Computational Geometry Column 60

Adrian Dumitrescu* Minghui Jiang†

October 29, 2014

Abstract

This column is devoted to maximal empty axis-parallel rectangles amidst a point set. In particular, among these, maximum-area rectangles are of interest.

Keywords: Largest empty box, largest empty hypercube, discrepancy of a set of points, van der Corput point set, Halton-Hammersley point set, approximation algorithm, data mining.

1 Introduction

Given an axis-parallel rectangle R in the plane containing n points, the problem of computing a maximum-area empty axis-parallel sub-rectangle contained in R is one of the oldest problems in computational geometry. For instance, this problem arises when a rectangular-shaped facility is to be located within a similarly shaped region which has a number of forbidden areas, or in cutting out a rectangular piece from a metal sheet with some defective spots to be avoided [24]. In higher dimensions, finding an empty box (i.e., an empty axis-parallel hyperrectangle) with the maximum volume has applications in data mining, in finding large gaps in a multi-dimensional data set [15].

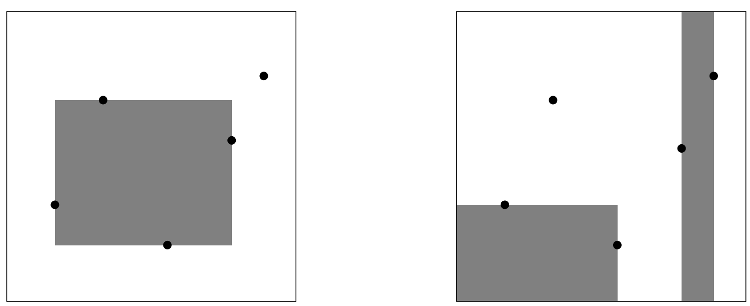

Figure 1: A maximal empty rectangle supported by one point on each side (left), and two maximal empty rectangles supported by both points and sides of $[0,1]^2$ (right).

Since 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 an open axis-parallel hyperrectangle contained in U_d and containing no points in S, 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.

*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`

2 The complexity of computing a maximum empty box

Several algorithms have been proposed over time for the Maximum Empty Box problem in the plane; see [12] and the references therein. The fastest one, due to Aggarwal and Suri [2], 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 1-dimensional variant, the so-called Maximum Gap problem [25, pp. 260–262]: given n points on the line, find the largest gap between two adjacent points.

Open question 1. *Can the largest-area empty axis-parallel rectangle be computed in $O(n \log n)$ time?*

For the Maximum Empty Box problem in higher dimensions, i.e., for $d \geq 3$, the only approach currently known for computing the largest empty box exactly is by examining *all* candidates, i.e., all maximal empty boxes. As observed previously, 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 $k = k(S)$ denote the number of maximal empty boxes in U_d, amidst the n points. A result of Kaplan, Rubin, Sharir, and Verbin [20] implies[1] 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 [4, 5] also reported an output-sensitive algorithm, running in $O(k \log^{d-2} n)$ time, for any fixed $d \geq 3$.

It is not hard to construct a point set for which the number k of maximal empty boxes is $\Omega(n^d)$, for any fixed $d \geq 2$; see [20, 4, 12]. Naamad, Lee, and Hsu [24] showed that for $d = 2$, the maximum number of maximal empty rectangles is $O(n^2)$, and that this bound is tight. Datta and Soundaralakshmi [11] conjectured that the maximum number of maximal empty boxes is $O(n^d)$ for each fixed d. The conjecture was later confirmed, indirectly, by Kaplan et al. [20]; see [13, p. 479].

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 be inefficient in the worst case. In fact, Backer and Keil [4, 5] proved that Maximum Empty Box is NP-hard when the dimension d is part of the input, and Giannopoulos, Knauer, Wahlström, and Werner [16] further showed that the problem is W[1]-hard with the dimension d as the parameter. By a standard result in parameterized complexity theory [22, Section 6.3], the W[1]-hardness of the problem implies [16, Corollary 1] that the existence of an exact algorithm running in $n^{o(d)}$ time is unlikely, i.e., unless the so-called Exponential Time Hypothesis (ETH) fails, i.e., unless 3-SAT can be solved in $2^{o(n)}$ time. However, as it is the case with the output-sensitive algorithm of Kaplan et al. [20], and the algorithm of Backer and Keil [5], such algorithms would be much faster when there are only a few maximal empty boxes. Indeed, when d is fixed, Dumitrescu and Jiang [13] proved that the expected number of maximal empty boxes amidst n random points in

[1] For each point p in $\mathbb{R}^d$, let p' be the corresponding point in $\mathbb{R}^{2d}$ by splitting the coordinate x of p along the ith axis into two coordinates x and $-x$ in the $(2i-1)$th and $(2i)$th axes, respectively, $1 \leq i \leq d$. Then a maximal empty box amidst a set S of n points in $\mathbb{R}^d$ becomes a maximal empty orthant amidst the set $S' = \{p' \mid p \in S\}$ of n points in $\mathbb{R}^{2d}$. This connection between maximal empty boxes and maximal empty orthants [20, p. 989] and an algorithm for enumerating all k maximal empty orthants determined by a set of n points in $\mathbb{R}^d$ in $O((n+k)\log^{d-1} n)$ time [20, pp. 994–995] together imply an algorithm for enumerating all k maximal empty boxes amidst n points in $\mathbb{R}^d$ in $O((n+k)\log^{2d-1} n)$ time. This running time is higher than the $O((n+k)\log^{d-1} n)$ time bound misattributed to [20] in [13, pp. 478–479].

the unit hypercube $[0,1]^d$ in $\mathbb{R}^d$ is

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

It is interesting to note the connections [13, Section 2] between the expected number of maximal empty boxes given above, of the order $\Theta(n \log^{d-1} n)$, and two other well-known expected numbers (all under the assumption of a fixed $d \geq 2$): (i) the expected number of maximal points in a set of n random points in $\mathbb{R}^d$, $\Theta(\log^{d-1} n)$, due to Bentley, Kung, Schkolnick, and Thompson [6], and (ii) the expected number of direct dominance pairs in a set of n random points in $\mathbb{R}^d$, $\Theta(n \log^{d-1} n)$, due to Klein [21].

Thus, for fixed d, the expected running times of the output-sensitive algorithms in [20, 5] on random points are nearly linear, i.e., $O(n \log^{O(d)} n)$, although their worst-case running times are $\Omega(n^d)$.

Open question 2. *Can the largest-volume empty box in $\mathbb{R}^d$ for some fixed $d \geq 3$ be computed in $O(n^{d-\delta})$ time for some constant $\delta > 0$? In particular, can the largest-volume empty box in $\mathbb{R}^3$ be computed in $O(n^{3-\delta})$ time for some constant $\delta > 0$?*

Open question 3. *Can the largest-volume empty box in $\mathbb{R}^d$ for any fixed $d \geq 3$ be computed in $O((n+k)\log^c n)$ time for some absolute constant c independent of d?*

In terms of approximation, Dumitrescu and Jiang [12] proposed an algorithm that finds an empty box whose volume is at least $1-\varepsilon$ times the optimal in $O\big((\frac{8ed}{\varepsilon^2})^d \cdot n \log^d n\big)$ time. A faster algorithm for the related Maximum Empty Cube problem, finds an empty axis-parallel hypercube whose volume is at least $1-\varepsilon$ times the optimal in $O(d^2\varepsilon^{-1} \cdot n \log n + (4d\varepsilon^{-1})^{d+1} \cdot n^{1/d} \log n)$ time [12]. As noted in [16, Section 1.5], the running times of these two approximation algorithms are FPT with both d and $1/\varepsilon$ as parameters (since $\log^d n \leq n \cdot f(d)$ for some function f). Moreover, when both d and ε are fixed, the running times become $O(n \log^d n)$ for boxes, and $O(n \log n)$ for cubes, respectively.

Open question 4. *Can a $(1-\varepsilon)$-approximation of the largest-volume empty box in $\mathbb{R}^d$ for any fixed $d \geq 3$ and $\varepsilon > 0$ be computed in $O(n \log^c n)$ time for some absolute constant c independent of d and ε?*

3 Bounds on the volume of a maximum empty box

Besides the problem of computing a largest empty box for a given instance, there is the question of determining its minimum volume over all n-element point sets. 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 [26] proved that $A_d(n) = \Theta\left(\frac{1}{n}\right)$ for any fixed $d \geq 2$. This generic bound was further refined by Dumitrescu and Jiang [12].

For the lower bound, we clearly have $A_d(n) \geq \frac{1}{n+1}$ for each d, by slicing the hypercube with n parallel hyperplanes, each incident to one of the n points. This trivial bound can be improved using the following observation that relates $A_d(n)$ to $A_d(b)$ for fixed d and b:

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

Proof. Partition U_d into smaller boxes by parallel hyperplanes through every $(b+1)$th of the n points along any one of the d axes. Then each of the $(1+o(1))\frac{n}{b+1}$ boxes contains at most b points in its interior, and hence contains an empty box whose volume is at least $A_d(b)$ times its volume. In particular, the largest of these empty boxes has volume at least

$$\frac{A_d(b)}{(1+o(1))\frac{n}{b+1}} = [(b+1)A_d(b) - o(1)] \cdot \frac{1}{n} \ . \qquad \square$$

Clearly $A_d(1) = 1/2$ for any $d \geq 2$. For $d = 2$, it is known [12] that $A_2(2) = \frac{3-\sqrt{5}}{2} = 0.3819\ldots$, $A_2(4) = 1/4$, and that $A_2(6) \geq 3 - 2\sqrt{2} = 0.1715\ldots$. In particular, by Lemma 1 and the easy inequality $A_d(n) \geq A_2(n)$ for any $d \geq 2$, the equality $A_2(4) = 1/4$ implies that $A_d(n) \geq A_2(n) \geq (5/4 - o(1)) \cdot \frac{1}{n}$. Similarly, the lower bound on $A_2(6)$ implies a slightly better lower bound $A_d(n) \geq (1.201\ldots - o(1)) \cdot \frac{1}{n}$. We next determine the exact value of $A_2(3)$ which yields the best lower bound on $A_d(n)$ that we know.

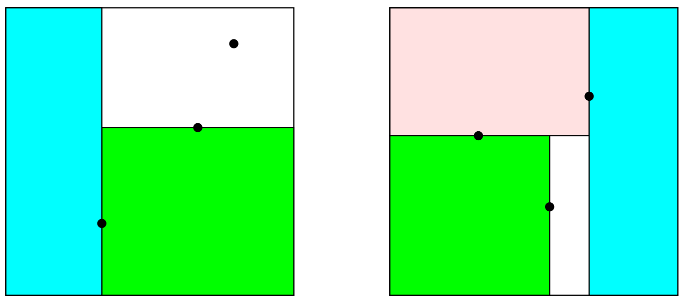

Figure 2: Illustration for $A_2(3)$.

Lemma 2. *Let $y = 0.5549\ldots$ be the solution of the cubic equation $y^3 - 2y^2 - y + 1 = 0$ in the interval $(0,1)$ and let $x = y^2 = 0.3079\ldots$. Then $A_2(3) = x$.*

Proof. We first prove the lower bound $A_2(3) \geq x$. Let (x_1, y_1), (x_2, y_2), and (x_3, y_3) be three points inside $U_2 = [0,1] \times [0,1]$, where $x_1 \leq x_2 \leq x_3$. By symmetry, we only need to consider two cases:

Case 1 (Fig. 2 left): $y_1 \leq y_2 \leq y_3$ and $y_2 \geq 1/2$. The rectangle $(0, x_1) \times (0, 1)$ has area x_1 and the rectangle $(x_1, 1) \times (0, y_2)$ has area $(1 - x_1)y_2 \geq (1 - x_1)/2$. One of the two areas is at least $1/3 > x$.

Case 2 (Fig. 2 right): $y_2 \leq y_1 \leq y_3$ and $x_2 \geq y_1$. The rectangle $(x_3, 1) \times (0, 1)$ has area $1 - x_3$, the rectangle $(0, x_2) \times (0, y_1)$ has area $x_2 y_1 \geq y_1^2$, and the rectangle $(0, x_3) \times (y_1, 1)$ has area $x_3(1 - y_1)$. If $x_3 \leq 1 - x$, then the first area is at least $1 - (1 - x) = x$. If $y_1 \geq y$, then the second area is at least $y^2 = x$. On the other hand, if $x_3 \geq 1 - x$ and $y_1 \leq y$, then the third area is at least $(1 - x)(1 - y) = x$; here the equality $(1 - x)(1 - y) = x$ follows from $y^3 - 2y^2 - y + 1 = 0$ and $x = y^2$. Thus one of the three areas is at least x.

In summary, we have shown that $A_2(3) \geq x$. The upper bound $A_2(3) \leq x$ is realized by the three points (x, y), (y, x), and $(1 - x, 1 - x)$, as in Fig. 2 right. $\square$

Therefore, for any $d \geq 2$, we have

$$A_d(n) \geq A_2(n) \geq (4A_2(3) - o(1)) \cdot \frac{1}{n} = (1.23\ldots - o(1)) \cdot \frac{1}{n}. \qquad (1)$$

From the other direction, for any $d \geq 2$, an upper bound of

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

where p_i is the ith prime, was established in [12] using the Halton-Hammersley generalizations [17, 18] of the van der Corput point set [8, 9]; the proof is similar to that used for bounding the geometric discrepancies of these sets [7, 23].

Further reducing the gap between the lower and upper bounds on $A_d(n)$ remains an interesting problem. A promising approach for improving the lower bound in (1) is to determine the exact values of $A_d(n)$ for small n:

Open question 5. *Is there an algorithm for computing $A_d(n)$? Is there one running in polynomial time? What are the exact values of $A_2(n)$ for $n = 5, 6, \ldots$?*

For the upper bound, it is not at all clear whether the dependence on d in the upper bound is necessary:

Open question 6. *Is the dependence on d necessary in the upper bound on $A_d(n)$ as given by (2), or is $A_d(n) \leq \frac{c}{n}$, where $c > 0$ is an absolute constant?*

A preliminary question to the previous question is the following:

Open question 7. *Given d points in the unit hypercube $[0,1]^d$, is there always an empty box of volume c, where $c > 0$ is an absolute constant, or does $A_d(d)$ tend to zero with the dimension?*

Note that if $A_d(n) \leq c/n$ for some absolute constant c, then for the same c we would have $A_d(d) \leq c/d$ and hence $A_d(d)$ would tend to zero as d goes to infinity.

4 Related problems

Augustine et al. [3] and Kaplan et al. [19] studied a related problem, that of finding the largest-area empty rectangle containing a query point. The authors of [19] showed how to construct in nearly linear time a suitable data structure that takes nearly linear space, so that given a query point q, the largest-area empty rectangle containing q can be computed in $O(\log^4 n)$ time.

Open question 8. *What results can be obtained for the largest-volume empty box query problem in higher dimensions?*

We conclude with another, perhaps only loosely related, however highly entertaining question proposed by Freedman [27, p. 345] in the 1960s; see also [10, p. 113]; more recently, the problem was popularized by Winkler [1, 28, 29, 30]. Let S be a set of n points in the unit square $[0,1]^2$, one of which is the origin. Freedman [27] asked whether one can construct n pairwise interior-disjoint axis-aligned empty rectangles such that the lower left corner of each rectangle is a point in S, and the rectangles jointly cover an area of at least 1/2. Pulleyblank and Winkler conjectured that this is true. Dumitrescu and Tóth [14] showed how to construct a rectangle packing with rectangles anchored at the points in S that jointly cover an area about 0.09. To see why 1/2 is the largest possible constant one can hope for, choose S to be a set of n equally spaced points along the diagonal $[(0,0),(1,1)]$, and then the total area of the rectangles in a packing is at most $\frac{1}{2} + \frac{1}{2n}$.

Open question 9. *Is it always possible to jointly cover an area of at least 1/2 in this setting?*

References

[1] Ponder this challenge: puzzle for June 2004; `http://domino.research.ibm.com/comm/wwwr_ponder.nsf/Challenges/June2004.html`.

[2] A. Aggarwal and S. Suri, Fast algorithms for computing the largest empty rectangle, in: *Proc. 3rd Annual Symposium on Computational Geometry*, 1987, pp. 278–290.

[3] 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, 2010, `http://arxiv.org/abs/1004.0558v2`.

[4] J. Backer and M. Keil, The bichromatic rectangle problem in high dimensions, in: *Proc. 21st Canadian Conference on Computational Geometry*, 2009, pp. 157–160.

[5] J. Backer and M. Keil, The mono- and bichromatic empty rectangle and square problems in all dimensions, in: *Proc. 9th Latin American Symposium on Theoretical Informatics*, 2010, pp. 14–25.

[6] J. L. Bentley, H. T. Kung, M. Schkolnick, and C. D. Thompson, On the average number of maxima in a set of vectors and applications, *Journal of the ACM*, **25** (1978), 536–543.

[7] B. Chazelle, *The Discrepancy Method: Randomness and Complexity*, Cambridge University Press, 2000.

[8] J. G. van der Corput, Verteilungsfunktionen I., *Proc. Nederl. Akad. Wetensch.*, **38** (1935), 813–821.

[9] J. G. van der Corput, Verteilungsfunktionen II., *Proc. Nederl. Akad. Wetensch.*, **38** (1935), 1058–1066.

[10] H. T. Croft, K. J. Falconer, and R. K. Guy, *Unsolved Problems in Geometry*, volume II of Unsolved Problems in Intuitive Mathematics, Springer, New York, 1991.

[11] A. Datta and S. Soundaralakshmi, An efficient algorithm for computing the maximum empty rectangle in three dimensions, *Information Sciences*, **128** (2000), 43–65.

[12] A. Dumitrescu and M. Jiang, On the largest empty axis-parallel box amidst n points, *Algorithmica*, **66(2)** (2013), 225–248.

[13] A. Dumitrescu and M. Jiang, Maximal empty boxes amidst random points, *Combinatorics, Probability and Computing*, **22** (2013), 477–498.

[14] A. Dumitrescu and C. D. Tóth, Packing anchored rectangles, *Combinatorica*, to appear. Also in *Proc. 23rd ACM-SIAM Sympos. on Discrete Algorithms*, 2012, pp. 294–305.

[15] J. Edmonds, J. Gryz, D. Liang, and R. Miller, Mining for empty spaces in large data sets, *Theoretical Computer Science*, **296(3)** (2003), 435–452.

[16] P. Giannopoulos, C. Knauer, M. Wahlström, and D. Werner, Hardness of discrepancy computation and ε-net verification in high dimension, *Journal of Complexity*, **28** (2012), 162–176.

[17] J. H. Halton, On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals, *Numer. Math.*, **2** (1960), 84–90.

[18] J. M. Hammersley, Monte Carlo methods for solving multivariable problems, *Ann. New York Acad. Sci.*, **86** (1960), 844–874.

[19] 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 Symposium on Discrete Algorithms*, 2012, pp. 338–355.

[20] H. Kaplan, N. Rubin, M. Sharir, and E. Verbin, Efficient colored orthogonal range counting, *SIAM Journal on Computing*, **38** (2008), 982–1011.

[21] R. Klein, Direct dominance of points, *International Journal of Computer Mathematics*, **19** (1986), 225–244.

[22] D. Marx, Parameterized complexity and approximation algorithms, *Computer Journal*, **51** (2008), 60–78.

[23] J. Matoušek, *Geometric Discrepancy: An Illustrated Guide*, Springer, 1999.

[24] A. Namaad, D. T. Lee, and W.-L. Hsu, On the maximum empty rectangle problem, *Discrete Applied Mathematics*, **8** (1984), 267–277.

[25] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*, Springer, New York, 1985.

[26] G. Rote and R. F. Tichy, Quasi-Monte-Carlo methods and the dispersion of point sequences, *Mathematical and Computer Modelling*, **23** (1996), 9–23.

[27] W. Tutte, *Recent Progress in Combinatorics: Proceedings of the 3rd Waterloo Conference on Combinatorics*, May 1968, Academic Press, New York, 1969.

[28] P. Winkler, Packing rectangles, pp. 133-134, in *Mathematical Mind-Benders*, A. K. Peters Ltd., Wellesley, MA, 2007.

[29] P. Winkler, Puzzled: rectangles galore, *Communications of the ACM*, **53(11)** (2010), 112.

[30] P. Winkler, Puzzled: solutions and sources, *Communications of the ACM*, **53(12)** (2010), 128.