

New results on the coarseness of bicolored point sets

J. M. Díaz-Báñez* R. Fabila-Monroy† P. Pérez-Lantero‡
I. Ventura*

April 1, 2019

Abstract

Let S be a 2-colored (red and blue) set of n points in the plane. A subset I of S is an island if there exists a convex set C such that $I = C \cap S$. The *discrepancy* of an island is the absolute value of the number of red minus the number of blue points it contains. A *convex partition* of S is a partition of S into islands with pairwise disjoint convex hulls. The discrepancy of a convex partition is the discrepancy of its island of minimum discrepancy. The *coarseness* of S is the discrepancy of the convex partition of S with maximum discrepancy. This concept was recently defined by Bereg et al. [CGTA 2013]. In this paper we study the following problem: Given a set S of n points in general position in the plane, how to color each of them (red or blue) such that the resulting 2-colored point set has small coarseness? We prove that every n -point set S can be colored such that its coarseness is $O(n^{1/4}\sqrt{\log n})$. This bound is almost tight since there exist n -point sets such that every 2-coloring gives coarseness at least $\Omega(n^{1/4})$. Additionally, we show that there exists an approximation algorithm for computing the coarseness of a 2-colored point set, whose ratio is between $1/128$ and $1/64$, solving an open problem posted by Bereg et al. [CGTA 2013]. All our results consider k -separable islands of S , for some k , which are those resulting from intersecting S with at most k halfplanes.

1 Introduction

Let S be a finite set of n elements, and $\mathcal{Y} \subseteq 2^S$ be a family of subsets of S . The tuple (S, \mathcal{Y}) is called a *range space*. If the range space arises from point

*Departamento de Matemática Aplicada II, Universidad de Sevilla, Seville, Spain. {dbanez,iventura}@us.es. Partially supported by project FEDER MEC MTM2009-08652, and the ESF EUROCORES programme EuroGIGA -ComPoSe IP04-MICINN Project EUI-EURC-2011-4306.

†CINVESTAV, Instituto Politécnico Nacional, Mexico. Email: ruyfabila@math.cinvestav.edu.mx. Partially supported by grant 153984 (CONACyT, Mexico).

‡Escuela de Ingeniería Civil en Informática, Universidad de Valparaíso, Chile. Email: pablo.perez@uv.cl. Partially supported by grant CONICYT, FONDECYT/Iniciación 11110069 (Chile).

sets and geometric objects, then (S, \mathcal{Y}) is called a *geometric range space*. A *coloring* of S is a mapping $\mathcal{X} : S \rightarrow \{-1, +1\}$. We think of the elements of S mapped to -1 as being colored *blue* and the elements of S mapped to $+1$ as being colored *red*. Let R be the red elements of S and B its blue elements. For $Y \subseteq S$, let $\mathcal{X}(Y) := \sum_{y \in Y} \mathcal{X}(y)$. The *discrepancy* of Y is defined as $\text{disc}(Y) = |\mathcal{X}(Y)|$, that is, the absolute value of the number of red minus the number of blue points that Y contains. The *discrepancy* of the family \mathcal{Y} is defined as $\text{disc}(\mathcal{Y}) := \min_{\mathcal{X}} \max_{Y \in \mathcal{Y}} \text{disc}(Y)$.

Geometric discrepancy theory has applications in statistics, clustering, optimization, and computer graphics. See the textbooks [2, 5, 8, 10, 11] for problems and results in geometric discrepancy.

Assume from now on that S is a set of n points in general position in the plane. A subset I of S is called an *island* if there is a convex set C on the plane such that $I = C \cap S$ [3]. A *convex partition* of S is a partition of S into islands, with pairwise disjoint convex hulls.

Given a coloring of S , the *discrepancy* of a convex partition $\Pi = \{S_1, S_2, \dots, S_k\}$ of S , denoted by $\text{disc}(\Pi)$, is the minimum of $\text{disc}(S_i)$ for $i = 1, \dots, k$. The *coarseness* of S , denoted by $\mathcal{C}(S)$, is defined as the maximum of $\text{disc}(\Pi)$ over all the convex partitions Π of S . This concept of coarseness was just recently defined by Bereg et al. [4], as a parameter to measure how well blended a finite set R of red points and a finite set B of blue points are. The smaller $\mathcal{C}(R \cup B)$ the more blended R and B [4].

Suppose now that we have an n -point set S in the plane and want to color each of its elements (red or blue) such that the resulting 2-colored point set has high coarseness. The answer to this question is trivial: take a halving line, color the points to one side red, and color the other points blue; or even easier, color all points of the same color, say red.

Then we post the following question: *What is the smallest coarseness of S over all colorings of S ?*

In this paper we show that for every n -point set S in the plane there exists a coloring of S such that the coarseness of S is upper bounded by $O(n^{1/4} \sqrt{\log n})$ (Theorem 10). We also show that there exist point sets such that all colorings give coarseness at least $\Omega(n^{1/4})$ (Theorem 4). We prove the upper bound by showing that the discrepancy of a convex partition is closely related to the discrepancy of a certain class of islands of S , which we call *k-separable* islands.

Given a finite point set S in the plane and a coloring of S , computing the coarseness of S is believed to be NP-hard [4]. We also show for the first time that there exists a polynomial-time approximation algorithm. Its approximation ratio is between $1/128$ and $1/64$, depending on $\text{disc}(S) = ||R| - |B||$ (Theorem 14). Specifically, the approximate value of the coarseness that we provide is at least $\max\left\{\frac{\mathcal{C}(S)}{128}, \frac{\mathcal{C}(S)}{64} - \text{disc}(S)\right\}$ and at most $\mathcal{C}(S)$. With this result, we solve an open problem posted by Bereg et al. [4].

2 Visiting the discrepancy theory

In this section we recall some definitions and results from discrepancy theory.

The *primal shatter function* $\pi_{\mathcal{Y}}(m)$ of (S, \mathcal{Y}) is a function of m . It is defined as the maximum number of subsets into which a subset of S , of at most m elements, can be split (or “shattered”) by all the elements of \mathcal{Y} . Formally:

$$\pi_{\mathcal{Y}}(m) := \max_{A \subset S, |A| \leq m} |\{Y \cap A : Y \in \mathcal{Y}\}|$$

The *dual shatter function* $\pi_{\mathcal{Y}}^*(m)$ is obtained by exchanging the roles of the points in S with the sets in \mathcal{Y} . $\pi_{\mathcal{Y}}^*(m)$ is defined as the maximum number of equivalence classes on S defined by an m -element subfamily $\mathcal{Z} \subset \mathcal{Y}$, where two elements x and y of S are equivalent if they belong to the same sets of \mathcal{Z} .

The primal and dual shatter functions have been used to give tight and almost tight upper bounds on the discrepancy of range spaces, via the following theorems (see Chapter 5 of [10]).

Theorem 1 (Primal shatter function bound). *Let $d > 1$ and C be constants such that $\pi_{\mathcal{Y}}(m) \leq Cm^d$ for all $m \leq n$. Then $\text{disc}(\mathcal{Y})$ is upper bounded by $O(n^{1/2-1/2d})$.*

Theorem 2 (Dual shatter function bound). *Let $d > 1$ and C be constants such that $\pi_{\mathcal{Y}}^*(m) \leq Cm^d$ for all $m \leq |\mathcal{Y}|$. Then $\text{disc}(\mathcal{Y})$ is upper bounded by $O(n^{1/2-1/2d}\sqrt{\log n})$.*

For example, if \mathcal{H} is the family of halfplanes, then it is easy to see that $\pi_{\mathcal{H}}(m) = O(m^2)$. Thus the discrepancy of halfplanes is $O(n^{1/4})$. It is known that this bound is tight:

Lemma 3 ([1,6]) *For arbitrarily large values of n , there exist sets of n points in general position in the plane such that, given any coloring of S , a halfplane exists within which one color outnumbers the other by at least $Cn^{1/4}$, for some positive constant C .*

From Lemma 3 we prove the following Theorem:

Theorem 4 *For arbitrarily large values of n , there exist sets of n points in general position in the plane with coarseness at least $Cn^{1/4}$ for some positive constant C .*

Proof. Assume that S is a set of points as in Lemma 3, and consider any coloring of S . Thus, there exists a halfplane H such that $\text{disc}(S \cap H) \geq C'n^{1/4}$ for some positive constant C' . Suppose that the trivial convex partition $\{S\}$ has discrepancy at most $(C'/2)n^{1/4}$, as otherwise we are done with $C := C'/2$. Then we have that $\text{disc}(S \setminus H) \geq (C'/2)n^{1/4}$ and the convex partition of $\Pi := \{S \cap H, S \setminus H\}$ of S has discrepancy $\text{disc}(\Pi) \geq (C'/2)n^{1/4}$. Thus, the coarseness of S is at least $Cn^{1/4}$, with $C := C'/2$. \square

3 k -separable islands and convex partitions

An island I of S is k -separable if it can be separated from $S \setminus I$ with at most k halfplanes, that is, there exist halfplanes H_1, H_2, \dots, H_t ($1 \leq t \leq k$), such that $I = S \cap (H_1 \cap H_2 \cap \dots \cap H_t)$. We denote the family of all the k -separable islands of S with \mathcal{I}_k . For constant k , we upper bound the discrepancy of \mathcal{I}_k by using its dual shatter function. Namely, we show that $\pi_{\mathcal{I}_k}^*(m) = O(m^2)$. We should point out that Dobkin and Gunopulos [7] proved the same asymptotic upper bound, but our proof considers more details and explicitly gives the constant hidden in the big-O notation.

Lemma 5 *If k is a positive integer and S a set of n points in convex and general position in the plane then $\pi_{\mathcal{I}_k}^*(m) \leq 4km$.*

Proof. Assume that S is sorted clockwise around its convex hull. Note that any k -separable island must consist of at most k intervals of consecutive points of S in this order. Consider a family of m , k -separable islands. There are at most $2km$ points of S that are the endpoints of any such intervals. There are at most $2km$ regions into which the remaining points (which are not endpoints of any interval) can lie. Thus, in total there are at most $4km$ equivalence classes. \square

Lemma 6 *If k is a positive integer and S a set of n points in general position in the plane then $\pi_{\mathcal{I}_k}^*(m) \leq (k^2 + 4k)m^2$.*

Proof. Let \mathcal{F} be a family of m , k -separable islands on S . We first consider the points lying in the convex hull of some island I of \mathcal{F} . Note that the convex hull of I is a set of points in convex position. By Lemma 5 these points are in at most $4k(m-1)$ different equivalence classes (when considering the other $m-1$ islands in \mathcal{F}). Thus, in total there are at most $4km^2$ equivalence classes for points in the boundary of some island in \mathcal{F} . We now bound the number of equivalence classes for points not lying in the boundary of any island. Each such equivalence class is contained in a cell of the line arrangement defined by the following set of lines \mathcal{L} . For each island $I \in \mathcal{F}$, let L_I be the set of at most k lines that separate I from $S \setminus I$. Set $\mathcal{L} := \cup_{I \in \mathcal{F}} L_I$. The line arrangement defined by \mathcal{L} has at most $|\mathcal{L}|^2 = k^2m^2$ cells. The result thus follows. \square

Using the dual shatter function bound and Lemma 6 we obtain the following theorem:

Theorem 7 *Let k be a positive constant and S a set of n points in general position in the plane. The discrepancy of the family \mathcal{I}_k of the k -separable islands of S is upper bounded by $O(n^{1/4}\sqrt{\log n})$.*

Note that although k -separable islands have small discrepancy, this is not the case for islands in general. For example for any coloring of a set of n points in convex position in the plane there always exists an island with discrepancy at least $n/2$. It can be shown that in this case the primal and dual shatter function are equal to 2^m .

We now show that every convex partition must contain a 5-separable island. This follows immediately from:

Lemma 8 (Theorem 2 in [9]) *A collection of n compact, convex, and pairwise disjoint sets in the plane may be covered with n non-overlapping convex polygons with a total of not more than $6n - 9$ sides.*

Theorem 9 *Every convex partition Π of S has a 5-separable island.*

Proof. Let $\Pi := \{S_1, S_2, \dots, S_m\}$. Using Lemma 8, there exist non-overlapping convex polygons C_1, C_2, \dots, C_m with a total of no more than $6m - 9$ sides, such that for each $i = 1, \dots, m$ the convex hull of S_i is enclosed by C_i . Thus, one of these convex polygons has at most 5 sides and the enclosed island is thus a 5-separable island. \square

We arrive at our main result by combining Theorems 9 and 7.

Theorem 10 *For every set S of n points in general position in the plane there exists a coloring such that the coarseness of S is upper bounded by $O(n^{1/4}\sqrt{\log n})$.*

4 Approximation

Let $S = R \cup B$ be finite point set in the plane in general position, and let \mathcal{X} be a coloring of S , where R is the set of points colored red and B the set of points colored blue. Let $r := |R|$, $b := |B|$, and $D_k := \max_{I \in \mathcal{I}_k} \text{disc}(I)$. For every set $X \subseteq \mathbb{R}^2$, let \bar{X} denote the complement of X , that is, $\bar{X} = \mathbb{R}^2 \setminus X$.

In this section we show that the value of D_2 is a constant approximation for the coarseness of S . We start with some lemmas before arriving to the final result.

Lemma 11 *Let t be an integer. If there exists an island $I \in \mathcal{I}_1$ of S such that $\text{disc}(I) \geq t$, then there exists a convex partition Π of S such that*

$$\text{disc}(\Pi) \geq \max\{t/2, t - |r - b|\}.$$

Proof. We have that $\text{disc}(S \setminus I) \geq t - |r - b|$. Indeed,

$$\begin{aligned} \text{disc}(S \setminus I) &= \left| (r - |I \cap R|) - (b - |I \cap B|) \right| \\ &= \left| (|I \cap R| - |I \cap B|) - (r - b) \right| \\ &\geq \left| |I \cap R| - |I \cap B| \right| - |r - b| \\ &= \text{disc}(I) - |r - b| \\ &\geq t - |r - b|. \end{aligned}$$

If $t - |r - b| \geq t/2$ then for the convex partition $\Pi = \{I, S \setminus I\}$ we have that $\text{disc}(\Pi) \geq t - |r - b|$. Otherwise, if $t - |r - b| < t/2$, then $\text{disc}(S) = |r - b| > t/2$ which implies that $\text{disc}(\Pi) > t/2$ for the trivial convex partition $\Pi = \{S\}$. The result thus follows. \square

Lemma 12 *Let t be an integer. If there exists an island $I \in \mathcal{I}_2$ of S such that $\text{disc}(I) \geq t$, then there exists a convex partition Π of S such that*

$$\text{disc}(\Pi) \geq \max \{t/8, t/4 - |r - b|\}.$$

Proof. If $I \in \mathcal{I}_1$ then the result follows from Lemma 11. Thus consider that $I \in \mathcal{I}_2 \setminus \mathcal{I}_1$, and let H_1 and H_2 be two halfplanes such that $I = S \cap (H_1 \cap H_2)$. Let $I' := S \cap (\overline{H_1} \cap H_2)$, $I'' := S \cap (H_1 \cap \overline{H_2})$, and $I''' := S \cap (\overline{H_1} \cap \overline{H_2})$. Refer to Figure 1.

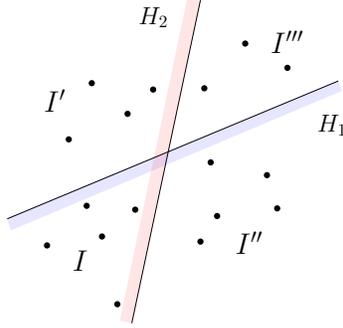


Figure 1: Proof of Lemma 12: The island I belongs to $\mathcal{I}_2 \setminus \mathcal{I}_1$, then it is the intersection of S with two halfplanes H_1 and H_2 .

If $\text{disc}(I') \leq t/2$ then the island $I \cup I' \in \mathcal{I}_1$ satisfies $\text{disc}(I \cup I') \geq t/2$ and by Lemma 11 there exists a convex partition Π_1 such that

$$\text{disc}(\Pi_1) \geq \max \{t/4, t/2 - |r - b|\}. \quad (1)$$

The same happens if $\text{disc}(I'') \leq t/2$. Otherwise, if $\text{disc}(I') > t/2$ and $\text{disc}(I'') > t/2$, then we proceed as follows. If $\text{disc}(I''') \geq t/4$ then the convex partition $\Pi_2 = \{I, I', I'', I'''\}$ satisfies

$$\text{disc}(\Pi_2) \geq t/4. \quad (2)$$

Otherwise, we have that the island $I' \cup I''' \in \mathcal{I}_1$ satisfies $\text{disc}(I' \cup I''') > t/4$ and then, by Lemma 11, there exists a convex partition Π_3 such that

$$\text{disc}(\Pi_3) \geq \max \{t/8, t/4 - |r - b|\}. \quad (3)$$

Combining equations (1-3) the result follows. \square

Lemma 13 $D_3 \leq 4D_2$, and $D_{k+1} \leq 2D_k$ for $k \geq 3$.

Proof. Let an island $I \in \mathcal{I}_3$ such that $D_3 = \text{disc}(I)$. If $I \in \mathcal{I}_2$ then $D_3 = D_2$ since $\mathcal{I}_2 \subseteq \mathcal{I}_3$. Otherwise, we have $I \in \mathcal{I}_3 \setminus \mathcal{I}_2$. Then, let H_1 , H_2 , and H_3 be three halfplanes such that $I = S \cap (H_1 \cap H_2 \cap H_3)$. Let $I' := S \cap (H_1 \cap H_2 \cap \overline{H_3})$

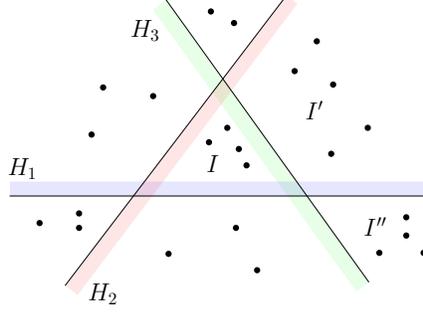


Figure 2: Proof of Lemma 13: The island I belongs to $\mathcal{I}_3 \setminus \mathcal{I}_2$, then it is the intersection of S with three halfplanes H_1 , H_2 , and H_3 .

and $I'' := S \cap (\overline{H_1} \cap \overline{H_3})$, and observe that $I \cup I'$, $I' \cup I''$, and I'' belong to \mathcal{I}_2 . Refer to Figure 2.

If $\text{disc}(I') \leq D_3/2$ then we have $D_3/2 \leq \text{disc}(I \cup I') \leq D_2$ which implies $D_3 \leq 2D_2$. Otherwise, if $\text{disc}(I') > D_3/2$, we proceed as follows. If $\text{disc}(I'') \geq D_3/4$ then we have $D_3 \leq 4D_2$. Otherwise, if $\text{disc}(I'') < D_3/4$, then $\text{disc}(I' \cup I'') > D_3/4$ implying $D_3 < 4D_2$. Then we have proved $D_3 \leq 4D_2$.

To prove the other part of the lemma, let $I \in \mathcal{I}_{k+1}$ such that $\text{disc}(I) = D_{k+1}$. If $I \in \mathcal{I}_k$ then $D_{k+1} = D_k$ since $\mathcal{I}_k \subseteq \mathcal{I}_{k+1}$. Otherwise, if $I \in \mathcal{I}_{k+1} \setminus \mathcal{I}_k$, let H_1, H_2, \dots, H_{k+1} be $k+1$ halfplanes such that $I = S \cap (H_1 \cap H_2 \cap \dots \cap H_{k+1})$. Refer to Figure 3.

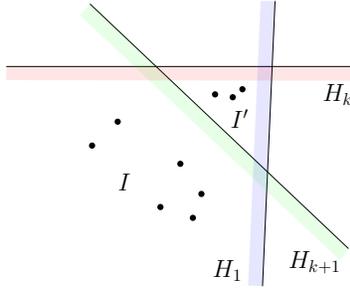


Figure 3: Proof of Lemma 13: The island I belongs to $\mathcal{I}_{k+1} \setminus \mathcal{I}_k$, then I is the intersection of S with $k+1$ halfplanes H_1, \dots, H_{k+1} . The island I' is such that $I \cup I'$ is the intersection of S with the halfplanes H_1, \dots, H_k .

Assume w.l.o.g. that the edges of the polygon $H_1 \cap H_2 \cap \dots \cap H_{k+1}$ in clockwise order belong to the boundary of H_1, H_2, \dots, H_{k+1} , respectively. Let $I' := S \cap (H_k \cap \overline{H_{k+1}} \cap H_1) \in \mathcal{I}_3$, and observe that $I \cup I' \in \mathcal{I}_k$. If $\text{disc}(I') \geq D_{k+1}/2$ then $D_{k+1} \leq 2 \text{disc}(I') \leq 2D_3 \leq 2D_k$. Otherwise, we have that $\text{disc}(I \cup I') > D_{k+1}/2$ which implies that $D_{k+1} < 2D_k$. \square

Theorem 14 $\max \left\{ \frac{\mathcal{C}(S)}{128}, \frac{\mathcal{C}(S)}{64} - |r - b| \right\} \leq \max \left\{ \frac{D_2}{8}, \frac{D_2}{4} - |r - b| \right\} \leq \mathcal{C}(S)$.

Proof. Observe that $\max\{\frac{D_2}{8}, \frac{D_2}{4} - |r - b|\} \leq \mathcal{C}(S)$ follows from Lemma 12. Since any convex partition of S has a 5-separable island, and using Lemma 13, we have that $\mathcal{C}(S) \leq D_5 \leq 2D_4 \leq 4D_3 \leq 16D_2$. With these facts the result follows. \square

Theorem 15 *There exists a polynomial time constant-approximation algorithm for computing $\mathcal{C}(S)$.*

Proof. The value of D_2 , equal to the discrepancy of a 2-separable island of maximum discrepancy, can be computed in $O(n^3 \log n)$ time [7]. The result then follows from Theorem 14. \square

5 Conclusions

We proved that the discrepancy of the family of all k -separable islands is upper bounded by $O(n^{1/4} \sqrt{\log n})$, by showing that its dual shatter function $\pi_{\mathcal{I}_k}^*(m)$ is upper bounded by $O(m^2)$. It is known that the dual shatter function bound can be tight for some range spaces (see [10]). It is not hard to see that the primal shatter function of the k -separable islands of point sets in convex position is lower bounded by $\Omega(m^k)$. So the primal shatter function bound can be arbitrarily worse than the dual shatter function bound in this case. It is also interesting to note that the discrepancy of 1-separable islands (or halfplanes) is upper bounded by $O(n^{1/4})$. We leave the exact (asymptotic) computation of the discrepancy of k -separable islands as an open problem.

Using the fact that every convex partition of a point set S has an island (in this case a 5-separable island) of low discrepancy, we showed that every n -point set in general position in the plane can be two-colored so that the coarseness is upper bounded by $O(n^{1/4} \sqrt{\log n})$. However, Theorem 8 provides more information; for any positive constant $c < 1$ there exists a positive integer k_c (depending only on c), so that in every convex partition of S into m islands at least cm of them are k_c -separable (and thus have small discrepancy). We think that computing the exact asymptotic value of the above bound on the coarseness of point sets is an interesting (and hard) open problem.

We further showed the first approximation for computing the coarseness of a colored point set, which is believed to be NP-hard. Our approximation is based on a known algorithm. Proving the hardness of this problem remains open, and giving improved approximations as well.

6 Acknowledgments

The problems studied here were introduced and partially solved during a visit to University of Valparaiso funded by project Fondecyt 11110069 (Chile). The authors would like to thank an anonymous referee for helpful comments.

References

- [1] J. R. Alexander. Geometric methods in the study of irregularities of distribution. *Combinatorica*, 10(2):115–136, 1990.
- [2] J. R. Alexander, J. Beck, and W. W. L. Chen. Geometric discrepancy theory and uniform distribution. In *Handbook of Discrete and Computational Geometry*, pages 185–207. CRC Press, 1997.
- [3] C. Bautista-Santiago, J. M. Díaz-Báñez, D. Lara, P. Pérez-Lantero, J. Urrutia, and I. Ventura. Computing optimal islands. *Oper. Res. Lett.*, 39(4):246–251, 2011.
- [4] S. Bereg, J. M. Díaz-Báñez, D. Lara, P. Pérez-Lantero, C. Seara, and J. Urrutia. On the coarseness of bicolored point sets. *Comput. Geom.*, 46(1):65–77, 2013.
- [5] B. Chazelle. The discrepancy method in computational geometry. In *Handbook of Discrete and Computational Geometry*, pages 983–996. CRC Press, 2004.
- [6] B. Chazelle, J. Matoušek, and M. Sharir. An elementary approach to lower bounds in geometric discrepancy. *Discrete & Computational Geometry*, 13:363–381, 1995.
- [7] D. P. Dobkin and D. Gunopulos. Concept learning with geometric hypotheses. In *Proceedings of the eighth annual conference on Computational learning theory*, COLT’95, pages 329–336, New York, NY, USA, 1995. ACM.
- [8] M. Drmota and R. F. Tichy. Sequences, Discrepancies and Applications. In *Lecture Notes in Mathematics*, volume 1651, pages 983–996. Springer, 1997.
- [9] H. Edelsbrunner, A. D. Robison, and X. Shen. Covering convex sets with non-overlapping polygons. *Discrete Mathematics*, 81(2):153–164, 1990.
- [10] J. Matoušek. *Geometric Discrepancy: An Illustrated Guide*. Springer-Verlag, 1999.
- [11] J. Pach and P. K. Agarwal. *Combinatorial geometry*. Wiley-Interscience series in discrete mathematics and optimization. Wiley, 1995.