Long non-crossing configurations in the plane (Draft)

Noga Alon *   Sridhar R †   Subhash Suri ‡

Abstract

It is shown that for any set of $2n$ points in general position in the plane there is a non-crossing perfect matching of n straight line segments whose total length is at least $2/\pi$ of the maximum possible total length of a (possibly crossing) perfect matching on these points. The constant $2/\pi$ is best possible and a non-crossing matching whose length is at least as above can be found in polynomial time. Similar results are obtained for the problem of finding a long non-crossing Hamilton path and a long non-crossing spanning tree for a given set of points in the plane.

1 Introduction

A *geometric graph* is a pair $G = (V, E)$, where V is a finite set of points in general position in the plane and E is a family of closed straight line-segments whose end-points lie in V. The elements of V are called *vertices* and these of E are called *edges*. The *length* of G, denoted by $L(G)$, is the sum of Euclidean lengths of all edges of G. G is *non-crossing* if the interiors of all its edges are pairwise disjoint.

Several results in Combinatorial and Computational Geometry deal with the extremal values of the length $L(G)$ of a geometric graph G of a prescribed type on a given set of vertices in the plane. The best known example of a problem of this type is the (Euclidean) *travelling salesman problem* in which one is interested in finding a Hamilton cycle (or path) G of minimum possible length on a

*Department of Mathematics, Raymond and Beverly Sackler Faculty of Excat Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv, ISRAEL and Bellcore, Morristown, NJ 07960, USA. Research supported in part by a United States Israel BSF Grant

†Department of Computer Science, UC Berkeley

‡Bellcore, Morristown, NJ 07960, USA

given set of points in the plane. It is well known that this problem is NP-hard ([5]), and there are numerous results on approximating the optimal solution and on the properties of such a solution. See, e.g., [4].

In this paper we consider the problem of finding a *non-crossing* geometric graph G of a prescribed type on a given set of points in the plane, whose length $L(G)$ is *maximum*. We consider this problem when the desired graph G is a perfect matching, a Hamilton path, or a spanning tree. Somewhat surprisingly, in all these cases there is always a non-crossing G of the desired type whose length $L(G)$ is at least some absolute constant times the maximum possible length of a possibly crossing graph G of the corresponding type on the same set of points. We note that this is not true for some other, similar, classes of graphs G. For example, the maximum possible length of a non-crossing Hamilton *cycle* on a set of n evenly distributed points on the unit cycle is smaller than the length of the cycle, whereas the length of the maximum possible (crossing) Hamilton cycle on this set of points is $\Omega(n)$, showing that the ratio between the two is unbounded as n grows.

Our main results are summarized in the following three theorems. All the graphs mentioned in their statements (i.e., the perfect matchings, the Hamilton paths and the spanning trees) are geometric graphs and hence their edges are straight line segments.

Theorem 1.1 *Let V be a set of $2n$ points in general position in the plane and let L_M denote the maximum possible length of a (possibly crossing) perfect matching on V. Then there is a non-crossing perfect matching on V whose length is at least $\frac{2}{\pi}L_M$. Such a non-crossing perfect matching can be found in polynomial time (in n, for a given set V as above). Moreover, the constant $2/\pi$ is the best possible constant (for which the assertion of the theorem holds for every n).*

Theorem 1.2 *Let V be a set of n points in general position in the plane and let L_P denote the maximum possible length of a (possibly crossing) Hamilton path on V. Then there is a non-crossing Hamilton path on V whose length is at least $\frac{1}{\pi}L_P$. Such a non-crossing path can be found in polynomial time.*

Theorem 1.3 *Let V be a set of n points in general position in the plane and let L_T denote the maximum possible length of a (possibly crossing) spanning tree on V. Then there is a non-crossing spanning tree on V whose length is at least $\frac{n}{2n-2}L_T$ ($> \frac{1}{2}L_T$). Such a non-crossing tree can be found in polynomial time.*

Note that the three theorem above supply efficient algorithms that approximate the maximum possible length of a non-crossing perfect matching, Hamilton path or spanning tree on a given set of points. We do not know if finding the exact value of the corresponding maximum is NP-hard for any of these problems, but suspect that this is the case for all of them. This, however, remains open.

2 Perfect matchings

A well known Olympiad problem (see, e.g., [3],[2]) asserts that for any two sets X and Y in the plane, where $|X| = |Y| = n$ and $X \cup Y$ is in general position, there is a non-crossing perfect matching that matches the points of X with these of Y. In the proofs of Theorems 1.1 and 1.2 we need the following slightly stronger result for a special case of this fact.

Lemma 2.1 *Let l be a strainght line in the plane, let X be a set of n points in the right side of l and let Y be a set of n points in its left side. Suppose, further, that $X \cup Y$ is in general position. Then there is a non-crossing perfect matching $x_i y_i$ $(1 \leq i \leq n)$ that matches X and Y so that for every i, $1 \leq i < n$, either the two open line segments $x_i x_{i+1}, x_i y_{i+1}$ or the two open line segments $y_i x_{i+1}, y_i y_{i+1}$ do not intersect the convex hull of $\{x_{i+1}, \ldots, x_n, y_{i+1}, \ldots, y_n\}$. Given the points in X and in Y, such a matching can be found in time $O(n \log n)$.*

Proof Without loss of generality, suppose l is a vertical line. For any set of points Z, let $C(Z)$ denote its convex hull. Put $X_1 = X$, $Y_1 = Y$ and observe that the two convex hulls $C(X_1)$ and $C(Y_1)$ are disjoint. Let l_1 be a common tangent to these hulls, so that both of them lie below it. Obviously l_1 contains a point of X_1 and a point of Y_1. Denote these points by x_1 and y_1 and define $X_2 = X - x_1$, $Y_2 = Y - y_1$. Let l_2 be a common tangent to $C(X_2)$ and $C(Y_2)$ so that both hulls lie below it. Let x_2 be the unique intersection point of l_2 and $C(X_2)$ and let y_2 be the unique intersection point of l_2 and $C(Y_2)$. Define $X_3 = X_2 - x_2$, $Y_3 = Y_2 - y_2$ and continue in the same manner. This procedure defines our matching $x_i y_i$. It is not difficult to check that it is non-crossing and satisfies the assertion of the lemma. Moreover, it can be found in time $O(n \log n)$, as shown in [7]. $\square$

Next we prove the following simple statement.

Lemma 2.2 *Let $z_1, \ldots, z_{2n}$ be a set of n points on a line appearing in this order from left to right on the line and let y be a point on the line segment joining z_n and z_{n+1}. Then the total length of any perfect matching on the above points is at most the sum of distances of y to the $2n$ points z_j. Equality holds for any matching that matches each point to the left of y to a point to its right and only for these matchings.*

Proof Given a perfect matching on the points z_j, one can replace each edge $z_i z_j$ of the matching by the two line segments $z_i y$ and $z_j y$ whose total length is at least that of $z_i z_j$, by the (one dimensional) triangle inequality. Since this procedure replaces the matching by the $2n$ segments $y z_j$ and since each application of the triangle inequality is strict unless the matched points lie in opposite sides of y the desired result follows. □

The next proposition establishes the first part of Theorem 1.1

Proposition 2.3 *Let V be a set of $2n$ points in general position in the plane and let L_M denote the maximum possible length of a (possibly crossing) perfect matching on V. Then there is a non-crossing perfect matching on V whose length is at least $\frac{2}{\pi} L_M$. Such a non-crossing perfect matching can be found in time $o(n^{5/2} \log n)$. Moreover, for any $\epsilon > 0$ a non-crossing matching of length at least $(1-\epsilon)\frac{2}{\pi} L_M$ can be found in time $O(n \log n / \sqrt{\epsilon})$.*

Proof Let PM be a (possibly crossing) perfect matching of length L_M on the points of V. For a line l in the plane and for $v \in V$, let $P_l(v)$ denote the (orthogonal) projection of v on l. Observe that for any line segment $s = uv$ of length g, if the slope of l is chosen uniformly and randomly in the range $[0, \pi)$, then the expected length of the projection of s on l, i.e., the expected length of the segment $P_l(u) P_l(v)$ is:

$$\frac{1}{\pi} g \int_0^{\pi} |\cos \alpha| \, d\alpha = \frac{2}{\pi} g.$$

Therefore, by linearity of expectation, the expected total length of the projection of the matching PM on l is $\frac{2}{\pi} L_M$. It follows that there is a line l so that the projection of PM on l is of length at least $\frac{2}{\pi} L_M$. Let $z_1, \ldots, z_{2n}$ be the projections of the points of V on l and suppose these appear in this order on l. (It is easy to see that we may assume these are all distinct). Let y be a point on the line segment $z_n z_{n+1}$. By Lemma 2.2 the length of the projection of PM on l is at most the sum of the distances between y and the points z_i, and this last sum is the length of any perfect matching between the points $z_1, \ldots, z_n$ and $z_{n+1}, \ldots, z_{2n}$. Let x_i be the point of V whose projection on l is z_i

$(1 \leq i \leq n)$, and let y_i be the point of V whose projection on l is z_{n+i} $(1 \leq i \leq n)$. By Lemma 2.1 there is a non-crossing matching M that matches the points x_i to the points y_j. The projection of this matching on l matches the points z_j with $j \leq n$ to the points z_k with $k > n$. Thus the length of this projection is at least $\frac{2}{\pi} L_M$ and since the length of M is at least the length of its projection we conclude that $L(M) \geq \frac{2}{\pi} L_M$, as needed.

In order to convert the proof above to an algorithm observe that one only has to find a non-crossing matching between X and Y for any possible partition $V = X \cup Y$ with $|X| = |Y| = n$ so that X and Y can be separated by a line. It is well known (see [6]) that there are at most $o(n^{3/2})$ such partitions and all of them can be found in time $o(n^{3/2})$. For each partition we can find the corresponding non-crossing matching in time $O(n \log n)$, by Lemma 2.1. The matching of maximum length found is the one to be chosen by the algorithm.

If a non-crossing matching of length at least $(1-\epsilon)\frac{2}{\pi} L_M$ suffices it is enough to choose $b/\sqrt{\epsilon}$ lines l_i of evenly distributed slopes in $[0, \pi)$, where b is an absolute constant. As shown above there is always a line l so that the length of the projection of PM on l is at least $\frac{2}{\pi} L_M$. Let l_i be the line among our lines whose slope is closest to that of l. It is easy to check that for sufficiently large b the projection of L_M on l_i is at least $(1-\epsilon)\frac{2}{\pi} L_M$, (since $\cos t = 1 - O(t^2)$ for small t). Hence it suffices to find a non-crossing matching for the appropriate $O(1/\sqrt{\epsilon})$ pairs (X, Y) obtained by projecting V on the lines l_i, yielding a total running time of $O(n \log n/\sqrt{\epsilon})$. This completes the proof of the proposition. $\Box$

In order to complete the proof of Theorem 1.1 it remains to show that the constant $2/\pi$ is best possible. To this end we prove the following.

Lemma 2.4 *The maximum length of a non-crossing perfect matching on $2n$ evenly distributed points on a cycle of radius 1 is the length of any perfect matching consisting of pairwise parallel edges, which is*

$$C_{2k} = 4(\cos\frac{\pi}{4k} + \cos\frac{3\pi}{4k} + \ldots + \cos\frac{(2k-1)\pi}{4k})$$

for $n = 2k$ and

$$C_{2k+1} = 2(1 + 2(\cos\frac{\pi}{2k+1} + \cos\frac{2\pi}{2k+1} + \ldots + \cos\frac{k\pi}{2k+1}))$$

for $n = 2k+1$.

Proof The n points split the cycle into n equal circular arcs. Define the *span* of a segment joining two of the points to be the number of such arcs in the shorter semi-cycle defined by its two ends. Thus, the span is always an integer between 1 and n. Observe that the span of any segment which is part of a non-crossing perfect matching is always an odd integer since it must split the other points into two even sets. Let M be a non-crossing perfect matching on our points. For each odd i, $i \leq n$, let N_i denote the number of segments in M whose span is at least i.

Claim: For each admissible i, $N_i \leq n - i + 1$.

Proof: Since every two segments of span n cross we may and will assume that $i < n$. If $N_i \leq 1$ there is nothing to prove. Otherwise, one can easily check that there are two segments p_1p_2 and p_3p_4 in our matching, each of span at least i, with the following properties. $p_1, \ldots, p_4$ appear in this order (clockwise) on the cycle and there is no other segment of span at least i in M that has an endpoint among the points on the cycle from p_1 to p_2 (clockwise) or from p_3 to p_4. (To see this, consider the "rightmost" and "leftmost" segments of span at least i in M). There are at least $i-1$ points on each of the open circular arcs p_1p_2 and p_3p_4 and hence all the endpoints of the segments of M of span at least i lie in a set of $2n - 2(i-1)$ points, implying the assertion of the claim.

Returning to the proof of the lemma, observe that for a perfect matching on our points consisting of parallel edges there are precisely $n - i + 1$ segments of span at least i for each odd $i \leq n$. Let l_i denote the Euclidean length of a segment of span i. Note that the sequence l_i is increasing. Define $N_i = 0$ for all $i > n$ and observe that the number of segments of span precisely i in M is $N_i - N_{i+2}$ for all odd $i \leq n$. Therefore,

$$L(M) = \sum_{i \leq n, i \equiv 1 (mod\ 2)} (N_i - N_{i+2}) l_i = N_1 l_1 + \sum_{3 \leq i \leq n, i \equiv 1 (mod\ 2)} N_i (l_i - l_{i-2})$$

$$\leq n l_1 + \sum_{3 \leq i \leq n, i \equiv 1 (mod\ 2)} (n - i + 1)(l_i - l_{i-2}) = \sum_{1 \leq i \leq n, i \equiv 1 (mod\ 2)} (n - i + 1)(l_i - l_{i-2}),$$

where the last inequality follows from the fact that $l_i \geq l_{i-2}$ and from the claim. However, the right hand side of the last inequality is precisely the length of a perfect matching consisting of parallel edges (which is given analytically in the assertion of the lemma). This completes the proof. □

It is worth noting that the last lemma can be proved in a somewhat simpler way, as observed by R. Adin [1]. However, the proof presented above can be easily applied to deduce similar bounds for the other problems considered here and is thus more useful.

By the definition of an integral as a limit of the corresponding Riemann sums, one can easily conclude that

$$\lim_{k\mapsto\infty}\frac{C_{2k}}{4k}=\lim_{k\mapsto\infty}\frac{C_{2k+1}}{4k+2}=\int_0^{\pi}|\cos\alpha|\ d\alpha=\frac{2}{\pi}.$$

Since the length of a maximum (crossing) perfect matching on our points is $2n$ it follows, by Lemma 2.4, that $\frac{2}{\pi}$ is indeed the best possible constant in Theorem 1.1, completing its proof. □

3 Hamilton paths

In this section we prove Theorem 1.2. Let V be a set of n points in general position in the plane and let H be a possibly crossing Hamilton path on V of maximum possible length $L(H)=L_P$. We must show that there is a non-crossing Hamilton path of length at least $\frac{1}{\pi}L_P$. The proof, which resembles the one presented in the previous section, is somewhat simpler in case n is even. In this case one can first complete H to a Hamilton cycle H' by adding to it the missing edge. Clearly $L(H')\geq L(H)$ and since n is even H' is a union of two perfect matchings on V. It follows that there is a perfect matching M on V of length $L(M)\geq L(H)/2$. By the argument given in the proof of Proposition 2.3 there is a line l in the plane so that the length of the projection of M on l is at least

$$\frac{2}{\pi}L(H)/2=\frac{1}{\pi}L(H).$$

Put $n=2k$, let $z_1,\ldots,z_{2k}$ denote the projections of the points in V on l and assume they appear in this order on l from left to right. Let X be the set of points of V whose projections are $z_1,\ldots,z_k$ and put $Y=V-X$. By the argument given in the previous section, the length of any perfect matching that matches the points of X with these of Y is at least the length of the projection of M on l, i.e., it is at least $\frac{1}{\pi}L(H)$. In order to obtain a non-crossing Hamilton path on V of the desired length it thus suffices to show that there is such a path that contains a perfect matching as above. To do so observe that by Lemma 2.1 there is a numbering of the points in X and Y and a non-crossing perfect matching x_iy_i, $(1\leq i\leq k)$ so that for every $1\leq i<k$, either

(i) x_ix_{i+1},x_iy_{i+1} do not intersect the convex hull of $\{x_{i+1},\ldots x_k,y_{i+1},\ldots,y_k\}$, or

(ii) y_ix_{i+1},y_iy_{i+1} do not intersect the convex hull of $\{x_{i+1},\ldots x_k,y_{i+1},\ldots,y_k\}$.

This enables us to complete the perfect matching x_iy_i into a non-crossing Hamilton path. Indeed, assuming that a non-crossing Hamilton path on $\{x_{i+1},\ldots y_k\}$ has already been defined so that its starting point is either x_{i+1} or y_{i+1} it can be extended by joining its starting point to x_i in case (i) holds or to y_i in case (ii) holds. The arguments of the pervious section can be easily applied to convert the last proof into an algorithm for finding a path with the required properties in time $o(n^{5/2} \log n)$, or a non-crossing path of length at least $\frac{1-\epsilon}{\pi} L_P$ in time $O(n \log n / \sqrt{\epsilon})$. This completes the proof of Theorem 1.2 for even n.

The proof for odd n is similar, but contains an additional complication. Suppose $n = 2k + 1$, and let us call any matching of k edges on a set of $2k + 1$ points, a *near perfect matching*. An easy modification of Lemma 2.2 is the following result, whose simple proof, which is analogous to that of Lemma 2.2, is omitted.

Lemma 3.1 *Let $z_1, \ldots, z_{2k+1}$ be a set of $2k + 1$ points on a line appearing in this order on the line. Then the total length of any near perfect matching on the above points is at most the sum of distances of z_{k+1} to the other $2k$ points z_j. Equality holds for any matching that matches each point to the left of z_{k+1} to a point to its right and only for these matchings.*

Suppose, now, that V is a set of points in general position in the plane, $|V| = 2k + 1$, and let H be a (possibly crossing) Hamilton path of maximum length $L(H) = L_P$ on V. Obviously H is a union of two near-perfect matchings on V and hence the length of one of them is at least $L(H)/2$. Thus there is a line l so that the length of the projection of some near perfect matching on l is at least $\frac{1}{\pi} L_P$. Let $z_1, \ldots z_{2k+1}$ be the projections of the points of V on l written according to their order on l, let $X \subset V$ be the set of points projected to $z_1, \ldots, z_k$ and let $Y \subset V$ be the set of points projected to $z_{k+1}, \ldots, z_{2k+1}$. Let z be the unique point in $V - (X \cup Y)$. The length of any matching that matches the points of X to those of Y is at least $\frac{1}{\pi} L_P$, and hence it suffices to prove that there is a non-crossing Hamilton path on V containing such a matching. This can be shown by applying Lemma 2.1 to the two sets X and Y. As before, it is easy to see that the matching guaranteed by the Lemma can be extended to a Hamilton path on $X \cup Y$. Moreover, since we may assume that the additional point z lies on the line separating X and Y (which intersects all the edges in the matching) it is not difficult to see that there is an edge of the path so that the line segments joining its two ends to z do not intersect the path. Therefore, one can replace such an edge by the two edges joining its ends to z to get, by the triangle inequality, a Hamilton path on V whose length is at least $\frac{1}{\pi} L_P$. This clearly yields an efficient algorithm as before, completing the proof of Theorem 1.2. □

It is worth noting that evenly distributed points on a cycle show easily that the constant $\frac{1}{\pi}$ in Theorem 1.2 cannot be replaced by any constant larger than $\frac{2}{\pi}$. We do not know the best possible constant in this theorem.

4 Spanning trees

The proof of Theorem 1.3 is simpler than these of the previous two theorems, and follows as a special case of a more general result that holds in arbitrary metric spaces. If V is a set of points in a metric space, and T is a graph on V, the *length* of T, denoted by $L(T)$, is the sum of lengths of the edges of T, where the length of an edge is the distance between its two end-points. If $v \in V$, the *star at* v, denoted by S_v, is the spanning tree on V consisting of all edges $\{vu : u \in V - v\}$. Obviously, if the metric space is the Euclidean plane and V is a set of points in general position, then each such star S_v is non-crossing. Therefore, Theorem 1.3 follows from the following result.

Proposition 4.1 *Let V be a set of n points in an arbitrary metric space and let T be a spanning tree of maximum length on V. Then there is a $v \in V$ so that*

$$L(S_v) \geq \frac{n}{2n-2} L(T). \tag{1}$$

Moreover, this estimate is best possible for every even n.

Proof To see that the above estimate (if true) is best possible for every $n = 2k$ let p_1 and p_2 be two points in a metric space the distance between which is 1, and let V consist of k copies of p_1 and k copies of p_2. Then the total length of any star on a member of V is $k = n/2$, whereas the maximum possible length of a spanning tree on V is $n - 1$. The ratio between these two quantities is $n/(2n-2)$, as desired. (Note that by a small perturbation one can give an example with almost the same ratio without repeated points).

It remains to show that there is always a star satisfying (1). For every $v \in V$, define a (multi) graph R_v on V by letting its set of edges consist of all edges of T incident with v together with two additional edges vx and vy for every edge xy of T with $x, y \neq v$. Thus R_v is a (multi) graph with $2(n-1) - deg(v)$ edges, where $deg(v)$ is the degree of v in T, and all these edges are incident with

v. Moreover, by the triangle inequality $L(R_v) \geq L(T)$ for all v. In addition, the number of copies of the edge uv in R_v is precisely $deg(u)$.

Let $l(uv)$ denote the length of the edge uv and let V_2 denote the set of all $\binom{n}{2}$ (unordered) pairs of members of V. Then, by the previous paragraph

$$nL(T) \leq \sum_{v \in V} L(R_v) = \sum_{uv \in V_2} (deg(u) + deg(v)) l(uv)$$

$$= \sum_{v \in V} deg(v) L(S_v) \leq (\sum_{v \in V} deg(v)) \; Max_{v \in V} L(S_v) = (2n-2) \; Max_{v \in V} L(S_v).$$

Therefore,

$$Max_{v \in V} L(S_v) \geq \frac{n}{2n-2} L(T),$$

completing the proof. □

It is not too difficult to extend Lemma 2.4 and deduce from that extension that the constant $\frac{1}{2}$ in Theorem 1.3 cannot be replaced by any constant larger than $\frac{2}{\pi}$. We do not know the best possible constant in this theorem. Our techniques can be used to obtain, for a given set V of n points in general position in the plane, a non-crossing spanning tree on V of length $\Omega(L_T)$ in linear time. We omit the details.

References

[1] R. Adin, Private communication.

[2] J. Akiyama and N. Alon, *Disjoint simplices and geometric hypergraphs*, in : Combinatorial Mathematics; Proc. of the Third International Conference , New York, NY 1985 (G. S. Blum, R. L. Graham and J. Malkevitch, eds.), Annals of the New York Academy of Sciences, Vol. 555 (1989), 1-3.

[3] L. C. Larson, *Problem Solving Through Problems*, Springer Verlag, New York 1983, 200-201.

[4] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan and D. B. Shmoys (editors), *The Traveling Salesman Problem*, Wiley, 1985.

[5] C. Papadimitriou, *Title*, Details.

[6] J. Pach, B. Steiger and E. Szemerédi,

[7] Algorithm for matching