

ON THE NUMBER OF MAXIMUM EMPTY BOXES AMIDST n POINTS*

Adrian Dumitrescu[†] Minghui Jiang[‡]

January 20, 2017

Abstract

We revisit the following problem (along with its higher dimensional variant): Given a set S of n points inside an axis-parallel rectangle U in the plane, find a maximum-area axis-parallel sub-rectangle that is contained in U but contains no points of S .

(I) We prove that the number of maximum-area empty rectangles amidst n points in the plane is $O(n \log n 2^{\alpha(n)})$, where $\alpha(n)$ is the extremely slowly growing inverse of Ackermann's function. The previous best bound, $O(n^2)$, is due to Naamad, Lee, and Hsu (1984).

(II) For any $d \geq 3$, we prove that the number of maximum-volume empty boxes amidst n points in \mathbb{R}^d is always $O(n^d)$ and sometimes $\Omega(n^{\lfloor d/2 \rfloor})$. This is the first superlinear lower bound derived for this problem.

(III) We discuss some algorithmic aspects regarding the search for a maximum empty box in \mathbb{R}^3 . In particular, we present an algorithm that finds a $(1 - \varepsilon)$ -approximation of the maximum empty box amidst n points in $O(\varepsilon^{-2} n^{5/3} \log^2 n)$ time.

Keywords: Maximum empty box, Davenport-Schinzel sequence, approximation algorithm, data mining.

1 Introduction

Given an axis-parallel rectangle U in the plane containing n points, MAXIMUM EMPTY RECTANGLE is the problem of computing a maximum-area axis-parallel empty sub-rectangle contained in U . This problem is one of the oldest in computational geometry, with multiple applications, e.g., in facility location problems [32]. In higher dimensions, finding the largest empty box has applications in data mining, such as finding large gaps in a multi-dimensional data set [22].

A *box* in \mathbb{R}^d , $d \geq 2$, is an open axis-parallel hyperrectangle $(a_1, b_1) \times \cdots \times (a_d, b_d)$ with $a_i < b_i$ for $1 \leq i \leq d$. Due to the fact that the volume ratio of any box inside another box is invariant under scaling, the problem can be reduced to the case when the enclosing box is a hypercube. Given a set S of n points in the unit hypercube $U_d = [0, 1]^d$, $d \geq 2$, an *empty box* is a box empty of points in S and contained in U_d , and MAXIMUM EMPTY BOX is the problem of finding an empty box with the *maximum* volume. Note that an empty box of maximum volume must be *maximal* with respect to inclusion. Some planar examples of maximal empty rectangles are shown in Fig. 1. All rectangles and boxes considered in this paper are axis-parallel.

*An extended abstract of this paper appeared in the Proceedings of the 32nd International Symposium on Computational Geometry, June 2016, Leibniz International Proceedings in Informatics (LIPIcs) series, Schloss Dagstuhl; DOI: 10.4230/LIPIcs.SoCG.2016.36.

[†]Department of Computer Science, University of Wisconsin–Milwaukee, USA. Email: dumitres@uwm.edu

[‡]Department of Computer Science, Utah State University, Logan, USA. Email: mjiang@cc.usu.edu

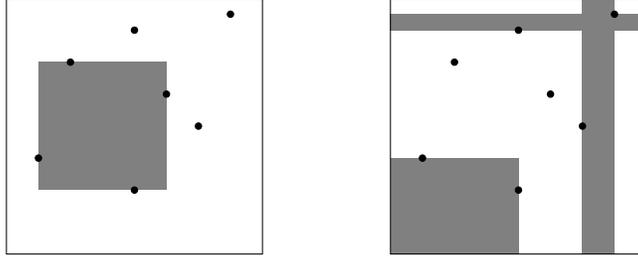


Figure 1: A maximal empty rectangle supported by one point on each side (left), and three maximal empty rectangles supported by both points and sides of $[0, 1]^2$ (right).

By slicing the hypercube with n parallel hyperplanes, each incident to one of the n points, the largest slice gives an empty box of volume at least $\frac{1}{n+1}$. For a fixed dimension d , one can exhibit point sets for which the largest empty box is $O(\frac{1}{n})$, and so the above estimate is the best possible, apart from constant factors (that depend on d) [35, 18, 5]. Knowing what to hope for in regard to the maximum volume, the next natural question is: how many such extremal boxes can there be?

The study of extremal problems on triangle areas determined by n points was initiated in a series of papers by Erdős and Purdy in the early 1970s [23]. It is known for instance that for n points in the plane, the maximum area of a triangle can occur at most n times [11].

In Section 3 we show that the number of maximum empty rectangles amidst n points is $O(n \log n 2^{\alpha(n)})$, where $\alpha(n)$ is the extremely slowly growing inverse of Ackermann's function. The previous best upper bound, due to Naamad, Lee, and Hsu from 1984 [32], is $O(n^2)$; this quadratic upper bound has been revisited numerous times [2, 3, 4, 6, 12, 15, 28, 35] during the last 30 years without any progress being made.

In Section 4 we show that the number of maximum empty boxes amidst n points in \mathbb{R}^d is always $O(n^d)$ and sometimes $\Omega(n^{\lfloor d/2 \rfloor})$. For $d \geq 4$, this is the first superlinear lower bound derived for the empty box problem.

In Section 5 we discuss two algorithmic applications: (i) Given n points inside a hyperrectangle U in \mathbb{R}^3 , a $(1 - \varepsilon)$ -approximation of the maximum volume of an empty box contained in U can be computed in $O(\varepsilon^{-2} n^{5/3} \log^2 n)$ time. (ii) If the data structure of Kaplan et al. [28] for the maximum empty rectangle containing a query point has a semi-dynamic extension that allows the deletion of points, with $o(n)$ amortized time per point deletion, then there is a subcubic-time exact algorithm for the MAXIMUM EMPTY BOX problem in \mathbb{R}^3 .

2 Background and related work

Bounds on the volume of a maximum empty box. Given a set S of n points in the unit hypercube $U_d = [0, 1]^d$, where $d \geq 2$, let $A_d(S)$ be the maximum volume of an empty box contained in U_d , and let $A_d(n)$ be the minimum value of $A_d(S)$ over all sets S of n points in U_d .

Rote and Tichy [35] proved that $A_d(n) = \Theta(\frac{1}{n})$ for any fixed $d \geq 2$. From one direction, for any $d \geq 2$, we have

$$A_d(n) < \left(2^{d-1} \prod_{i=1}^{d-1} p_i \right) \cdot \frac{1}{n}, \quad (1)$$

where p_i is the i th prime, as shown in [35, 18] using Halton-Hammersley generalizations [25, 26] of the van der Corput point set [13, 14]; see also [30, Ch. 2.1].

From the other direction, by slicing the hypercube with n parallel hyperplanes, each incident to one of the n points, the largest slice yields the lower bound $A_d(n) \geq \frac{1}{n+1}$ for each d . This trivial estimate can be improved using the following observation [18, 20] that relates $A_d(n)$ to $A_d(b)$ for fixed d and b :

Lemma 1. [20] *For any $d \geq 2$ and $b \geq 2$, $A_d(n) \geq [(b+1)A_d(b) - o(1)] \cdot \frac{1}{n}$.*

In particular, with $b = 4$ in Lemma 1, the following bound¹ was obtained in [18]:

$$A_d(n) \geq A_2(n) \geq (5A_2(4) - o(1)) \cdot \frac{1}{n} = (1.25 - o(1)) \cdot \frac{1}{n}.$$

By exploiting the above observation of Lemma 1 in a more subtle and fruitful way, Aistleitner, Hinrichs, and Rudolf [5] recently proved that $A_d(\lfloor \log d \rfloor) = \Omega(1)$. It follows that the dependence on d in the volume bound is necessary, i.e., the maximum volume grows with the dimension d . The following lower bound is derived in [5]:

$$A_d(n) \geq \frac{\log d}{4(n + \log d)}. \quad (2)$$

Further, one can show that their proof leads to an efficient algorithm that computes such a box. A broader analysis of this approach can be found in [21].

Complexity of computing a maximum empty box. Several algorithms have been proposed in the 1980s for the MAXIMUM EMPTY BOX problem in the plane; see [18, 20] and the references therein. The fastest one, due to Aggarwal and Suri [4], runs in $O(n \log^2 n)$ time and $O(n)$ space. On the other hand, a lower bound of $\Omega(n \log n)$ in the algebraic computation tree model already holds for the one-dimensional variant [34, pp. 260–262]: Given n points in a line, find the largest gap between two consecutive points. In contrast, Boland and Urrutia [10] showed that the largest axis-aligned rectangle contained in a simple polygon with n vertices can be found in $O(n \log n)$ time.

For $d \geq 3$, the only known approach for computing a largest empty box is by examining *all* candidates, i.e., all maximal empty boxes. As noted, an empty box of maximum volume must be maximal with respect to inclusion. Given a set S of n points in the unit hypercube $U_d = [0, 1]^d$, $d \geq 2$, let $k = k(S)$ and $\kappa = \kappa(S)$ denote the number of *maximal* empty boxes contained in U_d , and that of empty boxes of *maximum* volume contained in U_d , respectively. Clearly $\kappa(S) \leq k(S)$.

Chazelle et al. [12] show that in the plane, *all* largest empty rectangles can be reported in $O(k + n \log n)$ time (note the dependence on k , not on κ , in all time-bounds). Nandy and Bhattacharya [33] show that in \mathbb{R}^3 , *all* maximum empty boxes can be reported in $O(k + n^2 \log n)$ time. A result of Kaplan, Rubin, Sharir, and Verbin [29] implies an output-sensitive algorithm for MAXIMUM EMPTY BOX running in $O((n+k) \log^{2d-1} n)$ time, for any fixed $d \geq 2$. Subsequently, Backer and Keil [8, 9] reported an output-sensitive algorithm running in $O(k \log^{d-2} n)$ time, for any fixed $d \geq 3$.

One can easily exhibit point sets for which the number k of maximal empty boxes is $\Omega(n^d)$, for any fixed $d \geq 2$; see [29, 8, 18]. Naamad, Lee, and Hsu [32] showed that for $d = 2$, the maximum number of maximal empty rectangles is $O(n^2)$, and that this bound is tight. Kaplan et al. [29] proved (indirectly, see [19, p. 479]) that the maximum number of maximal empty boxes is $O(n^d)$ for each fixed d , confirming an earlier conjecture of Datta and Soundaralakshmi [15].

Since the maximum number of maximal empty boxes is $\Theta(n^d)$ for each fixed d , any algorithm that computes a maximum-volume empty box by enumerating all maximal empty boxes is bound to

¹A weaker bound with $b = 3$ was inadvertently labeled as an improvement over this bound in [20].

be inefficient in the worst case. In fact, Backer and Keil [8, 9] proved that MAXIMUM EMPTY BOX is NP-hard when the dimension d is part of the input, and Giannopoulos, Knauer, Wahlström, and Werner [24] further showed that the problem is W[1]-hard with the dimension d as the parameter. Further, the W[1]-hardness of the problem implies [24, Corollary 1] that the existence of an exact algorithm running in $n^{o(d)}$ time is unlikely. On the other hand, the output-sensitive algorithms of Kaplan et al. [29] and Backer and Keil [9] would run faster when there are only a few maximal empty boxes, and indeed, when d is fixed, Dumitrescu and Jiang [19] proved that the expected number of maximal empty boxes amidst n random points in the unit hypercube $[0, 1]^d$ in \mathbb{R}^d is

$$(1 \pm o(1)) \frac{(2d-2)!}{(d-1)!} n \ln^{d-1} n.$$

So while for fixed d , the expected running times of the output-sensitive algorithms in [9, 29] on random point sets are $O(n \log^{O(d)} n)$, their worst-case running times remain $\Omega(n^d)$.

In terms of approximation, Dumitrescu and Jiang [18] gave an algorithm that finds an empty box whose volume is at least $1 - \varepsilon$ times the optimal in $O\left(\left(\frac{8ed}{\varepsilon^2}\right)^d \cdot n \log^d n\right)$ time; a faster approximation algorithm for the related MAXIMUM EMPTY CUBE problem was also provided. This approximation algorithm is fixed parameter tractable with both d and $1/\varepsilon$ as parameters [24, Section 1.5].

Notation. As customary, Θ, O, Ω notation is used to describe the asymptotic growth of functions. Given a set S of n points in the unit hypercube $U_d = [0, 1]^d$, where $d \geq 2$, let $X_d(n)$ be the maximum value of $\kappa(S)$ over all sets S of n points in U_d .

3 The number of empty rectangles of maximum area

As mentioned earlier, the number of maximal empty rectangles is $O(n^2)$, and this bound can be attained. In contrast, we prove (see Theorem 1) that the number of empty rectangles of maximum area is linear or nearly linear. Refer for brevity to any empty rectangle of maximum area as an *maximum empty rectangle*; and *maximum empty box* for $d \geq 3$. The arguments we need are based on Davenport-Schinzel sequences.

A sequence a_1, a_2, \dots over the alphabet $\Sigma = \{1, 2, \dots, m\}$ is called an (m, s) -Davenport-Schinzel sequence of order s , if (i) it has no two consecutive elements which are the same, and (ii) it has no alternating subsequence of length $s + 2$, i.e., there are no indices $1 \leq i_1 < i_2 < \dots < i_{s+2}$ such that

$$a_{i_1} = a_{i_3} = a_{i_5} = \dots = a, \quad a_{i_2} = a_{i_4} = a_{i_6} = \dots = b,$$

where $a \neq b$. Let $\lambda_s(m)$ denote the maximum length of an (m, s) -Davenport-Schinzel sequence [16]; see also [36] and [31, Ch. 7]. Obviously, we have $\lambda_1(m) = m$, and one can verify that $\lambda_2(m) = 2m - 1$, for every m . It was shown by Hart and Sharir [27] that $\lambda_3(m) = O(m\alpha(m))$ and $\lambda_4(m) = O(m2^{\alpha(m)})$, where $\alpha(m)$ is the extremely slowly growing inverse of Ackermann's function, and that this estimate is asymptotically tight. They also proved that $\lambda_s(m)$ is only slightly superlinear in m , for every fixed $s \geq 4$. For the currently best bounds of this type, see [1].

Theorem 1. *We have $X_2(n) = \Omega(n)$ and $X_2(n) = O(\lambda_4(n) \log n) = O(n \log n 2^{\alpha(n)})$, where $\alpha(n)$ is the (slowly growing) inverse of Ackermann's function.*

3.1 Proof of Theorem 1

Lower bound. We start with a construction that yields a lower bound of $X_2(n) \geq n+1$. Consider the line $x+y = 2\sqrt{c}$ and the hyperbola $xy = c$ inside $U = [0, 1]^2$, where $1/4 \leq c' \leq c < 1$. They are

disjoint when $c' < c$, and tangent at $(\sqrt{c'}, \sqrt{c'})$ when $c' = c$. For any c' and c , $1/4 \leq c' \leq c < 1$, we can obtain a sequence $S = S(c', c)$ of points inside U on the line $x + y = 2\sqrt{c'}$ as follows. Start at the right end $(1, c)$ of the hyperbola, shoot a ray to the left until it hits the line at $p_1 = (2\sqrt{c'} - c, c)$. While the x -coordinate of p_i is greater than that of the left end $(c, 1)$ of the hyperbola, let the ray continue upward until it hits the hyperbola, and then to the left until it hits the line again at p_{i+1} . Refer to Fig. 2 for the limiting case when $c \rightarrow c' = 1/4$.

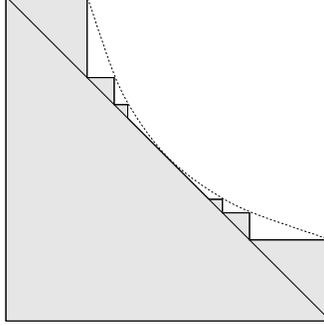


Figure 2: A hyperbola construction for the linear lower bounds on $X_2(n)$.

Fix $1/2 \leq c < 1$. As c' continuously increases from $1/4$ to c , the number of points p_i in the sequence S gradually increases from 1 and tends to ∞ . For each $n \geq 1$, there is a range $I_n \subset [1/4, c)$ of c' such that S contains exactly n points inside U on the line $\ell : x + y = 2\sqrt{c'}$. Moreover, as c' increases continuously in this range, the n points move continuously on this line. By continuity, there is a unique value of c' in the range I_n for each $n \geq 1$, such that the n points are placed symmetrically about the center $(\sqrt{c'}, \sqrt{c'})$ of the line segment $\ell \cap U$. Observe that all (final) point coordinates are at least c . With these n points in U , there are exactly $n + 1$ maximum empty rectangles with area c , all having lower-left corners at $(0, 0)$ and upper-right corners on the hyperbola.

We next give a more refined construction which yields a better, but still linear, lower bound of $X_2(n) \geq 2n + 2$. Again consider the line $x + y = 2\sqrt{c'}$ and the hyperbola $xy = c$ inside U , and refer to Fig. 2.

Fix $c' = 1/4$. Then the line $x + y = 2\sqrt{c'}$ is the diagonal $x + y = 1$ of U . As c continuously decreases from 1 to c' , the number of points p_i in the sequence S gradually increases from 1 and tends to ∞ . For each $n \geq 1$, there is a range of c such that S contains exactly n points inside U on the diagonal $x + y = 1$. Moreover, as c decreases continuously in this range, the n points move continuously on this diagonal. There is a particular value of c , $1/4 < c \leq 1/2$, such that the n points are placed symmetrically about the center $(1/2, 1/2)$ of the diagonal. With these n points in U , there are exactly $2n + 2$ maximum empty rectangles with area c , among which $n + 1$ have lower-left corners at $(0, 0)$, and the other $n + 1$ have upper-right corners at $(1, 1)$, symmetric about the diagonal $x + y = 1$.

Upper bound. We proceed by induction on n and develop a recurrence on $X(n) = X_2(n)$. Consider a *balanced partition* of the points by a vertical line into two sets, with $\lfloor n/2 \rfloor$ and $\lceil n/2 \rceil$ points, respectively: $S = L \cup R$ (points on the line are assigned as needed to any one of the two parts). We may suppose that the vertical line is the y -axis. Let U_- and U_+ , respectively, be the two subrectangles of U on the left and the right side of the y -axis.

Following the terminology in [12], a *support* of a maximum empty rectangle R is either a point in S or an edge of U that limits the expansion of R in that direction: leftward, rightward, upward or

downward. Every maximum empty rectangle has 4 supports. Every maximum empty rectangle is either contained in U_- , or contained in U_+ , or crosses the y -axis. Let Y denote the set of maximum empty rectangles crossing the y -axis.

A rectangle in Y can have either (i) 2 supports on the left side of the y -axis and 2 supports on the right side of the y -axis; or (ii) 3 supports on one side of the y -axis and 1 support on the other side of the y -axis.

Naamad et al. [32] showed that the number of maximal empty rectangles with at least one edge support is $O(n)$. Thus we immediately get:

Lemma 2. *The number of maximum empty rectangles in Y with at least one edge support is $O(n)$.*

We next establish upper bounds for the remaining types of maximum empty rectangles in Y , with 4 points of support (a point of support is also referred to as *defining point* in [28]). For the remainder of this section we only consider this type of maximum empty rectangles.

Chazelle et al. [12] showed that the number of maximal empty rectangles with 3 points of support on one side of an axis and one point of support on the other side of that axis is $O(n)$; see also [28, p. 346]. Thus we also have:

Lemma 3. *The number of rectangles in Y with 3 points of support on one side of the y -axis and one point of support on the other side of the y -axis is $O(n)$.*

The key argument is provided by the following.

Lemma 4. *The number of rectangles in Y with two points of support on each side of the y -axis is $O(\lambda_4(n))$.*

Proof. By symmetry, it suffices to show that the number h of maximum empty rectangles with bottom and left supports in L and with top and right supports in R is $O(\lambda_4(n))$.

Consider a maximum empty rectangle H amidst the points $L \cup R$ in U . Recall that U_- and U_+ are the two subrectangles of U left and right of the y -axis. Then the bottom-left corner of H is the bottom-left corner of a maximal empty rectangle amidst L in U_- with the y -axis as the right support, and the top-right corner of H is the top-right corner of a maximal empty rectangle amidst R in U_+ with the y -axis as the left support. By the result of Naamad et al. [32] mentioned earlier (amidst any n points in a rectangle, the number of maximal empty rectangles with at least one edge support is $O(n)$), we obtain l candidate bottom-left corners p_1, \dots, p_l on the left side of the y -axis, and r candidate top-right corners q_1, \dots, q_r on the right side of the y -axis, where $l + r = O(n)$, such that each maximum empty rectangle under our consideration has some p_i as the bottom-left corner and some q_j as the top-right corner.

Order the candidate corners on the left such that $y(p_1) \geq \dots \geq y(p_l)$, and corners with the same y -coordinate are ordered from left to right. Similarly, order the candidate corners on the right such that $y(q_1) \geq \dots \geq y(q_r)$, and corners with the same y -coordinate are ordered from left to right.

We next construct a sequence w over the alphabet $\Sigma = \{p_1, \dots, p_l, q_1, \dots, q_r\}$ of $m = l + r$ symbols; recall that $m = O(n)$. Initialize w to an empty string. For each $i = 1, \dots, l$, first append the symbol p_i to w , then for each $j = 1, \dots, r$, if there is a maximum empty rectangle with p_i as the bottom-left corner and with q_j as the top-right corner, append the symbol q_j to w . The length of the string w thus constructed is $l + h$; recall that h is the number of maximum empty rectangles under consideration. No two consecutive symbols in w are the same, because each symbol p_i occurs exactly once in w , and each symbol q_j occurs at most once after p_i and before p_{i+1} . We claim that w has no alternating subsequence of length 6, i.e., no subsequence of the form $\alpha\beta\alpha\beta\alpha\beta$; here α, β are elements of $\{q_1, \dots, q_r\}$, since each p_i occurs exactly once. Equivalently, we claim that if

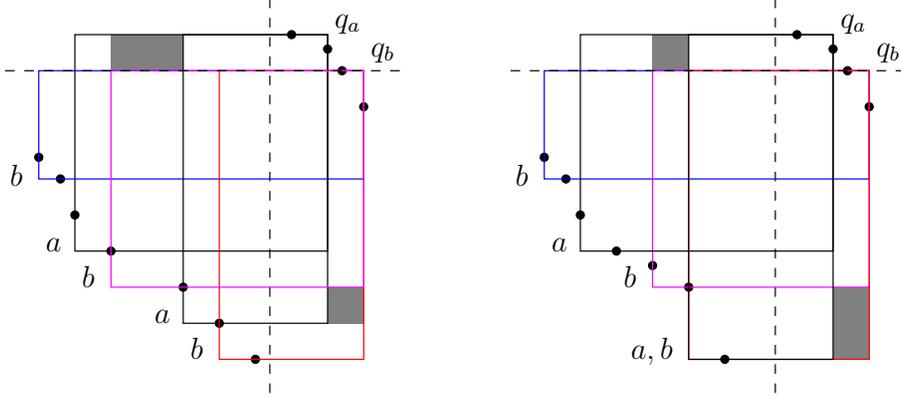


Figure 3: Extending the union of two maximum empty rectangles selected when reading a subsequence of the form $babab$ with $a < b$. The figure shows two examples.

$1 \leq a < b \leq r$, then w has no subsequence of the forms $q_a q_b q_a q_b q_a q_b$ or $q_b q_a q_b q_a q_b q_a$. If such a subsequence exists, we extract out of it a subsequence of length five of the form $q_b q_a q_b q_a q_b$, where $a < b$.

Suppose for contradiction that w has a subsequence $q_b q_a q_b q_a q_b$, where $a < b$, so $y(q_a) \geq y(q_b)$. By our construction of w , the corner q_a cannot have both coordinates greater than those of q_b . Suppose the contrary, that $x(q_a) > x(q_b)$ and $y(q_a) > y(q_b)$. Then the maximum empty rectangle corresponding to the first occurrence of the symbol q_a in $q_a q_b q_a q_b$ would contain the corner q_b and hence a point of R with the same y -coordinate as q_b , contradicting its emptiness assumption.

Label each corner p_i that is the bottom-left corner of some maximum empty rectangle with top-right corner at q_a (respectively, q_b) with the index a (respectively b). Then we have a sequence of bottom-left corners p_i on the left side of the y -axis and below the horizontal line through q_b , with non-decreasing y -coordinates and non-increasing x -coordinates, as illustrated in Fig. 3, where each corner is labeled with either a or b or ab (note that two maximum empty rectangles with different top-right corners may have the same point in L as the bottom support), and concatenation of all these labels is the index sequence $babab$ corresponding to the point sequence $q_b q_a q_b q_a q_b$.

Indeed: (i) since when constructing w , i goes from 1 to l , the sequence of corners has non-increasing y -coordinates; (ii) suppose for contradiction that the sequence of corners does not have non-decreasing x -coordinates, i.e., for some $i < i'$ we have $p_i, p_{i'}$ as lower-left corners corresponding to a subsequence $q_b q_a$ in w , with $x(p_{i'}) < x(p_i)$ and $y(p_{i'}) < y(p_i)$; then, as in a previous argument, the maximum empty rectangle corresponding to the occurrence of q_a in $q_b q_a$ would contain the corner q_b and hence a point of L with the same y -coordinate as q_b , contradicting its emptiness assumption; consequently, the sequence of corners has non-decreasing x -coordinates.

Note that the shaded area above the horizontal line through q_b is contained in the maximum empty rectangle with bottom-left corner labeled with the first a , and that the shaded area on the right side of the y -axis is contained in the maximum empty rectangle with bottom-left corner labeled with the third b . Thus they are both empty. The reader may reflect at this point on the significance of the occurrence of a subsequence $babab$, with $a < b$. The first b in $babab$ was used only to ensure that the a after it has smaller y -coordinate than q_b . The last b in $babab$ was used to ensure that the right shaded area is empty.

We are now in position to finalize the argument. Consider the two maximum empty rectangles with bottom-left corners labeled with the second b and the second a , respectively. By adding the two shaded areas, the union of the two maximum empty rectangles can be extended into the union

of two other empty rectangles with the same bottom-left corners and with the top-right corners swapped. Since the two pairs of rectangles have the same intersection, it follows that the total area of the two extended empty rectangles is larger than the total area of the two maximum empty rectangles, a contradiction.

Since w is an $(m, 4)$ -Davenport-Schinzel sequence, its length at most $\lambda_4(m)$. It then follows that $h \leq \lambda_4(m) = O(\lambda_4(n))$, concluding the proof of Lemma 4. \square

Let $Y(n)$ denote the total number of rectangles in Y . By Lemmas 2, 3, 4, we have $Y(n) = O(\lambda_4(n))$. By the Divide & Conquer approach we have

$$X(n) \leq X\left(\left\lfloor \frac{n}{2} \right\rfloor\right) + X\left(\left\lceil \frac{n}{2} \right\rceil\right) + Y(n),$$

thus $X(n)$ satisfies the following recurrence:

$$X(n) \leq X\left(\left\lfloor \frac{n}{2} \right\rfloor\right) + X\left(\left\lceil \frac{n}{2} \right\rceil\right) + O(\lambda_4(n)). \quad (3)$$

The solution of this recurrence is $X(n) = O(\lambda_4(n) \log n) = O(n \log n 2^{\alpha(n)})$, and this completes the proof of the planar upper bound in Theorem 1.

A refined upper bound on $X_2(n)$. Let A be the maximum area of an empty rectangle in U . We next show that the number of maximum empty boxes is $O(\lambda_4(n) \log(1/A))$. In particular, $A = \Omega(1/n)$ hence $\log(1/A) = O(\log n)$. When A is a constant, the bound is almost tight: the maximum number of maximum empty boxes is $\Omega(n)$ and $O(n 2^{\alpha(n)})$ in this case.

We can get this slightly stronger formulation of our current result because every maximum empty rectangle has area A and since its height is at most 1 its width is at least A . Thus we can group all maximum empty rectangles by their widths into about $\log(1/A)$ groups, such that the widths of all maximum empty rectangles in each group differ by a factor of at most 2, and then argue that the number of maximum empty rectangles in each group is at most $O(\lambda_4(n))$.

For each group, let the width of the maximum empty rectangles be between w and $2w$. Then we can use $1/w$ evenly-spaced vertical lines to stab all maximum empty boxes at least once and at most twice. Using Lemmas 3 and 4 with each vertical line as the y -axis yields the $O(\lambda_4(n))$ bound for each group.

4 The number of maximum empty boxes in \mathbb{R}^d

In this section we prove the following.

Theorem 2. *For any fixed $d \geq 3$, we have $X_d(n) = \Omega(n^{\lfloor d/2 \rfloor})$ and $X_d(n) = O(n^d)$.*

4.1 Proof of Theorem 2

Upper bound. The upper bound $X_d(n) = O(n^d)$ immediately follows from the same upper bound on the number of *maximal empty boxes*, as established (indirectly) in [29]. We next recall the main lines of reasoning in obtaining this result.

Motivated by a related problem, COLOURED ORTHOGONAL RANGE COUNTING, Kaplan, Rubin, Sharir and Verbin [29] gave estimates on the number of maximal empty boxes. In the context of this related problem, each input point in \mathbb{R}^d can be transformed into a point in \mathbb{R}^{2d} by splitting the coordinate x along the i th axis into two coordinates x and $-x$ in the $(2i-1)$ th and $(2i)$ th axes,

respectively, such that each maximal empty box in \mathbb{R}^d becomes a maximal empty orthant in \mathbb{R}^{2d} . Kaplan et al. then provided an upper bound of $O(n^{\lfloor d/2 \rfloor})$ on the maximum number of maximal empty orthants amidst n points in \mathbb{R}^d , which, by the transformation, implies an upper bound of $O(n^d)$ on the maximum number of maximal empty boxes amidst n points in \mathbb{R}^d . Moreover they also provided matching lower bounds of $\Omega(n^{\lfloor d/2 \rfloor})$ for maximal empty orthants and $\Omega(n^d)$ for maximal empty boxes. Thus the maximum number of maximal empty boxes amidst n points in \mathbb{R}^d is $\Theta(n^d)$. We refer to [19] for a historical account of related work on these bounds.

Since maximum empty boxes are also maximal empty boxes, the known upper bound for maximal empty boxes [29] immediately implies the upper bound $X_d(n) = O(n^d)$ for maximum empty boxes. In the following, we derive the lower bound $X_d(n) = \Omega(n^{\lfloor d/2 \rfloor})$.

Lower bound. Let $d \geq 4$ be even. Pair up the d dimensions into $d/2$ pairs, where each pair determines a 2-dimensional plane. Let $n = k \cdot (d/2)$. Associate k distinct points with each of the $d/2$ planes. Let $1/2 < c < 1$. In each of the $d/2$ planes, set the two coordinates of the k points associated with this plane as in the first planar construction in the proof of the lower bound in Theorem 1, such that there are $k + 1$ maximum empty rectangles of area $c^{2/d}$ amidst the k points in that plane. Set the other $d - 2$ coordinates of these k points to c . Note that $c^{2/d} > c$ and thus all point coordinates are at least c and less than 1.

Claim 1. *The maximum volume of an empty box is c . The number of maximum empty boxes is $(k + 1)^{d/2} + d$.*

To verify the claim, we first show that there are $(k + 1)^{d/2} + d$ maximal empty boxes of volume c . Since all coordinates of all points are at least c , a box that projects to the interval $(0, c)$ along any one of the d axes and to the whole interval $(0, 1)$ along the other $d - 1$ axes is empty. This box is maximal because all n points are interior to the unit box U_d , and there are points of coordinate c in each dimension. There are d boxes of this type.

Take any one of the $k + 1$ maximum empty rectangles from each of the $d/2$ planes, the product of these $d/2$ rectangles is a box. This box is empty of all n points because it is empty of the k distinct points associated with each of the $d/2$ planes. There are $(k + 1)^{d/2}$ boxes of this type, each of volume $(c^{2/d})^{d/2} = c$. For fixed d , the number of boxes is $\Omega(n^{d/2})$.

We next show that the maximum volume of an empty box is c , and hence these $(k + 1)^{d/2} + d$ boxes are all maximum empty boxes.

Let B be an arbitrary empty box of volume at least c . Then the projection of B to each axis is an interval of length at least c ; moreover, if the projection of B to one axis is an interval of length exactly c , then its projection to all other axes must have length 1. Since there is at least one point of coordinate c along each axis, B either projects to the interval $(0, c)$ along one axis and hence has to project to the whole interval $(0, 1)$ along all other axes, or projects to an interval that contains c along every axis. There are clearly d empty boxes of volume c of the first type. In the following, we assume that B belongs to the second type, that is, the projection of B to each axis contains c .

Let B_i and B_π denote the projection of B onto dimension i and plane π respectively. Consider one of the $d/2$ planes, say, the x_1x_2 plane π . Recall that the k points associated with this plane have coordinate c in the other $d - 2$ dimensions, and so their projections onto any axis x_i , $i \geq 3$, are contained in the interior of the corresponding interval B_i . On the other hand, their projections onto π are not contained in the interior of B_π , since otherwise B would not be empty.

Thus B_π is an empty rectangle in the plane π and so its area is at most $c^{2/d}$, which is the area of a maximum empty rectangle in this plane. By the same argument, the projection of B to each of the $d/2$ planes must have area at most $c^{2/d}$. Hence the volume of B is at most $(c^{2/d})^{d/2} = c$.

Since the volume of B is assumed to be at least c , its volume must be exactly c . Moreover, the projection of B to each of the $d/2$ planes must have area exactly $c^{2/d}$ and must be one of the $k+1$ maximum empty rectangles. The claim has been verified.

Let $d \geq 3$ be odd. Then $d-1 \geq 2$ is even. Take n points in U_{d-1} attaining the lower bound for even dimension, with maximum volume c , where $1/2 < c < 1$. Set the d th coordinate to c for all n points, now in U_d . Observe that the hyperplane $x_d = c$ divides $U_d = [0, 1]^d$ into two maximal empty boxes of volumes c and $1-c < c$, respectively, and that all other maximal empty boxes in U_d project to maximal empty boxes in U_{d-1} . For this point set the maximum volume of an empty box is again c , and there is exactly one extra maximum empty box, namely $(0, 1)^{d-1} \times (0, c)$, in addition to the maximum empty boxes lifted up from U_{d-1} . Thus $X_d(n) \geq X_{d-1}(n) + 1$.

Consequently, the lower bound $X_d(n) = \Omega(n^{\lfloor d/2 \rfloor})$ holds for any fixed $d \geq 3$, as required. The proof of Theorem 1 is now complete. \square

5 Algorithmic aspects

In this section we present an alternative approximation algorithm in \mathbb{R}^3 and a possible approach towards a faster exact algorithm, also in \mathbb{R}^3 .

5.1 An alternative $(1-\varepsilon)$ -approximation algorithm in \mathbb{R}^3

Given n points in $U_d = [0, 1]^3$ let $B_0 = \Pi_{i=1}^3 [a_i, b_i]$ be an (unknown) maximum empty box, and let v_0 denote its volume. Since $v_0 \geq 1/(n+1)$, the maximum extent of B_0 along one of the three coordinate axes, i.e., the length of its projection interval onto one of the three axes, is roughly at least $n^{-1/3}$. Guess (by repeating over the three axis-directions) that the maximum extent of B_0 is along the z -coordinate. Consider the subdivision of B_0 into horizontal slabs of height about $0.5 \varepsilon n^{-1/3}$.

Repeat over the $O(\varepsilon^{-2} n^{2/3})$ pairs of horizontal planes as top and bottom planes: project all the points in the corresponding slab onto a horizontal plane and use the planar exact algorithm by Aggarwal and Suri [4] to find a maximum empty rectangle in the projection plane in $O(n \log^2 n)$ time. Return the maximum volume of a box found in this way, i.e., as the product of the area of a maximum empty rectangle in the projection plane and the slab-extent along the z -coordinate. The relative error in the volume returned is at most ε and the resulting overall time is $O(\varepsilon^{-2} n^{2/3} n \log^2 n) = O(\varepsilon^{-2} n^{5/3} \log^2 n)$.

In comparison, the running time of the $(1-\varepsilon)$ -approximation algorithm from [18] is $O(\varepsilon^{-6} n \log^3 n)$. With an ε^{-4} factor reduction at the expense of an $n^{2/3} \log^{-1} n$ factor increase in the running time, our alternative approximation algorithm offers an attractive trade-off for input instances with small values of ε and moderate values of n .

5.2 Towards a faster exact algorithm in \mathbb{R}^3 (conditional result)

Augustine et al. [7] and Kaplan et al. [28] studied the problem of finding the largest-area empty rectangle contained in U and containing a query point. Kaplan et al. [28] showed how to construct in nearly linear time a data structure that takes nearly linear space, so that given a query point q , the largest-area empty sub-rectangle of U containing q can be computed in $O(\log^4 n)$ time.

We next show that a semi-dynamic version of this data structure allows solving the MAXIMUM EMPTY BOX problem in 3-space in subcubic time.

Given n points in $[0, 1]^3$, sort them by their z -coordinate; and set $z_0 = 0, z_{n+1} = 1$; hence we have $0 = z_0 \leq z_1 \leq \dots \leq z_n \leq 1 = z_{n+1}$, where $p_i = (x_i, y_i, z_i), i = 1, \dots, n$. Assume for simplicity that all z -coordinates are distinct (the number of phases done by the algorithm is essentially the number of distinct value of z_i). For $0 \leq i < j \leq n + 1$, let $\Gamma(i, j)$ be the planar data structure from [28] for the set of $j - i - 1$ points $P_{ij} = \{q_k = (x_k, y_k, 0), k = i + 1, \dots, j - 1\}$. The points q_k are the projections of the points in the open slab bounded by $z = z_i$ and $z = z_j$ onto a horizontal plane. Note that for $j = i + 1, P_{ij} = \emptyset$ (so these slabs are trivial to deal with).

The algorithm proceeds in n phases, numbered from 0 to n . Phase i finds the maximum volume of an empty box whose z -extent is $[z_i, z_j]$, over all $j = n + 1, \dots, i + 1$. To do this, it first computes $\Gamma(i, n + 1)$, and then by repeated updates obtains $\Gamma(i, n), \Gamma(i, n - 1), \dots, \Gamma(i, i + 2), \Gamma(i, i + 1)$.

The maximum volume of an empty box whose z -extent is $[z_i, z_j]$, where $i < j$, equals the area of a maximum empty rectangle for the planar set $P_{ij} \subset U$ times the height $z_j - z_i$. In each subsequent step of Phase i , the height of current slab is reduced vertically by removing its top point. The corresponding planar data structure $\Gamma(i, j)$ is updated with respect to the deletion of one point, say, in time at most $U(n)$. There are about $n - i$ updates (and point deletions) in Phase i . The data structure is queried for the maximum empty rectangle containing the planar (projection) point that was just deleted. The area of this rectangle is then multiplied with the height of the current new slab to obtain the volume of a maximum empty box in the new slab and then the process is repeated. The algorithm outputs the volume of a maximum empty box over all $\binom{n+2}{2}$ slabs.

The data structure is computed from scratch n times, once for each phase, and there are at most $O(n^2)$ deletions, each taking $U(n)$ time to process. Alternatively one can compute that data structure only once, i.e., $\Gamma(0, n + 1)$, but work with two copies, one being updated by deleting one point for each phase, and one being updated at each point deletion within a phase.

Suppose now that the above data structure can be made semi-dynamic with respect to point deletions: given a sequence of $m \leq n$ points in $[0, 1]^2$, construct the data structure and delete the points one by one in the given order, while updating the data structure after each deletion. Assume that this can be done in $U(n) = O(n^\gamma)$ (amortized) time per point deletion, where $\gamma < 1$. Recall that the data structure can be built from scratch in $T(n) = n \log^{O(1)} n$ time. Consequently, if $U(n) = O(n^\gamma)$, the overall running time for computing an empty box of maximum volume amidst n points in $[0, 1]^3$ becomes $T(n) = O(n^{2+\gamma})$, as desired.

Perhaps obtaining a dependence of the form $U(n) = O(\log^{O(1)} n)$ is too optimistic, due to the fact that computing a maximum empty box is *non-decomposable*; see [17]; a different variant of largest empty box is considered there, however, the conclusion remains the same.

6 Open problems

While we showed that the maximum number of rectangles of maximum area is linear or nearly linear in the plane, a significantly larger gap, between $\Omega(n^{\lfloor d/2 \rfloor})$ and $O(n^d)$, remains in higher dimensions. Moreover, the rôle of the parity of d remains unclear. We suspect that the logarithmic factor in the planar upper bound is only an artifact of our proof. These remarks suggests the following.

Problem 1. *Can an upper bound $X_2(n) = O(n 2^{\alpha(n)})$ be deduced? Is $X_2(n)$ superlinear in n ?*

Problem 2. *For fixed $d \geq 2$, what is the order of growth of $X_d(n)$? In particular, is $X_3(n) = \omega(X_2(n))$? Is $X_d(n) = \omega(X_{d-1}(n))$ for odd d ?*

In Section 5 we have briefly revisited the problem of computing or approximating a maximum empty box.

Problem 3. Can a maximum empty box in \mathbb{R}^d for some fixed $d \geq 3$ be computed in $O(n^{\gamma_d})$ time for some constant $\gamma_d < d$?

A possible approach in \mathbb{R}^3 is suggested by the following.

Problem 4. Is there a dynamic or semi-dynamic version—with respect to the deletion of points—of the data structure of Kaplan et al. [28] for finding the largest-area empty rectangle containing a query point?

Problem 5. Is there an efficient algorithm for enumeration (or counting) of all maximum empty boxes amidst n points in U_d ? Is there one for the planar version running in $O(n \log^{O(1)} n)$ time?

References

- [1] P. K. Agarwal, M. Sharir, and P. Shor. Sharp upper and lower bounds for the length of general Davenport-Schinzel sequences. *J. Combin. Theory, Ser. A*, 52:228–274, 1989.
- [2] A. Aggarwal, M. Klawe, S. Moran, P. Shor, and R. Wilber. Geometric applications of a matrix-searching algorithm. *Algorithmica*, 2:195–208, 1987.
- [3] A. Aggarwal and M. Klawe. Applications of generalized matrix searching to geometric algorithms. *Discrete Appl. Math.*, 27:3–23, 1990.
- [4] A. Aggarwal and S. Suri. Fast algorithms for computing the largest empty rectangle. In *Proc. 3rd Ann. Sympos. on Comput. Geom.*, pp. 278–290, 1987.
- [5] C. Aistleitner, A. Hinrichs, and D. Rudolf. On the size of the largest empty box amidst a point set. Preprint, <http://arxiv.org/abs/1507.02067>, 2015.
- [6] M. J. Atallah and G. N. Frederickson. A note on finding a maximum empty rectangle. *Discrete Appl. Math.*, 13(1):87–91, 1986.
- [7] J. Augustine, S. Das, A. Maheshwari, S. C. Nandy, S. Roy, and S. Sarvattomananda. Querying for the largest empty geometric object in a desired location. Preprint, <http://arxiv.org/abs/1004.0558>, 2010.
- [8] J. Backer and M. Keil. The bichromatic rectangle problem in high dimensions. In *Proc. 21st Canadian Conf. on Comput. Geom.*, pp. 157–160, 2009.
- [9] J. Backer and M. Keil. The mono- and bichromatic empty rectangle and square problems in all dimensions. In *Proc. 9th Latin American Sympos. on Theor. Informatics*, pp. 14–25, 2010.
- [10] R. Boland and J. Urrutia. Finding the largest axis aligned rectangle in a polygon in $O(n \log n)$ time. In *Proc. 13th Canadian Conf. on Comput. Geom.*, pp. 41–44, 2001.
- [11] P. Braß, G. Rote, and K. J. Swanepoel. Triangles of extremal area or perimeter in a finite planar point set. *Discrete Comput. Geom.*, 26:51–58, 2001.
- [12] B. Chazelle, R. Drysdale, and D. T. Lee. Computing the largest empty rectangle. *SIAM J. Comput.*, 15:300–315, 1986.
- [13] J. G. van der Corput. Verteilungsfunktionen I. *Proc. Nederl. Akad. Wetensch.*, 38:813–821, 1935.

- [14] J. G. van der Corput. Verteilungsfunktionen II. *Proc. Nederl. Akad. Wetensch.*, 38:1058–1066, 1935.
- [15] A. Datta and S. Soundaralakshmi. An efficient algorithm for computing the maximum empty rectangle in three dimensions. *Inform. Sci.*, 128:43–65, 2000.
- [16] H. Davenport and A. Schinzel. A combinatorial problem connected with differential equations. *Amer. J. Math.*, 87:684–694, 1965.
- [17] D. Dobkin and S. Suri. Maintenance of geometric extrema. *J. of ACM*, 38(2):275–298, 1991.
- [18] A. Dumitrescu and M. Jiang. On the largest empty axis-parallel box amidst n points. *Algorithmica*, 66(2):225–248, 2013.
- [19] A. Dumitrescu and M. Jiang. Maximal empty boxes amidst random points. *Combin. Probab. Comput.*, 22:477–498, 2013.
- [20] A. Dumitrescu and M. Jiang. Computational Geometry Column 60. *SIGACT News Bulletin*, 45(4):76–82, 2014.
- [21] A. Dumitrescu and M. Jiang. Perfect vector sets, properly overlapping partitions, and largest empty box. Preprint, <http://arxiv.org/abs/1608.06874>, 2016.
- [22] J. Edmonds, J. Gryz, D. Liang, and R. Miller. Mining for empty spaces in large data sets. *Theoret. Comp. Sci.*, 296(3):435–452, 2003.
- [23] P. Erdős and G. Purdy. Some extremal problems in geometry. *J. Combin. Theory, Ser. A*, 10:246–252, 1971.
- [24] P. Giannopoulos, C. Knauer, M. Wahlström, and D. Werner. Hardness of discrepancy computation and ε -net verification in high dimension. *J. Complexity*, 28:162–176, 2012.
- [25] J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.*, 2:84–90, 1960.
- [26] J. M. Hammersley. Monte Carlo methods for solving multivariable problems. *Ann. New York Acad. Sci.*, 86: 844–874, 1960.
- [27] S. Hart and M. Sharir. Nonlinearity of Davenport-Schinzel sequences and of generalized path compression schemes. *Combinatorica*, 6:151–177, 1986.
- [28] H. Kaplan, S. Mozes, Y. Nussbaum, and M. Sharir. Submatrix maximum queries in Monge matrices and Monge partial matrices, and their applications. In *Proc. 23rd ACM-SIAM Sympos. on Discrete Algorithms*, pp. 338–355, 2012.
- [29] H. Kaplan, N. Rubin, M. Sharir, and E. Verbin. Efficient colored orthogonal range counting. *SIAM J. Comput.*, 38:982–1011, 2008.
- [30] J. Matoušek. *Geometric Discrepancy: An Illustrated Guide*. Springer, 1999.
- [31] J. Matoušek. *Lectures on Discrete Geometry*. Springer, 2002.
- [32] A. Namaad, D. T. Lee, and W.-L. Hsu. On the maximum empty rectangle problem. *Discrete Appl. Math.*, 8:267–277, 1984.

- [33] S. Nandy and B. Bhattacharya. Maximal empty cuboids among points and blocks. *Computers Math. Applic.*, 36(3):11–20, 1998.
- [34] F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Springer, New York, 1985.
- [35] G. Rote and R. F. Tichy. Quasi-Monte-Carlo methods and the dispersion of point sequences. *Math. Comput. Modelling*, 23:9–23, 1996.
- [36] M. Sharir and P. K. Agarwal. *Davenport-Schinzel Sequences and Their Geometric Applications*. Cambridge University Press, Cambridge, 1995.