Matching Colored Points[1]

Richard Kaye
Rutgers University

Submitted in partial fulfillment of the requirements for a Masters of Science degree in Computer Science from Rutgers University

Abstract

Given a set of *n* points in the plane, each point colored either black or white. We show that it is always possible to connect $\frac{6}{7}n$ points with pairwise disjoint straight line segments that connect like colored points.

1 Introduction

Given a set of n points in the plane, b of them colored black and w of them white with no 3 points collinear, we define a non-crossing matching (hereinafter referred to as a matching) to be a set of edges such that a) the two endpoints of each edge have the same color, b) each point appears in at most one edge and c) no two edges intersect. We define the size of a matching as the number of points matched. For any such set of points, we can ask what is the largest matching possible (which we refer to as a maximum matching). For example, Figure 1 illustrates a set of 6 points where the size of a maximum matching is 4.

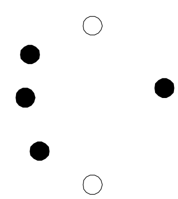

Figure 1

Define $g(n)$ to be the minimum number of points matched by a maximum matching among all possible configurations of n points. Section 2 describes previously known bounds on $g(n)$. Section 3 presents an improved lower bound on $g(n)$ and some new exact values of $g(n)$ for certain values of n. Section 4 describes an algorithm for constructing a matching with size equal to the lower bound presented in Section 3.

---

[1] The proof of the time bound on the algorithm in Section 4 was joint work with Adrian Dumitrescu.

2 Previous Results

As reported in [2], Aharoni and Saks considered this problem restricted to point sets with both w and b even and asked whether it was always true that $g(n) = n-2$. Eli Berger [2] produced an 18-point counter-example (Figure 2) which proves that $g(18) \le 14$. Notice that if all 6 white points are matched (there are 5 ways to do so) the 3 edges partition the circle into 4 parts. Each part contains an odd number of black points. So, at least 4 black points must remain unmatched. If 4 white points are matched, the circle is partitioned into 3 parts. In all cases, 2 of the 3 parts have an odd number of black points. Hence, 2 black points and 2 white points must remain unmatched. Therefore, no more than 14 of the 18 points can be matched. Later we will see that $g(18) = 14$.

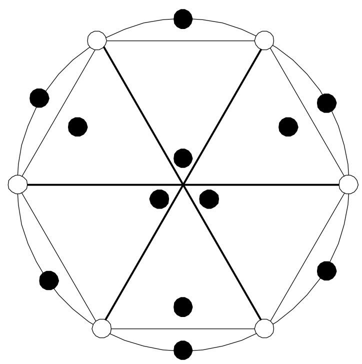

Figure 2*

Dumitrescu and Steiger [2] found the first bounds on $g(n)$ showing that

$$\frac{5}{6}n - O(1) \le g(n) \le \frac{155}{156}n + O(1) \qquad (1)$$

Their proof of the lower bound was based on the following observations:

Lemma 1 $g(5) = 4$

Proof Clearly $g(5) \le 4$ since 5 is an odd number. To show that it is always possible to match 4 points consider the split of the points by color. There are 3 possibilities: all 5 points are the same color, 4 points are one color and 1 is the other or 3 points are one color and 2 are the other. In the first 2 cases, sort the (at least 4) points of the more abundant color by their *x* coordinate. Match the first point with the second and the third with the fourth. Now, consider the 3-2 split. Without loss of generality assume 3 black points and 2 white points. Consider the line formed by extending the line segment between the 2 white points. Since no 3 points are collinear, one side of this line contains at least 2 black points. Match these 2 black points and the 2 white points. Thus $g(5) = 4$.

Notice that Lemma 1 leads to the conclusion that $g(n) \ge \frac{4}{5}n - O(1)$ as follows:

---

* In this and all subsequent figures, we can assume that all apparent collinearities are eliminated through small perturbations of the points involved.

Given n points, adopt a coordinate system so that no 2 points have the same x coordinate. Sweep a vertical line from right to left until 5 points have been encountered. Match 4 of these 5 points. Then continue sweeping from where you left off until an additional 5 points are encountered. Match 4 of these 5 points. Continue matching 4 out of every 5 points until less than 5 points are left. Clearly none of the matched edges will intersect since each group of 5 points is separated from the next by a vertical line. This procedure shows that $g(n) \geq \frac{4}{5}n - 4$.

Lemma 1 allows us to complete the proof that $g(18) = 14$ since 18 points can be split by a sweep line into groups of 5, 5, 5 and 3 points. Four out of each group of 5 can be matched along with 2 out of the final 3 points.

For Lemma 2, we will need the following definition. Define a star (see Figure 6 for an example) to be a set of 7 points with the following properties: a) the convex hull of the star must contain exactly 6 of the 7 points, b) the six points on the convex hull must alternate in color and c) the 7^{th} point, if colored white (resp. black) must be inside the triangle formed by the 3 black (white) points on the convex hull. We will refer to the 7^{th} (i.e. internal) point as the center of the star and to the points on the convex hull as external points.

Lemma 2

A. If a set of 7 points does not form a star, 6 of the 7 points can be matched.

B. If a set of 7 points does form a star then only 4 of the points can be matched.

Proof

A. There are 3 ways that a set of 7 points can fail to meet the criteria for a star

   a. The set could have 3, 4, 5 or 7 points on its convex hull.

   b. There could be 2 successive points on the convex hull with the same color.

   c. The center point could be outside the triangle formed by the 3 points with color different than that of the center point.

   Case a) If there are an odd number of points on the convex hull then there must be 2 successive points with the same color so we can treat this in case b. So, we only need to examine the configurations with 4 points on the convex hull that alternate in color.

   If all 3 internal points are the same color (say black) the two white points can be matched. The line drawn through the two white points will separate the black points into 3 on one side and two on the other or 4 on one side and 1 on the other. In either case, it is simple to match 4 black points.
   So it remains to consider the case where the internal points are split 2 of one color (say black) and 1 of the other (white).

   Now, the 4 black points either form a convex quadrilateral (Figure 3) or they form a triangle with a point inside (Figure 4). If they form a convex quadrilateral, and the internal white point is outside of this quadrilateral then match the internal white point with the external white point that is on the same side of the quadrilateral. Then match the 4 black points.

If the internal white point is inside the quadrilateral then match it with either of the 2 external white points. This edge will cross one edge of the quadrilateral. So 2 of the other quadrilateral edges can be chosen to match all 4 black points.

If the 4 black points form a triangle with a point inside, label the points as in Figure 4. Call the internal white point 7. If 7 falls inside {1, 2, 3, 6} then match 7 with 2, 6 with 1 and 5 with 3. If 7 falls inside {1, 4, 3, 5}, match 7 with 4, 1 with 6 and 5 with 3. If 7 falls inside {1, 5, 3, 6}, match 7 with 2, 6 with 1 or 3 depending on which one would create an edge that does not intersect the edge between 7 and 2 and match 5 with 1 or 3 whichever was not matched with 6.

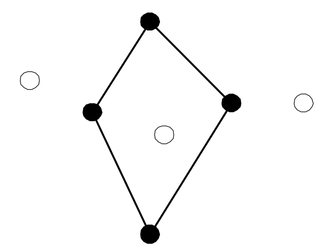

Figure 3

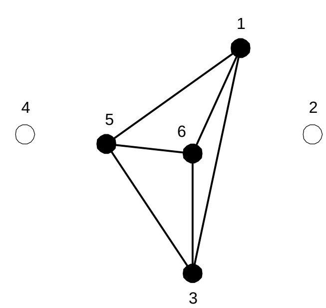

Figure 4

Case b) Find 2 consecutive points on the convex hull with the same color and match them. Since the other 5 points must all lie on the same side of the segment created by that match, it is always possible to match 4 of them by Lemma 1.

Case c) This case is illustrated in Figure 5 with a black internal point. Since the black internal point is outside the white triangle, it must be inside one of the triangles created by 2 white points and the black point that lies between them on the convex hull. Match the black internal point with that black external point and match those 2 white points. Finally match the other 2 black points.

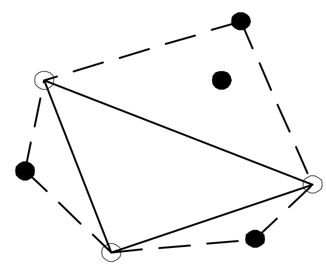

Figure 5

B. Assume without loss of generality that the internal point is black (see Figure 6). To match 6 points it would be necessary to match 4 black points and 2 white points. There are 3 possible pairs of 2 white points to match. However in each case, the line segment matching the 2 white points separates one black point from the other 3. So, it is impossible to match all 4 black points after matching any 2 white points.

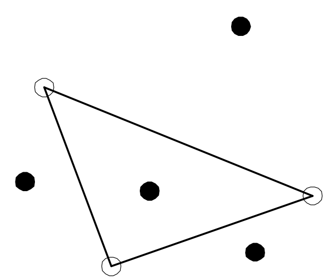

Figure 6

This completes the proof of Lemma 2. Dumitrescu and Steiger [2] used Lemmas 1 and 2 to prove that $g(12) = 10$. This result led directly to the lower bound in (1). Their proof is omitted here since Theorem 1 in Section 3 will lead to an alternative proof of that fact.

3 New Results

In [2], Dumitrescu and Steiger asked whether $g(14)$ is 10 or 12. It is clearly at least 10 since a vertical sweep line can split the points into groups of 5, 5 and 4 points and it is always possible to match 4, 4 and 2 points from each of the respective groups of points. They remarked that if $g(14) = 12$ it would lead to an immediate improvement on the lower bound in (1). However, Figure 7 illustrates that $g(14) = 10$. The impossibility of matching 12 of the 14 points in Figure 7 can be observed by noting the following. Since there are 5 (an odd number) white points and 9 (also an odd number) black points, matching 12 points would require matching 4 white points and 8 black points. Every line segment between a pair of white points separates an odd number of black points from the rest of the points. Therefore, matching 4 (2 pairs of) white points requires leaving at least 2 black points unmatched. So, 12 points cannot be matched in the given configuration.

Dumitrescu and Steiger also raised the question of whether $g(16) = 12$ or 14. By adding 2 black points near any of the internal black points in Figure 7 we find that $g(16) = 12$.

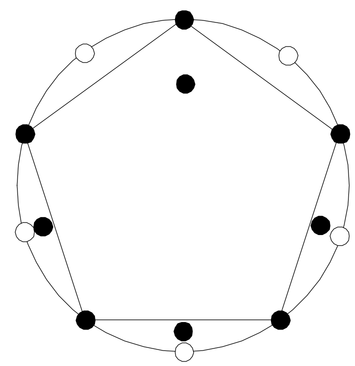

Figure 7

Despite the fact that $g(14) = 10$, we are able to improve Dumitrescu and Steiger's lower bound as follows:

Lemma 1 leads to the conclusion that $\frac{4}{5}n - O(1) \leq g(n)$ since a vertical sweep line can be used repeatedly to partition off groups of 5 points from which 4 can always be matched. Of course there's nothing special about the sweep line being vertical. A horizontal sweep line or one at a 45-degree angle would work just as well. In fact, there's no reason why the sweep line has to come from the same direction for each iteration. Eighty percent of any set of points can be matched by first sweeping from the right and matching 4 of the first 5 points encountered, then sweeping from above and matching 4 of the first 5 points encountered (from among those that had not previously been swept), then sweeping with a line angled at 45 degrees and so on.

Now remember, Lemma 2 showed that it is always possible to match 6 out of 7 points unless the 7 points are a star. Given a set of points *V*, we can define *V'* to be a 7-set of *V*, if *V'* consists of the first 7 points encountered by a line sweeping *V* from some direction. If it can be shown that for a sufficiently large set of points *S*, there is always a 7-set of *S* that is not a star, then we would have:

Theorem 1 Given *n* points in the plane, each point colored either black or white, it is always possible to match $\frac{6}{7}n - O(1)$ points.

Proof Theorem 1 follows directly from Lemma 3 as discussed above.

Lemma 3 Given a set of points, V, in the plane with $|V| \geq 12$, each point colored either black or white, there exists a 7-set of V that is not a star.

Proof Sweep a vertical line from right to left until 7 points have been encountered (Figure 8a). Call that set of 7 points S. If S does not form a star we are done, so assume that it does. Now, perform another sweep of the set, but this time instead of using a vertical line, use a sweep line that is at a small clockwise angle ε from vertical (see Figure 8b). If ε is small enough, then the first 7 points that this new sweep line encounters will also be S. Continue to increase ε and perform sweeps until ε gets big enough that the first 7 points encountered is some set other than S (Figure 9). Call this set T. Notice that S and T must have 6 points in common. Call the point that was in S but is not in T, v. When we consider the new sweep line just after it encounters all the points in T, we see that it separates v from the 6 other points in S. Therefore, v must be on the convex hull of S.

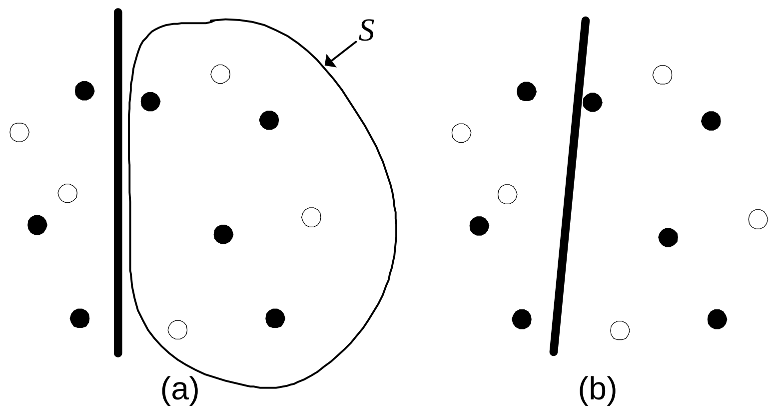

Figure 8

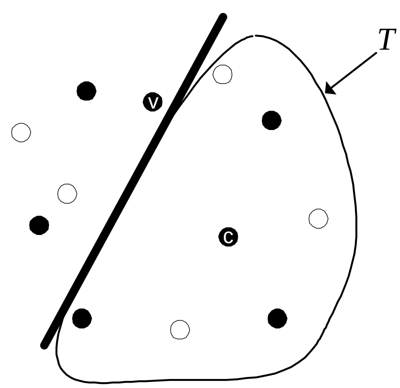

Figure 9

Now *T* must also be a star or else we are done. Since the center of *S* (call it *c*) is in *T*, we can ask the question, is *c* necessarily the center of *T* or could it be an external point in *T*? If *c* and *v* are the same color (say black) then *c* was inside the triangle formed by the 3 white points in *S*. Since those 3 white points are also in *T*, *c* is the center of *T*.

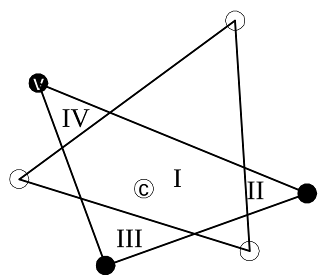

Figure 10

What if *c* and *v* are different colors (say *c* is white and *v* is black)? Then since *c* must be in the triangle formed by the 3 black points, it has four possible locations (labeled I-IV in Figure 10). If *c* is in I then it is inside the triangle formed by the 3 external white points. In that case, since the three external white points are all in *T*, *c* must be an internal point in *T*. If *c* is in II or III then it is inside the triangle formed by an external black point that is in *T* and its two white neighbors on the convex hull of S. Since all three of these points are in *T*, *c* is an internal point in *T*. It remains to consider the case where *c* is in IV. In that case, the set S*v* is convex (Figure 11) with the points 1, *c* and 2 comprising three consecutive white points on the convex hull.

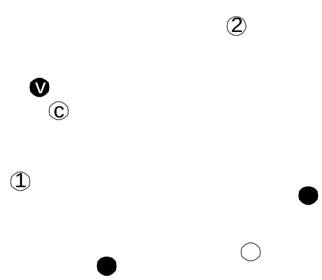

Figure 11

When a new black point is added to S*v* to form *T* one of these three white points must be pushed into the interior of *T* in order to make *T* a star. If 1 or 2 becomes the center point then the points on the convex hull would not alternate in color. Therefore, it must be *c* that is the center point.

So in all cases *c* is the center of *T*. This means that as we rotate the sweep line and keep encountering different 7-sets that in order for each 7-set to be a star, it must have the same center point as the previous 7-set.

But, it can not be the case that all 7-sets have the same center point. Clearly if the original set has 14 or more points, then the 7 set formed by sweeping from right-to-left and the 7-set formed by sweeping from left-to-right are disjoint. It can be shown that for as few as 12 points the center point in a right-to-left sweep must be different from the center point in a left-to-right sweep.

Therefore, at some point as we rotate the sweep line, the 7-set we encounter must not be a star. This completes the proofs of Lemma 3 and Theorem 1. Theorem 1 implies $g(n) \geq \frac{6}{7}n - O(1)$ thereby improving on the lower bound in (1).

4 An Algorithm for matching points

In this section, we will present an algorithm that computes a matching satisfying the lower bound from Section 3. The algorithm runs in $O(n^2)$ time using $O(n)$ space. However, for ease of exposition, we will first present a simpler algorithm and then modify it to achieve the desired time bound.

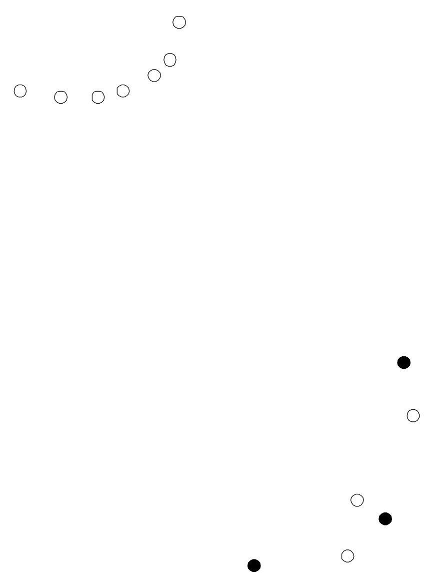

Figure 12

Algorithm Match1

Given: a set V of points each colored either white or black

Output: a matching $M \subseteq V \times V$ with at least $\frac{6}{7}n - 11$ points matched.

All angles in the algorithm are based on 0° being straight up (positive *y* direction), 90° being to the right, 180° being down and 270° being to the left.

```
1.  Set M = {}
2.  Sort V by the x coordinates of the points.
3.  While |V| ≥ 12
4.      Set S = {}
5.      Insert into S the 7 points in V with the largest x coordinates
6.      Set p = the leftmost point in S
7.      If S is a star then
8.          Compute the clockwise order of the 6 points on the convex hull of S
            starting with p. Let succ(p) denote the successor of p in this order.
9.      End if
10.     While S is a star
11.         θ = 360
12.         For each point x in V\S
13.             Compute the angle θ′ of the ray from x to p.
14.             If θ′ < θ
15.                 θ = θ′
16.                 x′ = x
17.             End if
18.         End For
19.         Set π to the angle of the ray from p to succ(p)
20.         If π < θ
21.             Set p = succ(p)
22.         else
23.             Set S = S - {p} + {x′}
24.             Set w = succ(p)
25.             Set p = x′
26.             Set succ(p) = w
27.         End if
28.     End While
29.     Match 6 of the 7 points in S
30.     Insert the 3 pairs of matched points into M
31.     Remove the 7 points in S from V
32. End While
```

To analyze the complexity of this algorithm, we note the following. Step 1 takes O(1) time and step 2 takes O(n log n) time. The while loop beginning at step 3 is executed O(n) times since 7 points are removed from V during each iteration of the loop. Steps 4-9 and 29-31 all run in O(1) time. Within the while loop at step 10, the for loop at step 12 takes O(n) time and all the other steps take O(1) time. So, if we can show that the while loop starting at step 10 can only be executed a fixed number of times during each iteration of the outer while loop then we will have our $O(n^2)$ time bound.

Unfortunately, that is not the case. Each time through the inner while loop, p is replaced with either it's successor on the convex hull or with a point from outside S. The former can only happen 5 times since there are only 6 points on the convex hull of S. As Figure 12 illustrates, however, the latter can happen any number of times before we find a group of 7 points that is not a star.

A careful study of Figure 12 provides the insight that allows us to adjust the algorithm to achieve the desired $O(n^2)$ time bound. If p is replaced 7 times in a row by a point outside of S without being replaced by its successor on the convex hull we have the situation depicted in Figure 12. Notice that there are 8 points of the same color that have each taken a turn as point p in the algorithm. These 8 points form a chain that separates the 6 other points in S from the rest of the points in V. At his point in the algorithm, we can match all 8 of the points in the chain into 4 pairs and match 4 of the 6 other points in S, thereby matching 12 of 14 points (see Figure 13). The rest of the points are all in the convex region above the chain and so the algorithm can move on to the next iteration of the outer while loop. The modified algorithm appears below.

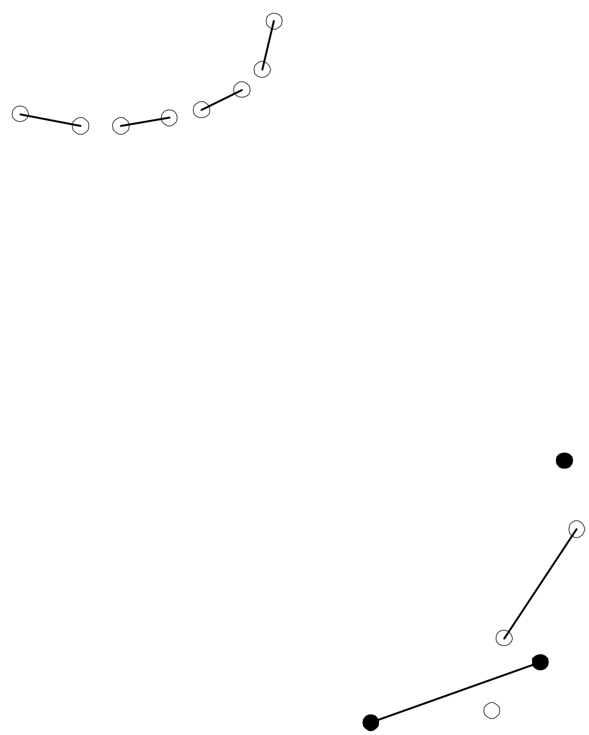

Figure 13

Algorithm Match2

Given: a set V of points each colored either white or black

Output: a matching $M \subseteq V \times V$ with at least $\frac{6}{7}n - 11$ points matched

All angles in the algorithm are based on 0° being straight up (positive *y* direction), 90° being to the right, 180° being down and 270° being to the left.

```
1.  Set M = {}
2.  Sort V by the x coordinates of the points.
3.  While |V| ≥ 12
4.      Set S = {}
5.      Insert into S the 7 points in V with the largest x coordinates
6.      Set p = the leftmost point in S
7.      If S is a star then
8.          Compute the clockwise order of the 6 points on the convex hull of S
            starting with p. Let succ(p) denote the successor of p in this order.
9.          Count = 1
10.     End if
11.     While S is a star and count < 8
12.         θ = 360
13.         For each point x in V\S
14.             Compute the angle θ′ of the ray from x to p.
15.             If θ′ < θ
16.                 θ = θ′
17.                 x′ = x
18.             End if
19.         End For
20.         Set π to the angle of the ray from p to succ(p)
21.         If π < θ
22.             Set p = succ(p)
23.             Count = 1
24.         else
25.             Set S = S - {p} + {x′}
26.             Set w = succ(p)
27.             Set p = x′
28.             Set succ(p) = w
29.             Count = Count + 1
30.         End if
31.     End While
32.     If S is not a star
33.         Match 6 of the 7 points in S
34.         Insert the 3 pairs of matched points into M
35.         Remove the 7 points in S from V
36.     Else
37.         Match the last 8 p’s into 4 pairs
```

```
38.           Insert the 4 pairs into M
39.           Remove the 8 points from V
40.           Match 4 of the 6 points in S\{p}
41.           Insert the 2 pairs into M
42.           Remove S\{p} from V
43.       End if
44. End While
```

Now the test at step 21 can only be true 5 times since each time it is true, p becomes the next point on the original convex hull of S and there are only 6 of these points. Additionally, if the test at step 21 is false 7 times consecutively, we break out of the inner while loop because count reaches 8. Since the while loop at step 21 can only be executed a fixed number of times during each iteration of the outer while loop, the algorithm runs in $O(n^2)$ time.

5 Acknowledgement

The proof of the time bound on the algorithm in Section 4 was joint work with Adrian Dumitrescu.

References

[1] D. Bienstock, Some provably hard crossing number problems, *Discrete and Computational Geometry* **6** (1991), 443-459.

[2] A. Dumitrescu, W. Steiger, On a matching problem in the plane, *Workshop on Algorithms and Data Structures*, 1999 (WADS '99), and in *Discrete Mathematics*, vol. 211 (2000), no. 1-3, 183-195.

[3] P. Erdös, L. Lovász, A. Simmons, E. G. Straus, Dissection graphs of planar point sets, in *A Survey of Combinatorial Theory*, (J. N. Srinastava et. al. eds.), North Holland, Amsterdam, 1973, 139-149.

[4] M. R. Garey, D. S. Johnson, Crossing number is NP-complete, *SIAM Journal on Algebraic and Discrete Methods* **4** (1983), 312-316.

[5] K. Jansen, G. Woeginger, The complexity of detecting crossingfree configurations in the plane, *BIT* **33** (1993) 580-595.

[6] L. Lovász, On the number of halving lines, *Ann. Univ. Sci. Budapest E*ötvös Sect. Math., ***14*** (1971), 107-108