Separated matchings on colored convex sets *

Viola Mészáros

Bolyai Institute, University of Szeged, 
Aradi vétanuk tere 1, 6720 Szeged, Hungary; viola@math.u-szeged.hu

Abstract

Erdős posed the following problem. Consider an equicolored point set of $2n$ points, $n$ points red and $n$ points blue, in the plane in convex position. We estimate the minimal number of points on the longest noncrossing path such that edges join points of different color and are straight line segments. The upper bound $\frac{4}{3}n + O(\sqrt{n})$ is proved [7], [5] and is conjectured to be tight. The best known lower bound is $n + \Omega(\sqrt{n})$ [5].

A separated matching is a matchings where no two edges cross geometrically and all edges can be crossed by a line. Here we give a class of configurations that allows at most $\frac{4}{3}n + O(\sqrt{n})$ points in the maximum separated matching. This underlines the importance of the separated matching conjecture [7], [5]. We also present a type of coloring such that the optimal coloring allows at most $\frac{4}{3}n + O(\sqrt{n})$ points in maximum separated matching. On the other hand, if the discrepancy (that is, the maximum difference in cardinality of color classes in any interval of consecutive points) is two or three, we show that the number of vertices in the maximum separated matching is at least $\frac{4}{3}n$.

1 Introduction

Erdős posed the following problem. There are $2n$ points in the plane in convex position. Without loss of generality we may assume that the points are on a circle $C$. An equicoloring of the points is a coloring where half of

---

*Work on this paper was partially supported by OTKA Grant K76099, by CIB (Centre Interfacultaire Bernoulli) and the NSF (National Science Foundation)
the points is colored red and the other half blue. An edge is a straight line segment between two points of different color. An alternating path is a path on which any two consecutive points have different colors. Note, if edges connect points of different color, any path will be alternating.

Erdős asked to determine or estimate the number of points on the longest noncrossing, alternating path for an arbitrary equicolored $2n$-element convex point set.

He and also others conjectured [7] that the next configuration was asymptotically extremal. Let us assume $n$ is divisible by four. Divide the circle into four intervals that consist of $\frac{n}{2}$ red, $\frac{n}{4}$ blue, $\frac{n}{2}$ red and $\frac{3n}{4}$ blue points, respectively. It is a short case analysis to check that in this construction the length of the longest noncrossing, alternating path is $\frac{3n}{2} + 2$.

On the other hand, the longest noncrossing, alternating path contains at least $n$ points. To see this, take a halving line $l$. At least half of the red points are on one side of $l$ on $C$ and consequently at least half of the blue points are on the other side of $l$ on $C$. We match the red points on one side to the blue point on the other side in a noncrossing way. This matching can be extended to a path which gives the lower bound $n$.

Jan Kynčl, János Pach and Géza Tóth disproved Erdős’ conjecture with a single construction in 2008 [7] and showed the $n + \Omega\left(\sqrt{n \log n}\right)$ lower and the $\frac{2}{3}n + O\left(\sqrt{n}\right)$ upper bound. At about the same time Abellanas et al. had a very similar construction for the same upper bound [2]. It is conjectured that this upper bound is asymptotically tight.

Hajnal and Mészáros improved the lower bound to $n + \Omega\left(\sqrt{n}\right)$ and gave a class of configurations for the $\frac{1}{3}n + O\left(\sqrt{n}\right)$ upper bound [5]. This result also underlines the importance of the separated matching conjecture [7] which is formulated as follows. Let $2k$ denote the number of alternations between the two colors in a $2n$-element point set on $C$. Then for any fixed $k$ and large $n$, any configuration admits a separated matching that contains at least $\frac{2k}{3} - 2n + o(n)$ points. (We describe separated matchings in the next section.)

In this paper we exhibit a new class of configurations that shows the $\frac{2}{3}n + O\left(\sqrt{n}\right)$ upper bound for the number of points in the largest separated matching. This class of configurations significantly differs from all known previous constructions. We also present a type of coloring such that among these colorings in the optimal one the maximum separated matching contains at most $\frac{4}{3}n + O\left(\sqrt{n}\right)$ points.

On the other hand, if we restrict the discrepancy (that is, the maximum difference in cardinality of color classes in any interval of $C$), we get an interesting result. For discrepancies two and three we prove the lower bound
So far no one was concerned with discrepancy since low discrepancy means many alternations among the two colors, and that alone guarantees a long noncrossing, alternating path. However, when we consider separated matchings, this is not the case.

2 Notations

First we introduce some basic definitions that will be necessary to describe the constructions and will be used in the proofs. Let our $2n$-element equicolored convex point set be denoted by $P$. An arc is an interval of points on $C \cap P$. The size of an arc is the number of its elements. In an arc the points are ordered, we always read the order in clockwise direction. A run is a maximal set of consecutive points on $C$ of the same color. The length of the run is the number of its elements.

A matching is a set of pairwise disjoint edges. Note, that the notion of matching is meant in geometrical sense, that is, no two edges cross in it. The size of a matching is defined as the total number of points participating in it, which is twice the number of edges. A separated matching is a matching where all edges can be crossed by a line.

This crossing yields a natural ordering of the edges of the matching. Separated matchings are closely related to noncrossing, alternating paths. Observe, that for every separated matching $S$ on a convex point set, there is a noncrossing, alternating path $R$ such that the vertex set of $R$ coincides with the vertex set of $S$ and all edges of $S$ are contained in $R$. We construct $R$ in the following way. The edges of $S$ will follow each other on $R$ in their natural ordering. Hence, every other edge of $R$ will belong to $S$. The remaining edges of $R$ will connect differently colored endpoints of consecutive edges of $S$. As a consequence of the properties of the separated matching we can always draw a noncrossing, alternating path $R$ in this way. We remark that $R$ is not unique. For every $S$ exist exactly two such paths depending on it which color will be chosen to be the color of the starting point of $R$.

The previous configurations contained long runs colored red or blue and at most two arcs consisting of alternating short runs of the two colors. Our class contains arbitrary many arcs of alternating short runs. The idea originates from the Kyncl-Pach-Tóth construction. We cut that construction into two pieces. We repeat the two pieces in arbitrary order an equal number of times along the circle.

We introduce some special arcs called blocks. They will be the building elements of our constructions. The first two types of blocks will be of the same size but they will contain a different number of points of the two color.
Figure 1: Two types of arcs

classes. However, altogether the number of red and blue points will be equal on a union of two blocks of different types. In the first two types of blocks the common size of blocks will be $3s$. The bloish block will consist of a red run of length $s$ and a blue run of length $2s$. We denote the buish block by $(s, 2s)$ block. The reddish block will consist of a red run of length $s$ followed by a mixed arc $M$. The mixed arc $M$ consists of $2s$ points alternating in color, see Figure [I]. Hence the reddish block will contain $2s$ red and $s$ blue points. We denote the reddish block by $(s, s(1, 1))$ block.

If needed we introduce notations $R$ and $B$ for red runs of size $s$, and for blue runs of size $2s$, respectively. We call $R$, $B$ and $M$ subblocks as they are the main building elements of blocks.

An $(as, bs)$ block consists of a red run of $as$ points and a blue run of $bs$ points. An $s(b, a)$ block consists of a red run of length $b$ followed by a blue run of length $a$ and this $a + b$ colored points are repeated $s$ many times. Consequently, an $s(b, a)$ block consists of $s$ consecutive arcs of size $a + b$ of the same coloring pattern. We call the unit of this pattern of $a + b$ points a period. Specifically, an $s(2, 1)$ block consists of the triple of two red and one blue point repeated $s$ many times. The period is two red points followed by a blue point.

We say that the discrepancy is $d$ if on any interval on the circle $C$ the difference between the cardinality of color classes is at most $d$. We will investigate the case of low discrepancy.

We say that an arc $A$ faces arcs $A_1, A_2, \ldots, A_n$ if all vertices of $A$ that participate in the separated matching $S$ have their pair in $S$ on one of the
arcs $A_1, A_2, \ldots, A_n$.

3 Constructions and Theorems

We will describe two main constructions and then we give another one by generalizing one of them.

The first construction is $C_1(s, t)$: Take $t$ consecutive $(s, 2s)$ blocks on $C$ followed by $t$ many $s(2, 1)$ blocks. Each block has size $3s$. Of course the last $t$ many blocks can be considered as one $st(2, 1)$ block.

The second construction is $C_1^+(a, b, s, t)$: In $C_1(s, t)$ instead of $(s, 2s)$ blocks we take $(as, bs)$ blocks and instead of the $s$ triples we take $s(b, a)$ blocks. Note that $C_1^+(1, 2, s, t) = C_1(s, t)$. Each block has size $(a + b)s$.

The third construction is a class of coloring $C_2(s, t)$: Take $t$ many $(s, 2s)$ blocks and $t$ many $(s, s(1, 1))$ blocks in arbitrary order along $C$. In other words, the same number of bluish and reddish blocks are placed along the circle in an arbitrary order.

Theorem 1. In $C_1(s, t)$ the size of every separated matching is at most $\frac{4}{3}n + O(s + t)$.

The upper bound is optimal if we disregard the remainder term. To see it, let us construct the following separated matching. If we match the the blue points of the first $t$ blocks (the bluish ones) with the red points of the last $t$ many blocks ($s$ many reddish triples), then we obtain a separated matching of size $\frac{4}{3}n$.

Theorem 2. In $C_1^+(a, b, s, t)$ the ratio of the size of the largest separated matching to the total number of points is

$$\max \left\{ \frac{2 \min \{a, b\}}{a + b}, \frac{\max \{a, b\}}{a + b} \right\} + O(\frac{s + t}{n}).$$

It follows that the order of magnitude of the size of the largest separated matching is at least $\frac{4}{3}n$. Equality occurs when $\max \{a, b\} = 2 \min \{a, b\}$. So $C_1(s, t)$ is optimal among $C_1^+(a, b, s, t)$.

Theorem 3. Let $C_2$ be any coloring from $C_2(s, t)$. Then the size of every separated matching in $C_2$ is at most $\frac{4}{3}n + O(s + t)$.

Theorem 4. Let $C_3$ be that coloring from $C_2(1000, t)$ where the reddish and bluish blocks alternate. Then size of the largest separated matching in $C_3$ is at least $1.34n$. 

5
Note that we refer to the $O(s + t)$ as remainder term. Since $s \cdot t = O(n)$, we can choose $s$ and $t$ so that $s, t = O(\sqrt{n})$ and the order of magnitude of $O(s + t)$ becomes negligible. This is how the reader should think about the first three theorems.

The fourth theorem is an exception, there we choose a setting where $s$ is a large constant and $t$ is $\epsilon \cdot n$. So $O(s + t)$ is very small but not negligible. The reason for choosing such a setting is that in $C_5$ the discrepancy of the coloring is constant (2000). At the same time the size of the optimal matching is very close to the conjectured value. In the fifth section we finish the paper with a few claims on coloring with low discrepancy.

4 Proofs

Take any separated matching $S$ in a coloring of $C$ from one of our theorems. Let the line $l$ be the axe, that is, a line that crosses all members of $S$. We think of $l$ as a horizontal line dividing $C$ into an upper and lower part. We can assume that the upper and lower part of $C$ consist of whole blocks by disregarding at most $O(s)$ edges of $S$.

We can order the edges of the matching according to their intersection with $l$. We can make a partition of $S$ so that in each class of the partition the matching goes between two subblocks. The previous partition determines $O(t)$ pairs of arcs facing each other on $C$. Furthermore, if an arc $A$ determined by the partition belongs to a mixed subblock, then we achieve that $A$ contains complete periods. This can be done by removing at most $O(t)$ many edges.

Let $S_0$ be the remainder of $S$. We call it the normalized matching. To prove the upper bounds of our theorems without loss of generality we may assume that we work with an arbitrary normalized matching.

**Proof of Theorem 1**: We will show that the ratio of points in $S_0$ to all $2n$ points on $C$ is at most $\frac{2}{3}$. Thanks to the simple structure of $C_1$ we partition the elements of the normalized matching $S_0$ into three classes. The first class contains edges with both endpoints in $(s, 2s)$ blocks. Edges with one endpoint in an $(s, 2s)$ block and the other endpoint in an $s(2, 1)$ block belong to the second class. The remaining edges with both endpoints in $s(2, 1)$ blocks we put in the third class. Note, some classes may be empty here. We can assume that borderlines between classes preserve complete blocks.

In the first and in the third class at most $\frac{2}{3}$ of the vertices are in $S_0$ because in both types of blocks $\frac{1}{3}$ of the points is of one color and $\frac{2}{3}$ is of the other color.

In the second class assume there are $L$ many $(s, 2s)$ blocks facing a mixed
coloring with period \((2,1)\). Let \(x\) denote the ratio of matched points in the red subblocks to the total number of points in the red subblocks. Let \(y\) be the same considering the blue subblocks. Hence, \(x \cdot (L \cdot s)\) points are matched out of the \(L \cdot s\) red points and \(y \cdot (L \cdot 2s)\) points are matched out of the \(L \cdot 2s\) blue points. The pairs of the \(x \cdot (L \cdot s)\) red points in \(S_0\) are blue points in the \((2,1)\)-periodic part. Hence, these pairs are contained in at least \(x \cdot (L \cdot s)\) many periods. Similarly, the pairs of \(y \cdot (L \cdot 2s)\) blue points come from \(y \cdot (L \cdot 2s)/2\) many periods (each period contains 2 red points). To prove the upper bound we can assume that the whole point set is the \(L(s,2s)\) blocks facing \(x \cdot (L \cdot s) + y \cdot (L \cdot 2s)/2\) many periods. Hence, its size equals \(3Ls + 3xLs + 3yLs\). The number of matched points is \(2[x \cdot (L \cdot s) + y \cdot (L \cdot 2s)]\). Their ratio is

\[
\frac{2xLs + 4yLs}{3Ls + 3xLs + 3yLs} = \frac{2}{3} \cdot \frac{x + 2y}{1 + x + y} \leq \frac{2}{3} \cdot \frac{x + 2y}{x + 2y} = \frac{2}{3}.
\]

This completes the proof of Theorem 1. \(\square\)

**Proof of Theorem 2.** Estimating the upper bound is the same as above with a small technical difficulty. We analogously partion \(S_0\) into three classes. We can bound the ratio of the number of matched points to the total number of points in each of the classes. In the case of edges from the first and third class the upper bound is

\[
\frac{2 \min\{a, b\}}{a + b}.
\]

In the case of edges from the second class the ratio is bounded above by

\[
f(x, y) = \frac{2}{a + b} \cdot \frac{ax + by}{1 + x + y}.
\]

It is not hard to see that \(f(x, y)\) is a quasiconvex function over the \([0,1] \times [0,1]\) domain, that is, its sublevel sets

\[
S_\alpha = \{(x, y) \in [0,1] \times [0,1] : f(x, y) \leq \alpha\}
\]

are convex for all \(\alpha\). So its maximum is attained in one of the vertices of its square domain:

\[
f(x, y) \leq \max\{f(0,0), f(0,1), f(1,0), f(1,1)\} = \max\left\{0, \frac{b}{a + b}, \frac{a}{a + b}, \frac{2}{3}\right\} = \max\left\{\frac{\max\{a, b\}}{a + b}, \frac{2}{3}\right\}.
\]

7
The overall ratio can be bounded by

$$\max \left\{ \frac{2 \min\{a, b\}}{a + b}, \frac{\max\{a, b\}}{a + b}, \frac{2}{3} \right\} = \max \left\{ \frac{2 \min\{a, b\}}{a + b}, \frac{\max\{a, b\}}{a + b} \right\}.$$ 

The final equality is straightforward to check.

Finally, we claim that the upper bound is optimal. To see that we construct two matchings. We can assume that $a \leq b$. In the first matching the blue points from $t(as, bs)$ will be matched with the red points from $ts(b, a)$. There will be $bs$ points matched in each block. The corresponding ratio is $\frac{b}{a+b}$. In the second matching the same types of blocks will be faced to each other. We pair up the half of the $(as, bs)$ blocks with the other half of $(as, bs)$ blocks. The matching will go between the pairs of blocks. The red subblock in a block will be matched to the blue subblock in its pair, and vice versa. Thus, in each block we can match $2as$ points. We can do the same inside the $s(b, a)$ blocks considering the periods in the same way as subblocks above. The corresponding ratio is $\frac{2a}{a+b}$.

This completes the proof of Theorem 2. □

**Proof of Theorem 3** Take a maximal normalized separated matching $S_0$ on $C_2$ (an arbitrary member of $C_2(s, t)$). We will show that the size of $S_0$ is at most $\frac{4}{3}n$.

In each block the ratio between the color classes is $2 : 1$. Hence, in each block we may call points of the major and the minor color class major and minor points, respectively. The number of minor points on $C_2$ is $\frac{2n}{3}$. Our proof will be an assignment: to each edge of $S_0$ we injectively assign a minor point. When we assign the point $p$ to edge $e$, we say mark $p$ for $e$.

If one endpoint of an edge $e$ is minor point and the other endpoint is major point, then mark the minor point for $e$. If both endpoints of $e$ are minor points, then mark the blue endpoint for $e$. Note, in this case the blue endpoint is in a mixed $M$ subblock. Non-marked minor points give the set of free vertices. Observe, the set of free vertices is changing during the procedure of marking new points.

If both endpoints of $e$ are major points, we distinguish two cases. If the major red point is in a mixed subblock, then we mark for $e$ the free blue point that is in the same period with the red endpoint of $e$. Otherwise, we call $e$ a bad edge. Note, in this case one endpoint of $e$ is in a blue $B$ subblock and the other endpoint of $e$ is in a red $R$ subblock which is in an $(s, s(1, 1))$ block. Bad edges are grouped according to blue subblocks. We can assume that $B$ is on the upper side of $l$. Take a blue subblock $B$ and consider the bad edges incident to it. Let $R$ be the red subblock pair of $B$, that is, $R$ and $B$ form a block together. We distinguish different cases.
Case 1: Subblock $R$ contains only free vertices. Let $k$ denote the number of bad edges incident to $B$. If the red endpoints of the $k$ bad edges are in the same red subblock, then $k \leq s$ and we can mark different elements of $R$ for each of the bad edges.

If the red endpoints come from different subblocks (of reddish blocks), then consider the mixed subblock $M$ in the block of the rightmost bad edge’s lower endpoint. Let $j$ be the number of vertices matched in $M$. Consequently, we have $s - j$ free vertices in $M$ and $s$ free vertices in $R$, altogether $2s - j$ free vertices. As $k + j \leq 2s$, we get $k \leq 2s - j$ as desired. Hence, we can mark a different free vertex for each bad edge in $B$.

Case 2: There is a non-free vertex in $R$. Let this non-free vertex be incident to edge $e$. Let $B'$ be the blue subblock of the low endpoint of $e$ and $R'$ the red subblock pair of $B'$, see Figure 2. Let $k$, $M$ and $j$ be defined as previously. Let $m$ be the number of vertices matched in $R'$. Therefore, the number of free vertices is $s - j$ on $M$ and $s - m$ on $R'$ which gives $2s - j - m$ free vertices altogether.

If $m > 0$ or no vertex in $M$ is matched to a vertex in $R$, then $j + k + m \leq 2s$, it follows that $k \leq 2s - j - m$ as desired. Hence, we can mark a different free vertex for each bad edge incident to $B$.

If $m = 0$ and there is a vertex in $M$ matched to a vertex in $R$, then $k \leq s$. In this case mark different vertices of $R'$ for each bad edge incident to $B$.

For each blue subblock $B$ we marked free vertices in the subblock pair of $B$ or in a subblock which was facing $B$ (that is, these points were in $R$, or in $M$ in case $M$ was facing $B$, or in $R'$ which necessarily faced $B$). Therefore, for each blue subblock the set of the possible free vertices was well defined and disjoint of the set of free vertices for any other blue subblock.
This completes the proof of Theorem 3. □

**Proof of Theorem 4.** This is a special case of Theorem 3. The constants in front of $s$ and $t$ are small as a consequence of the number of disregarded edges in the normalization procedure. The claim of the theorem is immediate. □

Although our goal is to investigate separated matchings, we mention that one can fix $a$ and $b$ in $C^+_i(a, b, s, t)$ in such a way that the run parameter of the coloring is $o(n)$ and at the same time the remainder term is $o(n)$, too. So we also gain new constructions for colored point set with short alternating paths.

## 5 Low discrepancy

When the discrepancy $d$ is rather small we found the following lower bounds for the separated matching. For $d = 1$ the coloring is alternating, hence all $2n$ points participate in the maximum separated matching.

The case of discrepancy at most 2 is little bit more technical.

**Theorem 5.** For any coloring with discrepancy $d = 2$ there is a separated matching of size at least $\frac{4d}{3}$. 

**Proof.** We describe a new way of visualizing of the colored point set: we introduce for each red point a unit up line segment and for each blue point we introduce a unit down line segment. This corresponds to the drawing scheme in [5]. (When the discrepancy is 1, then these up and down segments alternate.)

Actually, we will not choose a good axe. We can be given any axe that halves the number of runs and we will construct the separated matching of the desired size.

Since $d = 2$, there will be two types of runs: runs of length 1 and runs of length 2. Consequently, there will be at most two up and at most two down segments in each run, see Figure 3. Let us take a drawing for any case of $d = 2$ and halve the number of runs by taking an axe $t$. Then we pair up all the runs. The run $r$ will have pair run $r'$ if $r$ and $r'$ are on different sides of $t$ but for the same distance to $t$ regarding the number of runs. We make the separated matching $S$ so that each run will face only its pair in $S$. All runs of length 1 will be fully covered in $S$. Consider the runs of length 2. If a run $r$ of length 2 faces a run $r'$ of length 1, then $\frac{2}{3}$ of the vertices of $r$ and $r'$ will be in $S$. Otherwise, the run $r$ is also fully covered in $S$. Therefore, exists a separated matching of size at least $\frac{4d}{3}$.

□
For $d = 3$ we have the same result.

**Theorem 6.** For any coloring with $d = 3$ there is a separated matching of size at least $\frac{4n}{3}$.

The proof of this theorem is similar to the previous proof. Unfortunately, we need a more sophisticated pairing for runs. The detailed case analysis is sketched in the Appendix.

## 6 Open problems

It is not clear what role convexity plays in the problem. There are several lines of research on point sets other than the convex ones. Various open problems remain in the area.

**Problem 1.** Determine or estimate the number of vertices on the longest noncrossing, alternating path among $n$ red and $n$ blue points in general position.

Abellanas at al. proved that if the points are in general position and the color classes are separated by a line, then there is a noncrossing, alternating Hamiltonian path on the point set. If we do not assume that the color classes are separated by a line, the previous statement does not hold for $n \geq 8$. Even if the point set is in convex position, that is, its elements form a convex $2n$-gon.

By the existence of halving lines the result of Abellanas at al. gives the lower bound $n$ for the number of vertices on the longest noncrossing, alternating path in a point set in general position.
Based on the following result convexity might be an extremal case regarding the length of the longest noncrossing, alternating path. Pavel Valtr posed the following problem in 2007. Instead of convex position let the points be on a double-chain. A convex or a concave chain is a finite set of points in the plane lying on the graph of a strictly convex or a strictly concave function, respectively. A double-chain consists of a convex chain and a concave chain such that any line determined by any of the chains does not intersect the other chain.

Cibulka, Kynčl, Mészáros, Stolař, Valtr [4] proved if both chains contain at least one fifth of all the points, then there exists a Hamiltonian, non-crossing, alternating path. On the other hand, they showed that the above property does not hold for double-chains in which one of the chains contains at most $\approx \frac{1}{29}$ of all the points.

Instead of the double-chain we may consider the $2n$ points to be on two convex chains.

**Problem 2.** How long is the longest noncrossing, alternating path among $n$ red and $n$ blue points on two convex chains?

**Acknowledgement**

I would like to thank to my advisor Péter Hajnal for his fruitful suggestions and for all the help during this research.

**References**


Figure 4: The arc of three red points on the left faces the three arcs on the right, respectively. The number at a vertex on the right side shows the ratio of matched points in case the corresponding point is an endpoint of its run.

Figure 5: The right arc faces the left arc.

7 Appendix: coloring with discrepancy at most three

We do not give the full case analysis. We only give the pictures of the cases: the left and right hand side of the picture correspond to intervals of the colored point set that will be faced to each other. We always face one specific side to the other side where we mark the ratio of points that can be matched on the considered intervals, see Figure 4, Figure 5 and further pictures below. Some of the endpoints can vary on the sides and in that case we put the corresponding ratios to the alternate endpoints. On Figure 4 we merge more cases. There the left side contains a single interval while the right side contains three intervals. We face the left side to the intervals on the right, respectively.

Note that the number of points on the two intervals that face each other in the case analysis is at most 14.

The case analysis: