SERIE B — INFORMATIK

On Spanning Trees with Low Crossing Numbers*

Emo Welzl◇

B 92–02
January 1992

Abstract

Every set S of n points in the plane has a spanning tree such that no line disjoint from S has more than $O(\sqrt{n})$ intersections with the tree (where the edges are embedded as straight line segments). We review the proof of this result (originally proved by Bernard Chazelle and the author in a more general setting), point at some methods for constructing such a tree, and describe some algorithmic and combinatorial applications.

*Supported by the Deutsche Forschungsgemeinschaft, "Schwerpunktprogramm Datenstrukturen und effiziente Algorithmen", grant We 1265/1-2.
◇Institut für Informatik, Freie Universität Berlin, Takustr. 9, D-14195 Berlin, Germany, e-mail: emo@tcs.fu-berlin.de

1 Introduction.

Over the recent years there has been considerable progress in the simplex range searching problem. In the planar version of this problem we are required to store a set S of n points such that the number of points in any query triangle can be determined efficiently. One of the combinatorial tools developed for this problem are spanning trees with low crossing numbers.

Let S be set of n points in the plane. For a spanning tree on S and a line h, the *crossing number of* h in the tree is defined as $c_h = a + \frac{b}{2}$, where a is the number of edges $\{p, q\}$ in the tree with p and q on opposite sides of h, and b is the number of edges with exactly one endpoint on h. h *crosses* an edge, if that edge contributes to the crossing number of h. Note that an edge completely contained in the line h does not contribute to the crossing number. The *crossing number* of the tree is the maximal crossing number of any line.

In other words, a spanning tree with crossing number c ensures that no line (disjoint from S) intersects the straight line embedding of the tree in more than c edges. It has been shown in [16], that every set of n points allows a spanning tree with crossing number $O(\sqrt{n})$, which is tight. In Section 2 we review the proof of this result (which is treated in [16] in a more general setting, for arbitrary dimension, and for set systems of finite VC-dimension, see Section 5). We derive an explicit constant for the bound on the crossing number. The proof builds on a packing lemma for a pseudodistance on points in the presence of a set of lines (where the distance between two points is the number of separating lines), and on a reweighting technique, which has been applied to several seemingly unrelated problems, see [13, 18, 6, 26, 5].

Spanning trees are useful in a number of applications. The original motivation for introducing the concept in [31] was the triangle range searching problem which can be solved in $O(\sqrt{n} \log n)$ query time and linear space via spanning trees. This is close to the lower bound of $\Omega(\sqrt{n})$ for linear space data structures in the so-called arithmetic model [12]. Recently, this lower bound has actually been achieved in [26]. Several different algorithmic applications are described in [1, 20, 2, 4, 17, 3]. For example, spanning trees with low crossing numbers can be used for ray shooting among line segments in the plane (i.e., we want to preprocess line segments in the plane such that the first segment intersected by a query ray can be efficiently computed).

In Section 3 we indicate the application to triangle range searching, and we present two recent combinatorial results which can be easily derived from spanning trees with low crossing numbers [27, 28].

Section 4 indicates some of the building blocks of algorithms for constructing spanning trees with low crossing numbers. This will lead us to a randomized Monte-Carlo algorithm; however, we did not try to present the best known time bounds for construction. Finally in Section 5, we point at the generalizations to higher dimensions.

We tried to keep the paper largely self-contained, so that in particular in Sections 2 and 3 little foreknowledge should be required. Hence we start by reviewing some basics before we plunge into the rest of the paper.

Notation and basics. Let S be a set of n points in the plane, and let G be a set of ℓ lines in the plane. We say that S is in *general position*, if no three points lie on a common line, and no two points lie on a vertical line. G is in *general position*, if no three lines contain a common point, no two lines are parallel, and no line is vertical.

We denote by H_S the set of lines containing at least two points in S; if S is in general position, then $|H_S| = \binom{n}{2}$. $\tilde{H}_S$ is a *representative set* of lines for S, if whenever a line g (disjoint from S) partitions the set S into nonempty sets S' and S'' (on the respective sides of g), then there is a line h in $\tilde{H}_S$ which induces the same partitioning. It is an easy exercise to verify, that there is always a representative set of at most $\binom{n}{2}$ lines.

The *arrangement* $\mathcal{A}(G)$ *of* G is the partitioning of the plane induced by G into *vertices* (intersections of lines in G), *edges* (connected components on the lines in the complement of the vertices), and *cells* (connected components of the plane in the complement of the lines). Obviously, there are at most $\binom{\ell}{2}$ vertices, at most ℓ^2 edges, and a bound of $\binom{\ell}{2} + \ell + 1$ on the number of cells is also not too hard to prove; if G is in general position, then all three bounds are attained cf. [19].

We will use the *point/line duality* defined by: for a point $p = (a, b)$, the dual image p^* is the nonvertical line with equation $y = ax + b$, and for a nonvertical line g with equation $y = cx + d$, the dual image g^* is the point $(-c, d)$. This mapping preserves incidences between lines and points (i.e. p lies on g if and only if g^* lies on p^*), and it preserves the relative position between a point and a line (i.e. p lies above g if and only p^* lies above g^*).

For two nonvertical lines g and h, define the *double wedge of* g and h as the two open quadrants (defined by the two lines) which are disjoint from the vertical line through the common point of g and h; if g and h are parallel, then the double wedge degenerates to the strip between the two lines. Now a line g intersects the open line segment with endpoints p and q, if and only if g^* lies in the double wedge defined by p^* and q^*.

We frequently use the inequalities $1 + x \leq e^x$, for all real numbers x, and $\sum_{i=1}^{n} \frac{1}{\sqrt{i}} < 2\sqrt{n}$, for all positive integers n.

Conventions. All points and lines we consider in Sections 2, 3, and 4 are assumed to lie in the plane!

2 Proof of existence.

We want to prove that every set of n points in the plane allows a spanning tree such that no line has more than $O(\sqrt{n})$ crossings with the tree. Note that it suffices to concentrate on a representative set $\tilde{H}_S$ of lines: Let T be a spanning tree on S. Clearly, by definition, every line disjoint from S has a line in $\tilde{H}_S$ with the same (number of) crossings. If h contains points from S, then we consider two parallel lines h' and h'' on both sides of h, but sufficiently close so that all points in S (except those on h) have the same position relative to h' (and to h'') as to h. Then the respective crossing numbers satisfy $c_h = \frac{c_{h'}+c_{h''}}{2}$. That is, the maximum crossing number is attained by a line disjoint from S.

The $O(\sqrt{n})$ bound is asymptotically the best we can hope for. To see this for some positive integer n, choose a set G of $\ell = \lceil\sqrt{2n}\rceil$ lines in general position, and place n points into the cells of the arrangement, no two points in the same cell (which is possible, since $\binom{\ell}{2} + \ell + 1 \geq n$). Every edge of an (arbitrary) spanning tree will be crossed by at least one of the lines in G; thus there must be a line in G with at least $\frac{n-1}{\ell} = \Theta(\sqrt{n})$ crossings.

If we start the construction of our tree, then it looks like a good idea to begin with an edge $\{p, q\}$, such that p and q are separated by as few as possible lines in a representative set. To provide a bound on this number is our next step.

A packing lemma [11, 16]. Suppose we are given a set S of $n \geq 2$ points with diameter Δ (i.e. Δ is the maximal Euclidean distance between any two points in the set). Then there are two points at distance at most $\sigma = \frac{4\Delta}{\sqrt{n}}$. This can be easily seen by the fact that the closed disks of radius $\frac{\sigma}{2}$ centered at the points in S are contained in a 'large' disk of radius $\frac{3}{2}\Delta$ centered at an arbitrary point in S (this is true if $\sigma \leq \Delta$; otherwise the claim is trivial). If the small disks were pairwise disjoint, then they cover an area of $n\frac{\sigma^2\pi}{4} = 4\Delta^2\pi$ in the large disk of radius $\frac{3}{2}\Delta$, which is not possible. Hence two disks intersect, and the respective centers have distance at most σ.

We will use the same idea as just described to show that for any set S of n points, and any set G of ℓ lines there is always a pair of points separated by less than $\frac{2\ell}{\sqrt{n}}$ of the lines. To this end we introduce a *pseudodistance* δ_G for pairs of points (relative to G) by $\delta_G(p, q) = a + \frac{b}{2}$, where a is the number of lines in G which have p and q on opposite sides, and b is the number of lines which contain exactly one of the two points p and q. It is easily seen that δ_G is a pseudometric (i.e. it is symmetric and satisfies the triangle inequality).

For a point p and a real number σ, we let $D_G(p, \sigma)$ denote the set of vertices v in the arrangement of G with $\delta_G(p, v) \leq \sigma$. The sets $D_G(p, \sigma)$ will play the role of disks, and the cardinality of $D_G(p, \sigma)$ will play the role of area in our proof, and so we need a lower bound on this quantity in terms of σ.

Lemma 2.1 *If G is a set of ℓ lines in general position, and σ is an integer, $0 \leq \sigma \leq \lceil\frac{\ell}{2}\rceil$, then $|D_G(p, \sigma)| \geq \binom{\sigma+1}{2}$ for all points p disjoint from G.*

Proof. Choose a line g through p which intersects the same number of lines in G on both sides of p. Such a line exists, since we can take a directed line h through p and rotate it, while observing the number of intersections on h preceding p. After rotating h by π this will be the number of intersections succeeding p; so in between we must meet a situation as required for g (note that if ℓ is odd, then g must be parallel to one of the lines in G).

Now consider the intersections $q_1, q_2, \ldots, q_{\lfloor \ell/2 \rfloor}$ on g on one side of p, enumerated in such a way that $\delta_G(p, q_i) = i - \frac{1}{2}$ (if g passes through a vertex, we may perturb g with p to make sure that all lines in G create distinct intersections). Let us first assume that $\sigma \le \lfloor \frac{\ell}{2} \rfloor$. Then, for all $i \le \sigma$, q_i has at least $\sigma - i + 1$ vertices on its line at distance at most $\sigma - i + \frac{1}{2}$; all these vertices have distance at most σ from p (by the triangle inequality). If we collect vertices at distance at most σ in the same way on the other side of p, we obtain $2(\sigma + \cdots + 2 + 1) = 2\binom{\sigma+1}{2}$ such vertices, each of which may be counted at most twice. This gives the claimed bound for $\sigma \le \lfloor \frac{\ell}{2} \rfloor$.

If ℓ is odd, and $\sigma = \lceil \frac{\ell}{2} \rceil$, then the above procedure gives us a count of $2(\sigma + \cdots + 2)$ only. Now we recall that there is a line $h \in G$ parallel to g which contains at least two points at distance at most $\lceil \frac{\ell}{2} \rceil$; take the two vertices incident to the infinite edges on h. In this way we have again counted $2\binom{\sigma+1}{2}$ vertices, each vertex at most twice. The lemma is proved. ◻

The bound in Lemma 2.1 can be shown to be tight.

Lemma 2.2 *Let G be a set of ℓ lines, and let S be a set of $n \ge 2$ points. Then there are two distinct points p and q in S with $\delta_G(p, q) \le \frac{2\ell}{\sqrt{n}}$.*

Proof. Choose some positive integer k with the property that

$$(\lfloor \frac{2k\ell}{\sqrt{n}} \rfloor + 1)\lfloor \frac{2k\ell}{\sqrt{n}} \rfloor > \frac{(2k\ell)^2}{n} . \tag{1}$$

Replace each line h in G by two buckets of k parallel copies each, such that the 'original' h lies between these two buckets, and the two buckets are sufficiently close to h, so that there are no points from S within a bucket, and between a bucket and its original. So the only points from S between the two buckets are those which lie on h. The resulting set G' has $\ell' = 2k\ell$ lines, no point in S lies on a line in G' and for any pair $\{p, q\}$ of points in S, $\delta_{G'}(p, q) = 2k\delta_G(p, q)$. Then perturb the lines in G' to general position such that no line moves over a point in S; this does not change the pseudodistance $\delta_{G'}$ between points in S.

For $n \le 4$ the assertion of the lemma is trivial; so we have to consider only the case $n \ge 5$ and Lemma 2.1 applies to $\sigma = \lfloor \frac{\ell'}{\sqrt{n}} \rfloor$. We get

$$\sum_{p \in S} |D_{G'}(p, \sigma)| \ge n\binom{\sigma+1}{2} > n\frac{\ell'^2}{2n} > \binom{\ell'}{2} ,$$

(where property (1) proved to be useful). Since there are only $\binom{\ell'}{2}$ vertices, there must be two 'disks' $D_{G'}(p, \sigma)$ and $D_{G'}(q, \sigma)$, $p, q \in S$, $p \ne q$, which overlap in a vertex; by the triangle inequality their centers p and q have pseudodistance $\delta_{G'}(p, q)$ at most 2σ. Hence, $\delta_G(p, q) \le \frac{1}{2k} 2\lfloor \frac{\ell'}{\sqrt{n}} \rfloor \le \frac{2\ell}{\sqrt{n}}$, the bound claimed in the lemma. ◻

We need to extend Lemma 2.2 to sets of lines G where every line h has a positive real weight $w(h)$ associated. The pseudodistance $\delta_G(p, q)$ is now defined as $a + \frac{b}{2}$, where a is the sum of weights associated with lines separating p and q, and b is the sum of weights associated with lines which contain exactly one of the two points p and q.

Lemma 2.3 *Let G be a finite set of weighted lines with overall weight Δ, and let S be a set of $n \geq 2$ points. Then there are two distinct points p and q in S with $\delta_G(p,q) \leq \frac{2\Delta}{\sqrt{n}}$.*

Proof. Let k be some positive integer. Replace every line h in G by two buckets of $\lceil k \cdot w(h) \rceil$ unweighted lines each, in the same way as described in the previous proof. We obtain a set G' of at most $2k\Delta + 2\ell$ unweighted lines to which we can apply Lemma 2.2. It supplies us with two points p and q with $\delta_{G'}(p,q) \leq \frac{4k\Delta+4\ell}{\sqrt{n}}$. If ℓ is the number of lines in G, then

$$\delta_G(p,q) \leq \frac{\delta_{G'}(p,q)}{2k} \leq \frac{2\Delta}{\sqrt{n}} + \frac{2\ell}{k\sqrt{n}} \,.$$

In other words, for every $\epsilon > 0$ we find points p and q with $\delta_G(p,q) \leq \frac{2\Delta}{\sqrt{n}} + \epsilon$. Since there are only finitely many points, this implies the lemma. ◻

Construction by iterative reweighting [31, 16]. Using Lemma 2.2 we can easily show that for n points S and ℓ lines G the greedy algorithm (using δ_G as weight function on edges) constructs a spanning tree on S with weight at most $\sum_{i=2}^{n} \frac{2\ell}{\sqrt{i}} \leq 4\ell\sqrt{n}$. That is, the average crossing number of a line in G is $4\sqrt{n}$. We will show that by a different construction we can guarantee this bound (up to a low order term) for all lines.

Theorem 2.4 *Every set S of n points has a spanning tree with crossing number at most $4\sqrt{n} + O(n^{1/4}\sqrt{\log n})$.*

Proof. Let G_0 be a representative set of lines, $\ell = |G_0| \leq \binom{n}{2}$, and let $S_0 = S$. We start the construction of the spanning tree by choosing two points p and q in S_0 which are separated by the smallest number of lines in G_0 (i.e. no more than $\frac{2\ell}{\sqrt{n}}$). Next we put the edge $\{p,q\}$ into the edge set of our tree and remove p from the point set which gives $S_1 = S_0 - \{p\}$.

For the rest of the construction we need some means to ensure that no line gathers too many crossings. That is lines which have already many crossings with the edges constructed so far should cross a next edge less likely. We will achieve this by assigning weights to the lines. To be precise, a line which has c crossings so far will have multiplicity $(1+\mu)^c$ for $\mu > 0$ a parameter to be chosen later.

Hence, we continue our construction by multiplying by $1+\mu$ the weight of all lines in G_0 which separate p and q; this gives a new set G_1 of weighted lines with overall weight $\Delta_1 \leq \ell(1 + \frac{2\mu}{\sqrt{n}})$. Then we continue the construction with G_1 and S_1: we choose two points p_1 and q_1 which are separated by lines of overall minimal weight, add edge $\{p_1, q_1\}$ to the edge set, remove p_1, and multiply the weights of separating lines by $1+\mu$, and proceed as above.

After i steps we have a set G_i of weight

$$\Delta_i \leq \Delta_{i-1}(1 + \frac{2\mu}{\sqrt{n-(i-1)}}) \leq \ell \prod_{j=0}^{i-1}(1 + \frac{2\mu}{\sqrt{n-j}})$$

and a set S_i of $n-i$ points.

Step $n-1$ completes the construction of a spanning tree for S. What is the crossing number of this tree? Let c_h denote the number of crossings of line h in the tree. Then h is represented with weight $(1+\mu)^{c_h}$ in G_{n-1}, that is

$$\Delta_{n-1} = \sum_{h \in G_0} (1+\mu)^{c_h} \ .$$

However, we have also a bound of

$$\Delta_{n-1} \leq \ell \prod_{j=2}^{n} (1 + \frac{2\mu}{\sqrt{j}}) < \ell \, e^{\sum_{j=1}^{n}(2\mu/\sqrt{j})} \leq e^{4\mu\sqrt{n}+2\ln n}.$$

Hence, we may conclude that

$$c_h < \frac{1}{\ln(1+\mu)}(4\mu\sqrt{n} + 2\ln n) \ ,$$

for all lines h which implies $c_h < 4\sqrt{n} + O(n^{1/4}\sqrt{\log n})$ for the choice of μ which minimizes this bound (see Appendix). ◻

The theorem and its proof provide us with a number of immediate consequences. A spanning path is *simple*, if only line segments corresponding to consecutive edges on the path intersect.

Corollary 2.5 *Every set S of n points has a simple spanning path with crossing number at most $4\sqrt{n} + O(n^{1/4}\sqrt{\log n})$.*

Proof. The asymptotic bounds follow directly from Theorem 2.4, if we double the edges in a spanning tree of crossing number c, and consider an Eulerian walk in this graph, which has crossing number $2c$. We can now simply scan this walk and omit points which have occurred before. In this way the number of crossings with a line cannot increase. Let $p_0, p_1, \ldots, p_{n-1}$ be the resulting spanning path with crossing number at most $2c$. If two line segments $\overline{p_{i-1}p_i}$ and $\overline{p_{j-1}p_j}$, $1 \leq i < j-1 \leq n-2$ intersect then we replace the edges $\{p_{i-1}, p_i\}$ and $\{p_{j-1}, p_j\}$ by new edges $\{p_{i-1}, p_{j-1}\}$ and $\{p_i, p_j\}$ to obtain the spanning path

$$p_0, p_1, \ldots, p_{i-1}, p_{j-1}, p_{j-2}, \ldots, p_{i+1}, p_i, p_j, p_{j+1}, \ldots, p_{n-1} \ .$$

The crossing number of no line increases, and the *Euclidean* length decreases. Consequently, after a finite number of steps we have obtained a simple spanning path with crossing number at most $2c$.

In order to achieve the claimed constant we have to look at the proof of the theorem once more. We proceed as for the construction of a tree, except that we are more careful about the points we put into the sets S_i. We keep as an invariant, that the edges constructed so far give a set of vertex disjoint paths on S (some of which are just isolated vertices), and we let S_i contain all isolated vertices, and exactly one point of degree one from each path. In the next step, we choose two points p and q of minimal pseudodistance (with respect to the current weighted set of lines) in S_i. The addition of edge $\{p,q\}$ merges two connected components; we remove p and q from S_i, and add one of the two points of degree one in this component to the set, which gives us S_{i+1}. After the appropriate reweighting of the lines we continue the construction. The calculus of the analysis stays the same and gives the claimed bound. The constructed path can be converted into a simple one by the same procedure as described in the first paragraph of the proof. ◻

Corollary 2.6 *Every set S of n points has a matching of size k with crossing number at most $\frac{4k}{\sqrt{n}} + O(\sqrt{k \ln n/\sqrt{n}})$, for integers k, $\frac{1}{2}\sqrt{n}\ln n \le k \le \frac{n}{2}$, and with crossing number at most $\frac{2e\ln n}{\ln(\sqrt{n}\ln n/(2k))}$, for integers $k \le \frac{1}{2e}\sqrt{n}\ln n$.*

Proof. The construction of a matching works in the obvious way (referring to the notation in the proof of Theorem 2.4). We choose the edge of minimal pseudodistance, remove its two points from the current point set, and reweight the lines with new crossings. Now S_i has $n-2i$ points. After k steps we have a matching of required size. Via the overall weight Δ_k of G_k we get the following bound for the number of crossings of lines in G_0:

$$\sum_{h\in G_0}(1+\mu)^{c_h} = \Delta_k \le \ell\prod_{i=1}^{k}\Big(1+\frac{2\mu}{\sqrt{n-2(i-1)}}\Big) < \ell\, e^{\sqrt{2}\,\mu\sum_{j=0}^{k-1}\frac{1}{\sqrt{n/2-j}}}$$

$$< \ell\, e^{\sqrt{2}\,\mu 2(\sqrt{n/2}-\sqrt{n/2-k})} = \ell\, e^{2\mu(\sqrt{n}-\sqrt{n-2k})} < e^{4\mu k/\sqrt{n}+2\ln n}\ .$$

The last inequality uses that $\sqrt{n}-\sqrt{n-x\sqrt{n}} \le x$ for all x, $0\le x\le\sqrt{n}$.

It follows that $c_h\ln(1+\mu) < \frac{4\mu k}{\sqrt{n}} + 2\ln n$, and we obtain the bounds claimed in the corollary by the appropriate choice of μ (see Appendix). ◻

It is perhaps interesting to consider explicitly the bound for some values of k. For $k = n^{1/2-\epsilon}$, the lemma gives a bound of $O(\frac{1}{\epsilon})$; for $k=\sqrt{n}$, we obtain $O(\frac{\log n}{\log\log n})$; for $k=\sqrt{n}\ln n$, the crossing number does not exceed $O(\log n)$. The bounds for $k=\Omega(\sqrt{n}\log n)$ are asymptotically tight. It remains open whether there is always a matching of size $\sqrt{n}$ with constant crossing number.

The constant. We have not presented the best possible constant. Nevertheless, we briefly indicate the best bounds known to the author. Let us first observe that a lower bound of $\sqrt{n}-1$ for spanning trees can be obtained by a slight refinement of the lower bound construction in the beginning of the section. For a positive integer n choose a set G of $\ell = 2\lceil\sqrt{n}\rceil$ lines in general position. Then we assign colors to the cells such that no two adjacent cells (i.e. cells which share a common edge) have the same color. (Choose a fixed point o in one of the cells and color a cell red if for a point p in this cell $\delta_G(o,p)$ is odd, and color the cell blue otherwise.) We place n points in the cells of the larger color class – no two points in the same cell (which is possible since $\frac{1}{2}(\binom{\ell}{2}+\ell+1)\ge n$). Any two of these points are separated by at least 2 lines. Hence, the overall number of crossings between the set of ℓ lines and any spanning tree is at least $2(n-1)$; hence, there is always a line with at least $\frac{2(n-1)}{\ell} \geq \sqrt{n} - 1$ crossings.

Although the bound in Lemma 2.1 is tight, the bound can be improved to $|D_G(p,\sigma)| \geq 3\binom{\sigma+1}{2}$, if p has pseudodistance at least σ to every point in an infinite cell, and if $\sigma \leq \frac{\ell}{3}$; this follows from a result on k-sets proved in [21]. With this bound we can improve the estimates in Lemmas 2.2 and 2.3 to $\frac{2\ell}{\sqrt{3n}}$ and $\frac{2\Delta}{\sqrt{3n}}$ (up to low order terms). The bound in Theorem 2.4 improves to $(\frac{4}{\sqrt{3}} + o(1))\sqrt{n}$. So the optimal constant lies in the range between 1 and 2.31.

3 Applications.

We present three applications of spanning trees, paths, or matchings with low crossing numbers. The first is algorithmic, while the second and third are primarily of combinatorial interest. Nevertheless, the proofs reveal also algorithms for computing the structures whose existence we have proven.

Counting points in halfplanes [16]. Suppose we want to count the points below a nonvertical line from a given point set S, and we have to answer many such queries. Thus it pays off to prepare the points in a data structure.

The structure we use is a simple spanning path $p_1, p_2, \ldots, p_n$ of S with low crossing number c. The edges on the path are enumerated so that edge $\{p_i, p_{i+1}\}$ gets number i. For a nonvertical line h disjoint from S, let I^+ the set of indices of edges $\{p_i, p_{i+1}\}$ with p_i below h and p_{i+1} above h, and let I^- be the set of indices of edges $\{p_i, p_{i+1}\}$ with p_i above h and p_{i+1} below h. Then the number of points in S below h is given by $\sum_{i \in I^+} i - \sum_{i \in I^-} i$, if p_n lies above h, and $n + \sum_{i \in I^+} i - \sum_{i \in I^-} i$, if p_n lies below h. Thus, if we can determine the c_h crossings of line h with the path, then the number of points below h can be computed with c_h additions and subtractions. Here we can invoke a result from [14], which states that the edges of a simple path can be stored with $O(n)$ space, such that the first edge hit by a ray can be computed in $O(\log n)$ time. Clearly, this structure can be used to compute the intersections of a line with a path in $O(k \log n)$ time, where k is the number of intersections.

Theorem 3.1 *Every set S of n points can be stored in $O(n)$ space, such that the number of points in S below any query line can be computed in $O(\sqrt{n} \log n)$ time.* ◻

The structure can readily be used also for counting points in triangles within the same asymptotic time bounds.

Colorings with low discrepancy [27]. We want to color a set of n points in the plane by red and blue, such that every halfplane contains roughly the same number of red and blue points. How well can we achieve that goal? This type of questions are investigated in the field of discrepancy ([30], [9]).

For technical reasons we switch to colors -1 and $+1$. A coloring of a point set S is a mapping $\chi : S \to \{-1,+1\}$. The *discrepancy of* χ is defined as $\max_{h^*} |\chi(S \cap h^*)|$, where $\chi(A) = \sum_{p\in A} \chi(p)$, and the maximum is taken over all halfplanes h^*.

Theorem 3.2 *For every set S of n points there is a coloring χ with discrepancy at most $2\sqrt{2}\, n^{1/4}\sqrt{\ln n} + O(\log n)$.*

Proof. Assume that n is even (if not, we may ignore one point temporarily; the discrepancy grows at most by one by adding it back with an arbitrary color). Let M be a perfect matching on S with crossing number c. We consider the set $\mathcal{C}$ of all colorings χ with $\chi(p)+\chi(q) = 0$ for all $\{p,q\} \in M$. Note that every element of $\mathcal{C}$ has discrepancy at most c. We show that there is a better coloring in $\mathcal{C}$ by considering colorings randomly chosen from $\mathcal{C}$. We need the well-known Chernoff bound (see e.g. [30], [22]) in the following form: If X is the sum of k independent random $\{-1,+1\}$ variables — each variable attains -1 and $+1$ with equal probability —, then $\mathrm{Prob}(|X| > \lambda\sqrt{k}) < 2e^{-\lambda^2/2}$.

Let h be a nonvertical line disjoint from S with c_h crossings in M, and let h^- be the halfplane below h. Set

$$B_h = \{p \in S | p \in h^- \text{ and } h \text{ crosses the edge in } M \text{ containing } p\} .$$

Then $|B_h| = |c_h|$, $\chi(S \cap h^-) = \chi(B_h)$, and for a random χ in $\mathcal{C}$,

$$\mathrm{Prob}(|\chi(B_h)| > \lambda\sqrt{c_h}\,) < 2e^{-\lambda^2/2} \tag{2}$$

If $\lambda = 2\sqrt{\ln n}$ then the bound in (2) becomes $2n^{-2}$. Let $\tilde{H}_S$ be a representative set of lines with $|\tilde{H}_S| \leq \binom{n}{2} < n^2/2$. Thus there is a coloring χ_0 in $\mathcal{C}$ with $\chi_0(S \cap h^-) \leq 2\sqrt{c_h \ln n} \leq 2\sqrt{c \ln n}$ for all h in $\tilde{H}_S$; this coloring χ_0 is good for all (open or closed) halfplanes below lines. We have $|\chi(A)| = |\chi(S - A)|$ for all $\chi \in \mathcal{C}$ and all $A \subseteq S$, which takes care of halfplanes above lines. The lemma follows, since there is a perfect matching with $c = 2\sqrt{n} + O(n^{1/4}\sqrt{\log n})$, see Corollary 2.6. ◻

[8] proves a lower bound of $\Omega(n^{1/4-\epsilon})$, for any $\epsilon > 0$, for the discrepancy of colorings for halfplanes.

Mutually avoiding segments [28]. Two closed line segments are called *avoiding*, if the lines supporting the segments intersect outside both segments. The following result was first proved in [7]; the simple proof below was presented in [28].

Theorem 3.3 *Every set S of n points in general position allows $\frac{1}{8}\sqrt{n} - O(n^{1/4}\sqrt{\log n})$ mutually avoiding line segments with endpoints in S.*

Proof. Let $p_0, p_1, \ldots, p_{n-1}$ be a spanning path with crossing number $c-1$. For convenience add also the edge $\{p_{n-1}, p_0\}$ to obtain a spanning cycle with crossing number c. We show that among the n edges on this path there are $\lceil \frac{n}{2c+1} \rceil$ edges which define mutually avoiding line segments. To this end consider the graph which has the set L of line segments $\overline{p_{i-1}p_i}$, $i = 1, 2, \ldots, n-1$, and $\overline{p_{n-1}p_0}$, as vertices. Two vertices are adjacent, if their corresponding line segments are not avoiding. A line containing a line segments s in L intersects at most c of the line segments in $L - \{s\}$ (it's at most c including the adjacent segments on the cycle!). Consequently, our graph has at most cn edges. A graph with n vertices and cn edges has an independent set (i.e. a set of vertices where no two are adjacent) of cardinality $\lceil \frac{n}{2c+1} \rceil$ (the existence of a $\lceil \frac{n^2}{2m+n} \rceil$ size independent set in a graph with n vertices and m edges follows from Turan's theorem, cf. [10]). But an independent set in this graph corresponds to a set of mutually avoiding line segments; the theorem follows due to the bounds on c previously derived. □

It is not known whether there are point sets which do not allow a linear number of mutually avoiding line segments.

4 Construction.

The proof of existence of spanning trees with low crossing numbers in Theorem 2.4 describes an algorithm which can be implemented in polynomial time. A number of more efficient algorithms can be found in the literature [20, 2, 25, 23, 3]. We will present some of the basic ingredients of these algorithms, which will lead us to a randomized algorithm which computes in expected $O(n\sqrt{n \log n})$ time a spanning tree whose crossing number does not exceed $O(\sqrt{n \log n})$ with high probability.

The first step in making an algorithm more efficient is to reduce the number of lines which have to be considered in a construction.

Test sets. Given a set S of n points and two nonvertical lines g and h, we define $\delta^*_S(g, h) = a + \frac{b}{2}$, where a is the number of points from S in the double wedge defined by g and h, and b is the number of points from S which lie on exactly one of the lines g and h. Similar to δ on points, δ^* is a pseudometric on lines. In fact, if we denote by S^* the lines dual to the points in S, then $\delta^*_S(g, h) = \delta_{S^*}(g^*, h^*)$.

For a real number σ, we call a set H of lines a *σ-test set for S*, if for every line g disjoint from S, there is a line $h \in H$ with $\delta^*_S(g, h) \leq \sigma$.

Lemma 4.1 *Let S be a set of n points and let H be a σ-test set for S. If the maximal crossing number of a line in H in a spanning path on S is C, then the crossing number of this path (for all lines) is at most $C + 2\sigma$.*

Proof. For any two lines g and h, observe that if g crosses an edge which is not crossed by h, then one of the two endpoints of this edge has to lie in th double wedge of g and h, or on g. Since every point is incident to at most two edges on a *path*, we easily get that the respective crossing numbers c_g and c_h satisfy $|c_g - c_h| \leq 2\delta^*_S(g, h)$. The lemma is an immediate consequence of this fact. □

Lemma 4.2 *Let S be a set of n points and let σ be an integer with $0 \leq \sigma \leq n$.* **(i)** *There exists a σ-test set of at most $4(\frac{n}{\sigma})^2$ lines.* **(ii)** *If S is in general position, then,* *for every positive real λ, a set of lines obtained by connecting at least $(2+\lambda)(\frac{n}{\sigma})^2 \ln n$ random pairs of points in S is a σ-test set with probability at least $1 - n^{-\lambda}$.*

Proof. We prefer to dualize the scenario. In the dual environment statement (i) claims that for a set $G\,(= S^*)$ of $\ell\,(= n)$ lines, there exists a set Q of $4(\frac{\ell}{\sigma})^2$ points, such that every point p disjoint from G has a point $q \in Q$ with $\delta_G(p,q) \leq \sigma$. Choose Q as a maximal set of points, where any two points have pseudodistance δ_G greater than σ. Lemma 2.2 implies that Q contains at most $(\frac{2\ell}{\sigma})^2$ points, and the maximality of Q guarantees the desired property.

For a proof of (ii), we have to consider a set R of r random vertices in $\mathcal{A}(G)$, G a set of ℓ lines in general position. For any point p disjoint from G, a random vertex has pseudodistance at most σ from p with probability $|D_G(p,\sigma)|/\binom{\ell}{2} > (\frac{\sigma}{\ell})^2$ (use Lemma 2.1). Hence, the probability that all points in R have pseudodistance more than σ from p is less than

$$\left(1 - \left(\frac{\sigma}{\ell}\right)^2\right)^r \leq e^{-r\sigma^2/\ell^2} \ . \tag{3}$$

For $r \geq (2+\lambda)(\frac{\ell}{\sigma})^2 \ln \ell$, the expression in (3) is bounded by $\ell^{-2-\lambda}$. Let P be a set of $m = \binom{\ell}{2} + \ell + 1$ points, one in each cell of $\mathcal{A}(G)$. Then with probability at most $m\ell^{-2-\lambda} \leq \ell^{-\lambda}$ there is a point in P which has pseudodistance more than σ from all points in R (for $\ell \geq 2$, $m \leq \ell^2$). Since every point disjoint from G has a point in P at pseudodistance 0, the lemma is proved. □

The algorithm. Let G be a set of lines, and let p be a point. For a nonvertical line h (not necessarily in G), we say that *h sees p* (and *p sees h*) *in* $\mathcal{A}(G)$, if p lies on or above h, and the closed vertical segment connecting h and p is disjoint from all lines in $G - \{h\}$; (if p lies on h, then p sees h if and only if p lies on no line in $G - \{h\}$). Thus a point p which lies on a single line g in G sees g and no other line, and if p is contained in two or more lines in G, then p sees no line at all. Every point p sees at most one of the lines in G.

The algorithm proceeds now as follows. We assume that the set S of n points is in general position, and that $n \geq 2$. First we take a random sample T of n lines connecting points in S; this will be a σ-test set, for $\sigma \leq 2\sqrt{n \ln n}$, with probability $1 - n^{-2}$. Then we construct a set $F \subseteq T$ of $\tau \leq \sqrt{n \ln n}$ lines such that no line in $T - F$ sees more than $\kappa \leq 2e\sqrt{n \ln n}$ points from S in $\mathcal{A}(F)$ (the construction of F will be described below). We add to F a horizontal line h_0, which lies below all points in S. Each point p in S is projected vertically on a line from F directly below (or through) p; this gives a set S' of n projections. For $g \in F$, let S'_g be the points in S' which lie on g; if a point in S' lies on several lines in F, then we put it only in one set S'_g.

We add two extra vertical lines h^- and h^+ which lie to the left (right, respectively) of all points in S. On every line g connect all points in S'_g by a path along g, starting at the intersection of g with h^- and ending at the intersection of g with h^+. Connect these paths via edges on h^- and h^+ so that no line intersects more than two of these extra edges. Note that the resulting spanning path P' has crossing number $3+\tau$ at most ('3' accounts for crossings on h_0, h^-, and h^+). Now we consider the *vertical* edges connecting the points in $S-S'$ to their projections in S'. A line $g \in T-F$ crosses such a vertical edge only if it sees the upper endpoint in $\mathcal{A}(F)$, or it contains the lower endpoint.

For a line $g \in T$, consider a line g' parallel to and below g, but sufficiently close so that no point in $(S' \cup S)-g$ changes its relative position to g' (compared to g). For all lines $g \in T$, g' crosses at most $3+\tau$ edges in P'. If $g \in F$, then g' crosses no vertical edge, and if $g \in T-F$, then g crosses at most κ vertical edges.

In order to obtain a path on S we walk along P' with excursions along vertical edges, and we enumerate the points in S as we meet them on this walk. For any line $g \in T$, the primed version g' crosses at most $3+\tau+2\kappa$ edges, and since $\delta^*_S(g,g') \leq 1$ (recall that we assume S to be in general position), no line in T has crossing number exceeding $5+\tau+2\kappa$. Consequently, the crossing number of the path is at most $5+\tau+2\kappa+2\sigma$ (by Lemma 4.1), which is at most $5+(5+4e)\sqrt{n \ln n} = O(\sqrt{n \log n})$ with probability $1-n^{-2}$.

It remains to show how a set F obscuring many visibilities is constructed.

Obscuring sets.

Lemma 4.3 *Let S be a set of n points, and let G be a finite set of lines. For a random set R of r lines in G, and for a random line g in $G-R$, the expected number of points in S seen by g in $\mathcal{A}(R)$ is at most $\frac{n}{r+1}$.*

Proof. We employ backwards analysis, cf. [29]. Observe that g sees a point p in $\mathcal{A}(R)$ if and only if g sees p in $\mathcal{A}(R \cup \{g\})$. Thus the quantity we are interested in is the same as the expected number of points from S seen by a random line $g \in R'$ in $\mathcal{A}(R')$, with R' a random set of $r+1$ lines in G. Since every point in S sees at most one line in R', this number is bounded by $\frac{n}{r+1}$. ◻

We will use the lemma to make the following conclusion: If we choose r lines R at random, then with probability at least $\frac{1}{2}$ the expected number of points seen by a line in $G-R$ is at most $\frac{2n}{r+1}$; in this case at most $\frac{|G-R|}{e}$ lines see more than $\frac{2en}{r+1}$ points (we use Markov's inequality twice).

We start the construction of F by choosing a random sample R_0 of $r = \lfloor \sqrt{\frac{n}{\ln n}} \rfloor$ lines in $H_0 = T$. We determine the set $H_1 \subseteq H_0 - R_0$ of lines which see more than $\frac{2en}{r+1} \leq 2e\sqrt{n \ln n}$ points from S in $\mathcal{A}(R_0)$. If $|H_1| > |H_0|/e$ — which happens with probability less than $\frac{1}{2}$ —, then we choose a new sample R_0 from H_0 until $|H_1| \leq |H_0|/e$ holds. In the same way we produce a set R_1 of r lines in H_1, such that the set $H_2 \subseteq H_1 - R_1$ of lines which see more than $\frac{2en}{r+1}$ points in $\mathcal{A}(R_1)$ satisfies $|H_2| \leq |H_1|/e$. If we continue like this, we have exhausted all lines in T after at most $\lceil \ln \frac{|T|}{r} + 1 \rceil \leq \ln n$ steps (at least for n large enough), and the expected number of samples we took is at most twice this number. The union F of all R_i's constitutes a set of at most $r \ln n \leq \sqrt{n \ln n}$ lines, and no line in $T-F$ sees more than $2e\sqrt{n \ln n}$

points in $\mathcal{A}(F)$. (The constants can be decreased at the cost of a larger constant in the running time.)

If we are interested in the existence of F only, then we may choose '2' as 1.

Lemma 4.4 *Let S be a set of n points and let G be a set of ℓ lines. For every positive integer $r \leq \min\{n, \ell\}$, there is a set F of $r\lceil \ln \frac{\ell}{r} + 1 \rceil$ lines in G, such that no line in $G - F$ sees more than $\frac{e\,n}{r+1}$ points of S in $\mathcal{A}(F)$.* ◻

Time complexity. What is the time complexity of the construction of F? When we choose a random sample R of r lines then we construct the arrangement $\mathcal{A}(R)$ in $O(r^2)$ time, cf. [19]. Then, for every point in S, we determine the cell the point is contained in: We simply determine the line in R directly below a point p by looking at all lines (in $O(nr)$ time for all points). Then, for each line $g \in R$, we look at the points which have this line below and determine the respective edges of the arrangement directly below these points (this works again in $O(nr)$, if every point checks all edges on 'its' line). As we have located all points in their cells, we provide a list of points in each cell sorted by x-coordinate. Now we want to compute the number of points seen by a line $h \notin R$. We determine the cells intersected by h by threading the line through the arrangement in $O(r)$, cf. [19]. In each cell visited, we take the x-coordinates of the first and last point of h in the closure of this cell. h can see only points in this cell which have their x-coordinates in this range. In the sorted lists we can determine these points in $O(\log n + k')$, k' the number of points in this range. Similar to the proof of Lemma 4.3, we can show that the expected sum of all k' over all cells intersected by h is at most $\frac{2n}{r+1}$. So the expected time spent for a line h is $O(r \log n + \frac{n}{r+1})$. Altogether, if ℓ lines have to be checked, we spend time $O(nr + \ell(r \log n + \frac{n}{r+1})) = O(n\sqrt{\frac{n}{\log n}} + \ell(\sqrt{n \log n}))$. The expected number of times we have to handle such a set R is $O(\log n)$, and the number of lines to be checked decreases geometrically. Hence, the overall expected time for constructing F is $O(n\sqrt{n \log n})$. The spanning path can easily be obtained from the arrangement $\mathcal{A}(F)$ within this time bound.

Theorem 4.5 *There is a randomized algorithm which computes for any set of n points in general position a spanning path in expected $O(n\sqrt{n \log n})$ time, such that the crossing number does not exceed $O(\sqrt{n \log n})$ with probability $1 - n^{-2}$.* ◻

With some more sophistication, the algorithm can be tuned to have close to linear running time (see [25] for some of the ideas required). Test sets are used in most efficient constructions of spanning trees with low crossing numbers [25,23,2]. Efficient (deterministic) constructions of test sets are described in [26]. The idea of repeated sampling on 'bad' lines for the construction of obscuring sets is taken from [15].

A deterministic $O(n\sqrt{n}\log^2 n)$ algorithm which gives a spanning tree with $O(\sqrt{n})$ crossing number is described in [23]. [3] can produce a tree with crossing number $O(n^{1/2+\epsilon})$ in time $O(n^{1+\epsilon})$ for any $\epsilon > 0$, and they describe how such a tree can be maintained under a sequence of insertions and deletions. So-called simplicial partitions ([26], see Section 5) can be used to obtain a spanning tree with crossing number $O(\sqrt{n})$ in time $O(n^{1+\epsilon})$ for any $\epsilon > 0$ (where the constant in the crossing number depends on ϵ), [24].

5 Discussion.

The result on spanning trees generalizes to higher dimensions and other geometric objects: For every set of n points in d-space there is a spanning tree, such that no hyperplane intersects the straight line embedding of the tree in more than $O(n^{1-1/d})$ points, which is tight. The proof of the general result starts off by providing a higher-dimensional counterpart of Lemma 2.1, and then proceeds almost verbatim as in the planar case. Similarly, we can always find a tree which has $O(n^{1-1/d})$ crossings with any ball, if we define that a ball crosses an edge if exactly one endpoint of the edge lies in the ball.

For a set system $(X, \mathcal{R})$, $\mathcal{R} \subseteq 2^X$, we can also consider spanning trees on finite subsets A of X. We say that a set $R \in \mathcal{R}$ crosses an edge $\{x, y\}$ of the tree, if $|R \cap \{x, y\}| = 1$. Then it is possible to prove the existence of a spanning trees with crossing number $O(n^{1-1/d})$, where d is some combinatorial parameter associated with the set system (related to the VC-dimension); details can be found in [16].

An important extension of matchings with low crossing numbers, *simplicial partitions*, were introduced in [26]. In the planar version, for a set S of n points, such a partition consists of pairs (t_i, S_i), $i = 1, 2, \ldots, m$, where the t_i's are open triangles or line segments with $t_i \supseteq S_i$, and the S_i's form a partition of S. It is shown that for any r there is a simplicial partition such that $m = O(r)$, the cardinalities of the S_i's are roughly balanced ($|S_i| \leq \frac{2n}{m}$ for all i, to be precise), and no line intersects more than $O(\sqrt{m})$ of the t_i's. Note that perfect matchings with low crossing numbers are related to simplicial partitions with $m = \frac{n}{2}$. Simplicial partitions can be efficiently constructed, and they allow improvements in many algorithmic applications, [26].

We conclude by stating two open problems.

Problem 1 *Is there a constant C, such that every set of n points in the plane has a matching of size $\sqrt{n}$ whose straight line embedding is intersected in no more than C edges by any line disjoint from the points?*

Corollary 2.6 gives a bound of $O(\frac{\log n}{\log \log n})$ on C; a constant number of intersections can be guaranteed, if a matching of size $n^{1/2-\epsilon}$ is required, for any fixed $\epsilon > 0$.

Problem 2 *Given n points S and n nonvertical lines G in the plane, is there always a set F of $O(\sqrt{n})$ lines in G, such that no line in $G - F$ sees more than $\sqrt{n}$ points of S in $\mathcal{A}(F)$; a line $h \in G - F$ sees a point p in $\mathcal{A}(F)$ if p lies on or above h, and the closed vertical segment connecting p and h is disjoint from all lines in F?*

Lemma 4.4 gives a bound of $O(\sqrt{n} \log n)$ on the size of F.

References

[1] Pankaj Agarwal. Ray shooting and other applications of spanning trees with low stabbing number. In *Proc. 5th Annual ACM Symposium on Computational Geometry*, pages 315–325, 1989.

[2] Pankaj Agarwal. *Intersection and Decomposition Algorithms for Planar Arrangements.* Cambridge University Press, 1991.

[3] Pankaj Agarwal and Micha Sharir. Applications of a new space partitioning technique. Manuscript, 1991.

[4] Pankaj Agarwal, Marc van Kreveld, and Mark Overmars. Intersection queries for curved objects. In *Proc. 7th Annual ACM Symposium on Computational Geometry*, pages 41–50, 1991.

[5] Noga Alon and Daniel Kleitman. Piercing convex sets and the Hadwiger Debrunner (p, q)-problem. Manuscript, 1991.

[6] Noga Alon and Nimrod Megiddo. Parallel linear programming in fixed dimensions almost surely in constant time. In *Proc. 31st Annual IEEE Symposium on Foundations of Computer Science*, pages 574–582, 1990.

[7] Boris Aronov, Paul Erdös, Wayne Goddard, Daniel Kleitman, Michael Klugerman, Janos Pach, and Leonard Schulman. Crossing families. In *Proc. 7th Annual ACM Symposium on Computational Geometry*, pages 351–356, 1991.

[8] József Beck. Quasi-random 2-colorings of point sets. Technical Report 91-20, DIMACS, 1991.

[9] József Beck and William Chen. *Irregularities of Distributions.* Cambridge University Press, 1987.

[10] Béla Bollobás. *Extremal Graph Theory.* Academic Press, 1978.

[11] Bernard Chazelle. Tight bounds on the stabbing number of spanning trees in Euclidean space. Technical Report CS-TR-155-88, Princeton University, Department of Computer Science, 1988.

[12] Bernard Chazelle. Polytope range searching and integral geometry. *J. Amer. Math. Soc.*, 2:637–666, 1989.

[13] Bernard Chazelle and Joel Friedman. A deterministic view of random sampling and its use in geometry. In *Proc. 29th Annual IEEE Symposium on Foundations of Computer Science*, pages 539–549, 1988.

[14] Bernard Chazelle and Leonidas Guibas. Visibility and intersection problems in plane geometry. *Discrete Comput. Geom.*, 4:551–589, 1989.

[15] Bernard Chazelle, Micha Sharir, and Emo Welzl. Quasi-optimal upper bounds for simplex range searching and new zone theorems. In *Proc. 6th ACM Symp. on Comp. Geom.*, pages 23–33, 1990.

[16] Bernard Chazelle and Emo Welzl. Quasi-optimal range searching in spaces of finite VC-dimension. *Discrete Comput. Geom.*, 4:467–489, 1989.

[17] Siu Wing Cheng and Ravi Janardan. Space-efficient ray-shooting and intersection searching: Algorithms, dynamization, and applications. In *Proc. 2nd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 7–16, 1991.

[18] Kenneth Clarkson. Las vegas algorithms for linear and integer programming when the dimension is small. Manuscript, 1989.

[19] Herbert Edelsbrunner. Algorithms in combinatorial geometry. *Springer-Verlag, Heidelberg, Germany*, 1987.

[20] Herbert Edelsbrunner, Leonidas Guibas, John Herschberger, Raimund Seidel, Micha Sharir, Jack Snoeyink, and Emo Welzl. Implicitly representing arrangements of lines or segments. *Discrete Comput. Geom.*, 4:433–466, 1989.

[21] Herbert Edelsbrunner, N. Hasan, Raimund Seidel, and X. J. Shen. Circles through two points that always enclose many points. *Geom. Dedicata*, 32:1–12, 1989.

[22] Torben Hagerup and Christine Rüb. A guided tour of Chernoff bounds. *Inform. Process. Lett.*, 33:305–308, 1990.

[23] Jiří Matoušek. More on cutting arrangements and spanning trees with low crossing number. Technical Report B 90-02, Freie Universität Berlin, Fachbereich Mathematik, Institut für Informatik, 1990.

[24] Jiří Matoušek, 1991. Private communication.

[25] Jiří Matoušek. Range searching with efficient hierarchical cuttings. Manuscript, 1991.

[26] Jiří Matoušek. Spanning trees with low crossing number. *RAIRO Inform. Théor. Appl.*, 6:103–123, 1991.

[27] Jiří Matoušek, Emo Welzl, and Lorenz Wernisch. Discrepancy and ϵ-approximations for bounded VC-dimension. In *Proc. 32nd Annual IEEE Symposium on Foundations of Computer Science*, pages 424–430, 1991.

[28] János Pach. Drawing graphs. Talk presented at the Dagstuhl seminar on 'Computational Geometry', 1991.

[29] Raimund Seidel. Backwards analysis of randomized geometric algorithms. Manuscript, 1991.

[30] Joel Spencer. *Ten Lectures on the Probabilistic Method.* Society for Industrial and Applied Mathematics, 1987.

[31] Emo Welzl. Partition trees for triangle counting and other range searching problems. In *Proc. 4th Annual ACM Symposium on Computational Geometry*, pages 23–33, 1988.

Appendix: Optimal choice of reweighting factor $1 + \mu$.

We want to estimate $\min_{\mu>0} f(\mu)$ for

$$f(\mu) = \frac{1}{\ln(1+\mu)}(a\mu + b) \ , \tag{4}$$

with $a, b > 0$. The first derivative of f is

$$f'(\mu) = \frac{a\ln(1+\mu) - (a\mu+b)(1+\mu)^{-1}}{\ln^2(1+\mu)} \ .$$

So a local extremum (which obviously has to be a minimum) is achieved when

$$a\mu + b = a(1+\mu)\ln(1+\mu) \ , \tag{5}$$

or, equivalently, when

$$e^x(1-x) = 1 - c \ , \tag{6}$$

where we write x short for $\ln(1+\mu)$, and c short for b/a. Equality (6) has exactly one solution.

Let us first consider the case $c \le 1$. Then, for $x = \sqrt{c}$,

$$e^x(1-x) \ge (1+x)(1-x) = (1-c) \ ,$$

and, for $x = \sqrt{2c}$,

$$e^x(1-x) = 1 - \sum_{i=1}^{\infty} \frac{(i-1)x^i}{i!} < 1 - \frac{x^2}{2} = 1 - c \ .$$

Consequently, (6) is satisfied for some x in the range $\sqrt{c} \le x < \sqrt{2c}$; so the optimal μ has to be chosen such that

$$\sqrt{b/a} \le \ln(1+\mu_{opt}) < \sqrt{2b/a}, \text{ for } b \le a \ .$$

If we substitute (5) into (4), then we get for the optimal μ that $f(\mu) = a(1+\mu)$, and so

$$min_{\mu>0} f(\mu) < ae^{\sqrt{2b/a}} = a + O(\sqrt{ab}), \text{ for } b \le a \ ,$$

since $e^y \le 1 + (\frac{e^{\sqrt{2}}-1}{\sqrt{2}})y$ for $0 \le y \le \sqrt{2}$. (For $a = b$, we get $\min_{\mu>0} f(\mu) = ea$.)

If $c > 1$, then we rewrite (6) as

$$z(\ln z - 1) = c - 1 \ ,$$

where $z = \mu + 1$. We assume actually that c is sufficiently large, say, $c \ge e$. For $z = \frac{ec}{\ln c}$,

$$\frac{ec}{\ln c}(\ln ec - \ln\ln c - 1) = c(e - \frac{e\ln\ln c}{\ln c}) > c - 1 \ ,$$

and, for $z = \frac{c-1}{\ln c}$,

$$\frac{c-1}{\ln c}(\ln(c-1) - \ln\ln c - 1) < (c-1)(1 - \frac{\ln\ln c + 1}{\ln c}) < c - 1 \ .$$

Therefore, μ has to be chosen such that

$$\frac{b/a - 1}{\ln(b/a)} < 1 + \mu_{opt} < \frac{eb}{a\ln(b/a)} \ , \ \text{for } b \geq ea \ ,$$

which implies

$$\min_{\mu>0} f(\mu) < \frac{eb}{\ln(b/a)}, \ \text{for } b \geq ea \ .$$

□