher
Institut für Mathematik, Technische Universität Berlin, Germany
scheucher@math.tu-berlin.de

Pavel Valtr
Department of Applied Mathematics, Faculty of Mathematics and Physics,
Charles University, Prague, Czech Republic
Department of Computer Science, ETH Zürich, Switzerland

Abstract

For $d \in \mathbb{N}$, let $S$ be a finite set of points in $\mathbb{R}^d$ in general position. A set $H$ of $k$ points from $S$ is a $k$-hole in $S$ if all points from $H$ lie on the boundary of the convex hull $\text{conv}(H)$ of $H$ and the interior of $\text{conv}(H)$ does not contain any point from $S$. A set $I$ of $k$ points from $S$ is a $k$-island in $S$ if $\text{conv}(I) \cap S = I$. Note that each $k$-hole in $S$ is a $k$-island in $S$.

For fixed positive integers $d$, $k$ and a convex body $K$ in $\mathbb{R}^d$ with $d$-dimensional Lebesgue measure 1, let $S$ be a set of $n$ points chosen uniformly and independently at random from $K$. We show that the expected number of $k$-islands in $S$ is in $O(n^d)$. In the case $k = d + 1$, we prove that the expected number of empty simplices (that is, $(d+1)$-holes) in $S$ is at most $2^{d-1} \cdot d! \cdot (\frac{n}{d})^d$. Our results improve and generalize previous bounds by Bárány and Füredi [4], Valtr [19], Fabila-Monroy and Huemer [8], and Fabila-Monroy, Huemer, and Mitsche [9].

1 Introduction

For $d \in \mathbb{N}$, let $S$ be a finite set of points in $\mathbb{R}^d$. The set $S$ is in general position if, for every $k = 1, \ldots, d - 1$, no $k + 2$ points of $S$ lie in an affine $k$-dimensional subspace. A set $H$ of $k$ points from $S$ is a $k$-hole in $S$ if $H$ is in convex position and the interior of the convex hull $\text{conv}(H)$ of $H$ does not contain any point from $S$; see Figure 1 for an illustration in the plane. We say that a subset of $S$ is a hole in $S$ if it is a $k$-hole in $S$ for some integer $k$. 
14:2 Holes and Islands in Random Point Sets

Let \( h(k) \) be the smallest positive integer \( N \) such that every set of \( N \) points in general position in the plane contains a \( k \)-hole. In the 1970s, Erdős [6] asked whether the number \( h(k) \) exists for every \( k \in \mathbb{N} \). It was shown in the 1970s and 1980s that \( h(4) = 5 \), \( h(5) = 10 \) [11], and that \( h(k) \) does not exist for every \( k \geq 7 \) [12]. That is, while every sufficiently large set contains a 4-hole and a 5-hole, Horton constructed arbitrarily large sets with no 7-holes. His construction was generalized to so-called Horton sets by Valtr [18]. The existence of 6-holes in every sufficiently large point set remained open until 2007, when Gerken [10] and Nicolas [15] independently showed that \( h(6) \) exists; see also [20].

These problems were also considered in higher dimensions. For \( d \geq 2 \), let \( h_d(k) \) be the smallest positive integer \( N \) such that every set of \( N \) points in general position in \( \mathbb{R}^d \) contains a \( k \)-hole. In particular, \( h_2(k) = h(k) \) for every \( k \). Valtr [18] showed that \( h_d(k) \) exists for \( k \leq 2d + 1 \) but it does not exist for \( k > 2^{d-1}(P(d-1) + 1) \), where \( P(d-1) \) denotes the product of the first \( d - 1 \) prime numbers. The latter result was obtained by constructing multidimensional analogues of the Horton sets.

After the existence of \( k \)-holes was settled, counting the minimum number \( H_k(n) \) of \( k \)-holes in any set of \( n \) points in the plane in general position attracted a lot of attention. It is known, and not difficult to show, that \( H_3(n) \) and \( H_4(n) \) are in \( \Omega(n^2) \). The currently best known lower bounds on \( H_3(n) \) and \( H_4(n) \) were proved in [1]. The best known upper bounds are due to Bárány and Valtr [5]. Altogether, these estimates are

\[
n^2 + \Omega(n \log^{2/3} n) \leq H_3(n) \leq 1.6196n^2 + o(n^2)
\]

and

\[
n^2 + \Omega(n \log^{3/4} n) \leq H_4(n) \leq 1.9397n^2 + o(n^2).
\]

For \( H_5(n) \) and \( H_6(n) \), the best quadratic upper bounds can be found in [5]. The best lower bounds, however, are only \( H_5(n) \geq \Omega(n \log^{4/5} n) \) [1] and \( H_6(n) \geq \Omega(n) \) [21]. For more details, we also refer to the second author’s dissertation [17].

The quadratic upper bound on \( H_3(n) \) can be also obtained using random point sets. For \( d \in \mathbb{N} \), a convex body in \( \mathbb{R}^d \) is a compact convex set in \( \mathbb{R}^d \) with a nonempty interior. Let \( K \) be a convex body with \( d \)-dimensional Lebesgue measure \( \lambda_d(K) = 1 \). We use \( EH_{d,k}^n(n) \) to denote the expected number of \( k \)-holes in sets of \( n \) points chosen independently and uniformly at random from \( K \). The quadratic upper bound on \( H_3(n) \) then also follows from the following bound of Bárány and Füredi [4] on the expected number of \((d+1)\)-holes:

\[
EH_{d,d+1}^K(n) \leq (2d)^{2d^2} \cdot \binom{n}{d}
\]  

(1)
for any \( d \) and \( K \). In the plane, Bárány and Füredi [4] proved \( EH^{K}_{2,3}(n) \leq 2n^2 + O(n \log n) \) for every \( K \). This bound was later slightly improved by Valtr [19], who showed \( EH^{K}_{2,3}(n) \leq 4(\binom{n}{2}) \) for any \( K \). In the other direction, every set of \( n \) points in \( \mathbb{R}^d \) in general position contains at least \((n,d)\) \((d+1)\)-holes [4, 13].

The expected number \( EH^{K}_{2,4}(n) \) of 4-holes in random sets of \( n \) points in the plane was considered by Fabila-Monroy, Huemer, and Mitsche [9], who showed

\[
EH^{K}_{2,4}(n) \leq 18\pi D^2n^2 + o(n^2)
\]

for any \( K \), where \( D = D(K) \) is the diameter of \( K \). Since we have \( D \geq 2/\sqrt{\pi} \), by the Isodiametric inequality [7], the leading constant in (2) is at least 72 for any \( K \).

In this paper, we study the number of \( k \)-holes in random point sets in \( \mathbb{R}^d \). In particular, we obtain results that imply quadratic upper bounds on \( H_k(n) \) for any fixed \( k \) and that both strengthen and generalize the bounds by Bárány and Füredi [4], Valtr [19], and Fabila-Monroy, Huemer, and Mitsche [9].

2 Our results

Throughout the whole paper we only consider point sets in \( \mathbb{R}^d \) that are finite and in general position.

2.1 Islands and holes in random point sets

First, we prove a result that gives the estimate \( O(n^d) \) on the minimum number of \( k \)-holes in a set of \( n \) points in \( \mathbb{R}^d \) for any fixed \( d \) and \( k \). In fact, we prove the upper bound \( O(n^d) \) even for so-called \( k \)-islands, which are also frequently studied in discrete geometry. A set \( I \) of \( k \) points from a point set \( S \subseteq \mathbb{R}^d \) is a \( k \)-island in \( S \) if \( \text{conv}(I) \cap S = I \); see part (c) of Figure 1. Note that \( k \)-holes in \( S \) are exactly those \( k \)-islands in \( S \) that are in convex position. A subset of \( S \) is an island in \( S \) if it is a \( k \)-island in \( S \) for some integer \( k \).

**Theorem 1.** Let \( d \geq 2 \) and \( k \geq k + 1 \) be integers and let \( K \) be a convex body in \( \mathbb{R}^d \) with \( \lambda_d(K) = 1 \). If \( S \) is a set of \( n \geq k \) points chosen uniformly and independently at random from \( K \), then the expected number of \( k \)-islands in \( S \) is at most

\[
2^{d-1} \cdot \left( 2d^{2d-1} \binom{k}{[d/2]} \right)^{k-d-1} \cdot (k-d) \cdot \frac{n(n-1) \cdots (n-k+2)}{(n-k+1)^{k-d-1}}.
\]

which is in \( O(n^d) \) for any fixed \( d \) and \( k \).

The bound in Theorem 1 is tight up to a constant multiplicative factor that depends on \( d \) and \( k \), as, for any fixed \( k \geq d \), every set \( S \) of \( n \) points in \( \mathbb{R}^d \) in general position contains at least \( \Omega(n^d) \) \( k \)-islands. To see this, observe that any \( d \)-tuple \( T \) of points from \( S \) determines a \( k \)-island with \( k-d \) closest points to the hyperplane spanned by \( T \) (ties can be broken by, for example, taking points with lexicographically smallest coordinates), as \( S \) is in general position and thus \( T \) is a \( d \)-hole in \( S \). Any such \( k \)-tuple of points from \( S \) contains \( \binom{k}{d} \) \( d \)-tuples of points from \( S \) and thus we have at least \( \binom{k}{d} \) \( k \)-islands in \( S \).

Thus, by Theorem 1, random point sets in \( \mathbb{R}^d \) asymptotically achieve the minimum number of \( k \)-islands. This is in contrast with the fact that, unlike Horton sets, they contain arbitrarily large holes. Quite recently, Balogh, González-Aguilar, and Salazar [3] showed that the expected number of vertices of the largest hole in a set of \( n \) random points chosen independently and uniformly over a convex body in the plane is in \( \Theta(\log n/(\log \log n)) \).

For \( k \)-holes, we modify the proof of Theorem 1 to obtain a slightly better estimate.
Theorem 2. Let $d \geq 2$ and $k \geq d + 1$ be integers and let $K$ be a convex body in $\mathbb{R}^d$ with $\lambda_d(K) = 1$. If $S$ is a set of $n \geq k$ points chosen uniformly and independently at random from $K$, then the expected number $E_{H_{d,k}}^K(n)$ of $k$-holes in $S$ is in $O(n^d)$ for any fixed $d$ and $k$. More precisely,

$$E_{H_{d,k}}^K(n) \leq 2^{d-1} \cdot \left( \frac{k}{[d/2]} \right)^{k-d-1} \cdot \frac{n(n-1) \cdots (n-k+2)}{(k-d-1)! \cdot (n-k+1)^{k-d-1}}.$$

For $d = 2$ and $k = 4$, Theorem 2 implies $E_{H_{2,4}}^K(n) \leq 128 \cdot n^2 + o(n^2)$ for any $K$, which is a worse estimate than (2) if the diameter of $K$ is at most $8/(3\sqrt{\pi}) \approx 1.5$. However, the proof of Theorem 2 can be modified to give $E_{H_{2,4}}^K(n) \leq 12n^2 + o(n^2)$ for any $K$, which is always better than (2); see the final remarks in Section 3. We believe that the leading constant of Theorem 2 can be modified to give a worse estimate than (2) if the diameter of $K$ is not fixed. Due to space limitations, the proof of Theorem 4 is omitted.

Corollary 3. Let $d \geq 2$ be an integer and let $K$ be a convex body in $\mathbb{R}^d$ with $\lambda_d(K) = 1$. If $S$ is a set of $n$ points chosen uniformly and independently at random from $K$, then the expected number of $(d+1)$-holes in $S$ satisfies

$$E_{H_{d,d+1}}^K(n) \leq 2^{d-1} \cdot d! \cdot \left( \frac{n}{d} \right).$$

Corollary 3 is stronger than the bound (1) by Bárány and Füredi [4] and, in the planar case, coincides with the bound $E_{H_{2,3}}^K(n) \leq 4\binom{n}{3}$ by Valtr [19]. Very recently, Reitzner and Temesvari [16] proved an upper bound on $E_{H_{d,d+1}}^K(n)$ that is asymptotically tight if $d = 2$ or $d \geq 3$ and $K$ is an ellipsoid. In the planar case, their result shows that the bound $4\binom{n}{3}$ on $E_{H_{2,3}}^K(n)$ is best possible, up to a smaller order error term. No tight bounds on $E_{H_{d,d+1}}^K(n)$ are known if $d \geq 3$ and $K$ is not an ellipsoid.

We also consider islands of all possible sizes and show that their expected number is in $2\Theta(n^{(d-1)/(d+1)})$.

Theorem 4. Let $d \geq 2$ be an integer and let $K$ be a convex body in $\mathbb{R}^d$ with $\lambda_d(K) = 1$. Then there are constants $C_1 = C_1(d)$, $C_2 = C_2(d)$, and $n_0 = n_0(d)$ such that for every set $S$ of $n \geq n_0$ points chosen uniformly and independently at random from $K$ the expected number $E_d^K$ of islands in $S$ satisfies

$$2^{C_1} n^{(d-1)/(d+1)} \leq E_d^K \leq 2^{C_2} n^{(d-1)/(d+1)}.$$

Since each island in $S$ has at most $n$ points, there is a $k \in \{1, \ldots, n\}$ such that the expected number of $k$-islands in $S$ is at least $(1/n)$-fraction of the expected number of all islands, which is still in $2^{\Omega(n^{(d-1)/(d+1)})}$. This shows that the expected number of $k$-islands can become asymptotically much larger than $O(n^2)$ if $k$ is not fixed. Due to space limitations, the proof of Theorem 4 is omitted.

2.2 Islands and holes in $d$-Horton sets

To our knowledge, Theorem 1 is the first nontrivial upper bound on the minimum number of $k$-islands a point set in $\mathbb{R}^d$ with $d > 2$ can have. For $d = 2$, Fabila-Monroy and Huemer [8] showed that, for every fixed $k \in \mathbb{N}$, the Horton sets with $n$ points contain only $O(n^2)$
For $d \geq 2$, Valtr [18] introduced a $d$-dimensional analogue of Horton sets. Perhaps surprisingly, these sets contain asymptotically more than $O(n^d)$ $k$-islands for $k \geq d + 1$. For each $k$ with $d + 1 \leq k \leq 3 \cdot 2^{d-1}$, they even contain asymptotically more than $O(n^d)$ $k$-holes.

**Theorem 5.** Let $d \geq 2$ and $k$ be fixed positive integers. Then every $d$-dimensional Horton set $H$ with $n$ points contains at least $\Omega(n^{\min(2^{d-1}, k)})$ $k$-islands in $H$. If $k \leq 3 \cdot 2^{d-1}$, then $H$ even contains at least $\Omega(n^{\min(2^{d-1}, k)})$ $k$-holes in $H$.

### 3 Proofs of Theorem 1 and Theorem 2

Let $d$ and $k$ be positive integers and let $K$ be a convex body in $\mathbb{R}^d$ with $\lambda_d(K) = 1$. Let $S$ be a set of $n$ points chosen uniformly and independently at random from $K$. Note that $S$ is in general position with probability 1. We assume $k \geq d + 1$, as otherwise the number of $k$-islands in $S$ is trivially $\binom{n}{k}$ in every set of $n$ points in $\mathbb{R}^d$ in general position. We also assume $d \geq 2$ and $n \geq k$, as otherwise the number of $k$-islands is trivially $n - k + 1$ and 0, respectively, in every set of $n$ points in $\mathbb{R}^d$.

First, we prove Theorem 1 by showing that the expected number of $k$-islands in $S$ is at most

$$2^{d-1} \cdot \left(2d^{d-1} \left(\frac{k}{|d/2|}\right)\right)^{k-d-1} \cdot (k-d) \cdot \frac{n(n-1) \cdots (n-k+2)}{(n-k+1)^{k-d-1}},$$

which is in $O(n^d)$ for any fixed $d$ and $k$. At the end of this section, we improve the bound for $k$-holes, which will prove Theorem 2.

Let $Q$ be a set of $k$ points from $S$. We first introduce a suitable unique ordering $q_1, \ldots, q_k$ of points from $Q$. First, we take a set $D$ of $d + 1$ points from $Q$ that determine a simplex $\Delta$ with largest volume among all $(d + 1)$-tuples of points from $Q$. Let $q_1 q_2$ be the longest edge of $\Delta$ with $q_1$ lexicographically smaller than $q_2$ and let $a$ be the number of points from $Q$ inside $\Delta$. For every $i = 2, \ldots, d$, let $q_{i+1}$ be the furthest point from $D \setminus \{q_1, \ldots, q_i\}$ to $\text{aff}(q_1, \ldots, q_i)$. Next, we let $q_{d+2}, \ldots, q_{d+a+1}$ be the $a$ points of $Q$ inside $\Delta$ ordered lexicographically. The remaining $k - d - a - 1$ points $q_{d+a+2}, \ldots, q_{k}$ from $Q$ lie outside of $\Delta$ and we order them so that, for every $i = 1, \ldots, k - a - d - 1$, the point $q_{d+a+i+1}$ is closest to $\text{conv}(\{q_1, \ldots, q_{d+a+i}\})$ among the points $q_{d+a+i+1}, \ldots, q_{k}$. In case of a tie in any of the conditions, we choose the point with lexicographically smallest coordinates. Note, however, that a tie occurs with probability 0.

Clearly, there is a unique such ordering $q_1, \ldots, q_k$ of $Q$. We call this ordering the canonical $(k, a)$-ordering of $Q$. To reformulate, an ordering $q_1, \ldots, q_k$ of $Q$ is the canonical $(k, a)$-ordering of $Q$ if and only if the following five conditions are satisfied:

- (L1) The $d$-dimensional simplex $\Delta$, with vertices $q_1, \ldots, q_{d+1}$ has the largest $d$-dimensional Lebesgue measure among all $d$-dimensional simplices spanned by points from $Q$.
- (L2) For every $i = 1, \ldots, d - 1$, the point $q_{i+1}$ has the largest distance among all points from $\{q_1, \ldots, q_i\}$ to the $(i-1)$-dimensional affine subspace $\text{aff}(q_1, \ldots, q_i)$ spanned by $q_1, \ldots, q_i$. Moreover, $q_1$ is lexicographically smaller than $q_2$.
- (L3) For every $i = 1, \ldots, d - 1$, the distance between $q_{i+1}$ and $\text{aff}(q_1, \ldots, q_i)$ is at least as large as the distance between $q_{d+1}$ and $\text{aff}(q_1, \ldots, q_i)$. Also, the distance between $q_1$ and $q_2$ is at least as large as the distance between $q_{d+1}$ and any $q_i$ with $i \in \{1, \ldots, d\}$.
- (L4) The points $q_{d+2}, \ldots, q_{d+a+1}$ lie inside $\Delta$ and are ordered lexicographically.
- (L5) The points $q_{d+a+2}, \ldots, q_k$ lay outside of $\Delta$. For every $i = 1, \ldots, k - a - d - 1$, the point $q_{d+a+i+1}$ is closest to $\text{conv}(\{q_1, \ldots, q_{d+a+i}\})$ among the points $q_{d+a+i+1}, \ldots, q_k$. 

SoCG 2020
Figure 2 gives an illustration in $\mathbb{R}^2$. We note that the conditions (L2) and (L3) can be merged together. However, later in the proof, we use the fact that the probability that the points from $Q$ satisfy the condition (L2) equals $1/d!$, so we stated the two conditions separately.

![Diagram](image)

**Figure 2** An illustration of the canonical $(k,a)$-ordering of a planar point set $Q$. Here we have $k = 12$ points and $a = 4$ of the points lie inside the largest area triangle $\Delta$ with vertices $q_1, q_2, q_3$.

Before going into details, we first give a high-level overview of the proof of Theorem 1. First, we prove an $O(1/n^{d+1})$ bound on the probability that $\triangle$ contains precisely the points $p_{d+2}, \ldots, p_{d+a}$ from $S$ (Lemma 9), which means that the points $p_1, \ldots, p_{d+a}$ determine an island in $S$. Next, for $i = d + 2 + a, \ldots, k$, we show that, conditioned on the fact that the $(i - 1)$-tuple $(p_1, \ldots, p_{i-1})$ determines an island in $S$ in the canonical $(k,a)$-ordering, the $i$-tuple $(p_1, \ldots, p_i)$ determines an island in $S$ in the canonical $(k,a)$-ordering with probability $O(1/n)$ (Lemma 10). Then it immediately follows that the probability that $I$ determines a $k$-island in $S$ with the desired properties is at most $O\left(1/n^{a+1} \cdot (1/n)^{k-(d+1+a)}\right) = O(n^{d-k})$. Since there are $n \cdot (n - 1) \cdots (n - k + 1) = O(n^k)$ possibilities to select such an ordered subset $I$ and each $k$-island in $S$ is counted at most $k!$ times, we obtain the desired bound $O\left(n^k \cdot n^{d-k} \cdot k!\right) = O(n^d)$ on the expected number of $k$-islands in $S$.

We now proceed with the proof. Let $p_1, \ldots, p_k$ be points from $S$ in the order in which they are drawn from $K$. We use $\Delta$ to denote the $d$-dimensional simplex with vertices $p_1, \ldots, p_{d+1}$. We eventually show that the probability that $p_1, \ldots, p_k$ is the canonical $(k,a)$-ordering of a $k$-island in $S$ for some $a$ is at most $O(1/n^{k-d})$. First, however, we need to state some notation and prove some auxiliary results.

Consider the points $p_1, \ldots, p_d$. Without loss of generality, we can assume that, for each $i = 1, \ldots, d$, the point $p_i$ has the last $d - i + 1$ coordinates equal to zero. Otherwise we apply a suitable isometry to $S$. Then, for every $i = 1, \ldots, d$, the distance between $p_{i+1}$ and the $(i-1)$-dimensional affine subspace spanned by $p_1, \ldots, p_i$ is equal to the absolute value of the $i$th coordinate of $p_{i+1}$. Moreover, after applying a suitable rotation, we can also assume that the first coordinate of each of the points $p_1, \ldots, p_d$ is nonnegative.

Let $\Delta_0$ be the $(d - 1)$-dimensional simplex with vertices $p_1, \ldots, p_d$ and let $H$ be the hyperplane containing $\Delta_0$. Note that, according to our assumptions about $p_1, \ldots, p_d$, we have $H = \{(x_1, \ldots, x_d) \in \mathbb{R}^d : x_d = 0\}$. Let $B$ be the set of points $(x_1, \ldots, x_d) \in \mathbb{R}^d$ that satisfy the following three conditions:

(i) $x_1 \geq 0$,

(ii) $|x_i|$ is at most as large as the absolute value of the $i$th coordinate of $p_{i+1}$ for every $i \in \{1, \ldots, d - 1\}$, and

(iii) $|x_d| \leq d/\lambda_{d-1}(\Delta_0)$. 
See Figures 3a and 3b for illustrations in $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Observe that $B$ is a $d$-dimensional axis-parallel box. For $h \in \mathbb{R}$, we use $I_h$ to denote the intersection of $B$ with the hyperplane $x_d = h$.

**Figure 3** An illustration of the proof of Theorem 1 in (a) $\mathbb{R}^2$ and (b) $\mathbb{R}^3$.

Having fixed $p_1, \ldots, p_d$, we now try to restrict possible locations of the points $p_{d+1}, \ldots, p_k$, one by one, so that $p_1, \ldots, p_k$ is the canonical $(k, a)$-ordering of a $k$-island in $S$ for some $a$.

First, we observe that the position of the point $p_{d+1}$ is restricted to $B$.

**Lemma 6.** If $p_1, \ldots, p_{d+1}$ satisfy condition (L3), then $p_{d+1}$ lies in the box $B$.

**Proof.** Let $p_{d+1} = (x_1, \ldots, x_d)$. According to our choice of points $p_1, \ldots, p_d$ and from the assumption that $p_1, \ldots, p_d$ satisfy (L3), we get $x_1 \geq 0$ and also that $|x_i|$ is at most as large as the absolute value of the $i$th coordinate of $p_{i+1}$ for every $i \in \{1, \ldots, d-1\}$.

It remains to show that $|x_d| \leq d/\lambda_{d-1}(\Delta_0)$. The simplex $\Delta$ spanned by $p_1, \ldots, p_{d+1}$ is contained in the convex body $K$, as $p_1, \ldots, p_{d+1} \in K$ and $K$ is convex. Thus $\lambda_d(\Delta) \leq \lambda_d(K) = 1$. On the other hand, the volume $\lambda_d(\Delta)$ equals $\lambda_{d-1}(\Delta_0) \cdot h/d$, where $h$ is the distance between $p_{d+1}$ and the hyperplane $H$ containing $\Delta_0$. According to our assumptions about $p_1, \ldots, p_d$, the distance $h$ equals $|x_d|$. Since $\lambda_d(\Delta) \leq 1$, it follows that $|x_d| = h \leq d/\lambda_{d-1}(\Delta_0)$ and thus $p_{d+1} \in B$.

The following auxiliary lemma gives an identity that is needed later. We omit the proof, which can be found, for example, in [2, Section 1].

**Lemma 7 ([2]).** For all nonnegative integers $a$ and $b$, we have

\[
\int_0^1 x^a(1-x)^b \, dx = \frac{a! \, b!}{(a+b+1)!}.
\]

We will also use the following result, called the Asymptotic Upper Bound Theorem [14], that estimates the maximum number of facets in a polytope.
Let \( a \) be an integer satisfying \( 0 \leq a \leq k - d - 1 \) and let \( E_a \) be the event that \( p_1, \ldots, p_k \) is the canonical \((k, a)\)-ordering such that \( \{p_1, \ldots, p_{d+a+1}\} \) is an island in \( S \). To estimate the probability that \( p_1, \ldots, p_k \) is the canonical \((k, a)\)-ordering of a \( k \)-island in \( S \), we first find an upper bound on the conditional probability of \( E_a \), conditioned on the event \( L_2 \) that \( p_1, \ldots, p_d \) satisfy (L2).

**Theorem 8 (Asymptotic Upper Bound Theorem [14]).** For every integer \( d \geq 2 \), a \( d \)-dimensional convex polytope with \( N \) vertices has at most \( 2^{(N/d)} \) facets.

\begin{equation}
\text{For every integer } d \geq 2, \text{ a } d\text{-dimensional convex polytope with } N \text{ vertices has at most } 2^{(N/d)} \text{ facets.}
\end{equation}

**Lemma 9.** For every \( a \in \{0, \ldots, k - d - 1\} \), the probability \( \Pr[E_a \mid L_2] \) is at most

\begin{equation}
\frac{2^{d-1} \cdot d!}{(k - a - d - 1)! \cdot (n - k + 1)^{d+1}}.
\end{equation}

**Proof.** It follows from Lemma 6 that, in order to satisfy (L3), the point \( p_{d+1} \) must lie in the box \( B \). In particular, \( p_{d+1} \) is contained in \( I_h \cap K \) for some real number \( h \in [-d/\lambda_{d-1}(\Delta_0), d/\lambda_{d-1}(\Delta_0)] \). If \( p_{d+1} \in I_h \), then the simplex \( \Delta = \text{conv}\{p_1, \ldots, p_{d+a+1}\} \) has volume \( \lambda_d(\Delta) = \lambda_{d-1}(\Delta_0) \cdot |h|/d \) and the \( a \) points \( p_{d+2}, \ldots, p_{d+a+1} \) satisfy (L4) with probability

\begin{equation}
\frac{1}{a!} \cdot \left( \frac{\lambda_{d-1}(\Delta_0) \cdot |h|}{d} \right)^a,
\end{equation}

as they all lie in \( \Delta \subseteq K \) in the unique order.

In order to satisfy the condition (L5), the \( k - a - d - 1 \) points \( p_{d+a+1}, \ldots, p_{d+a+i} \), for \( i \in \{1, \ldots, k - a - d - 1\} \), must have increasing distance to \( \text{conv}\{p_1, \ldots, p_{d+a+i}\} \) as the index \( i \) increases, which happens with probability at most \( \frac{1}{(k - a - d - 1)^{d+1}} \). Since \( \{p_1, \ldots, p_{d+a+1}\} \) must be an island in \( S \), the \( n - d - a - 1 \) points from \( S \setminus \{p_1, \ldots, p_{d+a+1}\} \) must lie outside \( \Delta \). If \( p_{d+1} \in I_h \), then this happens with probability

\begin{equation}
(\lambda_d(K \setminus \Delta))^{n-d-a-1} = (\lambda_d(K) - \lambda_d(\Delta))^{n-d-a-1} = \left( 1 - \frac{\lambda_{d-1}(\Delta_0) \cdot |h|}{d} \right)^{n-d-a-1},
\end{equation}

as they all lie in \( K \setminus \Delta \) and we have \( \Delta \subseteq K \) and \( \lambda_d(K) = 1 \).

Altogether, we get that \( \Pr[E_a \mid L_2] \) is at most

\begin{equation}
\frac{d/\lambda_{d-1}(\Delta_0)}{a! \cdot (k - a - d - 1)!} \cdot \left( \frac{\lambda_{d-1}(\Delta_0) \cdot |h|}{d} \right)^a \left( 1 - \frac{\lambda_{d-1}(\Delta_0) \cdot |h|}{d} \right)^{n-d-a-1} dh.
\end{equation}

Since we have \( \lambda_{d-1}(I_h) = \lambda_{d-1}(I_k) \) for every \( h \in [-d/\lambda_{d-1}(\Delta_0), d/\lambda_{d-1}(\Delta_0)] \), we obtain \( \lambda_{d-1}(I_h \cap K) \leq \lambda_{d-1}(I_k) \) and thus \( \Pr[E_a \mid L_2] \) is at most

\begin{equation}
\frac{2 \cdot \lambda_{d-1}(I_0)}{a! \cdot (k - a - d - 1)!} \int_0^{d/\lambda_{d-1}(\Delta_0) \cdot h} \left( \frac{\lambda_{d-1}(\Delta_0) \cdot h}{d} \right)^a \left( 1 - \frac{\lambda_{d-1}(\Delta_0) \cdot h}{d} \right)^{n-d-a-1} dh.
\end{equation}

By substituting \( t = \frac{\lambda_{d-1}(\Delta_0) \cdot h}{d} \), we obtain

\begin{equation}
\Pr[E_a \mid L_2] \leq \frac{2 \cdot \lambda_{d-1}(I_0)}{a! \cdot (k - a - d - 1)! \cdot \lambda_{d-1}(\Delta_0)} \int_0^1 t^n (1 - t)^{n-d-a-1} dt.
\end{equation}
By Lemma 7, the right side in the above inequality equals

\[
\frac{2d \cdot \lambda_{d-1}(I_0)}{a! \cdot (k-a-d-1)! \cdot \lambda_{d-1}(\Delta_0)} \cdot \frac{a! \cdot (n-d-a-1)!}{(n-d)!} = \frac{2d \cdot \lambda_{d-1}(I_0)}{(k-a-d-1)! \cdot \lambda_{d-1}(\Delta_0)} \cdot \frac{(n-d-a-1)!}{(n-d)!}.
\]

For every \(i = 1, \ldots, d-1\), let \(h_i\) be the distance between the point \(p_{i+1}\) and the \((i-1)\)-dimensional affine subspace spanned by \(p_1, \ldots, p_i\). Since the volume of the box \(I_0\) satisfies

\[\lambda_{d-1}(I_0) = h_1(2h_2) \cdots (2h_{d-1}) = 2^{d-2} \cdot h_1 \cdots h_{d-1}\]

and the volume of the \((d-1)\)-dimensional simplex \(\Delta_0\) is

\[\lambda_{d-1}(\Delta_0) = \frac{h_1}{1} \cdot \frac{h_2}{2} \cdots \frac{h_{d-1}}{d-1} = \frac{h_1 \cdots h_{d-1}}{(d-1)!},\]

we obtain \(\lambda_{d-1}(I_0)/\lambda_{d-1}(\Delta_0) = 2^{d-2} \cdot (d-1)!\). Thus

\[
\Pr[E_a \mid L_2] \leq \frac{2^{d-1} \cdot d!}{(k-a-d-1)! \cdot \lambda_{d-1}(\Delta_0)} \cdot \frac{(n-d-a-1)!}{(n-d)!} \leq \frac{2^{d-1} \cdot d!}{(k-a-d-1)! \cdot (n-d) \cdots (n-d-a)} \leq \frac{2^{d-1} \cdot d!}{(k-a-d-1)! \cdot (n-k+1)^{a+1}},
\]

where the last inequality follows from \(a \leq k-d-1\).

For every \(i \in \{d+a+1, \ldots, k\}\), let \(E_{a,i}\) be the event that \(p_1, \ldots, p_k\) is the canonical \((k,a)\)-ordering such that \(\{p_1, \ldots, p_i\}\) is an island in \(S\). Note that in the event \(E_{a,i}\) the condition (L5) implies that \(\{p_1, \ldots, p_j\}\) is an island in \(S\) for every \(j \in \{d+a+1, \ldots, i\}\). Thus we have

\[L_2 \supseteq E_a = E_{a,d+a+1} \supseteq E_{a,d+a+2} \supseteq \ldots \supseteq E_{a,k}.\]

Moreover, the event \(E_{a,k}\) says that \(p_1, \ldots, p_k\) is the canonical \((k,a)\)-ordering of a \(k\)-island in \(S\). For \(i \in \{d+a+2, \ldots, k\}\), we now estimate the conditional probability of \(E_{a,i}\), conditioned on \(E_{a,i-1}\).

\begin{lemma}
For every \(i \in \{d+a+2, \ldots, k\}\), we have

\[
\Pr[E_{a,i} \mid E_{a,i-1}] \leq \frac{2^{d-1} \cdot \binom{k}{d/2}}{n-i+1}.
\]

\end{lemma}

**Proof.** Let \(i \in \{d+a+2, \ldots, k\}\) and assume that the event \(E_{a,i-1}\) holds. That is, \(p_1, \ldots, p_k\) is the canonical \((k,a)\)-ordering such that \(\{p_1, \ldots, p_{i-1}\}\) is an \((i-1)\)-island in \(S\).

First, we assume that \(\Delta\) is a regular simplex with height \(\eta > 0\). At the end of the proof we show that the case when \(\Delta\) is an arbitrary simplex follows by applying a suitable affine transformation.

For every \(j \in \{1, \ldots, d+1\}\), let \(F_j\) be the facet \(\text{conv}(\{p_1, \ldots, p_{d+1}\} \setminus \{p_j\})\) of \(\Delta\) and let \(H_j\) be the hyperplane parallel to \(F_j\) that contains \(p_j\). We use \(H_j^+\) to denote the halfspace determined by \(H_j\) such that \(\Delta \subseteq H_j^+\). We set \(\Delta^* = \cap_{j=1}^{d+1} H_j^+\); see Figures 4a and 4b for illustrations in \(\mathbb{R}^2\) and \(\mathbb{R}^3\), respectively. Note that \(\Delta^*\) is a \(d\)-dimensional simplex containing \(\Delta\). Also, notice that if \(x \notin \Delta^*\), then \(x \notin H_j^+\) for some \(j\) and the distance between \(x\) and the hyperplane containing \(F_j\) is larger than \(\eta\).
We show that the fact that \(p_1, \ldots, p_k\) is the canonical \((k, a)\)-ordering implies that every point from \(\{p_1, \ldots, p_k\}\) is contained in \(\Delta^*\). Suppose for contradiction that some point \(p \in \{p_1, \ldots, p_k\}\) does not lie inside \(\Delta^*\). Then there is a facet \(F_j\) of \(\Delta\) such that the distance \(\eta'\) between \(p\) and the hyperplane containing \(F_j\) is larger than \(\eta\). Then, however, the simplex \(\Delta'\) spanned by vertices of \(F_j\) and by \(p\) has volume larger than \(\Delta\), because

\[
\lambda_d(\Delta') = \frac{1}{d} \cdot \lambda_{d-1}(F_j) \cdot \eta' > \frac{1}{d} \cdot \lambda_{d-1}(F_j) \cdot \eta = \lambda_d(\Delta).
\]

This contradicts the fact that \(p_1, \ldots, p_k\) is the canonical \((k, a)\)-ordering, as, according to (L1), \(\Delta\) has the largest \(d\)-dimensional Lebesgue measure among all \(d\)-dimensional simplices spanned by points from \(\{p_1, \ldots, p_k\}\).

Let \(\sigma\) be the barycenter of \(\Delta\). For every point \(p \in \Delta^* \setminus \Delta\), the line segment \(\sigma p\) intersects at least one facet of \(\Delta\). For every \(j \in \{1, \ldots, d + 1\}\), we use \(R_j\) to denote the set of points \(p \in \Delta^* \setminus \Delta\) for which the line segment \(\sigma p\) intersects the facet \(F_j\) of \(\Delta\). Observe that each set \(R_j\) is convex and the sets \(R_1, \ldots, R_{d+1}\) partition \(\Delta^* \setminus \Delta\) (up to their intersection of \(d\)-dimensional Lebesgue measure 0); see Figure 5 for an illustration in the plane.

Consider the point \(p_i\). Since \(p_1, \ldots, p_k\) is the canonical \((k, a)\)-ordering, the condition (L5) implies that \(p_i\) lies outside of the polytope \(\text{conv}\{(p_1, \ldots, p_{i-1})\}\). To bound the probability \(\text{Pr}[E_{a,i} \mid E_{a,i-1}]\), we need to estimate the probability that \(\text{conv}\{(p_1, \ldots, p_i)\} \setminus \text{conv}\{(p_1, \ldots, p_{i-1})\}\) does not contain any point from \(S \setminus \{p_1, \ldots, p_i\}\), conditioned on \(E_{a,i-1}\).

We know that \(p_i\) lies in \(\Delta^* \setminus \Delta\) and that \(p_i \in R_j\) for some \(j \in \{1, \ldots, d + 1\}\).

Since \(p_i \notin \text{conv}\{(p_1, \ldots, p_{i-1})\}\), there is a facet \(\varphi\) of the polytope \(\text{conv}\{(p_1, \ldots, p_{i-1})\}\) contained in the closure of \(R_j\) such that \(\sigma p_i\) intersects \(\varphi\). Since \(S\) is in general position with probability 1, we can assume that \(\varphi\) is a \((d - 1)\)-dimensional simplex. The point \(p_i\) is contained in the convex set \(C_\varphi\) that contains all points \(c \in \mathbb{R}^d\) such that the line segment \(\sigma c\) intersects \(\varphi\). We use \(H(0)\) to denote the hyperplane containing \(\varphi\). For a positive \(r \in \mathbb{R}\), let \(H(r)\) be the hyperplane parallel to \(H(0)\) at distance \(r\) from \(H(0)\) such that \(H(r)\) is contained in the halfspace determined by \(H(0)\) that does not contain \(\text{conv}\{(p_1, \ldots, p_{i-1})\}\). Then we have \(p_i \in H(h)\) for some positive \(h \in \mathbb{R}\).

Since \(p_i \in K\) and \(\varphi \subseteq K\), the convexity of \(K\) implies that the simplex \(\text{conv}\{(\varphi \cup \{p_i\})\}\) has volume \(\lambda_d(\text{conv}\{(\varphi \cup \{p_i\})\}) \leq \lambda_d(K) = 1\). Since \(\lambda_d(\text{conv}\{(\varphi \cup \{p_i\})\}) = \lambda_{d-1}(\varphi) \cdot h/d\), we obtain \(h \leq d/\lambda_{d-1}(\varphi)\).
We denote by \( \text{conv} \) the convex hull of a finite set of points. Note that the parameters \( \eta \) and \( \tau \) coincide for \( d = 2 \), as then \( \tau = \frac{d^2-1}{d+1}\eta = \eta \).

The point \( p_i \) lies in the \((d-1)\)-dimensional simplex \( C_\varphi \cap H(h) \), which is a scaled copy of \( \varphi \). We show that

\[
\lambda_{d-1}(C_\varphi \cap H(h)) \leq d^{d-2} \cdot \lambda_{d-1}(\varphi).
\]

Let \( h_\varphi \) be the distance between \( H(0) \) and \( \sigma \) and, for every \( j \in \{1, \ldots, d+1\} \), let \( \overline{H}_j \) be the hyperplane parallel to \( F_j \) containing the vertex \( H_1 \cap \cdots \cap H_{j-1} \cap H_{j+1} \cap \cdots \cap H_{d+1} \) of \( \Delta^* \). We denote by \( \overline{H}_j^+ \) the halfspace determined by \( \overline{H}_j \) containing \( \Delta^* \). Since \( \Delta \) lies on the same side of \( H\) as \( \sigma \), we see that \( h_\varphi \) is at least as large as the distance between \( \sigma \) and \( F_J \), which is \( \eta/(d+1) \). Since \( p_i \) lies in \( \Delta^* \subseteq \overline{H}_j^+ \), we see that \( h \) is at most as large as the distance \( \tau \) between \( \overline{H}_j \) and the hyperplane containing the facet \( F_J \) of \( \Delta \). Note that \( \tau + \eta/(d+1) \) is the distance of the barycenter of \( \Delta^* \) and a vertex of \( \Delta^* \) and \( d\eta/(d+1) \) is the distance of the barycenter of \( \Delta^* \) and a facet of \( \Delta^* \). Thus we get \( \tau = \frac{d^2-1}{d^2+1} - \frac{d}{d^2+1} = \frac{d^2-1}{d^2+1} \eta \) from the fact that the distance between the barycenter of a \( d \)-dimensional simplex and any of its vertices is \( d \)-times as large as the distance between the barycenter and a facet. Consequently, \( h \leq \frac{d^2-1}{d^2+1} \eta \) and \( \frac{d}{d^2+1} \leq h_\varphi \), which implies \( h \leq (d^2-1)h_\varphi \). Thus \( C_\varphi \cap H(h) \) is a scaled copy of \( \varphi \) by a factor of size at most \( d^2 \). This gives \( \lambda_{d-1}(C_\varphi \cap H(h)) \leq d^{d-2} \cdot \lambda_{d-1}(\varphi) \).

Since the simplex \( \text{conv}(\varphi \cup \{p_i\}) \) is a subset of the closure of \( \text{conv}(\{p_1, \ldots, p_i\} \setminus \text{conv}(\{p_1, \ldots, p_{i-1}\})) \), the probability \( \Pr[\mathbf{E}_{a,i} \mid \mathbf{E}_{a,i-1}] \) can be bounded from above by the conditional probability of the event \( A_{i,\varphi} \) that \( p_i \in C_\varphi \cap K \) and that no point from \( S \setminus \{p_1, \ldots, p_i\} \) lies in \( \text{conv}(\varphi \cup \{p_i\}) \), conditioned on \( \mathbf{E}_{a,i-1} \). All points from \( S \setminus \{p_1, \ldots, p_i\} \) lie outside of \( \text{conv}(\varphi \cup \{p_i\}) \) with probability

\[
\left(1 - \frac{\lambda_d(\text{conv}(\varphi \cup \{p_i\}))}{\lambda_d(K \setminus \text{conv}(\{p_1, \ldots, p_{i-1}\}))}\right)^{n-i}.
\]
Since \( \lambda_d(K \setminus \text{conv}(\{p_1, \ldots, p_{i-1}\})) \leq \lambda_d(K) = 1 \), this is bounded from above by

\[
(1 - \lambda_d(\text{conv}(\{p_1\})))^{n-i} = \left(1 - \frac{\lambda_{d-1}(\varphi) \cdot h}{d}\right)^{n-i}.
\]

Since the sets \( C_\varphi \) partition \( K \setminus \text{conv}(\{p_1, \ldots, p_{i-1}\}) \) (up to intersections of \( d \)-dimensional Lebesgue measure 0) and since \( h \leq d/\lambda_{d-1}(\varphi) \), we have, by the law of total probability,

\[
\Pr[E_{a,i} \mid E_{a,i-1}] \leq \sum_{\varphi} \Pr[A_{i,\varphi} \mid E_{a,i-1}]
\]

\[
\leq \sum_{\varphi} \int_0^{d/\lambda_{d-1}(\varphi)} \lambda_{d-1}(C_\varphi \cap H(h)) \cdot \left(1 - \frac{\lambda_{d-1}(\varphi) \cdot h}{d}\right)^{n-i} \, dh.
\]

The sums in the above expression are taken over all facets \( \varphi \) of the convex polytope \( \text{conv}(\{p_1, \ldots, p_{i-1}\}) \). Using (3), we can estimate \( \Pr[E_{a,i} \mid E_{a,i-1}] \) from above by

\[
d^{2d-2} \cdot \sum_{\varphi} \frac{d}{\lambda_{d-1}(\varphi)} \cdot \int_0^{d/\lambda_{d-1}(\varphi)} \left(1 - \frac{\lambda_{d-1}(\varphi) \cdot h}{d}\right)^{n-i} \, dh.
\]

By substituting \( t = \frac{\lambda_{d-1}(\varphi) h}{d} \), we can rewrite this expression as

\[
d^{2d-2} \cdot \sum_{\varphi} \left(1 - \frac{\lambda_{d-1}(\varphi)}{d} \cdot t\right)^{n-i} \, dt = d^{2d-1} \cdot \sum_{\varphi} \int_0^1 (1 - t)^{n-i} \, dt.
\]

By Lemma 7, this equals

\[
d^{2d-1} \cdot \sum_{\varphi} \frac{0! \cdot (n-i)!}{(n-i+1)!} = \frac{d^{2d-1}}{n-i+1} \sum_{\varphi} 1.
\]

Since \( \text{conv}(\{p_1, \ldots, p_{i-1}\}) \) is a convex polytope in \( \mathbb{R}^d \) with at most \( i-1 \leq k \) vertices, Theorem 8 implies that the number of facets \( \varphi \) of \( \text{conv}(\{p_1, \ldots, p_{i-1}\}) \) is at most \( 2 \binom{k}{(d/2)} \). Altogether, we have derived the desired bound

\[
\Pr[E_{a,i} \mid E_{a,i-1}] \leq \frac{2d^{2d-1} \cdot \binom{k}{(d/2)}}{n-i+1}
\]

in the case when \( \Delta \) is a regular simplex.

If \( \Delta \) is not regular, we first apply a volume-preserving affine transformation \( F \) that maps \( \Delta \) to a regular simplex \( F(\Delta) \). The simplex \( F(\Delta) \) is then contained in the convex body \( F(K) \) of volume 1. Since \( F \) translates the uniform distribution on \( F(K) \) to the uniform distribution on \( K \) and preserves holes and islands, we obtain the required upper bound also in the general case. □

Now, we finish the proof of Theorem 1.

**Proof of Theorem 1.** We estimate the expected value of the number \( X \) of \( k \)-islands in \( S \). The number of ordered \( k \)-tuples of points from \( S \) is \( n(n-1) \cdots (n-k-1) \). Since every subset of \( S \) of size \( k \) admits a unique labeling that satisfies the conditions (L1), (L2), (L3), (L4), and (L5), we have
\[E[X] = n(n - 1) \cdots (n - k + 1) \cdot \Pr \left[ \bigcup_{a=0}^{k-d-1} E_{a,k} \right] \]
\[= n(n - 1) \cdots (n - k + 1) \cdot \sum_{a=0}^{k-d-1} \Pr[E_{a,k}],\]
as the events \(E_{a,k}, \ldots, E_{k-d-1,k}\) are pairwise disjoint.

The probability of the event \(L_2\), which says that the points \(p_1, \ldots, p_d\) satisfy the condition (L2), is \(1/d!\). Let \(P = \sum_{a=0}^{k-d-1} \Pr[E_{a,k} \mid L_2]\). For any two events \(E, E'\) with \(E \supseteq E'\) and \(\Pr[E] > 0\), we have \(\Pr[E'] = \Pr[E \cap E'] = \Pr[E' \mid E] \cdot \Pr[E]\). Thus, using \(L_2 \supseteq E_a = E_{a,d+a+1} \supseteq E_{a,d+a+2} \supseteq \cdots \supseteq E_{a,k}\), we get
\[E[X] = n(n - 1) \cdots (n - k + 1) \cdot \Pr[L_2] \cdot P = \frac{n(n - 1) \cdots (n - k + 1)}{d!}, \]
and
\[P = \sum_{a=0}^{k-d-1} \Pr[E_a \mid L_2] \cdot \prod_{i=d+a+2}^{k} \Pr[E_{a,i} \mid E_{a,i-1}].\]

For every \(a \in \{d + 2, \ldots, k - d - 1\}\), Lemma 9 gives
\[\Pr[E_a \mid L_2] \leq \frac{2^{d-1} \cdot d!}{(k - a - d - 1)! \cdot (n - k + 1)^{a+1}} \leq \frac{2^{d-1} \cdot d!}{(n - k + 1)^{a+1}}\]
and, due to Lemma 10,
\[\Pr[E_{a,i} \mid E_{a,i-1}] \leq \frac{2^{d^2-1} \cdot \binom{k}{d/2}}{n - i + 1}\]
for every \(i \in \{d + a + 2, \ldots, k\}\).

Using these estimates we derive
\[P \leq 2^{d-1} \cdot d! \cdot \left(\frac{2^{d^2-1} \cdot \binom{k}{d/2}}{d!}\right)^{k-d-1} \sum_{a=0}^{k-d-1} \frac{1}{(n - k + 1)^{a+1}} \cdot \prod_{i=d+a+2}^{k} \frac{1}{n - i + 1} \]
\[\leq 2^{d-1} \cdot d! \cdot \left(\frac{2^{d^2-1} \cdot \binom{k}{d/2}}{d!}\right)^{k-d-1} \sum_{a=0}^{k-d-1} \frac{1}{(n - k + 1)^{a+1}} \cdot \frac{1}{(n - k + 1)^{k-d-a-1}} \]
\[= 2^{d-1} \cdot d! \cdot \left(\frac{2^{d^2-1} \cdot \binom{k}{d/2}}{d!}\right)^{k-d-1} \cdot (k - d) \cdot \frac{1}{(n - k + 1)^{k-d}}.\]

Thus the expected number of \(k\)-islands in \(S\) satisfies
\[E[X] = \frac{n(n - 1) \cdots (n - k + 1)}{d!} \cdot P \]
\[\leq \frac{2^{d-1} \cdot d! \cdot \left(\frac{2^{d^2-1} \cdot \binom{k}{d/2}}{d!}\right)^{k-d-1} \cdot (k - d)}{d!} \cdot \frac{n(n - 1) \cdots (n - k + 1)}{(n - k + 1)^{k-d}} \]
\[= 2^{d-1} \cdot \left(\frac{2^{d^2-1} \cdot \binom{k}{d/2}}{d!}\right)^{k-d-1} \cdot (k - d) \cdot \frac{n(n - 1) \cdots (n - k + 2)}{(n - k + 1)^{k-d-1}}.\]
This finishes the proof of Theorem 1.
In the rest of the section, we sketch the proof of Theorem 2 by showing that a slight modification of the above proof yields an improved bound on the expected number $EH_{d,k}^n(n)$ of $k$-holes in $S$.

**Sketch of the proof of Theorem 2.** If $k$ points from $S$ determine a $k$-hole in $S$, then, in particular, the simplex $\Delta$ contains no points of $S$ in its interior. Therefore

$$EH_{d,k}^n(n) \leq n(n-1) \cdots (n-k+1) \cdot \Pr[\emptyset].$$

Then we proceed exactly as in the proof of Theorem 1, but we only consider the case $a=0$. This gives the same bounds as before with the term $(k-d)$ missing and with an additional factor $rac{1}{(k-d-1)!}$ from Lemma 9, which proves Theorem 2.

For $d=2$ and $k=4$, Theorem 2 gives $EH_{2,4}^n(n) \leq 128n^2 + o(n^2)$. We can obtain an even better estimate $EH_{2,4}^n(n) \leq 12n^2 + o(n^2)$ in this case. First, we have only three facets $\varphi$, as they correspond to the sides of the triangle $\Delta$. Thus the term $\left(2^{(\frac{k}{d-2})}\right)^{k-d-1} = 8$ is replaced by $3$. Moreover, the inequality (3) can be replaced by

$$\lambda_1(C_\varphi \cap H(h) \cap \Delta^*) \leq \lambda_1(\varphi),$$

since every line $H(h)$ intersects $R_j \subset \Delta^*$ in a line segment of length at most $\lambda_1(F_j) = \lambda(\varphi)$. This then removes the factor $d^{(2d-2)(k-d-1)} = 4$.

### 4 Proof of Theorem 5

Here, for every $d$, we state the definition of a $d$-dimensional analogue of Horton sets on $n$ points from [18] and show that, for all fixed integers $d$ and $k$, every $d$-dimensional Horton set $H$ with $n$ points contains at least $\Omega(n^{\min\{2d-1,k\}})$ $k$-islands in $H$. If $k \leq 3 \cdot 2^{d-1}$, then we show that $H$ contains at least $\Omega(n^k)$ $k$-holes in $H$.

First, we need to introduce some notation. A set $Q$ of points in $\mathbb{R}^d$ is in strongly general position if $Q$ is in general position and, for every $i=1,\ldots,d-1$, no $(i+1)$-tuple of points from $Q$ determines an $i$-dimensional affine subspace of $\mathbb{R}^d$ that is parallel to the $(d-i)$-dimensional linear subspace of $\mathbb{R}^d$ that contains the last $d-i$ axes. Let $\pi: \mathbb{R}^d \to \mathbb{R}^{d-1}$ be the projection defined by $\pi(x_1,\ldots,x_d) = (x_1,\ldots,x_{d-1})$. For $Q \subset \mathbb{R}^d$, we use $\pi(Q)$ to denote the set $\{\pi(q) \in \mathbb{R}^{d-1} : q \in Q\}$. If $Q$ is a set of $n$ points $q_0,\ldots,q_{n-1}$ from $\mathbb{R}^d$ in strongly general position that are ordered so that their first coordinates increase, then, for all $m \in \mathbb{N}$ and $i \in \{0,1,\ldots,m-1\}$, we define $Q_{i,m} = \{q_j \in Q : j \equiv i \pmod{m}\}$.

For two sets $A$ and $B$ of points from $\mathbb{R}^d$ with $|A|,|B| \geq d$, we say that $B$ is deep below $A$ and $A$ is high above $B$ if $B$ lies entirely below any hyperplane determined by $d$ points of $A$ and $A$ lies entirely above any hyperplane determined by $d$ points of $A$. For point sets $A'$ and $B'$ in $\mathbb{R}^d$ of arbitrarily size, we say that $B'$ is deep below $A'$ and $A'$ is high above $B'$ if there are sets $A \supset A'$ and $B \supset B'$ such that $|A|,|B| \geq d$, $B$ is deep below $A$, and $A$ is high above $B$.

Let $p_2 < p_3 < p_4 < \cdots$ be the sequence of all prime numbers. That is, $p_2 = 2$, $p_3 = 3$, $p_4 = 5$ and so on.

We can now state the definition of the $d$-dimensional Horton sets from [18]. Every finite set of $n$ points in $\mathbb{R}$ is a $1$-Horton. For $d \geq 2$, finite set $H$ of points from $\mathbb{R}^d$ in strongly general position is a $d$-Horton set if it satisfies the following conditions:
(a) the set $H$ is empty or it consists of a single point, or
(b) $H$ satisfies the following three conditions:

(i) if $d > 2$, then $\pi(H)$ is $(d - 1)$-Horton,
(ii) for every $i \in \{0, 1, \ldots, pd - 1\}$, the set $H_{i,pd}$ is $d$-Horton,
(iii) every $I \subseteq \{0, 1, \ldots, pd - 1\}$ with $|I| \geq 2$ can be partitioned into two nonempty subsets $J$ and $I \setminus J$ such that $\cup_{j \in J} H_{j,pd}$ lies deep below $\cup_{j \in I \setminus J} H_{j,pd}$.

Valtr [18] showed that such sets indeed exist and that they contain no $k$-hole with $k > 2^{d-1}(p_2p_3 \cdots p_d + 1)$. The 2-Horton sets are known as Horton sets. We show that $d$-Horton sets with $d \geq 3$ contain many $k$-islands for $k \geq d + 1$ and thus cannot provide the upper bound $O(n^d)$ that follows from Theorem 1. This contrasts with the situation in the plane, as 2-Horton sets of $n$ points contain only $O(n^2)$ $k$-islands for any fixed $k$ [8].

Let $d$ and $k$ be fixed positive integers. Assume first that $k \geq 2^{d-1}$. We want to prove that there are $\Omega(n^{2d-1})$ $k$-islands in every $d$-Horton set $H$ with $n$ points. We proceed by induction on $d$. For $d = 1$ there are $n - k + 1 = \Omega(n)$ $k$-islands in every 1-Horton set.

Assume now that $d > 1$ and that the statement holds for $d - 1$. The $d$-Horton set $H$ consists of $pd \in O(1)$ subsets $H_{i,pd}$, each of size at least $\lceil n/pd \rceil \in \Omega(n)$. The set $\{0, \ldots, pd - 1\}$ is ordered by a linear ordering $\prec$ such that, for all $i$ and $j$ with $i \prec j$, the set $H_{j,pd}$ is deep below $H_{i,pd}$; see [18]. Take two of sets $X = H_{a,pd}$ and $Y = H_{b,pd}$ such that $a \prec b$ are consecutive in $\prec$. Since $k \geq 2^{d-1}$, we have $\lceil k/2 \rceil \geq \lceil k/2 \rceil \geq 2^{d-2}$. Thus by the inductive hypothesis, the $(d - 1)$-Horton set $\pi(X)$ of size at least $\Omega(n)$ contains at least $\Omega(n^{2d-1})$ $\lceil k/2 \rceil$-islands. Similarly, the $(d - 1)$-Horton set $\pi(Y)$ of size at least $\Omega(n)$ contains at least $\Omega(n^{2d-1})$ $\lceil k/2 \rceil$-islands.

Let $\pi(A)$ be any of the $\Omega(n^{2d-1})$ $\lceil k/2 \rceil$-islands in $\pi(X)$, where $A \subseteq X$. Similarly, let $\pi(B)$ be any of the $\Omega(n^{2d-1})$ $\lceil k/2 \rceil$-islands in $\pi(Y)$, where $B \subseteq Y$. We show that $A \cup B$ is a $k$-island in $H$. Suppose for contradiction that there is a point $x \in H \setminus (A \cup B)$ that lies in $\text{conv}(A \cup B)$. Since $a$ and $b$ are consecutive in $\prec$, the point $x$ lies in $X \cup Y = H_{a,pd} \cup H_{b,pd}$. By symmetry, we may assume without loss of generality that $x \in X$. Since $x \notin A$ and $H$ is in strongly general position, we have $\pi(x) \in \pi(X) \setminus \pi(A)$. Using the fact that $\pi(A)$ is a $\lceil k/2 \rceil$-island in $\pi(X)$, we obtain $\pi(x) \notin \text{conv}(\pi(A))$ and thus $x \notin \text{conv}(A)$. Since $X$ is deep below $Y$, we have $x \notin \text{conv}(B)$. Thus, by Carathéodory’s theorem, $x$ lies in the convex hull of a $(d + 1)$-tuple $T \subseteq A \cup B$ that contains a point from $A$ and also a point from $B$.

Note that, for $U = (T \cup \{x\})$, we have $|U \cap A| \geq 2$, as $x \in A$ and $|T \cap A| \geq 1$. We also have $|U \cap B| \geq 2$, as $X$ is deep below $Y$ and $\pi(x) \notin \text{conv}(\pi(A))$. Thus the affine hull of $U \cap A$ intersects the convex hull of $U \cap B$. Then, however, the set $U \cap A$ is not deep below the set $U \cap B$, which contradicts the fact that $X$ is deep below $Y$.

Altogether, there are at least $\Omega(n^{2d-3}) \cdot \Omega(n^{2d-3}) \cdot \Omega(n^{2d-1})$ such $k$-islands $A \cup B$, which finishes the proof if $k$ is at least $2^{d-1}$. For $k < 2^{d-1}$, we use an analogous argument that gives at least $\Omega(n^{k/2}) \cdot \Omega(n^{k/2}) \cdot \Omega(n^k)$ $k$-islands in the inductive step.

If $d \geq 2$ and $k \leq 3 \cdot 2^{d-1}$ then a slight modification of the above proof gives $\Omega(n^{\min(2d-1,k)})$ $k$-islands which are actually $k$-holes in $H$. We just use the simple fact that every 2-Horton set with $n$ points contains $\Omega(n^2)$ $k$-holes for every $k \in \{2, \ldots, 6\}$ as our inductive hypothesis. This is trivial for $k = 2$ and it follows for $k \in \{3, 4\}$ from the well-known fact that every set of $n$ points in $\mathbb{R}^2$ in general position contains at least $\Omega(n^2)$ $k$-holes. For $k \in \{5, 6\}$, this fact can be proved using basic properties of 2-Horton sets (we omit the details). Then we use the inductive assumption, which says that every $d$-Horton set of $n$ points contains at least $\Omega(n^{\min(2d-1,k)})$ $k$-holes if $d \geq 2$ and $1 \leq k \leq 3 \cdot 2^{d-1}$. This finishes the proof of Theorem 5.
References


