Bichromatic Discrepancy via Convex Partitions

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

21st October 2009

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 for measuring how mixed the elements of \( S = R \cup B \) are. The discrepancy of a set \( X \subseteq S \) is \( ||X \cap R| - |X \cap B|| \). 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 of \( S \) is the minimum discrepancy of the sets \( S_i \). The discrepancy of \( S \) is the discrepancy of the convex partition of \( S \) with maximum discrepancy. We study the problem of computing the discrepancy of a bichromatic point set. We divide the study in general convex partitions for both general set of points and points in convex position, and also when the partition is given by a line. In this case we prove that this problem is 3SUM-hard.

1 Introduction

NEW!

In database management systems, clustering is often an initial stage for data classification [11]. Suppose you have a set of points classified according to two colors and you want to know if it is possible to divide the set into big monochromatic groups. In that case, you could run a clustering procedure to, for example, use the clusters as a training set in data classification. However, it is not possible if the colored points are mixed or uniformly distributed, and then we could say that the given data set is not suitable for getting a training set.

END NEW!

In this paper we study a parameter that measures how mixed two point sets are. In the rest of this paper \( S = R \cup B \) will always be a set of points on the plane in general position whose elements are colored either red (the elements of \( R \)), or blue (the elements of \( B \)). We will also assume that \( R \) and \( B \) are non-empty and have \( r \) and \( b \) elements respectively. Intuitively speaking \( R \) and \( B \) are well

\*University of Texas at Dallas, besp@utdallas.edu
†Departamento Matemática Aplicada II, Universidad de Sevilla, dbanez@us.es, Partially supported by Grant MEC MTM2006-03909
‡Instituto de Matemáticas, Universidad Nacional Autónoma de México, dlara@uxmcc2.iimas.unam.mx
§Departamento de Computación, Universidad de La Habana, pablo@matcom.uh.cu, Partially supported by Grant MAEC-AECI and MEC MTM2006-03909
¶Departament de Matemàtica Aplicada II, Universitat Politècnica de Catalunya, carlos.seara@upc.edu, Partially supported by Grants MEC MTM2006-01267 and DURSI 2005SGR00692
∥Instituto de Matemáticas, Universidad Nacional Autónoma de México, urrutia@matem.unam.mx, Partially supported by Grant MEC MTM2006-03909
mixed if for any convex region \( C \) of the plane the proportion of red elements of \( S \) is approximately \( \frac{r}{b+r} \), e.g. if \( r = 2b \), we would expect \( C \) to contain twice as many red points as blue. It is clear that we must be careful on how we define well mixed sets of points, as since \( S \) is in general position, we can always find many convex sets containing only two points of \( S \) with the same color. In this paper we introduce a parameter that seems like a good candidate to measure how well mixed a bicolored point set is, we call this parameter the discrepancy of \( S \).

For any set of points \( P \) on the plane, let \( CH(P) \) denote the convex hull of \( P \). Let \( S = R \cup B \) be a bicolored point set, and \( X \subseteq S \). The discrepancy of \( X \) is defined as \( \nabla(X) = ||X \cap R| - |X \cap B|| \).

We say that a partition \( \Pi = \{S_1, S_2, \ldots, S_k\} \) of \( S \) is a convex partition if \( CH(S_i) \cap CH(S_j) = \emptyset \) for all \( 1 \leq i < j \leq k \). The discrepancy of \( S \) with respect to \( \Pi \) is defined as \( d(S, \Pi) = \min_{i=1, \ldots, k} \nabla(S_i) \).

The discrepancy of \( S \), \( d(S) \), is defined as the largest \( d(S, \Pi) \) over all the convex partitions \( \Pi \) of \( S \).

Notice that if the discrepancy of a set is large, then there is a convex partition \( \Pi \) of \( S \) in which each element of \( \Pi \) has large discrepancy. If \( d(S) = 1 \) any convex partitioning of \( S \) has at least one element with discrepancy one. For example if \( S = R \cup B \) is separable, that is there is a line \( \ell \) that leaves all the elements of \( R \) on one of the half-planes it determines, and all the elements of \( B \) on the other, then the discrepancy of \( S \) is at least the minimum of \( r \) and \( b \).

If we restrict ourselves to convex partitions of \( S \) with exactly \( k \) elements, we obtain the \( k \)-discrepancy of \( S \), which will be denoted as \( d_k(S) \). When \( k = 1 \) then the partition \( \Pi \) has only one element, and thus \( d_1(S) = \nabla(S) = |r - b| \). If \( k = 2 \) then we have what we call linear discrepancy, that is the discrepancy obtained by partitions of \( S \) induced by lines that split \( S \) into two subsets.

Our concept of discrepancy has applications in Data Analysis and Clustering in sets of data of two classes, say red and blue. We can state that a red-blue dataset is not good for clustering when its discrepancy is low. Hence our concept can be used as a priori tester to a dataset for clustering. The extreme case is when \( d(S) = 1 \), in this case we say that \( S \) is locally balanced. Many of the results of this paper focus on the hardness of deciding if a bicolored point set is locally balanced or not.

The discrepancy between two objects is a measure of how different the objects are. In \([1, 13]\) authors study what is known as Combinatorial Discrepancy that is a concept of discrepancy for hypergraphs, that is, the problem of assigning weight \( +1 \) or \( -1 \) to the vertices of a given hypergraph in such a way the maximum weight of an edge (i.e. the absolute value of the sum of the weights of its vertices) is minimized. NEW! Another concept of discrepancy can be found in \([13]\), dealing with the “most uniform” way of distributing \( n \) points in the unit square according to some criteria.

END NEW! Geometric Discrepancy Theory \([4]\) studies how uniform nonrandom structures can be. For example, how to color \( n \) points in the plane to minimize the difference between the number of red points and the number of blue ones within any disk. In \([2, 6, 7]\) the concept of bichromatic discrepancy is considered by computing the object (e.g. box, triangle, strip, convex polygon, etc.) of maximum absolute difference between red and blue points inside it. In this paper we introduce a new parameter to measure the discrepancy of bicolored point sets.

The outline of this work is as follows. In section 2 we provide combinatoric results for measuring discrepancy for general convex partitions. We consider two cases, when the points are in convex position and when they are not. In section 3 we give combinatoric and hardness proofs on measuring discrepancy by using partitions by a line.
2 General Convex partitions

NEW!

In this section we address the problem of computing the discrepancy by using a general convex partition. A point set $P$ is in convex position if the elements of $P$ are the vertices of a convex polygon. We consider two cases, points in convex position and points in general position.

We begin our study by giving some useful properties that allow us to know the value of discrepancy for several special cases. Before that, we introduce some notation for a better comprehension of the proofs. Given a convex polygon $Q$, we say that a set of points $P$ is $Q$-consecutive if the elements of $P$ are vertices of $Q$ and they are consecutive along the boundary of $Q$. Given a bicolored point set $X$ with $r'$ red points, and $b'$ blue points, let $\nabla'(X) = r' - b'$. Furthermore, we say that $X$ is red (resp. blue) if $r' > b'$ (resp. $b' > r'$). Observe that $\nabla(X) = |\nabla'(X)|$. Additionally, if $\Pi = \{S_1, S_2, \ldots, S_k\}$ is a convex partition of $S$, we say that $\Pi$ is optimal if $d(S) = d(S, \Pi)$, and denote $r_i = |S_i \cap R|$ and $b_i = |S_i \cap B|$ for $i = 1 \ldots k$.

Lemma 2.1 Let $S = R \cup B$ then, $d(S) \geq 1$. Moreover, if $d(S) = 1$ then $|r - b| \leq 1$.

Proof. Suppose that $S = \{p_1, p_2, \ldots, p_{r+b}\}$ and let $\Pi = \{\{p_1\}, \{p_2\}, \ldots, \{p_{r+b}\}\}$. We have $d(S) \geq d(S, \Pi) = 1$. Moreover, if $d(S) = 1$ then $|r - b| = d_1(S) \leq d(S) = 1$.

Lemma 2.2 If $\Pi = \{S_1, S_2, \ldots, S_k\}$ is an optimal convex partition of $S$ such that $k \geq 2$, then there exist different indexes $i$ and $j$ such that $S_i$ is red and $S_j$ is blue.

Proof. Suppose that $S_i$ is blue for every index $i$ in the optimal convex partition $\Pi = \{S_1, S_2, \ldots, S_k\}$. Then $d_1(S) = b - r = \sum_{i=1}^{k} (b_i - r_i) = \sum_{i=1}^{k} \nabla(S_i) > \min_{i=1 \ldots k} \nabla(S_i) = d(S, \Pi) = d(S)$, which is a contradiction because $d_1(S) \leq d(S)$. Analogously we prove that $S_i$ cannot be red for every index $i$. The result follows.

Lemma 2.3 If $\Pi = \{S_1, S_2, \ldots, S_k\}$ is a convex partition of $S$ containing a red element and a blue element, then $d(S, \Pi) \leq \min\{r, b\}$.

Proof. Suppose w.l.o.g. that $S_1$ is red and $S_2$ is 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 2.4 If $r \geq 2b$ or $b \geq 2r$ then $d(S) = d_1(S) = |r - b|$.

Proof. Assume w.l.o.g. that $r \geq 2b$. By definition we have that $d_1(S) \leq d(S)$. Suppose now that $\Pi$ is an optimal convex partition of $S$ with cardinality bigger than one. By Lemma 2.2, $\Pi$ contains a red element and a blue element, thus $d(S, \Pi) \leq \min\{r, b\}$ by Lemma 2.3. Then $d(S) = d(S, \Pi) \leq \min\{r, b\} = b \leq r - b = d_1(S)$. This implies that $d(S) = d_1(S)$.

Lemma 2.5 Let $S = R \cup B$. If $R$ and $B$ are linearly separable, and $b \leq r < 2b$ or $r \leq b < 2r$, then $d(S) = \min\{r, b\}$.
Proof. Suppose w.l.o.g. that \( b \leq r < 2b \) and let \( \Pi \) be an optimal convex partition of \( S \). \( \Pi \) cannot have cardinality one because \( d_1(S) = r - b < b = d(S, \{R, B\}) \). Therefore, by Lemma 2.2, \( \Pi \) has a red element and a blue element implying, by Lemma 2.3, that \( d(S) = d(S, \Pi) \leq \min\{r, b\} = b \). Since \( d(S, \{R, B\}) = b \) then \( d(S) = b \).

From the two lemmas above we can conclude that if \( S \) is linearly separable then \( d(S) = \min\{r, b\} \) if the cardinality of the majority color in \( S \) does not double the minority one, and \( d(S) = |r - b| \), otherwise.

END NEW!

2.1 Point Sets in Convex Position

NEW!

In this subsection we show that the discrepancy can be easily determined when the points are in convex position. The following proposition gives us a simple property of convex partitions of point sets in convex position.

**Proposition 2.6** If \( S \) is in convex position then in any convex partition \( \Pi = \{S_1, S_2, \ldots, S_k\} \) of \( S \), there is at least one element \( S_i \) (1 \( \leq i \leq k \)) such that its elements are \( \text{CH}(S) \)-consecutive.

**Proof.** 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 \( \text{CH}(S) \)-consecutive. Then \( S \setminus S_1 \) is composed by at least two maximal \( \text{CH}(S) \)-consecutive chains. Let \( C \) be one of those chains and let \( \Pi_C = \{S_i \mid S_i \cap C \neq \emptyset\} \). It is clear that \( \Pi_C \subset \Pi \) and that \( \Pi_C \) is a convex partition of \( C \). By the induction hypothesis there is at least one \( S_i \in \Pi_C \) such that its elements are \( \text{CH}(C) \)-consecutive, and by definition of \( C \) they are also \( \text{CH}(S) \)-consecutive. \( \square \)

END NEW!

**Definition 2.7** Suppose that \( S \) is in convex position and that \( |r - b| \leq 1 \). We say that \( S \) is an alternating convex chain if we can label its elements \( p_1, p_2, \ldots, p_{r+b} \) in the counterclockwise order around \( \text{CH}(S) \) such that, for every \( 1 \leq i \leq r + b - 1 \), \( p_i \) and \( p_{i+1} \) have different color.

**Lemma 2.8** If \( S \) is in convex position then \( d(S) = 1 \) if and only if \( S \) is an alternating convex chain.

**Proof.** Suppose that \( d(S) = 1 \) and that \( S \) is not an alternating convex chain. By Lemma 2.1 \( |r - b| \leq 1 \). If \( r = b \) and for some \( i \) we have that \( p_i \) and \( p_{i+1} \) have the same color, the partition \( \Pi = \{S_1, S \setminus S_1\} \) where \( S_1 = \{p_i, p_{i+1}\} \) has discrepancy 2. If \( r = b + 1 \) and there is an \( i \) such that \( p_i \) and \( p_{i+1} \) are red points, then if \( S_1 = \{p_i, p_{i+1}\}, \nabla(S_1) = 2 \), \( \nabla(S \setminus S_1) = 3 \), and the partition \( \Pi = \{S_1, S \setminus S_1\} \) is such that \( d(S) \geq d(S, \Pi) = 2 \). It follows by contradiction that \( S \) is an alternating convex chain.

Suppose now that \( S \) is an alternating convex chain. Let \( \Pi = \{S_1, S_2, \ldots, S_k\} \) be any convex partition of \( S \). By Proposition 2.6 there is at least one \( S_i \) (1 \( \leq i \leq k \)) that is \( \text{CH}(S) \)-consecutive and thus an alternating convex chain. Then \( \nabla(S_i) \leq 1 \) implying that \( d(S, \Pi) \leq 1 \) for all \( \Pi \). Therefore \( d(S) = 1 \). \( \square \)
Theorem 2.9 If $S$ is in convex position then $d(S) = \max_{k=1,2,3} d_k(S)$.

Proof. Let $d = d(S)$, observe that in particular $0 \leq \nabla(S) \leq d$. Assume w.l.o.g. that $0 \leq \nabla'(S) \leq d$. Let $\Pi = \{S_1, S_2, \ldots, S_k\}$ be a convex partition of $S$ with minimum cardinality such that $d(S, \Pi) = d$. By definition, $\nabla(S_i) \geq d$ $(1 \leq i \leq k)$. Suppose that $k > 3$. Then $S$ has at least two elements, say $S_1$ and $S_2$, such that both of them contain only $CH(S)$-consecutive elements of $S$.

If any 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 because 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 $d(S, \Pi'') \geq d$. \hfill \square

This implies:

Theorem 2.10 The discrepancy of a bicolored point set in convex position can be computed in polynomial time.

2.2 Point Sets in General Position

NEW!

The problem of deciding the discrepancy of point sets in general position seems to be non-trivial. At this point, we are unable even to characterize point sets with discrepancy one.

In this section we explore bicolored point sets with discrepancy equal to one (locally balanced) and study the discrepancy of special configurations.

END NEW!

The following lemmas are straightforward.

Proposition 2.11 If $S$ has at most four elements and $d(S) = 1$ then $S$ is an alternating convex chain.

Proposition 2.12 There are two combinatorially different point configurations of a bichromatic point set $S$ with five points such that $d(S) = 1$ (Figure 1).

We now show configurations of points in general position that are locally balanced.

Proposition 2.13 For all $n \geq 4$ there are bichromatic point sets of size $n$, not in convex position, with $d(S) = 1$.

Proof. The proof is based on the following constructions. For $n = 2m+1$ let $Q$ be a regular convex polygon with $2m$ vertices such that their colors alternate blue and red along the boundary of $Q$. Let $S$ be the set of vertices of $Q$ plus a red point $p$ close to the center of $Q$ (Figure 2 a)). 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 $r \notin S_i$ and $S_i$ contains a set of consecutive vertices of $Q$. Then $\nabla(S_i) \leq 1$ and therefore $d(S, \Pi) \leq 1$. 

5
If $n = 2m + 2$, place two points $p$ and $q$ in the interior of $Q$ such that $p$ and $q$ are close enough to the middle of an edge $e$ of $Q$, and the line joining them is almost parallel to $e$. It is easy to see now that $d_2(S) = 1$ and that $d(S, \Pi) \leq 1$ for all the convex partitions $\Pi$ of $S$ (Figure 2 b)).

Figure 1: Two different point configurations with five points with $d(S) = 1$.

Constructing point sets with large discrepancy is straightforward. As we already mentioned in the introduction of this paper, if $R$ and $B$ are linearly separable then $d(S) \geq \min\{r, b\}$.

NEW!

Nevertheless the next result shows that although fixing the number of red and blue points, every value of discrepancy can be achieved:

**Proposition 2.14** If $b \leq r < 2b$ or $r \leq b < 2r$, then for every integer value $d$ such that $\max\{1, |r - b|\} < d \leq \min\{r, b\}$ there exists a set of $r$ red points and $b$ blue points not in convex position with discrepancy equal to $d$.

**Proof.** Let $d$ be an integer such that $\max\{1, |r - b|\} < d \leq \min\{r, b\}$ and suppose w.l.o.g. that $b \leq r \leq 2b$. Draw a circle $c$ centered at $o$ and let $\alpha$ be an arc of $c$ with amplitude $\frac{\pi}{2}$. Let $m$ be the point such that the midpoint of the segment $om$ coincides with the midpoint of $\alpha$. Put around $o$ (resp. $m$) and very close to $o$ (resp. to $m$) a set $X_1$ (resp. $X_2$) of $r - b + d - 2$ red points (resp. $d - 2$ blue points). Let $X_3$ be the set that results of distributing uniformly, along $\alpha$, $b - d + 2$ pairs of a red point $p$ and a blue point $q$ such that: $p$ and $q$ are very close to each other, $p$ is inside $c$, $q$ is outside $c$, and every line separates at most two pairs (Figure 3).

Let $S = X_1 \cup X_2 \cup X_3$. Now, if a convex set $Q$ is such that $S \cap Q$ is blue then $\nabla(S \cap Q) \leq d$. In fact, suppose $Q$ contains exactly $h$ blue points of $X_3$, then $Q$ includes at least $h - 2$ red points of $X_3$ and hence $\nabla(S \cap Q) \leq h - (h - 2) + (d - 2) = d$.

Figure 2: Point sets with $n$ points and $d(S) = 1$. a) $n$ is odd, b) $n$ is even.
Suppose that

\begin{equation}
\text{Proof.}
\end{equation}

\begin{proposition}
\label{pro:2.16}
\text{Proposition 2.16}
\end{proposition}

\begin{proof}
\text{Proof.}
\end{proof}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure3.png}
\caption{A bicolored point set that is not in convex position, have $r$ red points and $b$ blue points and its discrepancy is equal to $d$, where $d$ is such that $\max\{1, |r-b|\} < d \leq \min\{r, b\}$.}
\end{figure}

We discard to consider $d_1(S)$ because $d > r - b$, and we have by Lemma 2.2 that in any optimal convex partition $\Pi = \{S_1, S_2, \ldots, S_k\}$ of $S$ ($k \geq 2$) there is at least one blue $S_i$. Since $S_i = S \cap CH(S_i)$ then $\nabla(S_i) \leq d$ implying that $d(S, \Pi) \leq d$. Moreover, if the line $\ell$ is such that $\ell$ separates exactly two pairs of $X_3$ and splits $S$ into two subsets $S'$ and $S''$, then $d(S, \{S', S''\}) = \min\{r - b + d, d\} = d$. Hence $d(S) = d$.

Observe that in the above configuration there exists a circle separating the given bichromatic set and the “goodness” of the monochromatic clustering by using convex sets depends of the value of discrepancy $d$. The following results deal with nested even alternating convex chains. We say that an alternating convex chain is even if it contains an even number of points (Figure 4 a)). This kind of configuration allows to give examples with any value of discrepancy and whose points are “well mixed”.

\begin{proposition}
\label{pro:2.15}
\text{Proposition 2.15}
\end{proposition}

\begin{proof}
\text{Proof.}
\end{proof}

\begin{proposition}
\label{pro:2.16}
\text{Proposition 2.16}
\end{proposition}

\begin{proof}
\text{Proof.}
\end{proof}
Now we prove that the equality can hold. Let $m \geq 1$. A $m$-chessboard is a set $S$ of $2m^2$ red points and $2m^2$ blue points in the plane located at the centers of the cells of some $2m \times 2m$ squared regular grid $G_S$ and such that no two edge-adjacent cells of $G_S$ have centers with the same color. If $S$ is a $t$-chessboard then it is formed by $t$ nested even alternating convex chains, and thus $d(S) \leq t$. Let $\ell$ be a line that passes through a diagonal of $G_S$ and leaves the centers of the diagonal to the same half-plane defined by $\ell$ (Figure 4 b)). Then $\ell$ splits $S$ into two subsets $S'$ and $S''$ such that $d(S, \{S', S''\}) = t$. Thus $d(S) = t$.

![Figure 4: a) A configuration of 3 nested even alternating convex chains, b) a 3-chessboard and a line $\ell$ giving a partition with discrepancy 3.](image)

![Figure 5: A generalization of the chessboard.](image)

The idea of the chessboard can be generalized as in Figure 5 where $t$ nested even alternating chains can achieve different value of discrepancy. Note that there exists a line cutting the chains such that in each halfplane every chain has the same color (red or blue). We include some cases here. For instance, if $t$ is a constant then the discrepancy is $t$. If $S$ is composed by $t$ even alternating convex chains with $4t$ points each, then $d(S) = t = \sqrt[4t]{2} = \sqrt{t}$. If $S$ is a set of $n = 2^{2m}$ ($m \geq 1$) points distributed according $2^m = \log_2 n$ even alternating convex chains of length $2^{2m} - m = \frac{n}{\log_2 n}$ each, then $d(S) = \log_2 n$.

**END NEW!**

### 3 Partitions with a line

In this section we characterize sets with linear discrepancy one and show how to decide if the linear discrepancy of a bicolored point set $S$ is equal to $d$. We introduce the following notation.
Let $\Pi_{\ell^+}$ and $\Pi_{\ell^-}$ be the open half-planes bounded below and above respectively by a non vertical line $\ell$. Let $S_{\ell^+} = S \cap \Pi_{\ell^+}, S_{\ell^-} = S \cap \Pi_{\ell^-}$ and $\Pi_\ell = \{S_{\ell^+}, S_{\ell^-}\}$. The linear discrepancy of $S$ is $d_2(S) = \max_\ell d(S, \Pi_\ell)$ where the lines $\ell$ contain no point in $S$.

**Proposition 3.1** Let $S = R \cup B$ such that $r = b$ and $d_2(S) = 1$. Then the following properties hold:

1. The convex hull of $S$ is an alternating chain.
2. When projected on any line, the points of $S$ form a sequence such that no three consecutive points have the same color.
3. For every $p \in S$ 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 $\ell$ 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.

Property 2 in Proposition 3.1 is not sufficient to guarantee that $d(S) = 1$, e.g. see Figure 6 a). If $r \neq b$ properties 3 and 4 are not necessarily true, see Figure 6 b) and c). We now show that if $r = b$, Property 4 is sufficient.

The next result, proven in [5] will be useful:

**Theorem 3.2** Let $P$ and $Q$ be two disjoint convex polygons on the plane. Then there is at least one edge $e$ of $P$ or $Q$ such that the line $\ell_e$ containing $e$ separates the interior of $P$ from the interior of $Q$.

**Lemma 3.3** If $r = b$ then the following two conditions are equivalent: (i) $d_2(S) = 1$, (ii) for every line $\ell$ passing through two points of $S$ with the same color $\nabla(S_{\ell^+}) = \nabla(S_{\ell^-}) = 1$.

**Proof.** It is easy to prove that (i) implies (ii). We show here that (ii) implies (i). Suppose that $d_2(S) = d \geq 2$. We now show that there exists a line $\ell$ containing two points of the same color of $S$ such that $\{\nabla(S_{\ell^-}), \nabla(S_{\ell^+})\} = \{d, d - 2\}$.
Let $\ell_0$ be a line containing no elements of $S$ such that $d_2(S) = d(S, \Pi_{\ell_0}) = d$. Assume w.l.o.g. that $\ell_0$ is horizontal. Since $r = b$ we have that $d_2(S) = d(S, \Pi_{\ell_0}) = \nabla(S_{\ell_0^+}) = \nabla(S_{\ell_0^-}) = d$ 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 red and $S_{\ell_0^-}$ is 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 $\ell'$ passing trough $p$ that separates $P$ from $Q$. Then $p$ must be a red point, for otherwise by translating $\ell'$ up by a small distance, we obtain a partitioning $\Pi'$ of $S$ with discrepancy $d + 1$. Similarly any point $q$ in $Q$ such that there is a line trough $q$ that separates $P$ from $Q$ must be blue.

By Theorem 3.2 there is an edge $e$ of $P$ or $Q$, with vertices $p$ and $q$, such that the line $\ell_e$ containing $e$ separates $P$ from $Q$. If $e$ is an edge of $P$ then $p$ and $q$ are red by using the above observation. Thus we have that $\{\nabla(S_{\ell_e^+}), \nabla(S_{\ell_e^-})\} = \{d, d - 2\}$. A symmetric argument works when $e$ belongs to $Q$.

\[ \square \]

**Proposition 3.4** Let $S = R \cup B$, then we have,

\[
\max \left\{ 1, \left\lfloor \frac{|r - b|}{2} \right\rfloor \right\} \leq d_2(R \cup B) \leq \max \left\{ \left\lfloor \frac{|r - b|}{2} \right\rfloor, \min \{r, b\} \right\}.
\]

Furthermore, both bounds are tight.

**Proof.** Suppose w.l.o.g. that $r \geq b$. By the Ham Sandwich Cut Theorem [10] there exists a line $\ell$ passing 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 $b$. We only show the case where $r$ and $b$ are even. The other cases can be solved in a similar way.

If $r = 2a$ and $b = 2c$ then $\ell$ contains no point of $S$, and $\nabla(S_{\ell^+}) = \nabla(S_{\ell^-}) = a - c = \lfloor \frac{r - b}{2} \rfloor$. Thus $\lfloor \frac{r - b}{2} \rfloor = d(S, \Pi_{\ell}) \leq d_2(S)$.

If $|r - b| \geq 2$ then $d_2(S) \geq \lfloor \frac{|r - b|}{2} \rfloor \geq 1$ thus it is missing to prove that $d_2(S) \geq 1$ when $|r - b| \leq 1$. Suppose w.l.o.g. that $b \leq r \leq b + 1$. If there is a blue point $p$ in the convex hull of $S$ take a line $\ell$ 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 $d_2(S) \geq d(S, \Pi_{\ell}) = 1$. If no such point exists then there are two consecutive red points $p$ and $q$ in the convex hull of $S$, then take a line $\ell$ 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 $d_2(S) \geq d(S, \Pi_{\ell}) = 1$. This proves the lower bound.

We now show that this lower bound is tight. Suppose w.l.o.g. that $r > b$ and let $X$ be a set composed by $r$ 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 there is a line $\ell_e$ such that $d(X, \Pi_{\ell_e}) = 0$ and $\ell_e$ splits $Y$ into two subsets of cardinality $\lfloor \frac{|Y|}{2} \rfloor$ and $\lfloor \frac{|Y|}{2} \rfloor$ 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 $\ell$ we have that $d(X, \Pi_{\ell}) \in \{0, 1\}$ (by Lemma 2.8). 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 $\ell$ splits $Y$ in $x$ and $(r - b) - x$ points respectively. It is easy to prove that $\min\{x - 1, (r - b) - x + 1\} \leq \lfloor \frac{b - r}{2} \rfloor$. Then $d_2(S) = d(S, \Pi_{\ell}) = \lfloor \frac{b - r}{2} \rfloor$.

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 $d(S, \Pi_{\ell}) > b \Rightarrow d(S, \Pi_{\ell}) \leq \lfloor \frac{b - r}{2} \rfloor$ for every line $\ell$. Let $\ell$ be a line such that $d(S, \Pi_{\ell}) > b$. Then we have that $\nabla'(S_{\ell^+}) > 0$ and $\nabla'(S_{\ell^-}) > 0$. In fact, suppose that $\nabla'(S_{\ell^+}) < 0$, then $\nabla(S_{\ell^+}) = |S_{\ell^+} \cap R| - |S_{\ell^+} \cap B| \leq |S_{\ell^+} \cap R| \\ b \Rightarrow d(S, \Pi_{\ell}) \leq b$, a contradiction. Now, $\nabla'(S_{\ell^+}) > 0$ and $\nabla'(S_{\ell^-}) > 0$ imply that $d(S, \Pi_{\ell}) \leq \lfloor \frac{b - r}{2} \rfloor$. In fact, suppose the contrary, $\nabla(S_{\ell^+}) \geq \lfloor \frac{b - r}{2} \rfloor + 1$.
and \(\nabla(S_{\ell^-}) = (r - b) - \nabla(S_{\ell^+}) \geq \lfloor \frac{r-b}{2} \rfloor + 1\), thus \(r - b \geq 2\lfloor \frac{r-b}{2} \rfloor + 2\), 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 above how to build a set of points \(S\) with \(d_2(S) = \lfloor \frac{r-b}{2} \rfloor\).

\[\square\]

**Corollary 3.5** Let \(S = R \cup B\) such that \(|r - b| \geq 2\). If \(r \geq 3b\) or \(b \geq 3r\) then \(d_2(R \cup B) = \lfloor \frac{r-b}{2} \rfloor\).

**Proof.** Suppose that \(r \geq 3b\), then \(r - b \geq 2b = \frac{r-b}{2} \geq b = \lfloor \frac{r-b}{2} \rfloor \geq b\). Thus the upper and lower bounds of \(d_2(R \cup B)\) in Proposition 3.4 are equal.

\[\square\]

**3.1 Hardness**

**NEW!**

We start with a technical lemma whose proof is in the Appendix.

**Lemma 3.6** Let \(a, b\) and \(c\) be three distinct integers and \(M = \max\{|a|, |b|, |c|\}\). Let \(\varepsilon\) be a real positive value 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.

**END NEW!**

**Theorem 3.7** Given an integer \(d \geq 1\) it is 3SUM-hard to decide if \(d_2(S) = d\).

**Proof.** We will use a reduction from the 3SUM-problem similar to the 3SUM-hardness proof of the 3-POINTS-ON-LINE-problem [9]. Consider the set \(X = \{x_1, \ldots, x_n\}\) of \(n\) integer numbers (positive’s and negative’s) an instance of 3SUM-problem and assume w.l.o.g. that \(x_1 < \ldots < x_j < 0 < x_{j+1} < \ldots < x_n\) \((1 \leq j < n)\). 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 \(\varepsilon\) be a small 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 \((x_i + \varepsilon, x_i^3)\) (Figure 7). Since \(\varepsilon < \frac{1}{6M^2}\) we obtain by Lemma 3.6 that there exists 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. \(x_i + x_j + x_k = 0\)). Let \(S\) be the set of red an blue points as above. We have that \(d_2(S) \geq d\) because \(d_2(S, \Pi) = d\) for every line \(\ell\) separating exactly two distinct pairs \((p_i, q_i)\) and \((p_j, q_j)\). If \(d_2(S) > d\) 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 \(d_2(S) \neq d\).

\[\square\]

**Theorem 3.8** Computing the linear discrepancy of a bichromatic point set is 3SUM-hard and it can be done in \(O(n^2)\) time.

**Proof.** The hardness is an implication of Theorem 3.7 and we can use duality for computing \(d_2(S)\).

\[\square\]
3.2 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 number of blue points on one of its sides plus the number of the red ones on the other. The WS-problem can be solved in $O(n^2)$ time [12] or in $O(nk \log k + n \log n)$ time [8] where $k$ is the number of misclassified points. Recently an $O((n + k^2) \log n)$ expected time algorithm has been presented in [3]. We now prove that the WS-problem is 3SUM-hard.

Lemma 3.9 Let $S = R \cup B$ such that $r = b$. Solving the WS-problem for $S$ is equivalent to finding a line $\ell$ such that $d(S, \Pi_\ell) = d_2(S)$.

Proof. Let $\ell$ be any line such that $d(S, \Pi_\ell) = d_2(S)$. Since $r = b$ then $\nabla'(S_{\ell^+}) = -\nabla'(S_{\ell^-})$. Suppose w.l.o.g. that $d_2(S, \Pi_\ell) = \nabla'(S_{\ell^+}) = |S_{\ell^+} \cap R| - |S_{\ell^+} \cap B| > 0$. 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 $|S_{\ell^+} \cap R| - |S_{\ell^+} \cap B| = d_2(S)$ is maximum. \hfill $\Box$

The next result follows:

Theorem 3.10 The WS-problem is 3SUM-hard.

4 Appendix

Proof of Lemma 3.6

Suppose w.l.o.g. that $a < b < c$. For a given $\varepsilon > 0$ denote respectively as $s_a$, $s_b$ and $s_c$ the horizontal segments $[a - \varepsilon, a + \varepsilon] \times a^3$, $[b - \varepsilon, b + \varepsilon] \times b^3$ and $[c - \varepsilon, c + \varepsilon] \times c^3$. Let $\delta(b, \overline{cc})$ be the horizontal distance from $(b, b^3)$ to the line trough $(a, a^3)$ and $(c, c^3)$, then we have:
\[ \delta(b, \overline{ac}) = \left| b - \left( (b^3 - a^3) \frac{c-a}{c^3 - a^3} + a \right) \right| \]
\[ = b - a - \frac{(b-a)(b^2 + ab + a^2)}{c^2 + ac + a^2} \]
\[ = (b-a) \left( 1 - \frac{b^2 + ab + a^2}{c^2 + ac + a^2} \right) \]
\[ = (b-a) \frac{c^2 - b^2 + ac - ab}{c^2 + ac + a^2} \]
\[ = \frac{(b-a)(c-b)(a+b+c)}{c^2 + ac + a^2} \]
\[ = (b-a)(c-b) \frac{|a+b+c|}{|c^2 + ac + a^2|} \]

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 \( c-b \geq 1 \). 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.

5 Acknowledgements

The problems studied here, were introduced and partially solved during the Fourth Spanish Workshop on Geometric Optimization, June 2008, El Rocío, Huelva, Spain. The authors would like to thank other workshop participants for helpful comments.

References


