



Maximum Box Problem on Stochastic Points

Luis E. Caraballo¹ · Pablo Pérez-Lantero² · Carlos Seara³ · Inmaculada Ventura¹

Received: 24 January 2020 / Accepted: 10 October 2021 / Published online: 28 October 2021
© The Author(s) 2021

Abstract

Given a finite set of weighted points in \mathbb{R}^d (where there can be negative weights), the maximum box problem asks for an axis-aligned rectangle (i.e., box) such that the sum of the weights of the points that it contains is maximized. We consider that each point of the input has a probability of being present in the final random point set, and these events are mutually independent; then, the total weight of a maximum box is a random variable. We aim to compute both the probability that this variable is at least a given parameter, and its expectation. We show that even in $d = 1$ these computations are #P-hard, and give pseudo-polynomial time algorithms in the case where the weights are integers in a bounded interval. For $d = 2$, we consider that each point is colored red or blue, where red points have weight $+1$ and blue points weight $-\infty$. The random variable is the maximum number of red points that can be covered with a box not containing any blue point. We prove that the above two computations are also #P-hard, and give a polynomial-time algorithm for computing the probability that there is a box containing exactly two red points, no blue point, and a given point of the plane.

Keywords Random weighted points · Boxes · Red and blue points · #P-hardness

1 Introduction

The *maximum box* problem receives as input a finite point set in \mathbb{R}^d , where each point is associated with a positive or negative weight, and outputs an axis-aligned rectangle (i.e., box) such that the sum of the weights of the points that it contains is maximized [3]. We consider this problem on a recent uncertainty model in which each element of the input has assigned a probability. Particularly, each point has assigned its own and

A preliminary version of this work appeared in LATIN 2018 (Latin American Theoretical Informatics 2018, April 16–19, Buenos Aires, Argentina)

This work has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 734922.

Extended author information available on the last page of the article

independent probability of being present in the final (hence random) point set. Then, one can ask the following questions: What is the probability that for the final point set there exists a box that covers a weight sum greater than or equal to a given parameter? What is the expectation of the maximum weight sum that can be covered with a box? Uncertainty models come from real scenarios in which large amounts of data, arriving from many sources, have inherent uncertainty. In computational geometry, we can find several recent works on uncertain point sets such as: the expected total length of the minimum Euclidean spanning tree [5]; the probability that the distance between the closest pair of points is at least a given parameter [11]; the computation of the most-likely convex hull [16]; the probability that the area or perimeter of the convex hull is at least a given parameter [15]; the smallest enclosing ball [9]; the probability that a 2-colored point set is linearly separable [10]; the area of the minimum enclosing rectangle [17]; and Klee’s measure of random rectangles [20]. We deal with the *maximum box* problem in the above mentioned random model. The *maximum box* problem is a geometric combinatorial optimization problem, different from most of the problems considered in this random setting that are computations of some measure or structure of the extent of the geometric objects.

For $d = 1$, the maximum box problem asks for an interval of the line. If the points are uncertain as described above, then it is equivalent to consider as input a sequence of random numbers, where each number has two possible outcomes: zero if the number is not present and the actual value of the number otherwise. The output is the *subsequence of consecutive numbers* with maximum sum. We consider the simpler case when the subsequence is a partial sum, that is, it contains the first (or leftmost) number of the sequence. More formally: We say that a random variable X is *zero-value* if $X = v$ with probability ρ , and $X = 0$ with probability $1 - \rho$, for an integer number $v = v(X) \neq 0$ and a probability ρ . We refer to v as the value of X and to ρ as the probability of X . In any sequence of zero-value variables, all variables are assumed to be mutually independent. Let $\mathcal{X} = X_1, X_2, \dots, X_n$ be a sequence of n mutually independent zero-value variables, whose values are a_1, a_2, \dots, a_n , respectively. We study the random variable

$$S(\mathcal{X}) = \max\{0, X_1, X_1 + X_2, \dots, X_1 + \dots + X_n\},$$

which is the maximum partial sum of the random sequence \mathcal{X} . The fact that $\mathbb{E}[\max\{X, Y\}]$ is not necessarily $\max\{\mathbb{E}[X], \mathbb{E}[Y]\}$, even if X and Y are independent random variables, makes hard the computation of the expectation $\mathbb{E}[S(\mathcal{X})]$.

Kleinberg et al. [13] proved that the problem of computing $\Pr[X_1 + \dots + X_n > 1]$ is #P-complete, in the case where the values of the variables X_1, X_2, \dots, X_n are all positive. The proof can be straightforwardly adapted to also show that computing $\Pr[X_1 + \dots + X_n > z]$ is #P-complete, where the values of X_1, X_2, \dots, X_n are all positive, for any fixed $z > 0$. This last fact implies that computing $\Pr[S(\mathcal{X}) \geq z]$ for any fixed $z \geq 1$ is #P-hard. We show hardness results when the probabilities of X_1, X_2, \dots, X_n are the same, and their values are not necessarily positive. Namely, we prove (Sect. 2.1) that computing $\Pr[S(\mathcal{X}) \geq z]$ for any fixed $z \geq 1$, and computing the expectation $\mathbb{E}[S(\mathcal{X})]$ are both #P-hard problems, even if all variables of \mathcal{X} have the same less-than-one probability. When $a_1, a_2, \dots, a_n \in [-a..b]$ for bounded $a, b \in \mathbb{N}$,

we show (Sect. 2.2) that both $\Pr[S(\mathcal{X}) \geq z]$ and $\mathbb{E}[S(\mathcal{X})]$ can be computed in time polynomial in n , a , and b . For two integers $u < v$, we use $[u..v]$ to denote the set $\{u, u + 1, \dots, v\}$.

For $d = 2$, we consider the maximum box problem in the context of red and blue points, where red points have weight $+1$ and blue points weight $-\infty$. Let R and B be disjoint finite point sets in the plane with a total of n points, where the elements of R are colored red and the elements of B are colored blue. The maximum box problem asks for a box H such that $|H \cap R|$ is maximized subject to $|H \cap B| = \emptyset$. This problem has been well studied, with algorithms whose running times go from $O(n^2 \log n)$ [6], $O(n^2)$ [3], to $O(n \log^3 n)$ [2]. Let $S \subseteq R \cup B$ be the random point set where every point $p \in R \cup B$ is included in S independently and uniformly at random with probability $\pi(p) \in [0, 1]$. Let $\text{box}(S)$ denote the random variable equal to the maximum number of red points in S that can be covered with a box not covering any blue point of S .

We prove (Sect. 3.1) that computing the probability $\Pr[\text{box}(S) \geq k]$ for any given $k \geq 2$, and computing the expectation $\mathbb{E}[\text{box}(S)]$, are both #P-hard problems. We further show (Sect. 3.2) that given a point o of the plane, computing the probability that there exists a box containing exactly two red points of S , no blue point of S , and the point o can be solved in polynomial time. If we remove the restriction of containing o , this problem is also #P-hard. This fact is a direct consequence of the previous #P-hardness proofs.

In all running time upper bounds in this paper, in both algorithms and reductions, we assume a real RAM model of computation where each arithmetic operation on large-precision numbers takes constant time. Otherwise, the running times should be multiplied by a factor proportional to the bit complexity of the numbers, which is polynomial in n and the bit complexity of the input probability values [5, 11].

2 Weighted Points

2.1 Hardness

Theorem 1 *For any integer $z \geq 1$ and any $\rho \in (0, 1)$, the following problem is #P-hard: Given a sequence $\mathcal{X} = X_1, X_2, \dots, X_n$ of n zero-value random variables, each with probability ρ , compute $\Pr[S(\mathcal{X}) \geq z]$.*

Proof Let $z \geq 1$ be an integer, and $\rho \in (0, 1)$ be a probability. We show a Turing reduction from the #SubsetSum problem, which is known to be #P-complete [8]. Our reduction assumes an unknown algorithm (i.e., oracle) $\mathcal{A}(\mathcal{X})$ computing $\Pr[S(\mathcal{X}) \geq z]$ for any finite sequence \mathcal{X} of zero-value random variables, that will be called twice. #SubsetSum receives as input a set $\{a_1, \dots, a_n\} \subset \mathbb{N}$ of n numbers and a target $t \in \mathbb{N}$, and counts the number of subsets $J \subseteq [1..n]$ such that $\sum_{j \in J} a_j = t$. It remains #P-hard if the subsets J must also satisfy $|J| = k$, for given $k \in [1..n]$. Let $(\{a_1, \dots, a_n\}, t, k)$ be an instance of this #SubsetSum, in which we assume $t \leq a_1 + \dots + a_n$.

Let $m = \max\{z, 1 + a_1 + \dots + a_n\} > t$, and $\mathcal{X} = X_0, X_1, X_2, \dots, X_n$ be a sequence of $n + 1$ zero-value random variables, each with probability ρ , where the value of X_0 is $-km - t + z$, and the value of X_i is $m + a_i$ for every $i \in [1..n]$. Observe

that for $J \subseteq [1..n]$ we have

$$\sum_{j \in J} (m + a_j) = km + t \Leftrightarrow \left(\sum_{j \in J} a_j = t \text{ and } |J| = k \right).$$

Furthermore, $|J| > k$ implies $\sum_{j \in J} (m + a_j) > km + t$.

Let $J_{\mathcal{X}} = \{j \in [1..n] : X_j \neq 0\}$, and for any s , let

$$N_s = \left| J \subseteq [1..n] : |J| = k, \sum_{j \in J} a_j \geq s \right|.$$

Then, #SubsetSum asks for $N_t - N_{t+1}$. Call $\mathcal{A}(\mathcal{X})$ to compute $\Pr[S(\mathcal{X}) \geq z]$. Then:

$$\Pr[S(\mathcal{X}) \geq z] = \Pr[S(\mathcal{X}) \geq z, X_0 = 0] + \Pr[S(\mathcal{X}) \geq z, X_0 = -km - t + z]$$

where,

$$\begin{aligned} \Pr[S(\mathcal{X}) \geq z, X_0 = 0] &= \Pr[X_0 = 0] \cdot \Pr[S(\mathcal{X}) \geq z \mid X_0 = 0] \\ &= (1 - \rho) \cdot \Pr[|J_{\mathcal{X}}| \geq 1] = (1 - \rho) \cdot (1 - \Pr[|J_{\mathcal{X}}| = 0]) \\ &= (1 - \rho) \cdot (1 - (1 - \rho)^n), \end{aligned}$$

and

$$\begin{aligned} &\Pr[S(\mathcal{X}) \geq z, X_0 = -km - t + z] \\ &= \Pr[X_0 = -km - t + z] \cdot \Pr[S(\mathcal{X}) \geq z \mid X_0 = -km - t + z] \\ &= \rho \cdot \Pr \left[-km - t + z + \sum_{j \in J_{\mathcal{X}}} (m + a_j) \geq z \right] \\ &= \rho \cdot \Pr \left[\sum_{j \in J_{\mathcal{X}}} (m + a_j) \geq km + t \right] \\ &= \rho \cdot \left(\Pr \left[|J_{\mathcal{X}}| = k, \sum_{j \in J_{\mathcal{X}}} (m + a_j) \geq km + t \right] \right. \\ &\quad \left. + \sum_{i=k+1}^n \Pr \left[|J_{\mathcal{X}}| = i, \sum_{j \in J_{\mathcal{X}}} (m + a_j) \geq km + t \right] \right) \\ &= \rho \cdot \Pr \left[|J_{\mathcal{X}}| = k, \sum_{j \in J_{\mathcal{X}}} a_j \geq t \right] + \rho \cdot \sum_{i=k+1}^n \Pr[|J_{\mathcal{X}}| = i] \\ &= \rho \cdot N_t \cdot \rho^k \cdot (1 - \rho)^{n-k} + \rho \cdot \sum_{i=k+1}^n \binom{n}{i} \cdot \rho^i \cdot (1 - \rho)^{n-i}. \end{aligned}$$

Hence, we can compute N_t in polynomial time from the value of $\Pr[S(\mathcal{X}) \geq z]$. Consider now the random sequence $\mathcal{X}' = X'_0, X_1, X_2, \dots, X_n$, where X'_0 has value $-km - (t+1) + z$. Using arguments similar to those above, by calling $\mathcal{A}(\mathcal{X}')$ to compute $\Pr[S(\mathcal{X}') \geq z]$, we can compute N_{t+1} in polynomial time from this probability. Then, $N_t - N_{t+1}$ can be computed in polynomial time, plus the time of calling twice the oracle \mathcal{A} . This implies the theorem. \square

Theorem 2 *For any $\rho \in (0, 1)$, the following problem is #P-hard: Given a sequence $\mathcal{X} = X_1, \dots, X_n$ of n zero-value random variables, each with probability ρ , compute $\mathbb{E}[S(\mathcal{X})]$.*

Proof Let $\mathcal{X} = X_1, X_2, \dots, X_n$ be a sequence of zero-value random variables, each with probability ρ , and consider the sequence $\mathcal{X}' = X_0, X_1, \dots, X_n$, where X_0 is a zero-value random variable with value -1 and probability ρ . Let w be the sum of the positive values among the values of X_1, \dots, X_n . Then:

$$\mathbb{E}[S(\mathcal{X})] = \sum_{i=1}^w i \cdot \Pr[S(\mathcal{X}) = i] = \sum_{i=1}^w \Pr[S(\mathcal{X}) \geq i],$$

and

$$\begin{aligned} \mathbb{E}[S(\mathcal{X}')] &= \sum_{i=1}^w \Pr[S(\mathcal{X}') \geq i] \\ &= \sum_{i=1}^w (\Pr[S(\mathcal{X}') \geq i, X_0 = 0] + \Pr[S(\mathcal{X}') \geq i, X_0 = -1]) \\ &= \sum_{i=1}^w (\Pr[X_0 = 0] \cdot \Pr[S(\mathcal{X}') \geq i \mid X_0 = 0] \\ &\quad + \Pr[X_0 = -1] \cdot \Pr[S(\mathcal{X}') \geq i \mid X_0 = -1]) \\ &= \sum_{i=1}^w ((1 - \rho) \cdot \Pr[S(\mathcal{X}) \geq i] + \rho \cdot \Pr[S(\mathcal{X}) \geq i + 1]) \\ &= \sum_{i=1}^w (1 - \rho) \cdot \Pr[S(\mathcal{X}) \geq i] + \sum_{i=2}^{w+1} \rho \cdot \Pr[S(\mathcal{X}) \geq i] \\ &= (1 - \rho) \cdot \Pr[S(\mathcal{X}) \geq 1] + \sum_{i=2}^w \Pr[S(\mathcal{X}) \geq i]. \end{aligned}$$

Then, we have that

$$\mathbb{E}[S(\mathcal{X})] - \mathbb{E}[S(\mathcal{X}')] = \rho \cdot \Pr[S(\mathcal{X}) \geq 1].$$

Since computing $\Pr[S(\mathcal{X}) \geq 1]$ is #P-hard (Theorem 1), then computing $\mathbb{E}[S(\mathcal{X})]$ is also #P-hard via a Turing reduction. \square

2.2 Pseudo-Polynomial Time Algorithms

Let $\mathcal{X} = X_1, X_2, \dots, X_n$ be a sequence of n zero-value random variables, with values $a_1, a_2, \dots, a_n \in [-a..b] \subset \mathbb{Z}$ and probabilities $\rho_1, \rho_2, \dots, \rho_n$, respectively, for some $a, b \in \mathbb{N}$. We show that both $\Pr[S(\mathcal{X}) \geq z]$ and $\mathbb{E}[S(\mathcal{X})]$ can be computed in time polynomial in n, a , and b . Let $J = \{j \in [1..n] : a_j < 0\}$ and

$$w_0 = \sum_{j \in J} |a_j| = O(na) \text{ and } w_1 = \sum_{j \in [1..n] \setminus J} a_j = O(nb).$$

For every $t \in [1..n]$, let

$$S_t = X_1 + \dots + X_t, \quad M_t = \max\{0, S_1, S_2, \dots, S_t\}, \text{ and}$$

$$L_t = \{\Pr[M_t = k, S_t = s] : k \in [0..w_1], s \in [-w_0..w_1], k \geq s\}.$$

Observe that L_t has size $O(w_1(w_0 + w_1)) = O(nb(na + nb)) = O(n^2b(a + b))$ for every t , and that L_1 can be trivially computed. Using the dynamic programming algorithm design paradigm, we next show how to compute the values of $L_t, t \geq 2$, assuming that all values of L_{t-1} have been computed. Note that:

$$\Pr[M_t = k, S_t = s] = \Pr[M_t = k, S_t = s, X_t = 0] + \Pr[M_t = k, S_t = s, X_t = a_t],$$

where

$$\begin{aligned} \Pr[M_t = k, S_t = s, X_t = 0] &= \Pr[X_t = 0] \cdot \Pr[M_t = k, S_t = s \mid X_t = 0] \\ &= (1 - \rho_t) \cdot \Pr[M_{t-1} = k, S_{t-1} = s] \end{aligned}$$

and

$$\begin{aligned} \Pr[M_t = k, S_t = s, X_t = a_t] &= \Pr[X_t = a_t] \cdot \Pr[M_t = k, S_t = s \mid X_t = a_t] \\ &= \rho_t \cdot \Pr[M_t = k, S_t = s \mid X_t = a_t]. \end{aligned}$$

When $k = s$, we have for $a_t < 0$ that $\Pr[M_t = k, S_t = s \mid X_t = a_t] = 0$, since this event indicates that $S_t = X_1 + \dots + X_t$ is a maximum partial sum of X_1, \dots, X_t , but this cannot happen because any maximum partial sum ends in a positive element. For $a_t > 0$ we have

$$\begin{aligned} \Pr[M_t = k, S_t = s \mid X_t = a_t] &= \Pr[M_{t-1} \leq k, S_{t-1} = s - a_t] \\ &= \sum_{i=s-a_t}^k \Pr[M_{t-1} = i, S_{t-1} = s - a_t]. \end{aligned}$$

When $k > s$, M_t does not count the element a_t , hence $M_{t-1} = M_t$. Then

$$\Pr[M_t = k, S_t = s \mid X_t = a_t] = \Pr[M_{t-1} = k, S_{t-1} = s - a_t].$$

Modeling each set L_t as a 2-dimensional table (or array), note that each value of L_t can be computed in $O(k - (s - a_t)) = O(w_1)$ time, and hence all values of L_t can be computed in $O(w_1) \cdot O(n^2b(a + b)) = O(n^3b^2(a + b))$ time. Finally, once all the values of L_n have been computed in $O(n) \cdot O(n^3b^2(a + b)) = O(n^4b^2(a + b))$ time, we can compute $\Pr[S(\mathcal{X}) \geq z]$ as

$$\Pr[S(\mathcal{X}) \geq z] = \sum_{k=z}^{w_1} \Pr[S(\mathcal{X}) = k] = \sum_{k=z}^{w_1} \sum_{s=-w_0}^k \Pr[M_n = k, S_n = s]$$

in $O(w_1(w_0 + w_1)) = O(n^2b(a + b))$ time, and $\mathbb{E}[S(\mathcal{X})]$ as

$$\mathbb{E}[S(\mathcal{X})] = \sum_{z=1}^{w_1} \Pr[S(\mathcal{X}) \geq z]$$

in $O(w_1) = O(nb)$ time. As a consequence, we get the following result.

Theorem 3 *Let $\mathcal{X} = X_1, X_2, \dots, X_n$ be a sequence of n zero-value random variables, with values $a_1, a_2, \dots, a_n \in [-a..b] \subset \mathbb{Z}$ and probabilities $\rho_1, \rho_2, \dots, \rho_n$, respectively, for some $a, b \in \mathbb{N}$. Then, both $\Pr[S(\mathcal{X}) \geq z]$ and $\mathbb{E}[S(\mathcal{X})]$ can be computed in time polynomial in n, a , and b .*

3 Red and Blue Points

3.1 Hardness

Given a graph $G = (V, E)$, a subset $V' \subseteq V$ is an *independent set* of G if no pair of vertices of V' define an edge in E . Let $N(G)$ denote the number of independent sets of G . The problem **#IndSet** of counting the number of independent sets in a graph is #P-complete, even if the graph is planar, bipartite, and with maximum degree 4 [18]. We show in what follows a one-to-many Turing reduction from **#IndSet** to the problem of computing $\Pr[\text{box}(S) \geq k]$, for any given $k \geq 2$. The proof uses techniques similar to that of Kamousi et al. [11] to prove that counting vertex covers in weighted unit-disk graphs is #P-hard, and that of Vadhan [18] to prove that counting weighted matchings in planar bipartite graphs is #P-hard.

Let $G = (V, E)$ be the input of **#IndSet**, where G is a planar, bipartite graph with maximum degree 4. Let $n = |V|$ and $m = |E| = O(n)$. For an overview of the whole proof, refer to Fig. 1.

For any subset $V' \subseteq V$ and any edge $e = \{u, v\} \in E$, we say that V' *1-covers* edge e if exactly one of u and v belongs to V' . We also say that V' *2-covers* e if both u and v belong to V' . Let $C_{i,j}$ denote the number of subsets of V that 1-cover exactly i edges and 2-cover exactly j edges. Then,

$$N(G) = \sum_{i=0}^m C_{i,0}.$$

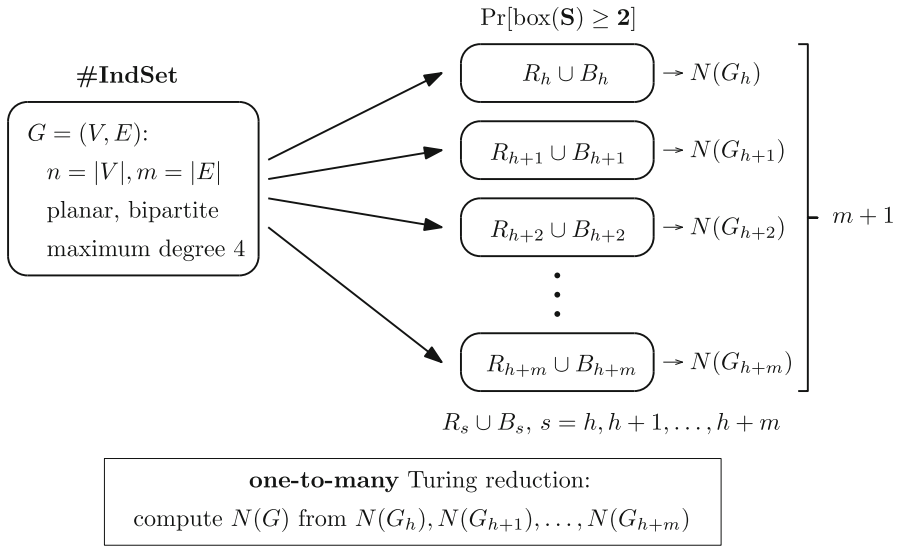


Fig. 1 Given an instance $G = (V, E)$ of **#IndSet**, for every s in the set $\{h, h + 1, \dots, h + m\}$ of $m + 1$ integers polynomially bounded in n , we construct the random colored point set $R_s \cup B_s$. Let G_s be the graph obtained from G by subdividing each edge with exactly s new vertices. Assuming that an oracle computes $\Pr[\text{box}(S) \geq 2]$, where $S \subseteq R_s \cup B_s$ is the random sample, we show that $N(G_s)$ can be computed in constant time from this probability. From all of the values $N(G_h), N(G_{h+1}), \dots, N(G_{h+m})$ already computed, we show that $N(G)$ can be computed in polynomial time

For $s \geq 1$, let $G_s = (V_s, E_s)$ be the graph obtained from G by adding exactly s intermediate vertices on each edge of E . Let $\{f_i\}_{i=1}^\infty$ be the Fibonacci sequence, with $f_1 = f_2 = 1$ and $f_i = f_{i-1} + f_{i-2}$ for $i \geq 3$. Let $\alpha_i = f_{i+1}/f_{i+2}$ for $i \geq 0$. The next lemma relates the number $N(G_s)$ of independent sets of G_s to the values $C_{i,j}$ in G .

Lemma 4 *We have*

$$N(G_s) = (f_{s+2})^m \cdot \sum_{0 \leq i+j \leq m} C_{i,j} \cdot (\alpha_s)^i \cdot (1 - \alpha_s)^j.$$

Proof Any independent set $V'_s \subseteq V_s$ of G_s induces the subset $V'_s \cap V$ of V , which is not necessarily an independent set of G because it may 2-cover some edges. Let $V' \subseteq V$ be any subset of V that 1-covers i edges and 2-covers j edges. For any edge $e \in E$, let p_e denote the path induced by the s vertices added to e when constructing G_s from G . An independent set of G_s inducing V' can be obtained by starting with V' and adding vertices in the following way. For every edge $e = \{u, v\} \in E$:

- (1) if V' neither 1-covers nor 2-covers e , then add any independent set of p_e .
- (2) if V' 1-covers e , say $u \in V'$, then add any independent set of p_e not containing the extreme vertex of p_e adjacent to u in G_s .
- (3) if V' 2-covers e , then add any independent set of p_e with no extreme vertex.

It is well known that the number of independent sets of a path of length ℓ is exactly $f_{\ell+3}$ [18]. Since p_e has length $s - 1$ for every e , the number of choices for cases (1), (2),

and (3) are f_{s+2} , f_{s+1} , and f_s , respectively. Therefore, the number of independent sets of G_s inducing a subset of V that 1-covers i edges and 2-covers j edges is precisely $C_{i,j} \cdot (f_{s+1})^i \cdot (f_s)^j \cdot (f_{s+2})^{m-i-j}$. Hence, the number $N(G_s)$ of independent sets of G_s satisfies

$$\begin{aligned} N(G_s) &= \sum_{0 \leq i+j \leq m} C_{i,j} \cdot (f_{s+1})^i \cdot (f_s)^j \cdot (f_{s+2})^{m-i-j} \\ &= (f_{s+2})^m \cdot \sum_{0 \leq i+j \leq m} C_{i,j} \cdot \left(\frac{f_{s+1}}{f_{s+2}}\right)^i \cdot \left(\frac{f_s}{f_{s+2}}\right)^j \\ &= (f_{s+2})^m \cdot \sum_{0 \leq i+j \leq m} C_{i,j} \cdot \left(\frac{f_{s+1}}{f_{s+2}}\right)^i \cdot \left(1 - \frac{f_{s+1}}{f_{s+2}}\right)^j \\ &= (f_{s+2})^m \cdot \sum_{0 \leq i+j \leq m} C_{i,j} \cdot (\alpha_s)^i \cdot (1 - \alpha_s)^j, \end{aligned}$$

which completes the proof. □

Lemma 5 *Let T be a set of $m + 1$ integers, each bounded by a polynomial in n . If we know the value of $N(G_s)$ for every $s \in T$, then the number $N(G)$ can be computed in time polynomial in n .*

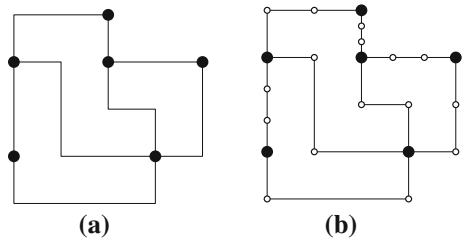
Proof For every $s \in T$, the value of $(f_{s+2})^m$ can be computed in $O(\log s + \log m) = O(\log n)$ time, and the value of α_s also in $O(\log s) = O(\log n)$ time. Let $b_s = N(G_s)/(f_{s+2})^m$ for every $s \in T$. Consider the polynomial

$$P(x) = \sum_{0 \leq i+j \leq m} C_{i,j} \cdot x^i \cdot (1 - x)^j = a_0 + a_1x + a_2x^2 + \dots + a_mx^m,$$

of degree m , whose coefficients a_0, a_1, \dots, a_m are linear combinations of the terms $C_{i,j}$. By Lemma 4, and using the known values of b_s and α_s for every $s \in T$, we have $m + 1$ evaluations of $P(x)$ of the form $b_s = P(\alpha_s)$, each corresponding to the linear equation $b_s = a_0 + a_1 \cdot \alpha_s + a_2 \cdot \alpha_s^2 + \dots + a_m \cdot \alpha_s^m$ with variables the coefficients a_0, a_1, \dots, a_m . The main matrix of this system of $m + 1$ linear equations is Vandermonde, with parameters α_s for every $s \in T$. All α_s are distinct (refer to [18] or Appendix A for completeness), so the determinant of the main matrix is non-zero, and the system has a unique solution a_0, a_1, \dots, a_m , which can be computed in time polynomial in n . Finally, observe that for $j = 0$, the coefficient of the polynomial $P(x) = C_{i,j} \cdot x^i \cdot (1 - x)^j = C_{i,0} \cdot x^i$ is $C_{i,0}$. Furthermore, for $j > 0$, all the coefficients of the polynomial

$$\begin{aligned} P(x) &= C_{i,j} \cdot x^i \cdot (1 - x)^j \\ &= C_{i,j} \cdot x^i \cdot \left(\binom{j}{0} - \binom{j}{1}x + \binom{j}{2}x^2 - \dots + (-1)^j \binom{j}{j}x^j \right) \end{aligned}$$

Fig. 2 a An embedding of G . **b** The embedding of G_s for $s = 2$: two intermediate vertices are added to each edge of G so that all polyline bends are covered



sum up to zero. Indeed, it suffices to note that $P(1) = 0$. Hence, we obtain

$$a_0 + a_1 + a_2 + \dots + a_m = \sum_{i=0}^m C_{i,0} = N(G).$$

which shows that $N(G)$ can be computed in time polynomial in n . □

In polynomial time, the graph $G = (V, E)$ can be embedded in the plane using $O(n^2)$ area in such a way that its vertices are at integer coordinates, and its edges are drawn so that they are polylines made up of line segments of the form $x = i$ or $y = j$, for integers i and j [19] (see Fig. 2a). Let $h = O(n)$ be the maximum number of bends of the polylines corresponding to the edges.

For $s = h, h + 1, \dots, h + m$, we embed the graph G_s in the following way. We embed the graph G as above; scale the embedding by factor $2(s + 1)$; and for each edge of G , add s intermediate vertices to the polyline of the edge so that they have even integer coordinates and cover all bends of the polyline (see Fig. 2b). Then, each edge of G_s is represented in the embedding by a vertical or horizontal segment. Let the point set $R_0 = R_0(s) \subset \mathbb{Z}^2$ denote the vertex set of the embedding, and color these points in red. By translation if necessary, we can assume $R_0 \subset [0..N]^2$ for some $N = O(n^2)$. Let $B_0 = B_0(s)$ be the following set of blue points: For each horizontal or vertical line ℓ through a point of R_0 , and each two consecutive points $p, q \in R_0$ in ℓ such that the vertices p and q are not adjacent in G_s , we add a blue point in the segment pq connecting p and q , in order to “block” this segment, so that the blue point has one odd coordinate. In this way, blue points blocking horizontal segments have odd x -coordinates and even y -coordinates; and blue points blocking vertical segments have even x -coordinates and odd y -coordinates. Hence, a blue point cannot block at the same time a horizontal and a vertical segment defined by two red points. Note that $|B_0| = O(|R_0|) = O(n + m \cdot s) = O(n^2)$. Now, a horizontal or vertical segment connecting two points p and q of $R_0 \cup B_0$ represents an edge of G_s if and only if $p, q \in R_0$ and the segment does not contain any other point of $R_0 \cup B_0$ in its interior (see Fig. 4).

We perturb $R_0 \cup B_0 \subset [0..N]^2$ to obtain a point set with rational coordinates by applying the function $\lambda : [0..N]^2 \rightarrow \mathbb{Q}^2$, where

$$\lambda(p) = \left(x(p) + \frac{x(p) + y(p)}{4N + 1}, y(p) + \frac{x(p) + y(p)}{4N + 1} \right),$$

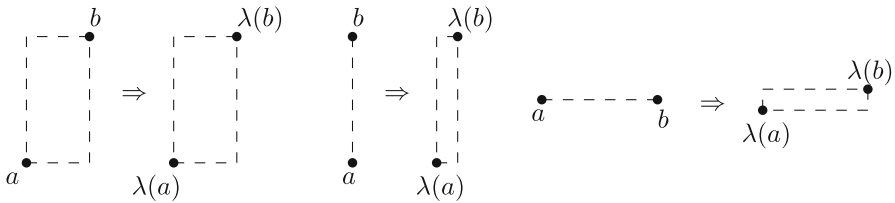


Fig. 3 The way in which points are perturbed using function λ

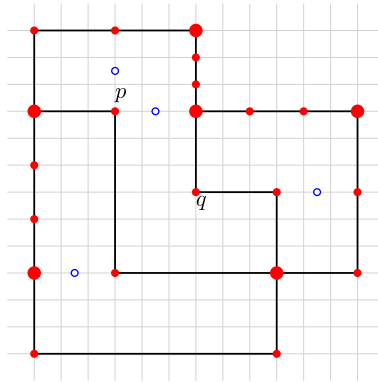


Fig. 4 The point set $R_0 \cup B_0$ for the embedding $G_s, s = 2$, depicted in Fig. 2b. The grid lines are equally spaced in two units, then their intersection points have even coordinates. After the perturbation with the function λ (see Fig. 3), the extra blue points $\{p + (1/2, 1/2), p + (\delta, -\delta) \mid p \in R_s\}$ are included in B_s to avoid that the box $D(\lambda(p), \lambda(q))$ contains no blue point, for any two points $p, q \in R_0$ which do not belong to the same edge of the embedding. See for example the points p and q , denoted in the figure (Color figure online)

to every $p \in R_0 \cup B_0$, where $x(p)$ and $y(p)$ denote the x - and y -coordinates of p , respectively. Similar perturbations can be found in [1,4], and refer to Fig. 3. Since λ is injective [4], let λ^{-1} denote the inverse of λ . For $A \subset [0..N]^2$, let $\lambda(A) = \{\lambda(p) \mid p \in A\}$, and for $A' \subset \lambda([0..N]^2)$ let $\lambda^{-1}(A') = \{\lambda^{-1}(p) \mid p \in A'\}$. Let $\delta = 1/(4N + 2)$, and define the sets

$$R_s = \lambda(R_0) \text{ and } B_s = \lambda(B_0) \cup \{p + (1/2, 1/2), p + (\delta, -\delta) \mid p \in R_s\}.$$

To simplify the notation, let $R = R_s$ and $B = B_s$. Note that $|R| = O(n^2)$ and $|B| = O(n^2)$. For two points a and b , let $D(a, b)$ be the box with the segment ab as a diagonal. The proof of the next technical lemma is deferred to Appendix B.

Lemma 6 *For any different $p, q \in R$, the vertices $\lambda^{-1}(p)$ and $\lambda^{-1}(q)$ are adjacent in G_s if and only if the box $D(p, q)$ contains no point of B .*

Theorem 7 *Given $R \cup B$, it is #P-hard to compute $\Pr[\text{box}(S) \geq k]$ for every integer $k \geq 2$, and it is also #P-hard to compute $\mathbb{E}[\text{box}(S)]$.*

Proof Let $k = 2$. Assume that there exists an algorithm (i.e., oracle) \mathcal{A} that computes $\Pr[\text{box}(S) \geq 2]$. Consider the planar bipartite graph $G = (V, E)$, with maximum

degree 4, the input of **#IndSet**. Let $T = \{h, h+1, \dots, h+m\}$. For each $s \in T$ we create the graph G_s , embed G_s in the plane, and build the colored point set $R \cup B = R_s \cup B_s$ from this embedding. For each red point $p \in R$ we set its probability $\pi(p)$ to $1/2$, and for each blue point $q \in B$ we set $\pi(q) = 1$. Note from Lemma 6 that there does not exist any box containing more than two red points of R and no blue point from B . Then, we have $\Pr[\text{box}(S) \geq 2] = \Pr[\text{box}(S) = 2]$, where $S \subseteq R \cup B$ is the random subset of $R \cup B$. Furthermore,

$$\begin{aligned} \Pr[\text{box}(S) = 2] &= \Pr[\lambda^{-1}(S \cap R) \text{ is not an independent set in } G_s] \\ &= 1 - \Pr[\lambda^{-1}(S \cap R) \text{ is an independent set in } G_s] \\ &= 1 - \frac{N(G_s)}{2^{|R|}} \\ N(G_s) &= 2^{|R|} \cdot (1 - \Pr[\text{box}(S) \geq 2]). \end{aligned}$$

Then, for each $s \in T$ we can compute $N(G_s)$ by calling \mathcal{A} once. By Lemma 5, we can compute $N(G)$ from the $m + 1$ computed values of $N(G_s)$ for each $s \in T$. Hence, it is **#P**-hard to compute $\Pr[\text{box}(S) \geq 2]$ via a Turing reduction from **#IndSet**. To show that computing $\mathbb{E}[\text{box}(S)]$ is also **#P**-hard, for each $s \in T$ consider the above point set $R \cup B$ and note that

$$\begin{aligned} \mathbb{E}[\text{box}(S)] &= 1 \cdot \Pr[\lambda^{-1}(S \cap R) \text{ is an independent set in } G_s, S \cap R \neq \emptyset] + \\ &\quad 2 \cdot \Pr[\lambda^{-1}(S \cap R) \text{ is not an independent set in } G_s] \\ &= \frac{N(G_s) - 1}{2^{|R|}} + 2 \cdot \left(1 - \frac{N(G_s)}{2^{|R|}}\right) = 2 - \frac{N(G_s) + 1}{2^{|R|}} \\ N(G_s) &= 2^{|R|} \cdot (2 - \mathbb{E}[\text{box}(S)]) - 1. \end{aligned}$$

Let now $k \geq 3$. For each $s \in T$, the graph G_s can be colored with two colors, 0 and 1, because it is also a bipartite graph. Each red point in R corresponds to a vertex in G_s . Then, for each red point $p \in R$ with color 0 we add new $\lfloor \frac{k}{2} \rfloor - 1$ red points close enough to p (say, at distance much smaller than δ), and for each red point $q \in R$ with color 1 we add new $\lceil \frac{k}{2} \rceil - 1$ red points close enough to q . Let $R' = R'(s)$ be the set of all new red points, and assign $\pi(u) = 1$ for every $u \in R'$. In this new colored point set $R \cup R' \cup B$, there is no box containing more than k red points and no blue point. Furthermore, every box containing exactly k red points and no blue point contains two points $p, q \in R$ such that $\lambda^{-1}(p)$ and $\lambda^{-1}(q)$ are adjacent in G_s ; and for every $p, q \in R$ such that $\lambda^{-1}(p)$ and $\lambda^{-1}(q)$ are adjacent in G_s such a box containing p and q exists. Then, when taking $S \subseteq R \cup R' \cup B$ at random, we also have

$$\begin{aligned} \Pr[\text{box}(S) \geq k] &= \Pr[\text{box}(S) = k] \\ &= \Pr[\lambda^{-1}(S \cap R) \text{ is not an independent set in } G_s] \\ &= 1 - \frac{N(G_s)}{2^{|R|}}. \end{aligned}$$

Hence, computing $\Pr[\text{box}(S) \geq k]$ is also **#P**-hard for any $k \geq 3$. □

3.2 Two-Point Boxes

From the proof of Theorem 7, note that it is also #P-hard to compute the probability that in $S \subseteq R \cup B$ there exists a box that contains exactly two red points p, q and no blue point; and that this box can be restricted to be the minimum box $D(p, q)$ having p and q as opposed vertices. In this section, we present a polynomial-time algorithm to compute such a probability when the box is further restricted to contain a given point $o \notin R \cup B$ of the plane in the interior. We assume general position, that is, there are no two points of $R \cup B \cup \{o\}$ with the same x - or y -coordinate. We further assume w.l.o.g. that o is the origin of coordinates.

Given a fixed $X \subseteq R \cup B$, and $S \subseteq R \cup B$ chosen at random, let $E(X) = E(X, S)$ be the event that there exist two red points $p, q \in S \cap X$ such that the box $D(p, q)$ contains the origin o , no other red point in $S \cap X$, and no blue in $S \cap X$. Then, our goal is $\Pr[E(R \cup B)]$.

Theorem 8 *Given $R \cup B$, $\Pr[E(R \cup B)]$ can be computed in polynomial time.*

Proof Let $X \subseteq R \cup B$, and define $X^+ = \{p \in X \mid y(p) > 0\}$ and $X^- = \{p \in X \mid y(p) < 0\}$. Given points $q \in X^+$ and $r \in X^-$, define the events

$$U_q(X, S) = \left[q = \arg \min_{p \in X^+ \cap S} \{y(p)\} \right] \text{ and } W_r(X, S) = \left[r = \arg \max_{p \in X^- \cap S} \{y(p)\} \right].$$

Let $U_q(X) = U_q(X, S)$ and $W_r(X) = W_r(X, S)$. Using the formula of the total probability, we have:

$$\begin{aligned} \Pr[E(X)] &= \sum_{q \in X^+} \Pr[E(X) \mid U_q(X)] \cdot \Pr[U_q(X)] \\ &= \sum_{q \in X^+} \Pr[E(X) \mid U_q(X)] \cdot \left(\pi(q) \cdot \prod_{p \in X^+; y(p) < y(q)} (1 - \pi(p)) \right). \end{aligned}$$

To compute $\Pr[E(X) \mid U_q(X)]$, we assume $x(q) > 0$. The case where $x(q) < 0$ is symmetric. If $q \in B$, then observe that when restricted to the event $U_q(X)$ any box $D(p', q')$ defined by two red points $p', q' \in S \cap X$, containing the origin o and no other red point in $S \cap X$, where one between p' and q' is to the right of q , will contain q . Hence, we must “discard” all points to the right of q , all points between the horizontal lines through q and o because they are not present, and q itself. Then, we have that:

$$\Pr[E(X) \mid U_q(X)] = \Pr[E(X_q)],$$

where $X_q \subset X$ contains the points $p \in X$ such that $x(p) < x(q)$ and either $y(p) > y(q)$ or $y(p) < 0$. If $q \in R$, we expand $\Pr[E(X) \mid U_q(X)]$ as follows:

$$\Pr[E(X) \mid U_q(X)] = \sum_{r \in X^-} \Pr[E(X) \mid U_q(X), W_r(X)] \cdot \Pr[W_r(X)]$$

$$= \sum_{r \in X^-} \Pr[E(X) \mid U_q(X), W_r(X)] \cdot \left(\pi(r) \cdot \prod_{p \in X^-: y(p) > y(r)} (1 - \pi(p)) \right).$$

There are now three cases according to the relative positions of q and r .

Case 1: $x(r) < 0 < x(q)$. Let $Y_{q,r} \subset X$ contain the points $p \in X$ (including q) such that $x(r) < x(p) \leq x(q)$ and either $y(p) < y(r)$ or $y(p) \geq y(q)$. If $r \in R$, then $\Pr[E(X) \mid U_q(X), W_r(X)] = 1$. Otherwise, if $r \in B$, given that $U_q(X)$ and $W_r(X)$ hold, any box $D(p', q')$ defined by two red points p', q' of $S \cap X$, containing the origin o and no other red point in $S \cap X$, where one between p' and q' is not in $Y_{q,r}$, will contain q or r in the interior. Then

$$\Pr[E(X) \mid U_q(X), W_r(X)] = \Pr[E(Y_{q,r}) \mid U_q(Y_{q,r})].$$

Similar arguments are given in the next two cases.

Case 2: $0 < x(q) < x(r)$. We have

$$\Pr[E(X) \mid U_q(X), W_r(X)] = \Pr[E(X_q \cup \{q\}) \mid U_q(X_q \cup \{q\})].$$

Case 3: $0 < x(r) < x(q)$. If $r \in R$, then

$$\Pr[E(X) \mid U_q(X), W_r(X)] = \Pr[E(Z_{q,r} \cup \{r\}) \mid W_r(Z_{q,r} \cup \{r\})],$$

where $Z_{q,r} \subset X$ contains the points $p \in X$ such that $x(p) < x(r)$ and either $y(p) < y(r)$ or $y(p) > y(q)$. Note that the event $[E(Z_{q,r} \cup \{r\}) \mid W_r(Z_{q,r} \cup \{r\})]$ is symmetric to the event $[E(X) \mid U_q(X)]$, thus its probability can be computed similarly. Otherwise, if $r \in B$, we have

$$\Pr[E(X) \mid U_q(X), W_r(X)] = \Pr[E(Z_{q,r})].$$

Note that in the above recursive computation of $\Pr[E(X)]$, for $X = R \cup B$, there is a polynomial number of subsets X_q , $Y_{q,r}$, and $Z_{q,r}$; each of such subsets can be encoded in constant space (i.e., by using a constant number of coordinates). Then, we can use dynamic programming, with a polynomial-size table, to compute $\Pr[E(R \cap B)]$ in time polynomial in n . \square

4 Discussion and Open Problems

For fixed $d \geq 1$, the maximum box problem for non-probabilistic points can be solved in $O(n^d)$ time [3]. This fact, combined with the Monte Carlo method and known techniques, can be used to approximate the expectation of the total weight of the maximum box on probabilistic points, in polynomial time and with high probability of success. That is, we provide a FPRAS, which is explained in Appendix C. Approximating the probability that the total weight of a maximum box is at least a given parameter is an open question of this paper. To this end, we give in Appendix D a FPRAS for

approximating this probability, but only in the case where the points are colored red or blue, each with probability $1/2$, and we look for the box covering the maximum number of red points and no blue point (i.e. red points have weight $+1$ and blue points weight $-\infty$).

For $d = 2$ and red and blue points, there are several open problems: For example, to compute $\Pr[\text{box}(S) \geq k]$ (even for $k = 3$) in $d = 2$ when the box is restricted to contain a fixed point. Other open questions are the cases in which the box is restricted to contain a given point as vertex, or has some side contained in a given axis-parallel line. This two latter variants can be solved in $n^{O(k)}$ time (see Appendix E), which means that they are polynomial-time solvable for fixed k . This contrasts with the original question that is #P-hard for every $k \geq 2$.

For red and blue points in $d = 1$, both $\Pr[\text{box}(S) \geq k]$ and $\mathbb{E}[\text{box}(S)]$ can be computed in polynomial time by using standard dynamic programming techniques. This implies that for $d = 2$ and a fixed direction, computing the probability that there exists a strip (i.e., the space between two parallel lines) perpendicular to that direction and covering at least k red points and no blue point can be done in polynomial time. If the orientation of the strip is not restricted, then such a computation is #P-hard for every $k \geq 3$ (see Appendix F).

Acknowledgements L.E.C. and I.V. are supported by MTM2016-76272-R (AEI/FEDER, UE). L.E.C. is also supported by Spanish Government under grant agreement FPU14/04705. P.P.L. was partially supported by projects CONICYT FONDECYT/Regular 1160543 (Chile), DICYT 041933PL Vicerrectoría de Investigación, Desarrollo e Innovación USACH (Chile), and Programa Regional STICAMSUD 19-STIC-02. C.S. is supported by PID2019-104129GB-I00/ MCIN/ AEI/ 10.13039/501100011033 and Gen. Cat. DGR 2017SGR1640.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Fibonacci Numbers

Lemma 9 *Let $\{f_n\}_{n=1}^\infty$ be the Fibonacci sequence, with $f_1 = f_2 = 1$ and $f_n = f_{n-1} + f_{n-2}$ for $n \geq 3$. Then, the numbers f_i/f_{i+1} for $i \geq 1$ are all different.*

Proof Let $1 \leq i < j$ be integers such that $f_i/f_{i+1} = f_j/f_{j+1}$. Assume also that i is minimum over all possible pairs (i, j) satisfying this property. If $i = 1$, then $f_j = f_{j+1}$ because $f_1 = f_2 = 1$. Since f_2, f_3, f_4, \dots are all different, this is a contradiction. Otherwise, if $i > 1$,

$$\begin{aligned} f_i \cdot f_{j+1} &= f_j \cdot f_{i+1} \\ f_i \cdot (f_j + f_{j-1}) &= f_j \cdot (f_i + f_{i-1}) \end{aligned}$$

$$f_{i-1}/f_i = f_{j-1}/f_j.$$

Then, the pair $(i - 1, j - 1)$ satisfies the property, which is a contradiction because (i, j) is such that i is minimum. Hence, the lemma follows. \square

B Proof of Lemma 6

Lemma 6. *For any different $p, q \in R$, the vertices $\lambda^{-1}(p)$ and $\lambda^{-1}(q)$ are adjacent in G_s if and only if the box $D(p, q)$ contains no point of B .*

Proof (\Rightarrow) Let $p, q \in R$ be red points such that vertices $p_0 = \lambda^{-1}(p)$ and $q_0 = \lambda^{-1}(q)$ are adjacent in G_s . We have either $x(p_0) = x(q_0)$ or $y(p_0) = y(q_0)$. We will prove that $D(p, q) \cap B$ is empty by assuming $x(p_0) = x(q_0) = a$; the case where $y(p_0) = y(q_0)$ is similar. Further assume w.l.o.g. that $y(p_0) < y(q_0)$. Since the segment p_0q_0 contains no other point of $R_0 \cup B_0$, then $D(p, q)$ does not contain points in $\lambda(R_0) \cup \lambda(B_0)$ different from p and q (refer to Lemma 7 in [4]). Then, we need to prove that $D(p, q)$ does not contain any blue point of the form $\lambda(u_0) + (1/2, 1/2)$ or $\lambda(u_0) + (\delta, -\delta)$, for $u_0 \in R_0$. Assume that there exists $u_0 \in R_0$ such that $u = \lambda(u_0) + (1/2, 1/2) \in D(p, q) \cap B$. We must have $x(p) \leq x(u) \leq x(q)$, that is

$$a + \frac{a + y(p_0)}{4N + 1} \leq x(u_0) + \frac{x(u_0) + y(u_0)}{4N + 1} + \frac{1}{2} \leq a + \frac{a + y(q_0)}{4N + 1}. \tag{1}$$

The left inequality implies

$$a < x(u_0) + \frac{2N}{4N + 1} + \frac{1}{2} < x(u_0) + 1,$$

that is, $a - 1 < x(u_0)$. The right inequality implies

$$x(u_0) + \frac{1}{2} < a + \frac{2N}{4N + 1} < a + \frac{1}{2},$$

that is, $x(u_0) < a$. Since both a and $x(u_0)$ are integers, $a - 1 < x(u_0) < a$ is a contradiction. Hence, such a point u_0 does not exist. Assume now that there exists $u_0 \in R_0$ such that $u = \lambda(u_0) + (\delta, -\delta) \in D(p, q) \cap B$. Then, Eq. (1) translates to

$$a + \frac{a + y(p_0)}{4N + 1} \leq x(u_0) + \frac{x(u_0) + y(u_0)}{4N + 1} + \delta \leq a + \frac{a + y(q_0)}{4N + 1}. \tag{2}$$

The left inequality again implies $a - 1 < x(u_0)$, and the right one implies $x(u_0) < a + 1/2 - \delta < a + 1$. Then, we must have $x(u_0) = a$, and Eq. (2) simplifies to

$$\frac{y(p_0)}{4N + 1} \leq \frac{y(u_0)}{4N + 1} + \frac{1}{4N + 2} \leq \frac{y(q_0)}{4N + 1}.$$

This implies $y(u_0) < y(q_0)$, and then $y(p_0) = y(u_0)$ (i.e., $u_0 = p_0$) because the segment p_0q_0 is empty of points of $R_0 \cup B_0$ in its interior. Then, we have $y(u) = y(p) - \delta < y(p) < y(q)$ which contradicts $u \in D(p, q)$. Hence, such a point u_0 does not exist.

(\Leftarrow) Let $p, q \in R$ be red points such that vertices $p_0 = \lambda^{-1}(p)$ and $q_0 = \lambda^{-1}(q)$ are not adjacent in G_s . We will prove that $D(p, q) \cap B$ is not empty. We have several cases:

- (a) $x(p_0) = x(q_0)$ or $y(p_0) = y(q_0)$, and segment p_0q_0 contains a point $u_0 \in R_0 \cup B_0$. Consider $x(p_0) = x(q_0)$, if $y(p_0) = y(q_0)$ then the proof is similar. Assume $y(p_0) < y(q_0)$ w.l.o.g., let $u = \lambda(u_0)$, and note that $x(p_0) = x(u_0) = x(q_0)$ and $y(p_0) < y(u_0) < y(q_0)$. If $u_0 \in B_0$, then

$$x(p_0) + \frac{x(p_0) + y(p_0)}{4N + 1} < x(u_0) + \frac{x(u_0) + y(u_0)}{4N + 1} < x(q_0) + \frac{x(q_0) + y(q_0)}{4N + 1},$$

and

$$y(p_0) + \frac{x(p_0) + y(p_0)}{4N + 1} < y(u_0) + \frac{x(u_0) + y(u_0)}{4N + 1} < y(q_0) + \frac{x(q_0) + y(q_0)}{4N + 1}.$$

which imply $x(p) < x(u) < x(q)$ and $y(p) < y(u) < y(q)$. Hence, $u \in D(p, q) \cap B$. Otherwise, if $u_0 \in R_0$, then let $v = u + (\delta, -\delta) \in B$. Similarly as above, we have $x(p) < x(u) = x(v) - \delta < x(v)$ and $y(v) = y(u) - \delta < y(u) < y(q)$. Then,

$$\begin{aligned} x(v) &= x(u_0) + \frac{x(u_0) + y(u_0)}{4N + 1} + \delta = x(q_0) + \frac{x(q_0) + y(u_0)}{4N + 1} + \frac{1}{4N + 2} \\ &< x(q_0) + \frac{x(q_0) + y(q_0) - \frac{4N+1}{4N+2}}{4N + 1} + \frac{1}{4N + 2} \\ &= x(q_0) + \frac{x(q_0) + y(q_0)}{4N + 1} = x(q), \end{aligned}$$

and

$$\begin{aligned} y(p) &= y(p_0) + \frac{x(p_0) + y(p_0)}{4N + 1} = y(p_0) + \frac{x(u_0) + y(p_0)}{4N + 1} \\ &< y(u_0) + \frac{x(u_0) + y(u_0) - \frac{4N+1}{4N+2}}{4N + 1} \\ &= y(u_0) + \frac{x(u_0) + y(u_0)}{4N + 1} - \frac{1}{4N + 2} = y(u) - \delta = y(v), \end{aligned}$$

which imply $x(p) < x(v) < x(q)$; $y(p) < y(v) < y(q)$; and $v \in D(p, q) \cap B$.

(b) (Up to symmetry) $x(p_0) < x(q_0)$ and $y(p_0) < y(q_0)$. Let $u = p + (1/2, 1/2) \in B$. Then,

$$\begin{aligned} x(u) &= x(p) + \frac{1}{2} = x(p_0) + \frac{x(p_0) + y(p_0)}{4N + 1} \\ &\quad + \frac{1}{2} < x(q_0) + \frac{x(q_0) + y(q_0)}{4N + 1} = x(q), \end{aligned}$$

and

$$\begin{aligned} y(u) &= y(p) + \frac{1}{2} = y(p_0) + \frac{x(p_0) + y(p_0)}{4N + 1} \\ &\quad + \frac{1}{2} < y(q_0) + \frac{x(q_0) + y(q_0)}{4N + 1} = y(q). \end{aligned}$$

Hence, $x(p) < x(u) < x(q)$; $y(p) < y(u) < y(q)$; and $u \in D(p, q) \cap B$.

(c) (Up to symmetry) $x(p_0) < x(q_0)$ and $y(p_0) > y(q_0)$. Let $v = p + (\delta, -\delta) \in B$. Then,

$$\begin{aligned} x(v) &= x(p) + \delta = x(p_0) + \frac{x(p_0) + y(p_0)}{4N + 1} + \delta \\ &< x(p_0) + \frac{2N}{4N + 1} + \frac{1}{4N + 2} < x(q_0) < x(q), \end{aligned}$$

and

$$\begin{aligned} y(v) &= y(p) - \delta = y(p_0) + \frac{x(p_0) + y(p_0)}{4N + 1} - \delta \\ &> y(p_0) - \frac{1}{4N + 2} \geq y(q_0) + 1 - \frac{1}{4N + 2} > y(q_0) + \frac{2N}{4N + 1} \\ &> y(q_0) + \frac{x(q_0) + y(q_0)}{4N + 1} = y(q). \end{aligned}$$

Hence, $x(p) < x(v) < x(q)$; $y(p) > y(v) > y(q)$; and $v \in D(p, q) \cap B$.

The result thus follows. \square

C Random Approximation of the Expectation

In this section, let $P \subset \mathbb{R}^d$ be an n -point set, for fixed $d \geq 1$, with weights $w : P \rightarrow \mathbb{R} \setminus \{0\}$, and probabilities $\pi : P \rightarrow (0, 1]$. Let $S \subseteq P$ be the random sample where each $p \in P$ is included in S with probability $\pi(p)$. We assume the model in which deciding whether a given p is in the sample S can be done in $O(1)$ time, then any random sample of $J \subseteq P$ can be generated in $O(|J|)$ time.

Let $\text{box}(S)$ denote the total weight of the maximum box of S . We show, given $\varepsilon, \delta \in (0, 1)$, how to compute in time polynomial in n, ε^{-1} , and δ^{-1} , the estimate $\tilde{\mu}$

of $\mu = \mathbb{E}[\text{box}(S)]$ such that

$$\Pr[(1 - \varepsilon)\mu \leq \tilde{\mu} \leq (1 + \varepsilon)\mu] > 1 - \delta. \tag{3}$$

Let $\mathcal{Q} = \{p \in P \mid w(p) > 0\}$, $m = |\mathcal{Q}|$, and q_1, q_2, \dots, q_m denote the elements of \mathcal{Q} such that $w(q_1) \leq w(q_2) \leq \dots \leq w(q_m)$. For $j = 1, \dots, m$, let $w_j = w(q_j)$, $\mathcal{Q}_j = \{q_1, \dots, q_j\}$, and μ_j be defined as follows:

$$\mu_j = \mathbb{E}[\text{box}(S) \mid S \subseteq \mathcal{Q}_j \cup (P \setminus \mathcal{Q}), q_j \in S],$$

that is, μ_j is the expectation of $\text{box}(S)$ conditioned on q_j being the element of heaviest weight in S . By the formula of the total probability, we have that:

$$\begin{aligned} \mu &= \sum_{j=1}^m \Pr[S \subseteq \mathcal{Q}_j \cup (P \setminus \mathcal{Q}), q_j \in S] \cdot \mu_j \\ &= \sum_{j=1}^m \left(\pi(q_j) \prod_{i=j+1}^m (1 - \pi(q_i)) \right) \mu_j. \end{aligned}$$

Then, to approximate μ we compute the approximation $\tilde{\mu}_j$ of μ_j , for $j = 1, \dots, m$. We do this by using the Monte-Carlo method, within the proof of the next lemma:

Lemma 10 *Let $\varepsilon, \delta \in (0, 1)$, $j \in [1..m]$, and $N = \lceil (4j/\varepsilon^2) \ln(2/\delta) \rceil$. Let S_1, S_2, \dots, S_N be N random samples of $\mathcal{Q}_j \cup (P \setminus \mathcal{Q})$, each containing q_j , and let $\tilde{\mu}_j$ be the average of $\text{box}(S_1), \text{box}(S_2), \dots, \text{box}(S_N)$. Then, we have that*

$$\Pr[(1 - \varepsilon)\mu_j \leq \tilde{\mu}_j \leq (1 + \varepsilon)\mu_j] > 1 - \delta. \tag{4}$$

Furthermore, $\tilde{\mu}_j$ can be computed in $N \cdot (O(j + n - m) + O(n^d)) = O((n^{d+1}/\varepsilon^2) \log(1/\delta))$ time.

Proof A standard Chernoff bound [14] asserts that if X_1, \dots, X_N are i.i.d. random variables over a bounded domain $[0, U]$ with expectation $\alpha = \mathbb{E}[X_i]$, then the average $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ satisfies:

$$\Pr[(1 - \varepsilon)\alpha \leq \bar{X} \leq (1 + \varepsilon)\alpha] > 1 - 2e^{-N(\alpha/U)\varepsilon^2/4}. \tag{5}$$

By letting $X_i = \text{box}(S_i)$ for $i = 1, \dots, N$, we have that $\alpha = \mu_j$, $\bar{X} = \tilde{\mu}_j$, and $U = w_1 + w_2 + \dots + w_j \leq j \cdot w_j$. Furthermore, since $q_j \in S_i$ for all i , we must have $w_j \leq X_i \leq U$ (because there is a small enough box containing only point q_j , and X_i is the weight of a maximum box of S_i), which implies $w_j \leq \mu_j \leq U$. For $N \geq (4/\varepsilon^2)(U/\alpha) \ln(2/\delta) = (4/\varepsilon^2)(U/\mu_j) \ln(2/\delta)$, Eqs. (5) and (4) hold. This is ensured by the definition of N , and because $w_j \leq \mu_j$ and $U \leq j \cdot w_j$. That is,

$$(4/\varepsilon^2)(U/\mu_j) \ln(2/\delta) \leq (4/\varepsilon^2)(U/w_j) \ln(2/\delta) \leq (4j/\varepsilon^2) \ln(2/\delta) \leq N.$$

Since for any $d \geq 1$, the maximum box problem on n points in \mathbb{R}^d can be solved in $O(n^d)$ time [3] (after sorting the points), the running time follows. \square

We use Lemma 10 and the union bound to compute $\tilde{\mu}$ to satisfy Eq. (3). Namely, within $O(m \cdot (n^{d+1}/\varepsilon^2) \log(m/\delta)) = O((n^{d+2}/\varepsilon^2) \log(n/\delta))$ time we compute the estimates $\tilde{\mu}_1, \tilde{\mu}_2, \dots, \tilde{\mu}_m$, such that for all $j \in [1..m]$ we have Eq. (4) in the following way:

$$\Pr[(1 - \varepsilon)\mu_j \leq \tilde{\mu}_j \leq (1 + \varepsilon)\mu_j] > 1 - \delta/m.$$

Let

$$\tilde{\mu} = \sum_{j=1}^m \Pr[S \subseteq Q_j \cup (P \setminus Q), q_j \in S] \cdot \tilde{\mu}_j,$$

which by the union bound and Bernoulli's inequality satisfies $(1 - \varepsilon)\mu \leq \tilde{\mu} \leq (1 + \varepsilon)\mu$ with probability at least $(1 - \delta/m)^m \geq 1 - \delta$. Hence, $\tilde{\mu}$ satisfies Eq. (3).

D Random Approximation of the Probability

In this section, let $P \subset \mathbb{R}^d$, $d \geq 1$, be an n -point set of two-colored points. Let $P = R \cup B$, where R is the set of red points and B the set of blue points. We assume that P is in general position, which means that no two points of P belongs to the same axis-parallel hyperplane. Let $S \subseteq P$ be a random sample in which each element of P is included in S independently with probability $1/2$, and let $\text{box}(S)$ denote the maximum number of red points in S that can be covered with a box, without covering any blue point.

Given an integer $z \geq 1$, we show how to approximate the probability $f = \Pr[\text{box}(S) \geq z]$. Namely, we show how to compute in $O(n^{O(\min\{z, n-z\})} \cdot \varepsilon^{-2})$ time a value \tilde{f} that satisfies

$$\Pr[(1 - \varepsilon)f \leq \tilde{f} \leq (1 + \varepsilon)f] \geq 3/4. \quad (6)$$

Note that the running time is polynomial if $\min\{z, n - z\} = O(1)$.

The idea of the algorithm is to reduce the computation of the probability to count the number of satisfiable assignments of some DNF formula. Given n boolean variables x_1, x_2, \dots, x_n , a DNF formula on these variables is a disjunction of clauses, where each clause is a conjunction of variables or negations of variables (e.g. $(x_1 \wedge \overline{x_2} \wedge x_3) \vee (x_2 \wedge x_4) \vee (\overline{x_3} \wedge x_4)$). The key observation for the reduction is that there exists a box covering at least z red points of S , without covering any blue points, if and only if there exists a similar box covering exactly z points. This follows from the fact that a box with more than z red points can be shrunk to cover exactly z of them.

For every $p \in P$, let $x_p \in \{0, 1\}$ be the boolean, or indicator, (random) variable such that $x_p = 1$ if and only if $p \in S$. For every set $Q \subseteq R$ of cardinality z , let C_Q

denote the clause

$$C_Q = \left(\bigwedge_{p \in Q} x_p \right) \wedge \left(\bigwedge_{q \in B \cap bb(Q)} \bar{x}_q \right),$$

where $bb(Q)$ denotes the minimum box covering Q . That is, C_Q stands for the event that $Q \subseteq S$ and the elements of Q , all of them red, can be covered with a box without covering any blue point of S (because $bb(Q)$ is forced to be empty of elements of B). Let $F = \bigvee_{Q \in 2^R: |Q|=z} C_Q$ be the DNF formula consisting of the disjunction of C_Q over all subsets $Q \subseteq R$ with $|Q| = z$, which has n variables and $m = \binom{n}{z}$ clauses.

Let N be the number of satisfying assignments to F . Then, note that $f = N/2^n$. Using the algorithm of Karp, Luby, and Madras [12], we can find in $O(nm \cdot \varepsilon^{-2}) = O(n^{O(\min\{z, n-z\})} \cdot \varepsilon^{-2})$ time a value \tilde{N} such that $\Pr[(1 - \varepsilon)N \leq \tilde{N} \leq (1 + \varepsilon)N] \geq 3/4$. This implies that \tilde{f} satisfies Eq. (6).

E Anchored Boxes

In this section, we consider the computation of $\Pr[\text{box}(S) \geq k]$ when the box is restricted to contain the origin o of coordinates as bottom-left vertex. We assume that $R \cup B$ is in general position and k is an integer. For simplicity, we further assume that $R \cup B$ is contained in the first quadrant, $\pi(b) = 1$ for every $b \in B$, the elements $b_1, b_2, \dots, b_{|B|}$ of B form a staircase (i.e., $x(b_1) < x(b_2) < \dots < x(b_{|B|})$ and $y(b_1) > y(b_2) > \dots > y(b_{|B|})$), and $x(b_1) < x(r) < x(b_{|B|})$ for every $r \in R$. We show that $\Pr[\text{box}(S) \geq k]$ can be computed in $n^{O(k)}$ time. If all the simplifications are dropped, or it is considered that the box has a side in a given axis-parallel line, the computation (although more detailed) can also be done within this time.

For every $i \in \{1, 2, \dots, |B| - 1\}$, let $D_i = D(o, b_i)$, and let H_i be the smallest box covering o, b_i , and b_{i+1} . Let $D_{|B|} = D(o, b_{|B|})$. Let $E(i, Q)$ denote the event such that for $S \subseteq R$ chosen at random, at least one of $H_i, H_{i+1}, \dots, H_{|B|-1}$ contains k red points (or more) of S subject to $|D_i \cap S| = |Q|$. Then, $[\text{box}(S) \geq k] = E(1, \emptyset)$. We compute $\Pr[E(i, Q)]$ recursively (see Fig. 5). We have that:

$$\begin{aligned} & \Pr[E(i, Q)] \\ &= \Pr\left[E(i, Q) \mid |(H_i \setminus D_i) \cap S| \geq k - |Q|\right] \cdot \Pr\left[|(H_i \setminus D_i) \cap S| \geq k - |Q|\right] \\ & \quad + \Pr\left[E(i, Q) \mid |(H_i \setminus D_i) \cap S| < k - |Q|\right] \cdot \Pr\left[|(H_i \setminus D_i) \cap S| < k - |Q|\right] \\ &= 1 - \Pr\left[|(H_i \setminus D_i) \cap S| < k - |Q|\right] \\ & \quad \left(1 - \Pr\left[E(i, Q) \mid |(H_i \setminus D_i) \cap S| < k - |Q|\right]\right). \end{aligned}$$

Note that $\Pr[|(H_i \setminus D_i) \cap S| < k - |Q|]$ can be computed in $n^{O(k)}$ time since we need to consider all subsets of size less than k of $(H_i \setminus D_i) \cap R$. Also note that $\Pr[E(i, Q) \mid |(H_i \setminus D_i) \cap S| < k - |Q|] = 0$ if $i = |B| - 1$. Otherwise, if

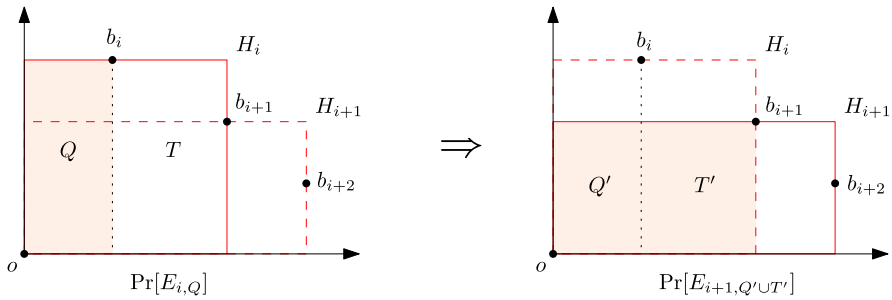


Fig. 5 Recursive computation of $\Pr[E(i, Q)]$

$i < |B| - 1$, then:

$$\begin{aligned} & \Pr\left[E(i, Q) \mid |(H_i \setminus D_i) \cap S| < k - |Q|\right] \\ &= \sum_{\substack{T \subseteq (H_i \setminus D_i) \cap R \\ |T| < k - |Q|}} \Pr\left[E(i, Q) \mid (H_i \setminus D_i) \cap S = T\right] \\ &= \sum_{\substack{T \subseteq (H_i \setminus D_i) \cap R \\ |T| < k - |Q|}} \Pr\left[E(i + 1, Q' \cup T')\right], \end{aligned}$$

where $Q' = Q \cap D_{i+1}$ and $T' = T \cap D_{i+1}$. The sum has $n^{O(k)}$ terms, and using dynamic programming on a table of size $n^{O(k)}$, we can compute $\Pr[\text{box}(S) \geq k] = \Pr[E(1, \emptyset)]$ in $n^{O(k)}$ time and space.

F Covering with a Strip

Let R and B be disjoint finite point sets in the plane with a total of n points, where the elements of R are colored red and the elements of B are colored blue. Let $S \subseteq R \cup B$ be the random point set where every point $p \in R \cup B$ is included in S independently and uniformly at random with probability $\pi(p) \in [0, 1]$. Let $\text{strip}(S)$ denote the random variable equal to the maximum number of red points in S that can be covered with a strip without covering any blue point.

Theorem 11 *Given $R \cup B$, it is #P-hard to compute the probability $\Pr[\text{strip}(S) \geq k]$ for every integer $k \geq 3$, and it is also #P-hard to compute $\mathbb{E}[\text{strip}(S)]$.*

Proof Let $k = 3$, and the bipartite graph $G = (V_1 \cup V_2, E)$ be an instance of #IndSet. Assume w.l.o.g. $V_1 = \{-1, -2, \dots, -N\}$, $V_2 = \{1, 2, \dots, N\}$, and every edge of E connects a vertex of V_1 and a vertex of V_2 . For every $i \in V_1$ and $j \in V_2$, define the following points on the curve $y = x^3$:

$$p_i = \left(i, i^3\right), q_j = \left(\frac{j}{N}, \left(\frac{j}{N}\right)^3\right), \text{ and } s_{i,j} = \left(-i - \frac{j}{N}, \left(-i - \frac{j}{N}\right)^3\right).$$

Note that $p_i, q_j,$ and $s_{i,j}$ are collinear for every i, j . Furthermore, for $i' \in V_1$ and $j' \in V_2$ such that $i' \neq i$ or $j' \neq j$ we have $s_{i,j} \neq s_{i',j'}$. Consider the next sets of red points:

$$R_1 = \{p_1, p_2, \dots, p_N\}, R_2 = \{q_1, q_2, \dots, q_N\}, \text{ and } R_{1,2} = \{s_{i,j} \mid \{i, j\} \in E\},$$

in which the only triplets of points that are collinear are those of the form $(p_i, q_j, s_{i,j})$. There exist rational numbers $\delta > 0$ such that the following set

$$B = \{(x, y + \delta), (x, y - \delta) \mid (x, y) \in R_1 \cup R_2 \cup R_{1,2}\}$$

of blue points ensures that every triangle with vertices in $R_1 \cup R_2 \cup R_{1,2}$ and positive area contains in the interior a vertex from B [7]. Note that we are excluding the degenerate triangles (those with zero area) with vertices at $p_i, q_j,$ and $s_{i,j}$ for some i, j . Such a value of δ can be computed in polynomial time. For $\varepsilon > 0,$ let us define

$$s'_{i,j} = s_{i,j} + (\varepsilon, 0) \text{ and } R'_{1,2} = \{s'_{i,j} \mid \{i, j\} \in E\}.$$

In polynomial time, we can also compute a rational value for ε such that the set $R = R_1 \cup R_2 \cup R'_{1,2}$ of red points ensures that a triangle with vertices at elements of R contains in the interior a point of B if and only if the triangle does not have vertices at $p_i, q_j,$ and $s'_{i,j}$ for some i, j . Furthermore, for every $i \in V_1$ and $j \in V_2$ such that $\{i, j\} \in E,$ there exists a (very thin) strip containing $p_i, q_j,$ and $s'_{i,j}$ and no other point of $R \cup B$. These conditions imply that three red points of R can be covered with a strip that does not cover any blue point if and only if they are $p_i, q_j,$ and $s'_{i,j}$ for some i, j . For every $u \in R_1 \cup R_2$ assign $\pi(u) = 1/2,$ and for every $v \in R'_{1,2} \cup B$ assign $\pi(v) = 1.$ When taking $S \subseteq R \cup B$ at random, we have $\text{strip}(S) \geq 3$ if and only if $V(S \cap (R_1 \cup R_2))$ is not an independent set in $G,$ where for $U \subseteq R_1 \cup R_2$ we denote $V(U) = \{x(u) \mid u \in U\} \subseteq V_1 \cup V_2.$ Then,

$$\begin{aligned} \Pr[\text{strip}(S) \geq 3] &= \Pr[\text{strip}(S) = 3] \\ &= \Pr[V(S \cap (R_1 \cup R_2)) \text{ is not an independent set in } G] \\ &= 1 - \Pr[V(S \cap (R_1 \cup R_2)) \text{ is an independent set in } G] \\ &= 1 - \frac{N(G)}{2^{|R_1 \cup R_2|}} = 1 - \frac{N(G)}{2^{|V|}} \\ N(G) &= 2^{|V|} \cdot (1 - \Pr[\text{strip}(S) \geq 3]). \end{aligned}$$

Hence, computing $\Pr[\text{strip}(S) \geq 3]$ is #P-hard via a reduction from #IndSet. To show that computing $\mathbb{E}[\text{strip}(S)]$ is also #P-hard, note that $\text{strip}(S) \in \{2, 3\}$ and

$$\begin{aligned} \mathbb{E}[\text{strip}(S)] &= 2 \cdot \Pr[V(S \cap (R_1 \cup R_2)) \text{ is an independent set in } G] \\ &\quad + 3 \cdot \Pr[V(S \cap (R_1 \cup R_2)) \text{ is not an independent set in } G] \\ &= 2 \cdot \left(\frac{N(G)}{2^{|V|}}\right) + 3 \cdot \left(1 - \frac{N(G)}{2^{|V|}}\right) \end{aligned}$$


$$N(G) = 2^{|V|} \cdot (3 - \mathbb{E}[\text{strip}(S)]).$$

When $k \geq 4$, as in the proof of Theorem 7, for each red point $s \in R'_{1,2}$ we can add $k - 3$ new red points with probability 1 close enough to s . The proof then follows similar arguments. \square

References

- Alliez, P., Devillers, O., Snoeyink, J.: Removing degeneracies by perturbing the problem or perturbing the world. *Reliable Comput.* **6**(1), 61–79 (2000)
- Backer, J., Keil, J.M.: The mono- and bichromatic empty rectangle and square problems in all dimensions. In: *LATIN'10*, pp. 14–25 (2010)
- Barbay, J., Chan, T.M., Navarro, G., Pérez-Lantero, P.: Maximum-weight planar boxes in $O(n^2)$ time (and better). *Inf. Process. Lett.* **114**(8), 437–445 (2014)
- Caraballo, L.E., Ochoa, C., Pérez-Lantero, P., Rojas-Ledesma, J.: Matching colored points with rectangles. *J. Comb. Optim.* **33**(2), 403–421 (2017)
- Chan, T.M., Kamousi, P., Suri, S.: Stochastic minimum spanning trees in Euclidean spaces. In: *SoCG'11*, pp. 65–74 (2011)
- Cortés, C., Díaz-Báñez, J.M., Pérez-Lantero, P., Seara, C., Urrutia, J., Ventura, I.: Bichromatic separability with two boxes: a general approach. *J. Algorithms* **64**(2–3), 79–88 (2009)
- Czyzowicz, J., Kranakis, E., Urrutia, J.: Guarding the convex subsets of a point-set. In: *12th Canadian Conference on Computational Geometry*, pp. 47–50. Fredericton, New Brunswick (2000)
- Faliszewski, P., Hemaspaandra, L.: The complexity of power-index comparison. *Theoret. Comput. Sci.* **410**(1), 101–107 (2009)
- Feldman, D., Munteanu, A., Sohler, C.: Smallest enclosing ball for probabilistic data. In: *SoCG'14*, pp. 214–223 (2014)
- Fink, M., Hershberger, J., Kumar, N., Suri, S.: Hyperplane separability and convexity of probabilistic point sets. *J. Comput. Geom.* **8**(2), 32–57 (2017)
- Kamousi, P., Chan, T.M., Suri, S.: Closest pair and the post office problem for stochastic points. *Comput. Geom.* **47**(2), 214–223 (2014)
- Karp, R.M., Luby, M., Madras, N.: Monte-Carlo approximation algorithms for enumeration problems. *J. Algorithms* **10**(3), 429–448 (1989)
- Kleinberg, J., Rabani, Y., Tardos, É.: Allocating bandwidth for bursty connections. *SIAM J. Comput.* **30**(1), 191–217 (2000)
- Motwani, R., Raghavan, P.: *Randomized Algorithms*. Cambridge University Press, New York (1995)
- Pérez-Lantero, P.: Area and perimeter of the convex hull of stochastic points. *Comput. J.* **59**(8), 1144–1154 (2016)
- Suri, S., Verbeek, K., Yıldız, H.: On the most likely convex hull of uncertain points. In: *ESA'13*, pp. 791–802 (2013)
- Tsirogiannis, C., Staals, F., Pellissier, V.: Computing the expected value and variance of geometric measures. *J. Exp. Algorithmics* **23**(2), 24:1–24:32 (2018)
- Vadhan, S.P.: The complexity of counting in sparse, regular, and planar graphs. *SIAM J. Comput.* **31**(2), 398–427 (2001)
- Valiant, L.G.: Universality considerations in VLSI circuits. *IEEE Trans. Comput.* **100**(2), 135–140 (1981)
- Yıldız, H., Foschini, L., Hershberger, J., Suri, S.: The union of probabilistic boxes: maintaining the volume. In: *ESA*, pp. 591–602 (2011)

Authors and Affiliations

Luis E. Caraballo¹ · Pablo Pérez-Lantero² · Carlos Seara³  ·
Inmaculada Ventura¹

✉ Carlos Seara
carlos.seara@upc.edu

Luis E. Caraballo
lcaraballo@us.es

Pablo Pérez-Lantero
pablo.perez.l@usach.cl

Inmaculada Ventura
iventura@us.es

¹ Departamento de Matemática Aplicada II, Universidad de Sevilla, Seville, Spain

² Facultad de Ciencia, Departamento de Matemática y Ciencia de la Computación, Universidad de Santiago de Chile (USACH), Santiago, Chile

³ Departament de Matemàtiques, Universitat Politècnica de Catalunya, Barcelona, Spain