On the Coarseness of Bicolored Point Sets

S. Bereg [*], J.M. Díaz-Báñez[†], D. Lara[‡], P. Pérez-Lantero[§], C. Seara[¶], J. Urrutia[||]

October 18, 2011

Abstract

Let R be a set of red points and B a set of blue points on the plane. In this paper we introduce a new concept, which we call *coarseness*, for measuring how *blended* the elements of $S = R \cup B$ are. For $X \subseteq S$, let $\nabla(X) = ||X \cap R| - |X \cap B||$ be the bichromatic discrepancy of X. We say that a partition $\Pi = \{S_1, S_2, \ldots, S_k\}$ of S is convex if the convex hulls of its members are pairwise disjoint. The *discrepancy* of a convex partition Π of S is the minimum $\nabla(S_i)$ over the elements of Π. The *coarseness* of S is the discrepancy of the convex partition of S with maximum discrepancy. We study the coarseness of bicolored point sets, and relate it to *well blended* point sets. In particular, we show combinatorial results on the coarseness of general configurations and give efficient algorithms for computing the coarseness of two specific cases, namely when the set of points is in convex position and when the measure is restricted to convex partitions with two elements.

1   Introduction

Let S be a set of n points on the plane in general position. A bicoloring of S is a partition of S into two disjoint subsets R and B such that $S = R \cup B$. A bicoloring of S will be denoted by $\{R, B\}$. The elements of R and B are called the blue and red points of S respectively. We will assume that R and B are non-empty and have r and b elements, respectively. If S is a point set, $CH(S)$ will denote the convex hull of S. A point set S is in *convex position* if the elements of S are the vertices of a convex polygon.

Given a bicoloring $\{R, B\}$ of S, the following problem is studied: How well blended are the elements of R and B? In this paper, we will define a new concept, which we call the *coarseness* of a bicoloring $\{R, B\}$ of S, which can be used to study the previous problem. We will focus mainly on how to calculate the coarseness of some families of bicolored point sets on the plane.

Intuitively speaking, given a bicoloring $\{R, B\}$ of S, R and B are well blended if, at a glance, the elements of R and B appear to be distributed uniformly on any convex region of the plane. Unfortunately, experience dictates that when one attempts to give a formal definition of well blended point sets, one always runs afoul, encountering numerous contradictions and counterexamples.

[*]University of Texas at Dallas, `besp@utdallas.edu`

[†]Universidad de Sevilla, `dbanez@us.es`

[‡]Universidad Nacional Autónoma de México, `dlara@uxmcc2.iimas`

[§]Escuela de Ingeniería Civil Informática, Universidad de Valparaíso, `pablo.perez@uv.cl`

[¶]Universidad Politécnica de Catalunya, `carlos.seara@upc.edu`

[||]Instituto de Matemáticas, Universidad Nacional Autónoma de México, `urrutia@matem.unam.mx`

For example, let us take a look at a seemingly easy problem. Let S be a set of n points on the real line. How can we find a bicoloring $\{R, B\}$ of S in such a way that R and B are well blended? One may expect the following random procedure to generate well blended bicolored point sets: With probability $\frac{1}{2}$ *randomly* color the elements of S red or blue. Unfortunately, a simple exercise in probability theory asserts that for any fixed k, as n grows, the probability converges to one that the coloring of S obtained above contains monochromatic blocks of length greater than or equal to k.

For the real line, we may argue that the colorings that best blend R and B are those in which the colors of the elements of S alternate. These colorings are not likely to be obtained randomly (the probability of alternating coloring is $\frac{1}{2^{n-1}}$ where $n = |R| + |B|$ and $|R| = |B|$). This example, however, suggests what appears to be a good parameter to study. We say that a bicolored set of points on the real line ℓ is well blended if in any interval I of ℓ the difference between the number of red and the number of blue points of S is at most one. This in fact, solves the problem of finding well blended bicolorings of point sets on the line.

The natural generalization of the above apparently good definition fails in the plane. Let S be a set of points in general position on the plane. A subset I of S is an *island* if there is a convex set C on the plane such that $I = C \cap S$. Given a bicoloring $\{R, B\}$ of S, the discrepancy of an island of S, denoted by $\nabla(I)$, is the difference between the number of red and blue points of I, that is, $\nabla(I) = ||I \cap R| - |I \cap B||$. An island I has *discrepancy k* if $\nabla(I) = k$. We might now try saying that a bicoloring $\{R, B\}$ of S is well blended if the discrepancy of any island of S is bounded by a constant.

For point sets in general position, this will never happen, since any bicolored point set S contains islands with logarithmic discrepancy. In order to see this, we refer to a well known result by Erdős-Szekeres [15] which states that any set with at least $\binom{2k-4}{k-2} + 1$ points in general position contains a subset of k points in convex position. Assume that $|S| \geq \binom{2k-4}{k-2} + 1$, and consider any bicoloring $\{R, B\}$ of S. Let S' be a subset of S in convex position with cardinality $|S'| \geq k$. Let I_1 be the island of S obtained by intersecting S with the convex hull of S'. If the discrepancy of I_1 is greater than or equal to $\frac{k}{4}$ we are done. Assume w.l.o.g. that the number of red points in S' is at least $\frac{k}{2}$. Then by removing from I_1 all of the red points in S', we obtain an island with discrepancy greater than or equal to $\frac{k}{4}$.

A more worrisome fact arises with point sets in convex position. Let S be a set of $2n$ points equally spaced on a circle. Color each element of S red or blue in such a way that the colors of the elements of S alternate when traversing the circle clockwise. At first glance, S appears to be well blended; however it does contain monochromatic islands with n elements!

Clearly we must be careful how we define well blended point sets. In this paper we introduce a reasonable parameter to detect well blended point sets, which we call the *coarseness* of S. With this new measure, the concept of *discrepancy* is generalized from an object to a partition of the point set. Intuitively speaking, if a bicoloring $\{R, B\}$ of S is not well blended, we should be able to split it into large convex blocks, each with large discrepancy; see Figure 1. We formalize the definition as follows.

Let $\{R, B\}$ be a bicoloring of S. For $X \subseteq S$, let $\nabla(X) = ||X \cap R| - |X \cap B||$ be the bichromatic discrepancy of X. We say that a partition $\Pi = \{S_1, S_2, \ldots, S_k\}$ of S is convex if $CH(S_i) \cap CH(S_j) = \emptyset$, $S_i, S_j \in \Pi$. The discrepancy of a convex partition Π of S, $d(S, \Pi)$, is the minimum $\nabla(S_i)$ over the elements of Π.
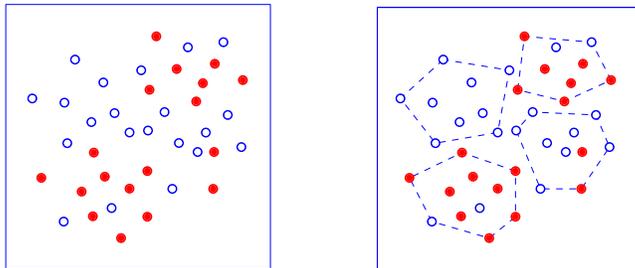
Figure 1: a) A point set S that is not well blended. b) A partition of S into large blocks, each with large discrepancy.

Definition 1. *The coarseness of $\{R, B\}$, $\mathcal{C}(\{R, B\})$, is the discrepancy of the convex partition of S with maximum discrepancy.*

Given r and b, let $\mathcal{C}(S, r, b)$ be the smallest coarseness taken over all the bicolorings $\{R, B\}$ of S such that $|R| = r$, and $|B| = b$. Given two colorings $\{R, B\}$ and $\{R', B'\}$ with $|R| = |R'|$ and $|B| = |B'|$, we say that $\{R, B\}$ is better than $\{R', B'\}$ if the coarseness of $\{R, B\}$ is smaller than or equal to the coarseness of $\{R', B'\}$.

Definition 2. *A bicoloring $\{R, B\}$ of S is well blended if the coarseness of $\{R, B\}$ is within a constant factor of $\mathcal{C}(S, r, b)$.*

Let us examine again the example of $2n$ points on a circle whose colors alternate. It is easy to see that in any convex partition of S, there must be an element, say S_i, containing a set of *consecutive* elements of S on the circle. The discrepancy of such S_i, $\nabla(S_i)$, is thus 0 or 1. It follows that the coarseness of this bicoloring of S is equal to 1, and thus it is well blended.

Let us consider another example. Let S be the set of $4m^2$ points with integer coordinates (i, j), $1 \leq i, j \leq 2m$ colored as follows: A point $(i, j) \in S$ is in B if $i + j$ is even, otherwise it is in R. Let $\{S_1, S_2\}$ be the partition of S such that S_1 contains the points $(i, j) \in S$ such that $i + j \leq 2m + 1$, and $S_2 = S \setminus S_1$; see Figure 2b). It is easy to see that $\nabla(S_1) = m + 1$, and $\nabla(S_2) = m$. Thus $\mathcal{C}(\{R, B\}) \geq m$.

Finally, let us briefly analyze bicolorings $\{R, B\}$ of point sets in which R and B are *linearly separable*; i.e., there is a line ℓ that leaves all the elements of R on one of the half-planes it determines, and all the elements of B on the other. In this case, the partition $\{R, B\}$ of S is convex, and thus the coarseness of $\{R, B\}$ is at least the minimum of r and b. Let us point out that in general the partition $\{R, B\}$ is not the optimal partition. To see this, let us consider R and B such that $r \geq 2b$ as in Figure 2b). In Section 2 we will see that the optimal partition in this case is $\Pi = \{S\}$, and thus $\mathcal{C}(\{\text{R, B}\}) = \nabla(S) = r - b$.

1.1 Our contribution

We believe that the problem of calculating $\mathcal{C}(S, r, b)$ is *NP*-hard, and focus mainly on calculating the coarseness of some families of bicolored point sets. Thus in the rest of this paper, we will assume that we are given a set of points S with a fixed bicoloring, $\{R, B\}$, and focus mainly on the algorithmic issues of calculating $\mathcal{C}(\{R, B\})$. To make our notation easier, instead of referring to a bicoloring $\{R, B\}$ of S, we will refer to S as $S = R \cup B$, and to $\mathcal{C}(\{R, B\})$ simply as $\mathcal{C}(S)$.
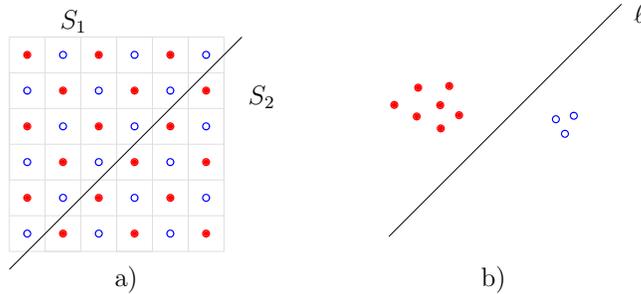
3

Figure 2: a) A bicolored grid with coarseness 3. b) The coarseness is not given by the separation line ℓ.

The remainder of the paper is structured as follows. In Section 2 we study some general properties of the coarseness of bicolored point sets. In Section 3 we focus on point sets in convex position, and in Section 4 we give properties for some specific configurations of well blended points. In Section 5 we study the *linear coarseness* of bicolored point sets; that is, the maximum discrepancy over all the convex partitionings of a bicolored point set into two subsets. We study some point sets with minimum linear coarseness and prove that computing the linear coarseness of point sets is 3SUM-hard.

1.2 Related work

It is important to observe that in our problem the cardinality of the convex partition is not fixed. A related problem is the so-called k-clustering problem [5]: Given a set S of n points (non-colored) in the plane and a fixed value k, compute a k-clustering (a partitioning of S into k islands) which minimizes any monotone function of the diameters or the radii of the clusters. The fact that any two clusters in an optimal solution can be separated by a line allows the k-clustering problem to be solved in polynomial time. If we restrict ourselves to convex partitions of S with exactly k elements, we obtain what we call *the k-coarseness* of S, denoted as $\mathcal{C}_k(S)$. When $k = 1$, the partition Π has only one element, and thus $\mathcal{C}_1(S) = \nabla(S) = |r - b|$. If $k = 2$ then we have what we call *linear coarseness*; that is, the coarseness obtained by partitions of S induced by lines that split S into two subsets. Clearly, the k-coarseness can be computed in polynomial time when k is a constant. Since for any set of n points in the plane there are $O(n^{6k-12})$ k-partitions into disjoint islands and all of them can be computed within the same complexity [5], the k-coarseness of a bicolored set of points can be found in $O(n^{6k-11})$ time, $k \geq 3$.

Note that our concept of coarseness is related to the study of uniform distributions [1, 19] and could be applied in data analysis and clustering as follows. We say that $S = R \cup B$ is *not good* for clustering when its coarseness is low.

The concept of discrepancy has been used before in combinatorial geometry. In [2, 21] a parameter known as the combinatorial discrepancy of hypergraphs is studied. The problem studied in these papers is that of assigning to each vertex of a hypergraph either weight $+1$ or -1 in such a way the maximum weight over all the edges of the hypergraph is minimized, where the weight of an edge is the absolute value of the sum of the weights of its vertices. Another concept of discrepancy is considered in [21], in which the authors study the problem of finding the most uniform way of distributing n points in the unit square according to some criteria. In geometric discrepancy theory [7] the following problem is addressed: How can n points on the plane be colored in such a way that the difference between the number of red points and blue points within any disk is minimized? Finally, we make reference to papers that consider the computation of geometric objects with

maximum discrepancy and their applications to computer graphics and machine learning. These papers [3, 11, 12, 16] consider the problem of computing a convex set Q such as a box, triangle, strip, convex polygon, etc., such that the discrepancy of the subset of S contained in Q is maximized.

Partitionings of bicolored point sets into monochromatic islands with disjoint convex hulls have been studied in [13]. If there is a partitioning Π of S into monochromatic islands with disjoint interiors such that every island has at least k elements, then the coarseness of S is at least k.

In some cases, e.g. bicolored point sets in convex positions with $2n$ elements whose colors alternate, any such partitioning has discrepancy 0 or 1. Unfortunately for some point sets with high coarseness, any partitioning of them into monochromatic islands with disjoint interiors has constant discrepancy. Consider the point set S consisting of two sets of points R and B which are linearly separable by a line ℓ and contained in a circle of radius one centered at a point at distance 3 from the origin. In addition, S contains a set S' of $2n$ points evenly distributed on the unit circle centered at the origin, and such that ℓ bisects the elements of S' evenly. It is not hard to see that any convex partitioning of S into monochromatic islands with disjoint interiors always has an island with one or two elements. However ℓ partitions S into two subsets each with discrepancy n.

2   Basic Properties

Let $S = R \cup B$, $|R| = r$, and $|B| = b$, and let $X \subseteq R \cup B$. We denote $|X \cap R| - |X \cap B|$ by $\nabla'(X)$. Observe that $\nabla(X) = |\nabla'(X)|$. We say that X is majority-red, for short, *m-red* (resp. *m-blue*) if $\nabla'(X) > 0$ (resp. $\nabla'(X) < 0$). Let $\Pi = \{S_1, S_2, \ldots, S_k\}$ be a convex partition of S. We say that Π is *optimal* if $\mathcal{C}(S) = d(S, \Pi)$. Let $r_i = |S_i \cap R|$ and $b_i = |S_i \cap B|$, for $i = 1, \ldots, k$.

The following lemmas list some basic properties of the coarseness of point sets.

Lemma 1. $\mathcal{C}(S) \geq 1$. If $\mathcal{C}(S) = 1$, then $|r - b| \leq 1$.

Proof. Suppose that $S = \{p_1, p_2, \ldots, p_{r+b}\}$. Let $\Pi = \{\{p_1\}, \{p_2\}, \ldots, \{p_{r+b}\}\}$. We have $\mathcal{C}(S) \geq d(S, \Pi) = 1$. Moreover, if $\mathcal{C}(S) = 1$ then $|r - b| = \mathcal{C}_1(S) \leq \mathcal{C}(S) = 1$. ∎

Lemma 2. If $\Pi = \{S_1, \ldots, S_k\}$ is an optimal convex partition of S and $k \geq 2$, then there are two elements S_i and S_j of Π such that S_i is m-red and S_j is m-blue.

Proof. Suppose that S_i is m-blue for every index i. Then

$$b - r = \sum_{i=1}^{k} (b_i - r_i) = \sum_{i=1}^{k} \nabla(S_i) > \min_{i=1}^{k} \nabla(S_i) = d(S, \Pi) = \mathcal{C}(S),$$

which contradicts that $\mathcal{C}_1(S) \leq \mathcal{C}(S)$. Therefore, it follows that Π contains m-blue and m-red elements. ∎

Lemma 3. If a convex partition $\Pi = \{S_1, \ldots, S_k\}$ of S contains an m-red and an m-blue element, then $d(S, \Pi) \leq \min\{r, b\}$.

Proof. Suppose w.l.o.g. that S_1 is m-red and S_2 is m-blue. Then

$$d(S, \Pi) \leq \nabla(S_1) = r_1 - b_1 \leq r_1 \leq r$$
$$d(S, \Pi) \leq \nabla(S_2) = b_2 - r_2 \leq b_2 \leq b.$$

Hence $d(S, \Pi) \leq \min\{r, b\}$. ∎

Lemma 4. *If $r \geq 2b$ or $b \geq 2r$, then $\mathcal{C}(S) = \mathcal{C}_1(S) = |r - b|$.*

Proof. Assume w.l.o.g. that $r \geq 2b$. We have that $\mathcal{C}_1(S) \leq \mathcal{C}(S)$. Suppose now that Π is an optimal convex partition of S with cardinality bigger than one. By Lemma 2, Π contains m-red and m-blue elements, thus $d(S, \Pi) \leq \min\{r, b\}$ by Lemma 3. Then $\mathcal{C}(S) = d(S, \Pi) \leq \min\{r, b\} = b \leq r - b = \mathcal{C}_1(S)$. This implies that $\mathcal{C}(S) = \mathcal{C}_1(S)$. ∎

Lemma 5. *If R and B are linearly separable, and $b \leq r < 2b$ or $r \leq b < 2r$, then $\mathcal{C}(S) = \min\{r, b\}$.*

Proof. Suppose w.l.o.g. that $b \leq r < 2b$ and let Π be an optimal convex partition of S. Π can not have cardinality one because $\mathcal{C}_1(S) = r - b < b = d(S, \{R, B\})$. Therefore by Lemma 2, Π has m-red and m-blue elements implying, by Lemma 3, that $\mathcal{C}(S) = d(S, \Pi) \leq \min\{r, b\} = b$. Since $d(S, \{R, B\}) = b$ then $\mathcal{C}(S) = b$. ∎

Remark 1. It follows easily from Lemmas 4 and 5 that if R and B are linearly separable, and $b \leq r < 2b$ or $r \leq b < 2r$, then $\mathcal{C}(S) = \min\{r, b\}$, otherwise $\mathcal{C}(S) = |r - b|$.

We now prove:

Lemma 6. *Let $\Pi = \{S_1, \ldots, S_k\}$ be a convex partition of S. Then $d(S, \Pi) \leq \frac{r+b}{k}$.*

Proof.

$$d(S, \Pi) = \min_{i=1}^{k} \nabla(S_i) \leq \frac{1}{k} \sum_{i=1}^{k} \nabla(S_i) = \frac{1}{k} \sum_{i=1}^{k} |r_i - b_i| \leq \frac{1}{k} \sum_{i=1}^{k} (r_i + b_i) = \frac{r+b}{k}$$

∎

Remark 2. It is worth noting that if $\mathcal{C}(S)$ is large with respect to the cardinality of S, then there exists an optimal convex partition of S with few elements. For point sets with small coarseness, the cardinality of an optimal convex partition can be small or large; see Figure 3.
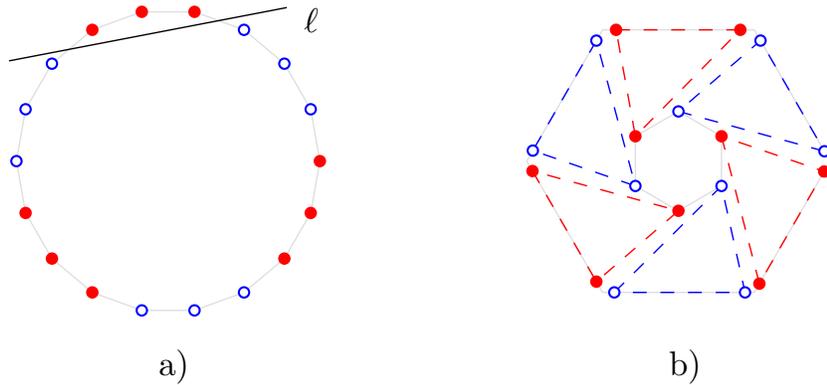

a)   b)

Figure 3: Two point sets with coarseness 3. In a) the coarseness is determined by the convex partition induced by line ℓ. In b) the coarseness is obtained by using the subsets determined by the triangles shown. These examples can be generalized to point sets with $n = tm$ points and coarseness t.

The following lemma states a relation between the members of a minimum-cardinality optimal convex partition. It shows that in an optimal convex partition with minimum cardinality, one can not join two or more members having the same majority color (i.e. m-red or m-blue) because the convex hull of the resulting member contains points of other members of the partition.

Lemma 7. *If Π is a minimum-size optimal convex k-partition and $S_i, S_j \in \Pi$ are both m-red (or m-blue), then $k \geq 3$. Moreover $CH(S_i \cup S_j) \cap S_l \neq \emptyset$ for every $l \neq i, j$.*

Proof. Suppose that $CH(S_i \cup S_j) \cap S_l = \emptyset$ for every $l = 1, \ldots, k$, $l \neq i, j$. Then, the partition $\Pi' = (\Pi \setminus \{S_i, S_j\}) \cup \{S_i \cup S_j\}$ is a convex partition of S so that $d(S, \Pi) \leq d(S, \Pi')$ because $\min\{\nabla(S_i), \nabla(S_j)\} < \nabla(S_i) + \nabla(S_j) = \nabla(S_i \cup S_j)$. This is a contradiction since the cardinality of Π' is $k - 1$. ∎

3 Point Sets in Convex Position

Let S be a point set in convex position. A subset of S is called *S-consecutive* if it is empty or its elements are consecutive vertices of $CH(S)$.

Lemma 8. *If S is in convex position, then any convex partition $\Pi = \{S_1, \ldots, S_k\}$ of S with $k > 1$ has at least two elements S_i and S_j which are S-consecutive.*

Proof. We assume that $S_i \neq \emptyset$ for every i. The proof is by induction on $r + b$. If $r + b = 1$ then it is trivial. Now suppose that $r + b > 1$. Let $\Pi = \{S_1, S_2, \ldots, S_k\}$ be any convex partition of S, and suppose that the elements of S_1 are not S-consecutive. Then $S \setminus S_1$ is composed of at least two maximal S-consecutive chains. Let C be one of these chains. If C is not an element of Π then C is partitioned into at least two sets (induced by Π) and the claim follows by induction. Otherwise each chain in $S \setminus S_1$ is an element of Π and the lemma follows. ∎

A point set S in convex position is called an *alternating convex chain* if we can label its elements $p_1, p_2, \ldots, p_{r+b}$ counterclockwise along $CH(S)$ so that for every $1 \leq i < r + b$, p_i and p_{i+1} do not have the same color (Figure 4).
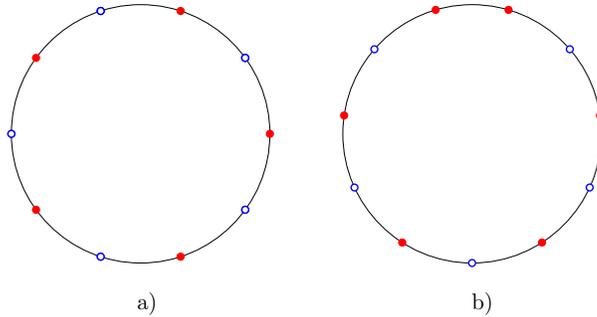

Figure 4: Alternating convex chains. a) 5 red points and 5 blue points, b) 6 red points and 5 blue points.

Lemma 9. *If S is in convex position then $\mathcal{C}(S) = 1$ if and only if S is an alternating convex chain.*

Proof. Suppose that $\mathcal{C}(S) = 1$ and that S is not an alternating convex chain. By Lemma 1, $|r - b| \leq 1$. If $r = b$ and for some i we have that p_i and p_{i+1} have the same color, then the partition $\Pi = \{\{p_i, p_{i+1}\}, S \setminus \{p_i, p_{i+1}\}\}$ has coarseness two. If $r = b + 1$ and there is an $1 \leq i < r + b$ such that p_i and p_{i+1} are blue points, then if $S_1 = \{p_i, p_{i+1}\}$, we have that $\nabla(S_1) = 2$, $\nabla(S \setminus S_1) = 3$, and then $\mathcal{C}(S) \geq d(S, \Pi) = 2$. Thus S is an alternating convex chain.

Suppose now that S is an alternating convex chain. By Lemma 8, any convex partition $\Pi = \{S_1, S_2, \ldots, S_k\}$ of S has at least two S-consecutive elements, and thus for at least one of them, say S_i, $\nabla(S_i)$ is 1 or 0. Then $d(S, \Pi) \leq 1$. ∎

Theorem 1. *If S is in convex position, then $\mathcal{C}(S) = \max_{k=1,2,3} \mathcal{C}_k(S)$.*

Proof. Let $d = \mathcal{C}(S)$, and observe that $0 \leq \nabla(S) \leq d$. Assume w.l.o.g. that $0 \leq \nabla'(S) \leq d$. Let $\Pi = \{S_1, S_2, \ldots, S_k\}$ be an optimal convex partition of S of minimum cardinality. By definition, $\nabla(S_i) \geq d$ $(1 \leq i \leq k)$. Suppose that $k > 3$. By Lemma 8, S has at least two S-consecutive elements, say S_1 and S_2. If either of S_1 or S_2, say S_1, is such that $\nabla'(S_1) \leq -d$, then $\nabla'(S \setminus S_1) = \nabla'(S) - \nabla'(S_1) \geq 0 + d = d$, and thus $\nabla(S \setminus S_1) = |\nabla'(S \setminus S_1)| \geq d$. This is a contradiction since the convex partition $\Pi' = \{S_1, S \setminus S_1\}$ has cardinality 2 and $d(S, \Pi') \geq d$. Suppose then that $\nabla'(S_1) \geq d$ and $\nabla'(S_2) \geq d$. Observe that $\nabla'((S \setminus S_1) \setminus S_2) = \nabla'(S) - \nabla'(S_1) - \nabla'(S_2) \leq d - d - d = -d$, and thus $\nabla((S \setminus S_1) \setminus S_2) = |\nabla'((S \setminus S_1) \setminus S_2)| \geq d$. This is a contradiction because $\Pi'' = \{S_1, S_2, S \setminus (S_1 \cup S_2)\}$ has cardinality 3 and $d(S, \Pi'') \geq d$. ∎

Our objective now is to prove that the coarseness of point sets in convex position can be computed in $O(n \log n)$ time. First we will show how to solve two problems on circular sequences of real values. These problems may also be of independent interest.

3.1 Two Maximum Weight Problems on Circular Sequences

Consider a set X of n points on a circle labeled clockwise with real numbers $x_0, \ldots, x_{n-1}$. An *interval* $[x_i, x_j]$ of X is the subset containing the points $x_i, x_{i+1}, \ldots, x_j$ with addition taken mod n. The weight $w[x_i, x_j]$ of $[x_i, x_j]$ is defined as $x_i + x_{i+1} + \cdots + x_j$. In this section the following problems will be solved:

Problem 1 (The Maximum Weight Interval of a Circular Sequence Problem, abbreviated MWI-Problem). *Find the interval of X with maximum weight.*

Problem 2 (The Max-Min Two Interval Problem, abbreviated MM2I-Problem). *Find two disjoint intervals $[x_i, x_j]$ and $[x_k, x_\ell]$ of X such that the minimum of $w[x_i, x_j]$ and $w[x_k, x_\ell]$ is maximized.*

We give an outline of how to solve the MWI-Problem, and a more detailed solution to the MM2I-Problem.

To start with, we see that the MWI-Problem is a small variation of the well known Bentley's maximum-weight interval problem [4]. More specifically, let $X = (x_0, \ldots, x_{n-1})$ be a *linear* sequence of n real values. An interval $[x_i, x_j]$ of X contains the elements $x_i, x_{i+1}, \ldots, x_j$, $i \leq j$. Bentley's problem is that of finding the interval of X with maximum weight. Observe that in Bentley's problem, $i \leq j$, whereas in the MWI-Problem this is not necessarily the case. It is well known that Bentley's problem can be solved in linear time.

Observe that if a solution $[x_i, x_j]$ to the MWI-Problem is such that $0 \leq i \leq j \leq n - 1$, then the solution obtained by solving Bentley's problem on $(x_0, \ldots, x_{n-1})$ is $[x_i, x_j]$. Otherwise, $[x_i, x_j]$ is the union of two disjoint intervals $[x_0, x_i]$ and $[x_j, x_{n-1}]$ maximizing $w[x_0, x_i] + w[x_j, x_{n-1}]$. This case can be solved in linear time; see [9] for more detail.

We show now how to solve the MM2I-Problem in $O(n \log n)$ time. In [8] the authors solve the following problem, which they call the *Range Maximum-Sum Segment Query Problem with Two Query Intervals*:

Problem 3 (RMSQ2-Problem). *Preprocess in linear time a given sequence $X = (x_1, \ldots, x_n)$ of real numbers such that for any $i \leq j \leq k \leq \ell$, the following query can be answered in constant time: Find the interval $[x_s, x_t]$ of maximum weight such that $i \leq s \leq j$ and $k \leq t \leq \ell$.*

Remark 3. If $i = j = k$ in the above problem, $[x_s, x_t]$ will be the interval of maximum weight contained in $[x_i, x_\ell]$ starting at x_i.

Let $X' = (x_0, \dots, x_{n-1}, x_n, \dots, x_{2n-1})$, where $x_{n+i} = x_i$, $i = 0, \dots, n-1$. Preprocess X' as in [8] to solve the RMSQ2-Problem.

Let $I_1 = [x_i, x_j]$ and $I_2 = [x_k, x_\ell]$ form an optimal solution to the MM2I-Problem, and suppose w.l.o.g. that $i \le j < k$. We solve now the MM2I-Problem in $O(\log n)$ time for a fixed value of i, $0 \le i \le n-1$.

For simplicity assume that $i = 0$. Let $I_1 = [x_0, x_j]$ and $I_2 = [x_k, x_\ell]$ be an optimal solution to our problem. Notice that there exists an index t such that $j \le t < k$. Let $I_1(t)$ be an interval of maximum weight contained in $[x_0, x_t]$ that starts at x_0. By Observation 3, $I_1(t)$ can be found in constant time. Let $I_2(t)$ be the interval of maximum weight contained in $[x_{t+1}, x_{n-1}]$. This can also be found in constant time. Observe that by the way we choose I_1 and I_2, we can assume that I_1 and I_2 are $I_1(t)$ and $I_2(t)$, respectively. Then the MM2I-Problem consists of finding a value of t so that $\min\{w(I_1(t)), w(I_2(t))\}$ is maximum.

Suppose now that for a given t, $w(I_1(t)) \le w(I_2(t))$. Then we can discard all indexes $t' < t$ since $\min\{w(I_1(t')), w(I_2(t'))\} \le w(I_1(t')) \le w(I_1(t)) = \min\{w(I_1(t)), w(I_2(t))\}$. The case when $w(I_1(t)) > w(I_2(t))$ is analogous. It now follows that we can search for t in a logarithmic number of steps. Since the RMSQ2-Problem can be used to obtain both $I_1(t)$ and $I_2(t)$ in constant time, t can be found in logarithmic time. We repeat this procedure for $i = 1, \dots, n-1$ by using the preprocessing done in X' and choosing I_1 and I_2 in the interval $[x_i, x_{i+n-1}]$ of X'. Thus we have proved:

Theorem 2. *The MM2I-Problem can be solved in $O(n \log n)$ time.*

3.2 Computing the Coarseness of Point Sets in Convex Position

Theorem 3. *The coarseness of a point set S in convex position can be computed in $O(n \log n)$ time and $O(n)$ space.*

Proof. Suppose w.l.o.g. that $b \le r$. By Lemma 4, if $r \ge 2b$, then $\mathcal{C}(S) = \mathcal{C}_1(S) = r - b$. Suppose then that this is not the case. By Theorem 1, we have to compute the maximum among $\mathcal{C}_1(S)$, $\mathcal{C}_2(S)$, and $\mathcal{C}_3(S)$. Assign weights to the elements of S as follows; red points are weighted $+1$, and blue points -1. We can now consider S as a *weighted circular sequence*.

Computing $\mathcal{C}_2(S)$: By Lemma 2, any optimal convex partition contains an m-red element and an m-blue element. Let $\Pi = \{S_1, S_2\}$ be a convex partition of S such that S_1 is m-blue and S_2 is m-red; see Figure 5a). We have that $\nabla(S_2) = r_2 - b_2 = (r - r_1) - (b - b_1) = r - b + b_1 - r_1 = r - b + \nabla(S_1) \ge \nabla(S_1)$, and thus $d(S, \Pi) = \nabla(S_1)$. Then $d(S, \Pi)$ is maximum if and only if $\nabla(S_1)$ is maximum. In this case, S_1 corresponds to an interval of S with minimum weight (i.e., S_2 has maximum weight). This is an instance of the MWI-Problem and can be solved in linear time.

Computing $\mathcal{C}_3(S)$: Assume that $\mathcal{C}(S) = \mathcal{C}_3(S)$. Let $\Pi = \{S_1, S_2, S_3\}$ be an optimal convex partition of S; see Figure 5 b). It is easy to see that if S_i is m-blue (resp. m-red) then S_{i+1} is m-red (resp. m-blue), $i = 1, 2$. Moreover $d(S, \Pi)$ is $\nabla(S_1)$ or $\nabla(S_3)$, otherwise $d(S, \Pi) \le d(S, \{S_1 \cup S_2 \cup S_3\}) = \mathcal{C}_1(S)$. If both S_1 and S_3 are m-blue, then $d(S, \Pi)$ is at most $\mathcal{C}_2(S)$. Then S_1 and S_3 are m-red, and thus computing $\mathcal{C}_3(S)$ reduces to the problem of finding two disjoint intervals in S such that the minimum weight of both is maximized. By Theorem 2 we can solve this problem in $O(n \log n)$ time. ∎
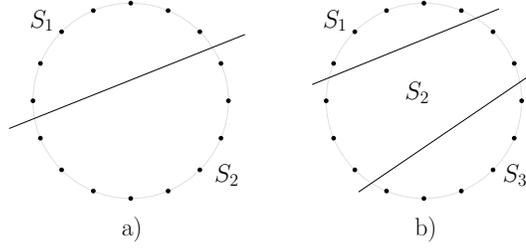
Figure 5: Points in convex position. a) Computing $\mathcal{C}_2(S)$; b) computing $\mathcal{C}_3(S)$.

4 Point Sets in General Position

The problem of determining the coarseness of point sets in general position seems to be nontrivial. We are unable even to characterize point sets whose coarseness is just one. Let us recall that by Remark 2, the cardinality of an optimal convex partition can be small or large, so we have no idea in advance of the number of elements in an optimal partition. In this section we study some particular families of point sets.

Proposition 1. *For all $n \geq 4$, there are bichromatic point sets of size n, not in convex position, with coarseness one.*

Proof. Let S be a point set consisting of the vertices of a regular polygon $\mathcal{P}$ with $2n$ vertices together with an extra point p close to the center of $\mathcal{P}$. Color the vertices of $\mathcal{P}$ red or blue in such a way that adjacent vertices receive different color, and color p red; see Figure 6a). Let $\Pi = \{S_1, S_2, \ldots, S_k\}$ be any convex partition of S. If $k = 1$ then $d(S, \Pi) = 1$. Suppose that $k > 1$. Then there is some $S_i \in \Pi$ ($1 \leq i \leq k$) such that $p \notin S_i$ and S_i contains a set of consecutive vertices of $\mathcal{P}$. Then, $\nabla(S_i) \leq 1$ and therefore $d(S, \Pi) \leq 1$.

For $n = 2m + 2$, let S consist of the vertices of $\mathcal{P}$, colored as before, plus two points p and q in the interior of $\mathcal{P}$ close enough to the middle of an edge e of $\mathcal{P}$ such that the line joining them is almost parallel to e. It is now easy to see that $\mathcal{C}_2(S) = 1$ and that $d(S, \Pi) \leq 1$ for all convex partitions Π of S; see Figure 6 b). $\blacksquare$
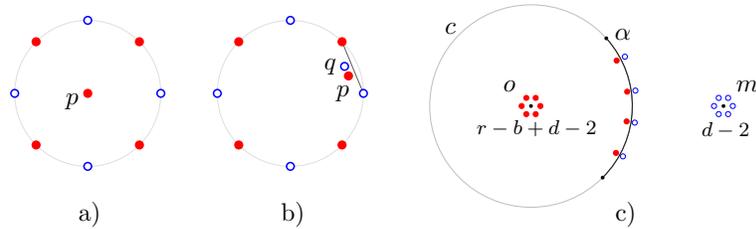

Figure 6: In a) (resp. b)), we show a point set with discrepancy one and an odd (resp. even) number of points. In c), we show a point set in general position with r red points and b blue points. Its coarseness is equal to d, where $\max\{1, |r - b|\} < d \leq b$.

Constructing point sets with large coarseness is straightforward. If R and B are linearly separable then $\mathcal{C}(S) \geq \min\{r, b\}$. Next we prove:

Proposition 2. *Let $b \leq r < 2b$, and d be an integer such that $\max\{1, |r - b|\} < d \leq b$. Then there exists a set of S with r red points and b blue points, not in convex position, such that $\mathcal{C}(S) = d$.*

10

Proof. We construct S as follows. Let c be a circle centered at the origin O, and let α be an arc of c of length $\frac{\pi}{2}$. Let m be the point such that the midpoint of the segment joining O to m is the midpoint of α. On a small circle c' centered at O place $r - b + d - 2$ uniformly spaced red points. In a similar way, place $d - 2$ uniformly spaced blue points on a small circle whose center is m. Finally, place a set T of $b - d + 2$ pairs of points p_i, q_i close enough to α such that each pair contains a red and a blue point as shown in Figure 6c), $i = 1, \ldots, b - d + 2$. The following is easy to prove: If Q is an m-blue subset of S, then $\nabla(Q) \le d$. It now follows that for any convex partition Π of S with at least two elements, $d(S, \Pi) \le d$, and thus $\mathcal{C}(S) \le d$. Since $\mathcal{C}_1(S) = d > r - b$, it follows that $\mathcal{C}(S)$ is not $\mathcal{C}_1(S)$.

Now choose two pairs of points p_i, q_i and p_j, q_j in T, and let ℓ be the line that passes through the midpoints of the segments determined by p_i, q_i and p_j, q_j. Let S' and S'' be the subsets of S determined by ℓ. Then $d(S, \{S', S''\}) = \min\{r - b + d, d\} = d$. Hence $\mathcal{C}(S) = d$. ∎

Let $C_1, \ldots, C_t$ be a family of sets of points such that each C_i is in convex position. We say that $C_1, \ldots, C_t$ is *nested* if the elements of C_{i+1} belong to the interior of $CH(C_i)$, $i = 1, \ldots, t - 1$. The following results deal with families of nested even alternating convex chains; that is, each C_i is an alternating convex chain containing an even number of points; see Figure 7 a).

Lemma 10. *Let $C_1, \ldots, C_t$ be a nested family of point sets in convex position, $S = C_1 \cup \cdots \cup C_t$, and let $\Pi = \{S_1, \ldots, S_k\}$ be a convex partition of S. Then there is an element S_i such that $S_i \cap C_j$ is C_j-consecutive for $j = 1, \ldots, t$.*

Proof. The proof is by induction on t. For $t = 1$ the result follows from Lemma 8. Suppose that $t > 1$ and let $\Pi = \{S_1, \ldots, S_k\}$ be any convex partition of S. If $S_i \subset C_1$ for some $i \in \{1, \ldots, k\}$, then the result follows again from Lemma 8. Let $\Pi_1 = \{S_1 \setminus C_1, \ldots, S_k \setminus C_1\}$. Π_1 is a convex partition of $S \setminus C_1$, and by induction there is a subset $S_i \in \Pi$ such that $S_i \setminus C_1 \in \Pi_1$ and $C_j \cap (S_i \setminus C_1)$ is C_j-consecutive for $j = 2, \ldots, t$. If $S_i \cap C_1$ is C_1-consecutive our result follows. Otherwise, it is easy to see that there is another element $S_l \in \Pi$ such that $S_l \subset C_1$. Our result follows. ∎

Proposition 3. *Let $C_1, \ldots, C_t$ be a family of nested even alternating convex chains, and $S = C_1 \cup \cdots \cup C_t$. Then $\mathcal{C}(S) \le t$. In some cases, $\mathcal{C}(S) = t$.*

Proof. Let $\Pi = \{S_1, \ldots, S_k\}$ be a convex partition of S. By Lemma 10, there is at least one $S_i \in \Pi$ such that $S_i \cap C_j$ is C_j-consecutive for $j = 1, \ldots, t$. Then $\nabla(S_i \cap C_j) \le 1$ for $j = 1, \ldots, t$. But $d(S, \Pi) \le \nabla(S_i) \le \sum_{j=1}^{t} \nabla(S_i \cap C_j) \le t$ and $\mathcal{C}(S) \le t$.

Let W_m be the set of $4m^2$ points with integer coordinates (i, j), $1 \le i, j, \le 2m$, and such that if $i + j$ is even the point is colored blue, otherwise it is colored red. We call such a point set an m-chessboard. Note that W_m is the union of m nested even alternating convex chains, and thus $\mathcal{C}(S) \le m$. Let ℓ be a line with slope $\frac{\pi}{2}$ that leaves $1 + \cdots + 2m - 1$ elements of W_m below it; see Figure 2 b). Then the partition of W_m induced by ℓ has coarseness t. It is clear that W_m can be perturbed a bit so that all of its points are in general position without changing our results. ∎

The idea of the t-chessboard can be generalized as in Figure 7b), in which there is a line ℓ such that in each of the half-planes defined by ℓ, all t chains have the same majority color. It results in interesting cases depending on the value of t. If S is formed by t even alternating convex chains with $4t$ points each, then $C(S) = t = \frac{\sqrt{4t^2}}{2} = \frac{\sqrt{n}}{2}$. If S is a set of $n = 2^{2^m}$ $(m \ge 1)$ points distributed in $t = 2^m = \log_2 n$ even alternating convex chains of length $2^{2^m - m} = \frac{n}{\log_2 n}$ each, then $C(S) = t = \log_2 n$.
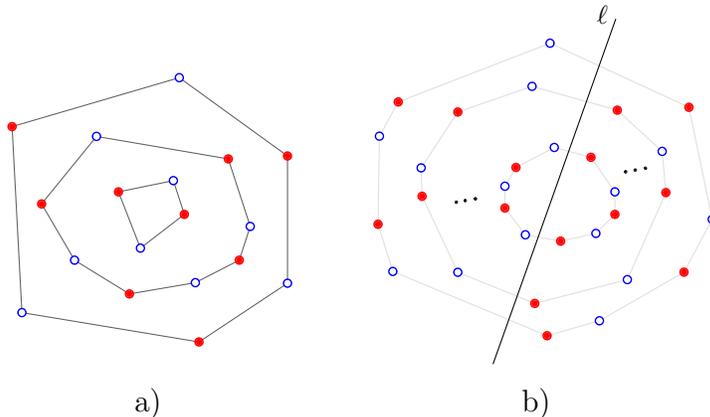
Figure 7: a) A configuration of 3 nested even alternating convex chains. b) A generalization of the m-chessboard.

5 Partitions With a Line

In this section we will characterize sets with linear coarseness one and show how to decide if the linear coarseness of a bicolored point set is equal to a given d. The following notation will be introduced.

Consider a non vertical line ℓ containing no points in S, and let $\Pi_{\ell+}$ and $\Pi_{\ell-}$ be the open half-planes bounded below and above respectively by ℓ. Let $S_{\ell+} = S \cap \Pi_{\ell+}$, $S_{\ell-} = S \cap \Pi_{\ell-}$, and $\Pi_\ell = \{S_{\ell+}, S_{\ell-}\}$. The *linear coarseness of* S is $\mathcal{C}_2(S) = \max_\ell d(S, \Pi_\ell)$.

The following proposition is straightforward.

Proposition 4. *Let $S = R \cup B$ such that $r = b$ and $\mathcal{C}_2(S) = 1$. Then the following properties hold:*

1. *The convex hull of S is an alternating chain.*

2. *When we project the points of S on any line, they form a sequence in which no three consecutive points have the same color.*

3. *For every point $p \in S$ on the convex hull of S, the angular ordering of the elements of $S \setminus \{p\}$ with respect to p is a sequence with alternating colors.*

4. *For every line ℓ passing through two points of the same color, say red, the number of red points in each of $S_{\ell+}$ and $S_{\ell-}$ is exactly one less than the number of blue points in $S_{\ell+}$ and $S_{\ell-}$, respectively.*

Item 2 in Proposition 4 is not sufficient to guarantee that $\mathcal{C}(S) = 1$; see, for example, Figure 8a). If $r \neq b$, items 3 and 4 in the proposition are not necessarily true; see Fig 8b) and c). We now show that if $r = b$, item *4* is sufficient.

The next result, proven in [10], will be useful:

Theorem 4. *Let P and Q be two disjoint convex polygons in the plane. Then there is at least one edge e of P or Q such that the line ℓ_e containing e separates the interior of P from the interior of Q. See Figure 9.*

Lemma 11. *If $r = b$, then the following two conditions are equivalent: (a) $\mathcal{C}_2(S) = 1$; (b) for every line ℓ passing through two points of S with the same color, $\nabla(S_{\ell+}) = \nabla(S_{\ell-}) = 1$.*
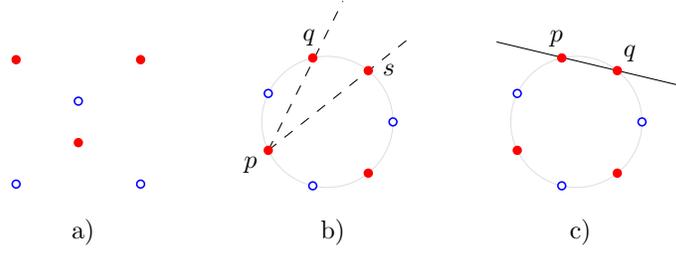
Figure 8: a) There are no three consecutive points of the same color in the projection of S on any line and $C_2(S) = 2$, b) the red points q and s are consecutive in the angular sorting of $S \setminus \{p\}$ with respect to p and $C_2(S) = 1$, c) the number of red and blue points in the half-plane above the line through p and q is zero and $C_2(S) = 1$ because S is an alternating convex chain.
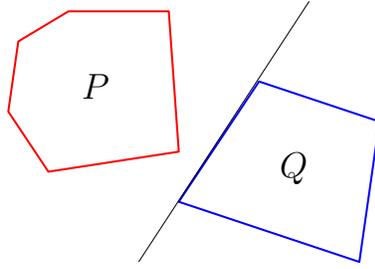

Figure 9: Two convex polygons P and Q and a separating line ℓ_e containing the edge e of Q.

Proof. It is easy to prove that (a) implies (b). We show here that (b) implies (a). Suppose that $C_2(S) = d \geq 2$. We now show that there exists a line ℓ containing two points of the same color of S such that $\{\nabla(S_{\ell^-}), \nabla(S_{\ell^+})\} = \{d, d-2\}$.

Let ℓ_0 be a line containing no elements of S such that $C_2(S) = d(S, \Pi_{\ell_0}) = d$. Assume w.l.o.g. that ℓ_0 is horizontal. Since $r = b$ we have that $C_2(S) = d(S, \Pi_{\ell_0}) = \nabla(S_{\ell_0^+}) = \nabla(S_{\ell_0^-}) = d$ and $\nabla'(S_{\ell_0^+}) = -\nabla'(S_{\ell_0^-})$. Assume w.l.o.g. that $\nabla'(S_{\ell_0^+}) > 0$ (i.e. $S_{\ell_0^+}$ is m-red and $S_{\ell_0^-}$ is m-blue).

Let P and Q be the polygons induced by the convex hulls of $S_{\ell_0^+}$ and $S_{\ell_0^-}$, respectively. Let p be a vertex of P such that there is a line ℓ' passing through p that separates P from Q. Then p must be a red point, for otherwise by translating ℓ' up by a small distance, we obtain a convex partition Π' of S with coarseness $d+1$. Similarly any point q in Q such that there is a line through q that separates P from Q must be blue.

By Theorem 4, there is an edge e of P or Q, with vertices p and q, such that the line ℓ_e containing e separates P from Q. If e is an edge of P, then by using the above observation it can be shown that p and q are red. Thus $\{\nabla(S_{\ell_e^+}), \nabla(S_{\ell_e^-})\} = \{d, d-2\}$. A symmetric argument works when e belongs to Q. ∎

The next proposition gives lower and upper bounds on the linear coarseness.

Proposition 5. $\max\left\{1, \lfloor \frac{|r-b|}{2} \rfloor\right\} \leq C_2(S) \leq \max\left\{\lfloor \frac{|r-b|}{2} \rfloor, \min\{r, b\}\right\}$. *Furthermore, both bounds are tight.*

Proof. Suppose w.l.o.g. that $r \geq b$. By the *Ham Sandwich Theorem* [18] there exists a line ℓ that passes through at most one red point and at most one blue point, such that $|S_{\ell^+} \cap R| = |S_{\ell^-} \cap R| = \lfloor \frac{r}{2} \rfloor$ and $|S_{\ell^+} \cap B| = |S_{\ell^-} \cap B| = \lfloor \frac{b}{2} \rfloor$. Four cases arise depending on the parities of r and b. Here we only show the case when r and b are even. The other cases can be solved in a similar way.

13

If $r = 2a$ and $b = 2c$ then ℓ contains no point of S, and $\nabla(S_{\ell+}) = \nabla(S_{\ell-}) = a - c = \lfloor \frac{r-b}{2} \rfloor$. Thus $\lfloor \frac{r-b}{2} \rfloor = d(S, \Pi_\ell) \leq \mathcal{C}_2(S)$.

If $|r - b| \geq 2$ then $\mathcal{C}_2(S) \geq \lfloor \frac{|r-b|}{2} \rfloor \geq 1$; thus what remains to be proved is that $\mathcal{C}_2(S) \geq 1$ when $|r-b| \leq 1$. Suppose w.l.o.g. that $b \leq r \leq b+1$. If there is a blue point p on the convex hull of S, take a line ℓ separating p from $S \setminus \{p\}$ and suppose that $p \in S_{\ell+}$; then $\nabla(S_{\ell+}) = 1$, $\nabla(S_{\ell-}) = r-b+1 \geq 1$, and $\mathcal{C}_2(S) \geq d(S, \Pi_\ell) = 1$. If no such p exists, there are two consecutive red points p and q in the convex hull of S. Take a line ℓ separating p and q from $S \setminus \{p, q\}$ and suppose that $p, q \in S_{\ell+}$; then $\nabla(S_{\ell+}) = 2$, $\nabla(S_{\ell-}) = b - r + 2 \geq 1$ and $\mathcal{C}_2(S) \geq d(S, \Pi_\ell) = 1$. This proves the lower bound.

Let us show now that this lower bound is tight. Suppose w.l.o.g. that $r > b$ and let X be a set of r red points and r blue points, and let Y be a set of $r - b$ red points. Put the elements of X on an alternating convex chain and the elements of Y in the interior of the convex hull of X in such a way that there is a line ℓ_e such that $d(X, \Pi_{\ell_e}) = 0$ and ℓ_e splits Y into two subsets of cardinality $\lfloor \frac{|Y|}{2} \rfloor$ and $\lceil \frac{|Y|}{2} \rceil$, respectively. Let $S = X \cup Y$ and observe that $d(S, \Pi_{\ell_e}) = \lfloor \frac{|Y|}{2} \rfloor = \lfloor \frac{r-b}{2} \rfloor$. For any line ℓ we have that $d(X, \Pi_\ell) \in \{0, 1\}$ (by Lemma 9). If $d(X, \Pi_\ell) = 0$, then $d(S, \Pi_\ell) = d(X \cup Y, \Pi_\ell) = d(Y, \Pi_\ell) \leq \lfloor \frac{|Y|}{2} \rfloor = \lfloor \frac{r-b}{2} \rfloor$. If $d(X, \Pi_\ell) = 1$ then $d(X \cup Y, \Pi_\ell) = \min\{x - 1, (r - b) - x + 1\}$ where x is such that ℓ splits Y into x and $(r - b) - x$ points, respectively. It is easy to prove that $\min\{x - 1, (r - b) - x + 1\} \leq \lfloor \frac{r-b}{2} \rfloor$. Then $\mathcal{C}_2(S) = d(S, \Pi_{\ell_e}) = \lfloor \frac{r-b}{2} \rfloor$.

To prove the upper bound, suppose w.l.o.g. that $r \geq b$ (i.e. $b = \min\{r, b\}$). We have to show that $d(S, \Pi_\ell) > b \Rightarrow d(S, \Pi_\ell) \leq \lfloor \frac{r-b}{2} \rfloor$ for every line ℓ. Let ℓ be a line such that $d(S, \Pi_\ell) > b$. Then we have $\nabla'(S_{\ell+}) > 0$ and $\nabla'(S_{\ell-}) > 0$. In fact, suppose that $\nabla'(S_{\ell+}) < 0$; then $\nabla(S_{\ell+}) = |S_{\ell+} \cap B| - |S_{\ell+} \cap R| \leq |S_{\ell+} \cap B| \leq b$. Thus $d(S, \Pi_\ell) \leq b$, a contradiction. Now $\nabla'(S_{\ell+}) > 0$ and $\nabla'(S_{\ell-}) > 0$ imply that $d(S, \Pi_\ell) \leq \lfloor \frac{r-b}{2} \rfloor$. In fact, suppose the contrary; $\nabla(S_{\ell+}) \geq \lfloor \frac{r-b}{2} \rfloor + 1$ and $\nabla(S_{\ell-}) = (r - b) - \nabla(S_{\ell+}) \geq \lfloor \frac{r-b}{2} \rfloor + 1$. Then $r - b \geq 2\lfloor \frac{r-b}{2} \rfloor + 2$, a contradiction. If $\lfloor \frac{r-b}{2} \rfloor \leq b$, the upper bound is tight if we take separable sets R and B. If $\lfloor \frac{r-b}{2} \rfloor > b$, we have shown how to build a set of points S with $\mathcal{C}_2(S) = \lfloor \frac{r-b}{2} \rfloor$. ∎

Corollary 1. *Let $|r - b| \geq 2$. If $r \geq 3b$ or $b \geq 3r$. Then $\mathcal{C}_2(S) = \lfloor \frac{|r-b|}{2} \rfloor$.*

Proof. Suppose that $r \geq 3b$. Then $r - b \geq 2b \Rightarrow \frac{r-b}{2} \geq b \Rightarrow \lfloor \frac{r-b}{2} \rfloor \geq b$. Thus the upper and lower bounds on $\mathcal{C}_2(S)$ in Proposition 5 are equal. ∎

5.1 Hardness

In this subsection we prove that computing the linear coarseness of a bicolored point set is 3SUM-hard [17]. We also give an optimal $O(n^2)$-time algorithm to calculate it.

Theorem 5. *Let d be an integer greater than or equal to 1. Then deciding if $\mathcal{C}_2(S) = d$ is 3SUM-hard.*

Proof. We will use a reduction from the 3SUM-problem similar to the 3SUM-hardness proof of the 3-POINTS-ON-LINE problem [17]. Consider the set $X = \{x_1, \ldots, x_n\}$ of n integer numbers, an instance of the 3SUM-problem, and assume w.l.o.g. that $x_1 < \cdots < x_j < 0 < x_{j+1} < \cdots < x_n$ $(1 \leq j < n)$. Let $M = \max\{|x_1|, |x_n|\}$. If $d = 1$, put a blue point in $(-2M, 0)$ and a red point in $(2M, 0)$. If $d > 2$, then for each $1 \leq i \leq d - 2$ put a red point in $(-2M - i + 1, 0)$ and a blue point in $(2M + i - 1, 0)$. Let ε be a real positive number such that $\varepsilon < \frac{1}{6M^2}$. For each $1 \leq i \leq n$ put a red point p_i in $(x_i - \varepsilon, x_i^3)$ and a blue point q_i in $(x_i + \varepsilon, x_i^3)$; see Figure 10. Since $\varepsilon < \frac{1}{6M^2}$, we can prove that there is a line separating three distinct pairs (p_i, q_i), (p_j, q_j), and (p_k, q_k) if and only if

(x_i, x_i^3), (x_j, x_j^3), and (x_k, x_k^3) are collinear (i.e., $x_i + x_j + x_k = 0$). See Appendix for more detail. Let S be the set of red and blue points as above. We have that $\mathcal{C}_2(S) \geq d$, because $d(S, \Pi_\ell) = d$ for every line ℓ separating exactly two distinct pairs (p_i, q_i) and (p_j, q_j). If $\mathcal{C}_2(S) > d$, then there is a line separating more than two pairs, implying that three elements in X sum to zero. Therefore, three elements in X sum to zero if and only if $\mathcal{C}_2(S) \neq d$. ∎
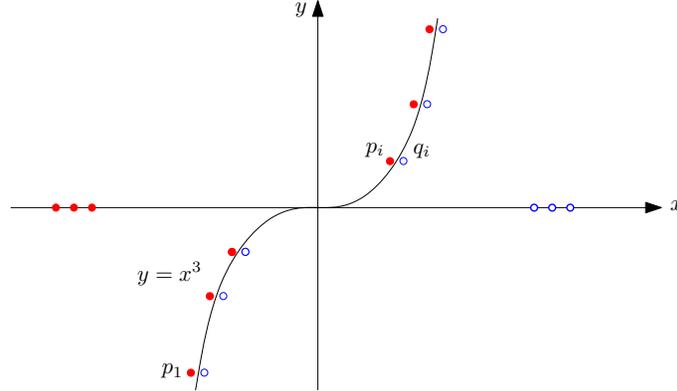

Figure 10: Reduction from the 3SUM-problem when $d = 5$.

Theorem 6. *Computing the linear coarseness of a bicolored point set S is 3SUM-hard and can be done in $O(n^2)$ time.*

Proof. The hardness is due to Theorem 5, and duality can be used to compute $\mathcal{C}_2(S)$. ∎

5.2 The Weak Separator Problem

Given a bichromatic set of points in the plane, the *Weak Separator Problem* (WS-problem) looks for a line that maximizes the sum of the number of blue points on one side of it and the number of red points on the other. The WS-problem can be solved in $O(n^2)$ [20] or in $O(nk \log k + n \log n)$ time, where k is the number of misclassified points [14]. An $O((n + k^2) \log n)$ expected-time algorithm was presented recently in [6]. We prove that the WS-problem is 3SUM-hard.

Lemma 12. *Let $S = R \cup B$ with $r = b$. Solving the WS-problem for S is equivalent to finding a line ℓ such that $d(S, \Pi_\ell) = \mathcal{C}_2(S)$.*

Proof. Let ℓ be any line such that $d(S, \Pi_\ell) = \mathcal{C}_2(S)$. Since $r = b$ then $\nabla'(S_{\ell+}) = -\nabla'(S_{\ell-})$. Suppose w.l.o.g. that $d(S, \Pi_\ell) = \nabla'(S_{\ell+}) = |S_{\ell+} \cap R| - |S_{\ell+} \cap B| > 0$. We have that $|S_{\ell+} \cap R| + |S_{\ell-} \cap B| = |S_{\ell+} \cap R| + |B| - |S_{\ell+} \cap B| = b + |S_{\ell+} \cap R| - |S_{\ell+} \cap B|$. Hence $|S_{\ell+} \cap R| + |S_{\ell-} \cap B|$ is maximum if and only if $|S_{\ell+} \cap R| - |S_{\ell+} \cap B| = \mathcal{C}_2(S)$ is maximum. ∎

Theorem 7. *The WS-problem is 3SUM-hard.*

6 Conclusions

In this paper we have presented a new parameter, coarseness, to measure how blended a bicolored set of points is. Firstly, we introduced a definition of discrepancy for a convex partition of the

bicolored point set and, the concept of coarseness then uses convex partitions of the points to determine if the set admits big blocks in such a way that a color dominates in each of them.

We provided useful combinatorial properties on the coarseness, and a complete characterization if R and B are linearly separable. As an interesting result, it was shown that the coarseness of points in convex position can be computed in $O(n \log n)$ time by using a reduction to instances of problems on circular sequences of weighted elements. The case in which the coarseness is induced by partitions with a straight line was also studied, and we gave exact combinatorial lower and upper bounds on the value of coarseness. Furthermore, we showed that computing this type of coarseness is 3SUM-hard [17]. Additionally, and as a consequence, we proved that the well-known *Weak Separator Problem* [6, 14, 20] is also 3SUM-hard.

In view of the results obtained in this paper, we believe that the problem of finding the coarseness of bicolored sets is NP-hard. An interesting problem is that of finding efficient algorithms to approximate the coarseness.

7   Acknowledgements

References

[1] J.R. Alexander, J. Beck and W.W.L. Chen. *Geometric discrepancy theory and uniform distribution.* Handbook of Discrete and Computational Geometry (second ed.), CRC Press, pp. 185–207, 1997.

[2] P.K. Agarwal and J. Pach. *Combinatorial Geometry.* Wiley-Interscience Series in Discrete Mathematics and Optimization, 16:267–290, 1995.

[3] C. Bautista, J.M. Díaz-Báñez, D. Lara, P. Pérez-Lantero, J. Urrutia and I. Ventura. *Computing maximal islands.* Operations Research Letters, 3(39), pp. 246–251, 2011.

[4] J.L. Bentley. *Programming pearls: algorithm design techniques.* Comm. ACM, Vol. 27, No. 9, pp. 865–873, 1984.

[5] V. Capoyleas, G. Rote, G. Woeginger. *Geometric clusterings.* Journal of Algorithms, 12, 341–356, 1991.

[6] T.M. Chan. *Low-dimensional linear programming with violations.* SIAM Journal on Computing, 34(4), 879–893, 2005.

[7] B. Chazelle. *The discrepancy method in computational geometry.* Handbook of Discrete and Computational Geometry, CRC Press 44, 983–996, 2004.

[8] K.-Y. Chen and K.-M. Chao. *On the range maximum-sum segment query problem.* R. Fleischer and G. Trippen Eds., ISAAC 2004, LNCS 3341, pp. 294–305, 2004.

[9] C. Cortés, J.M. Díaz-Báñez, P. Pérez-Lantero, C. Seara, J. Urrutia, and I. Ventura. *Bichromatic separability with two boxes: a general approach.* Journal of Algorithms, Vol. 64, pp. 79–88, 2009.

[10] J. Czyzowicz, E. Rivera Campo, J. Urrutia and J. Zaks. *Separating convex sets in the plane.* Discrete and Computational Geometry 7:189–195, 1992.

[11] D.P. Dobkin and D. Gunopulos. *Concept learning with geometric hypotheses*. In Proc. 8rd Annu. Conference on Comput. Learning Theory. ACM Press, 1995.

[12] D.P. Dobkin, D. Gunopulos, and W. Maass. *Computing the maximum bichromatic discrepancy, with applications to computer graphics and machine learning*. Journal of Computer and Systems Sciences, 52(3), 453–470, 1996.

[13] A. Dumitrescu, and J. Pach. *Partitioning colored point sets into monochromatic parts*. Int. J. Comput. Geom Appl., 12, no. 5, 401–412, 2002.

[14] H. Everett, J.M. Robert, M. Kreveld. *An optimal algorithm for computing ($\leq k$)-levels, with applications to separation and transversal problems*. Int. J. Comput. Geom. Appl., 6:247–261, 1996.

[15] P. Erdős and G. Szekeres. *A combinatorial problem in geometry*. Compositio Math., 2, 463–470, 1935.

[16] P. Fischer. *Sequential and parallel algorithms for finding a maximum convex polygon*. Computational Geometry: Theory & Applications, 7, 187–200, 1997.

[17] A. Gajentaan and M.H. Overmars. *On a class of $O(n^2)$ problems in computational geometry*. Computational Geometry: Theory & Applications, 5:165–185, 1995.

[18] J. Goodman and J. O'Rourke. Handbook of Discrete and Computational Geometry. Edited by CRC Press, 211, 1997.

[19] J. Han and M. Kamber. Data Mining: Concepts and techniques, Morgan Kaufmann Publishers, USA, 2006.

[20] M.E. Houle. *Algorithms for weak and wide separation of sets*. Proc. Int. Workshop on Discr. Alg. and Comp, 61–68, 1989.

[21] J. Matoušek. Geometric Discrepancy: An Illustrated Guide. Springer-Verlag, 1999.

A   Appendix

Lemma 13. *Let a, b and c be three distinct integers and $M = \max\{|a|, |b|, |c|\}$. Let ε be a real positive number such that $\varepsilon < \frac{1}{6M^2}$. Then there is no line that simultaneously intersects the horizontal segments $[a - \varepsilon, a + \varepsilon] \times a^3$, $[b - \varepsilon, b + \varepsilon] \times b^3$ and $[c - \varepsilon, c + \varepsilon] \times c^3$ unless the points (a, a^3), (b, b^3) and (c, c^3) are collinear.*

Proof. Suppose w.l.o.g. that $a < b < c$. For a given $\varepsilon > 0$, denote the horizontal segments $[a - \varepsilon, a + \varepsilon] \times a^3$, $[b - \varepsilon, b + \varepsilon] \times b^3$, and $[c - \varepsilon, c + \varepsilon] \times c^3$, respectively as s_a, s_b, and s_c. Let $\delta(b, \overline{ac})$ be the horizontal distance from (b, b^3) to the line through (a, a^3) and (c, c^3). Then we have:

$$
\begin{aligned}
\delta(b, \overline{ac}) &= \left| b - \left((b^3 - a^3) \frac{c - a}{c^3 - a^3} + a \right) \right| \\
&= \left| b - a - \frac{(b - a)(b^2 + ab + a^2)}{c^2 + ac + a^2} \right| \\
&= \left| (b - a) \left(1 - \frac{b^2 + ab + a^2}{c^2 + ac + a^2} \right) \right| \\
&= \left| (b - a) \left(\frac{c^2 - b^2 + ac - ab}{c^2 + ac + a^2} \right) \right| \\
&= \left| \frac{(b - a)(c - b)(a + b + c)}{c^2 + ac + a^2} \right| \\
&= (b - a)(c - b) \frac{|a + b + c|}{|c^2 + ac + a^2|}.
\end{aligned}
$$

If $a + b + c = 0$ then $\delta(b, \overline{ac}) = 0$, and thus (a, a^3), (b, b^3), and (c, c^3) are collinear, and for all $\varepsilon > 0$ the line through them intersects the segments s_a, s_b, and s_c.

Now suppose that $a + b + c \neq 0$ (i.e., $|a + b + c| \geq 1$). Since $a < b < c$, we have that $b - a \geq 1$ and that $c - b \geq 1$. Therefore

$$
\delta(b, \overline{ac}) \geq \frac{1}{|c^2 + ac + a^2|} \geq \frac{1}{|c|^2 + |a||c| + |a|^2} \geq \frac{1}{3M^2}.
$$

Note that for a given $\varepsilon > 0$ there is no line that intersects s_a, s_b, and s_c if and only if $2\varepsilon < \delta(b, \overline{ac})$. This can be ensured if $\varepsilon < \frac{1}{6M^2}$. Hence the result follows. $\blacksquare$