Tail Estimates for the Efficiency of Randomized Incremental Algorithms for Line Segment Intersection *

K. Mehlhorn
Max-Planck-Institut für Informatik
Saarbrücken

M. Sharir
School of Mathematical Sciences
Tel-Aviv University
and
Courant Institute of Mathematical Sciences
New York University

E. Welzl
Institut für Informatik, Fachbereich Mathematik
Freie Universität Berlin

27 May 1991
revised November 25 1992

Abstract

We give tail estimates for the efficiency of some randomized incremental algorithms for line segment intersection in the plane. In particular, we show that there is a constant C such that the probability that the running times of algorithms due to Mulmuley [Mul88] and Clarkson and Shor [CS89] exceed C times their expected time is bounded by $e^{-\Omega(m/(n \ln n))}$ where n is the number of segments, m is the number of intersections, and $m \geq n \ln n \ln^{(3)} n$.

1 Introduction

Randomized incremental algorithms have received considerable attention recently; cf. [CS89], [Mul88] and [BDS$^+$92]. They solve a large number of geometric problems, including the construction of Voronoi diagrams and convex hulls and the intersection of line segments, in optimal expected time and space. In this paper we discuss the randomized line segment intersection algorithms of Clarkson and Shor [CS89], Mulmuley [Mul88] and Boissonnat et al. [BDS$^+$92]

*Supported in part by GIF project No. I-135-113.6/89 and by the ESPRIT II Basic Research Actions Program of the EC under contract no. 7141(project ALCOM). A preliminary version of this paper was presented at the 3rd ACM-SIAM Symposium on Discrete Algorithms, 1992.

and prove a tail estimate for the running time of two algorithms in [Mul88] and [CS89] and for the space efficiency of the algorithms in [BDS$^+$92]. More precisely, we show that there is a constant C such that the probability that the running time (space efficiency) exceeds C times its expected value is $e^{-\Omega(m/(n \ln n))}$, where n is the number of line segments, m is the number of intersections, and $m \geq n \ln n \ln^{(3)} n$. The tail estimate is shown is Section 3. In Section 2 a simple probabilistic lemma is proven; it extends a lemma shown in [CMS92]. In the preliminary version of this paper we only claimed a tail estimate for the space efficiency of the algorithm in [BDS$^+$92]. Jirka Matoušek and Raimund Seidel have pointed out to us that our method also implies a tail estimate for the running time of two of the intersection algorithms. A tail estimate for the running time of Mulmuley's algorithm was also claimed in [MS92]. Unfortunately, the argument in [MS92] is flawed (personal communication by the authors).

2 A Probabilistic Lemma

Let $\mathbb{N}$ denote the set of nonnegative integers and let $\mathbb{R}_{\geq 0}$ denote the set of nonnegative reals. For functions $M : \mathbb{N} \to \mathbb{R}_{\geq 0}$ and $d : \mathbb{N} \to \mathbb{R}_{\geq 0}$ and integers n and r with $n \geq r \geq 0$, call a rooted tree T an *(n, r)-tree respecting M and d* if either $r = 0$ and T consists of a single node, or $r > 0$, the root of T has n subtrees each of which is an $(n-1,\ r-1)$-tree respecting M and d, and the n edges incident to the root are labeled with non-negative *weights* d_i so that $d_i \leq d(n)$ for $1 \leq i \leq n$ and $\sum_{1 \leq i \leq n} d_i \leq M(n)$.

For a path π in T, let $X = X_\pi$ be the sum of the weights of the edges along the path. The uniform distribution on the $n(n-1) \cdots (n-r+1)$ paths in T makes X a random variable with expectation

$$\mathbf{E}(X) \leq \sum_{0 \leq i \leq r-1} \frac{M(n-i)}{(n-i)} \ .$$

Lemma 1 *: For all $t > 0$ and $B \geq 0$:*

$$\mathbf{Prob}\ (X \geq B) \leq \exp\left(-tB + \sum_{0 \leq i \leq r-1} \frac{M(n-i)}{(n-i)d(n-i)}\ (e^{td(n-i)} - 1) \right) \ .$$

Remark: Lemma 1 is related to Azuma's inequality [ASE91, Section 7] for martingales but does not follow from that inequality. Note that one can easily derive a martingale from the tree T: Label each node v of T by $E[X_\pi | \pi$ goes through $v]$ and for i, $0 \leq i < r$, let Y_i be the label of a random node of depth i. Then $Y_0, Y_1, \ldots, Y_r$ is a martingale. The proof of Lemma 1 is an adaptation of the standard proof for Hoeffding's inequality, cf. [Hoe63]. The case $r = n$ and $d(i) = M(i)$ for all i was previously treated in [CMS92].

Proof: Let $I_0, \ldots, I_{r-1}$ be independent random variables where I_i, $0 \leq i \leq r-1$, is uniformly distributed on $[1..n-i]$. The variables $I_0, \ldots, I_{r-1}$ select a path π of T in a natural way: I_0 selects the first edge of π, I_1 selects the second edge of π, and so on. For $i \in [0..r-1]$ let X_i be the weight of the $(i+1)$-st edge on the path π. Then $X = \sum_{0\leq i\leq r-1} X_i$ and, for all $t > 0$,

$$\mathbf{Prob}(X \geq B) = e^{-tB}e^{tB}\mathbf{Prob}(e^{tX} \geq e^{tB})$$

$$\leq e^{-tB}\mathbf{E}(e^{tX}) = e^{-tB}\mathbf{E}\left(\prod_{0\leq i\leq r-1} e^{tX_i}\right),$$

using Markov's inequality in step two. We now prove for all j, $0 \leq j \leq r-1$, and all integers $i_0, i_1, \ldots, i_{r-1}$ with $i_l \in [1..n-l]$

$$\mathbf{E}\left(\prod_{j\leq i\leq r-1} e^{tX_i} | I_0 = i_0, \ldots, I_{j-1} = i_{j-1}\right)$$

$$\leq \exp\left(\sum_{j\leq i\leq r-1} \frac{M(n-i)}{(n-i)d(n-i)}(e^{td(n-i)} - 1)\right).$$

For $j = 0$ this is the claim of the lemma. We use backward induction on j. For $j = r$ both sides are equal to one. So assume $j \leq r-1$. We have

$$\mathbf{S} := \mathbf{E}\left(\prod_{j\leq i\leq r-1} e^{tX_i} | I_0 = i_0, \ldots, I_{j-1} = i_{j-1}\right)$$

$$= \sum_{1\leq l\leq n-j} \frac{1}{n-j} \cdot \mathbf{E}\left(\prod_{j\leq i\leq r-1} e^{tX_i} | I_0 = i_0, \ldots, I_{j-1} = i_{j-1}, I_j = l\right).$$

Let $d_1, \ldots, d_{n-j}$ be the weights of the edges emanating from the node corresponding to $i_0, i_1, \ldots, i_{j-1}$. Then $I_0 = i_0, \ldots, I_{j-1} = i_{j-1}, I_j = l$ implies $X_j = d_l$. Thus

$$\mathbf{S} = \sum_{1\leq l\leq n-j} \frac{e^{td_l}}{n-j} \mathbf{E}\left(\prod_{j+1\leq i\leq r-1} e^{tX_i} | I_0 = i_0, \ldots, I_{j-1} = i_{j-1}, I_j = l\right)$$

$$\leq \exp\left(\sum_{j+1\leq i\leq r-1} \frac{M(n-i)}{(n-i)d(n-i)}(e^{td(n-i)} - 1)\right) \cdot \sum_{1\leq l\leq n-j} \frac{e^{td_l}}{n-j},$$

where the inequality comes from the induction hypothesis. Since $d_l \leq d(n-j)$ for $1 \leq l \leq n-j$ and $\sum_{1\leq l\leq n-j} d_l \leq M(n-j)$, and by the convexity of the exponential function, the last expression is maximized when $\lfloor M(n-j)/d(n-j)\rfloor$ weights d_l are equal to $d(n-j)$, one weight is equal to $M(n-j) - \lfloor M(n-j)/d(n-j)\rfloor d(n-j)$ and the remaining weights are equal to zero. Let $x = M(n-j)/d(n-j) - \lfloor M(n-j)/d(n-j)\rfloor$. Then $0 \leq x < 1$ and hence, by the convexity of the exponential function, $\exp(x \cdot t \cdot d(n-j)) \leq 1 - x + x\exp(t \cdot d(n-j))$. It follows that

$$\sum_{1\leq l\leq n-j} \frac{e^{td_l}}{n-j} \leq \frac{\frac{M(n-j)}{d(n-j)} e^{td(n-j)} + n - j - \frac{M(n-j)}{d(n-j)}}{n-j}$$

$$= 1 + \frac{M(n-j)}{(n-j)\ d(n-j)}\ (e^{td(n-j)}\ - 1)$$

$$\leq \exp\left(\frac{M(n-j)}{(n-j)d(n-j)}(e^{td(n-j)} - 1)\right)\ ,$$

where the last inequality follows from $1 + y \leq e^y$ for all real y. This completes the induction step and the proof of the lemma. ∎

If $d(\cdot)$ is a non-decreasing function then the upper bound of Lemma 1 becomes more manageable.

Theorem 1 *Let $d(n)$ be a non-decreasing function of n and let*

$$A \geq \sum_{0 \leq i \leq r-1} \frac{M(n-i)}{n-i}\ .$$

Then

$$\mathbf{Prob}(X \geq B) \leq \left(\frac{e}{1 + B/A}\right)^{B/d(n)}$$

for all $B \geq 0$.

Proof: If $d(\cdot)$ is a nondecreasing function, then

$$\frac{e^{td(n-i)} - 1}{d(n-i)} \leq \frac{e^{td(n)} - 1}{d(n)}$$

since $(e^x - 1)/x$ is an increasing function of x for $x > 0$ (compute the derivative). This simplifies Lemma 1 to

$$\mathbf{Prob}\ (X \geq B) \leq \exp(-tB + A \cdot \frac{e^{td(n)} - 1}{d(n)})\ .$$

Put $t = (1/d(n)) \ln(B/A + 1)$. Then

$$\mathbf{Prob}\ (X \geq B) \leq \exp\left(-\frac{B}{d(n)} \ln \frac{A+B}{eA}\right)\ .$$

∎

3 Tail Estimates for the Efficiency of Randomized Line Segment Intersection Algorithms

Randomized incremental algorithms for line segment intersection were described by Clarkson and Shor [CS89], Mulmuley [Mul88], and Boissonnat et al. [BDS+92]. All these algorithms have expected running time $O(n \log n + m)$, where n is the number of segments and m is the number of intersections. We first discuss an algorithm due to Mulmuley [Mul88] and prove a tail estimate for its time complexity. At the end of the section we briefly discuss the other algorithms.

Here is a brief account of Mulmuley's algorithm (called second algorithm in [Mul88]). Let S be a set of non-vertical line segments in the plane. We assume for simplicity that the x–coordinates of all endpoints and all intersection points are distinct. For a subset $R \subseteq S$ let $\mathcal{T}(R)$ be the trapezoidal decomposition defined by the segments in R plus the endpoints of the segments in $S \setminus R$, cf. Figure 1. $\mathcal{T}(R)$ is obtained from the segments in R and the endpoints of the segments in $S \setminus R$ by emanating an upward and a downward vertical ray from each endpoint and each intersection point. The rays extend up to the next segment. We call this a *vertical bar*. When we talk about an *edge*, then we mean an edge in the arrangment of R – such an edge may be touched by several vertical bars. Mulmuley constructs $\mathcal{T}(S)$ incrementally starting with $\mathcal{T}(\emptyset)$ and adding the segments in S one by one in random order. For $R \subseteq S$ and $s \in S \setminus R$ the decomposition $\mathcal{T}(R \cup \{s\})$ is constructed from $\mathcal{T}(R)$ as follows. Starting at an endpoint of s walk along s through $\mathcal{T}(R)$. The vertical bar extended from this endpoint determines the first trapezoid of $\mathcal{T}(R)$ entered by s. If s leaves a trapezoid through a vertical (bar) boundary then the trapezoid entered by s can be determined in constant time since the vertical boundary of a trapezoid is incident to at most two other trapezoids (by our general position assumption). If s leaves through a segment (edge) boundary the situation is more involved. Assume for concreteness that s leaves a trapezoid T through its upper boundary contained in edge e and that e is part of the segment $t \in R$. The walk then proceeds from the intersection $s \cap t$ to one of the endpoints of e (this is either an endpoint of t or an intersection $t \cap s'$ for some $s' \in R$) and then along the other side of t back to $s \cap t$, cf. Figure 1. When the walk reaches $s \cap t$ at the other side of t the trapezoid entered by s is known and the walk proceeds along s. At the end of the walk all trapezoids intersected by s are known and $\mathcal{T}(R \cup \{s\})$ is readily constructed from $\mathcal{T}(R)$. It is not hard to see that the time for the walk dominates the time to construct $\mathcal{T}(R \cup \{s\})$ from $\mathcal{T}(R)$.

For $R \subseteq S$ and $s \in R \setminus S$ let $t(R, s)$ be the time needed to construct $\mathcal{T}(R \cup \{s\})$ from $\mathcal{T}(R)$. To be more concrete, we define $t(R, s)$ as the number of vertical segments touching s or edges incident to s in $\mathcal{T}(R \cup \{s\})$ (this accounts also two for every intersection of s with R). Clearly, this quantity is proportional to the insertion time for s. For the analysis of the total running time, we define a tree T_0 as follows. The nodes of depth i of T_0, $0 \leq i \leq n$, have exactly $n - i$ children. The nodes of depth i correspond to the subsets of S of size $n - i$ in a natural way. The root corresponds to S and if a node v corresponds to a subset $R \subseteq S$ then the children of v correspond to the sets $R - \{x\}$, where x ranges over R. The edge connecting the nodes corresponding to R and $R - \{x\}$ is labeled by $t(R - \{x\}, x)$. For a path π in T_0, let $X(\pi)$ be the sum of the edge labels on path π. Then $X(\pi)$ is the total running time when the elements of S are inserted in the order specified by π, i.e., a walk along π from leaf to root specifies a permutation of S and $X(\pi)$ is the running time of the algorithm for this insertion order. In other words, T_0 represents all possible "backward" executions of the insertion process [Sei91]. Let $X = X(\pi)$ be the random variable defined by the uniform distribution on the paths in T_0.

b

a

s

figure=decomposition.ps

Figure 1: A trapezoidal decomposition $\mathcal{T}(R)$ for a subset R of three segments (shown solid) of a set S of five segments. The segment s (shown dotted) and the segment with endpoints a and b do not belong to R. When s is added to R the indicated walk (shown dashed) through $\mathcal{T}(R)$ is performed.

Let

$$T(R) = \sum_{x \in R} t(R - \{x\}, x) \ ,$$

$$M(r) = \max\{T(R);\ R \subseteq S,\ |R| = r\} \ ,$$

and

$$d(r) = \max\{t(R - \{x\}, x);\ R \subseteq S, |R| = r, x \in R\} \ .$$

Then T_0 is an (n, n)-tree respecting M and d and hence Theorem 1 can be used to prove a tail estimate for the running time. In order to apply Theorem 1 we need bounds on $d(r)$ and $M(r)$. These bounds are provided by Lemmas 2 and 3.

Lemma 2 *Let $R \subseteq S$ and $r = |R|$. Then $d(r) \leq \beta n \alpha(n)$, where α is the functional inverse of Ackermann's function and β is some constant independent of r.*

Proof: This Lemma was already shown in [Mul88]. We enclose its proof for completeness. Let $\mathcal{A}(R)$ be the arrangement defined by the segments in R. Let $s \in R$ be arbitrary and let E be the set of edges in the arrangement $\mathcal{A}(R)$ having at least one vertex on s (i.e. incident to s). Recall that $t(R \setminus \{s\}, \{s\})$ is the number of vertical bars touching s or an edge incident to s in $\mathcal{T}(R)$. To count these vertical bars we distinguish two kinds of bars. The type I bars are incident to an endpoint of a segment in $S \setminus R$ and the type II bars are not.

There are clearly only at most $2n$ type I bars. To count the number of type II bars consider an arrangement $\mathcal{A}(R')$ obtained from $\mathcal{A}(R)$ as follows: Split every segment in $R \setminus \{s\}$ intersecting s at its intersection with s and move the two new endpoints slightly away from s. Then the number of type II bars is at most proportional to the complexity of the face of $\mathcal{A}(R')$ containing s. Since $|R'| \leq 2r$ this complexity is $O(r\alpha(r))$, cf. [GSS89, Theorem 3.1]. ∎

For $R \subseteq S$, let $m(R)$ be the number of pairs of intersecting segments in S. We have

Lemma 3 *(a) Let $R \subseteq S$. Then $T(R) \leq \gamma(n + m(R))$ where γ is a constant independent of r.*

(b) $M(r) \leq \gamma(n + \min\{r^2, m\})$ for all integers r, $0 \leq r \leq n$.

Proof: Part (b) follows immediately from part (a) and the fact that $m(R) \leq \min\{|R|^2, m\}$. For part (a) we say that an edge e of $\mathcal{A}(R)$ contributes to $t(R \setminus \{s\}, s)$ for $s \in R$ if $e \subseteq s$ or e is incident to s; the contribution of e is the number of vertical bars touching it (in this way the overall contribution to $t(R \setminus \{s\}, s)$ is at least $t(R \setminus \{s\}, s)$). Every edge contributes to $t(R \setminus \{s\}, s)$ for at most three segments $s \in R$; a vertical bar touches at most three edges.

Summing up, we have that $T(R)$ is bounded by 9 times the number of vertical bars; the number of vertical bars is $2(2n + m(R))$. ∎

Substituting the bounds of Lemmas 2 and 3 into Theorem 1 gives us our first tail estimate.

Theorem 2 *Let β and γ be defined as in Lemmas 2 and 3 and let*

$$A = 2\gamma(n \ln n + m \ln(n/\sqrt{m})).$$

Then for all $c \geq 0$

$$\mathbf{Prob}(X \geq cA) \leq \left(\frac{e}{1+c}\right)^{cA/(\beta n \alpha(n))}.$$

Proof: We have $M(r) \leq \gamma(n + \min\{r^2, m\})$ for all r and hence

$$\begin{aligned}
\sum_{1 \leq r \leq n} \frac{M(r)}{r} &\leq \gamma \left(\sum_{1 \leq r \leq n} \frac{n}{r} + \sum_{r \leq \sqrt{m}} r + \sum_{\sqrt{m} < r \leq n} \frac{m}{r} \right) \\
&\leq \gamma(nH_n + m + m(H_n - H_{\sqrt{m}})) \\
&\leq 2\gamma \left(n \ln n + m \ln \left(\frac{n}{\sqrt{m}} \right) \right).
\end{aligned}$$

The bound now follows directly from Theorem 1. ∎

The bound of Theorem 2 is quite good for small m ($m = O(n)$) and large m ($m \approx n^2$). In these cases the quantity A is of the same order as the expected running time. We will next derive a better bound for intermediate values of m. In the proof of Lemma 3 we bounded $m(R)$ by $\binom{|R|}{2}$. However, the expected value of $m(R)$ for a random subset of R of S of size $r = |R|$ is only $m\frac{r(r-1)}{n(n-1)}$ [CS89, Lemma 4.1]. The idea is now to prove a tail estimate for (a quantity related to) $m(R)$ and to argue that one can essentially replace $m(R)$ by its expected value in the bound for $M(r)$ without invalidating Theorem 1.

Theorem 3 *There are absolute constants C, $\delta > 0$ such that*

$$\mathbf{Prob}(X \geq Cm) \leq \exp(\frac{-\delta m}{n \ln n}) \quad \text{for } m \geq n \ln n \ln^{(3)} n \ .$$

Proof: Put $x = \ln n$. Let $a_1 = 4e^2$, $a_2 = 16e$, let γ be as in Lemma 3, and redefine $M(r)$ as $M(r) = 4\gamma(a_1 mr/n + a_2 m/x)$.

For r, $1 \leq r \leq n$, define the random variable Y_r on the paths of T_0 so that $Y_r(\pi) = 1$ if $\gamma(n + m(R)) > M(r)$ for the set R corresponding to the node of depth $n - r$ on path π, and 0 otherwise. Let $Y = \max_{1 \leq r \leq n} Y_r$.

Let T_1 be the following (n, n)–tree respecting M and d, where $d(r) = \beta n \alpha(n)$ and β is as in Lemma 2. Let v be any node of T_0, let R be the set corresponding to v, and let w be the node corresponding to v in T_1. If $\gamma(n+m(R)) \leq M(|R|)$, then the labels of the edges emanating from w in T_1 are identical to the labels of the edges emanating from v in T_0; if $\gamma(n + m(R)) > M(|R|)$, then the labels of the edges emanating from w are arbitrary, but respect M and d. Let X_1 be the random variable defined by the sum of the edge labels along the paths in T_1.

The following three claims imply the theorem.

Claim 1 $\mathbf{Prob}(X \geq B) \leq \mathbf{Prob}(X_1 \geq B) + \mathbf{Prob}(Y = 1)$ *for any* $B \geq 0$.

Claim 2 $\mathbf{Prob}(Y = 1) \leq \exp(-m/(nx))$.

Claim 3 *There is a constant C such that*

$$\mathbf{Prob}(X_1 \geq Cm) \leq \exp\left(-\Omega\left(\frac{m}{n\alpha(n)}\right)\right) \ .$$

We now prove the three claims in turn.

Proof of Claim 1: For paths π with $Y(\pi) = 0$ we have $X(\pi) = X_1(\pi)$. Thus

$$\begin{aligned}
\mathbf{Prob}(X \geq B) &\leq \mathbf{Prob}(Y = 1) + \mathbf{Prob}(X \geq B \quad \text{and} \quad Y = 0) \\
&= \mathbf{Prob}(Y = 1) + \mathbf{Prob}(X_1 \geq B \quad \text{and} \quad Y = 0) \\
&\leq \mathbf{Prob}(Y = 1) + \mathbf{Prob}(X_1 \geq B) \ .
\end{aligned}$$

∎

Proof of Claim 3: The tree T_1 respects $M(r) = 4\gamma(a_1 mr/n + a_2 m/x)$ and $d(r) = \beta n\alpha(n)$. Now apply Theorem 1 with $A = C_1 m \geq \sum_{1\leq r\leq n} M(r)/r$ and C_1 sufficiently large. ∎

Proof of Claim 2: This claim is the hardest to prove. We will first define a quantity $D(R)$ related to $m(R)$ and then show that $Y(\pi) = 1$ implies that $D(R)$ is large for some set R on the path π. We will then use Theorem 1 to bound the probability that $D(R)$ is large.

For a line segment $s \in S$, let $\deg(s)$ be the number of intersections between s and the other segments in S. For $R \subseteq S$, let $D(R) = \sum_{s\in R} \deg(s)$. Then, clearly, $D(R) \geq 2m(R)$; in fact, $D(R)$ counts all intersections between segments in R and segments in S, where intersections between two segments in R are counted twice.

Claim 4 *Let π be a path in T_0. If $Y(\pi) = 1$ then there is some $r = 2^l \leq n/2$, $l \in I\!N$, such that $D(R) \geq a_1 mr/n + a_2 m/x$, where R is the set corresponding to the node of depth $n - r$ on path π.*

Proof: If $Y(\pi) = 1$ then there is an r', $1 \leq r' \leq n$, such that $\gamma(n + m(R')) > M(r')$, where R' is the set corresponding to the node of depth $n - r'$ on π. Let $r = 2^{\lceil \log r' \rceil}$. Then $r \leq n/2$, since $\gamma(n + m(R')) \leq M(r')$ for all $r' \geq n/4$. Let R be the set corresponding to the node of depth $n - r$ on path π. Then $R' \subseteq R$ and hence $\gamma(n + m(R)) \geq \gamma(n + m(R')) > M(r') \geq M(r)/2$. The claim now follows from $D(R) \geq 2m(R)$, $m/x \geq n$, and $a_2 \geq 1$. ∎

Claim 5 *Let $1 \leq r \leq n$. Then*

$$\mathbf{Prob}(D(R) \geq a_1 mr/n + a_2 m/x) \;\leq\; \min\left\{ e^{-a_1 mr/n^2}, \left(\frac{xr}{4n}\right)^{a_2 m/(nx)} \right\},$$

where R is a random subset of S of size r.

Proof: For $r \geq 2n/a_1$ there is nothing to prove since $D(R) \leq 2m$ always holds. For $r < 2n/a_1 \leq n/2$, we use Theorem 1, as follows. Consider the following (n, r)–tree T. The nodes of T of depth i, $0 \leq i \leq r$, correspond to subsets of S of cardinality i; the correspondence is many to one (the correspondence between nodes and permutations of subsets is one to one). If node v of T corresponds to $R' \subseteq S$ then the $n - |R'|$ children of v correspond to the sets $R' \cup \{s\}$, where $s \in S - R'$. Also, the edge connecting R' and $R' \cup \{s\}$ is labeled with $\deg(s)$. In this way, the edge labels on a leaf-to-root path sum to $D(R)$, where R is the subset of S corresponding to the leaf. Also, with $d(i) = n$ and $M(i) = 2m$, the tree T respects d and M and, by symmetry, each subset $R \subseteq S$ with $|R| = r$ corresponds to the same number of leaves of T. We now apply Theorem 1 with $A = 4mr/n$. Note that $\sum_{0\leq i\leq r-1} M(n-i)/(n-i) = 2m \sum_{n-r+1\leq i\leq n} 1/i$ and $\sum_{n-r+1\leq i\leq n} 1/i \leq \int_{n-r}^{n} (1/i)di \leq \ln(n/(n-r)) \leq \ln(1 + r/(n-r)) \leq 2r/n$ where the last inequality follows from $r \leq n/2$ and $\ln(1 + x) \leq x$.

Let $B = a_1 mr/n + a_2 m/x$. Then

$$\begin{aligned}\mathbf{Prob}(D(R) \geq B) &\leq (e/(B/A))^{B/n} \\ &\leq \min\left\{exp(-a_1 mr/n^2), \left(\frac{xr}{4n}\right)^{a_2 m/(nx)}\right\},\end{aligned}$$

where the first bound follows from $B \geq a_1 mr/n$ and $a_1 = 4e^2$ and the second bound follows from $B \geq a_2 m/x$ and $a_2 = 16e$, and that this bound is relevant only if $xr/4n < 1$. ∎

We can now complete the proof of Claim 2. Let

$$f(r) = \begin{cases} e^{-a_1 mr/n^2} & \text{if} \quad r \geq n/x \\ (\frac{xr}{4n})^{a_2 m/(nx)} & \text{if} \quad r < n/x \end{cases}.$$

Then

$$\mathbf{Prob}(Y = 1) \leq \sum_{l=0}^{\lfloor \log n \rfloor - 1} f(2^l)$$

according to the two preceding claims. Next observe that this sum can be split into two subsums, the first is

$$\begin{aligned}\sum_{\substack{r \in [n/x, n/2] \\ r = 2^l \text{ for some } l}} f(r) &\leq e^{-(a_1 m/n^2)\cdot(n/x)} \cdot \sum_{i=0}^{\lfloor \log(x/2) \rfloor} e^{-(a_1 m/n^2)2^i} \\ &\leq e^{-a_1 m/(nx)} \cdot \log x\end{aligned}$$

and the second is

$$\begin{aligned}\sum_{l=0}^{\lfloor \log(n/x) \rfloor - 1} f(2^l) &= \left(\frac{x}{4n}\right)^{a_2 m/(nx)} \cdot \sum_{l=0}^{\lfloor \log(n/x) \rfloor - 1} (2^{a_2 m/(nx)})^l \\ &\leq \left(\frac{x}{4n}\right)^{a_2 m/(nx)} 2^{a_2 m/(nx) \cdot \lfloor \log(n/x) \rfloor} \\ &\leq \left(\frac{x}{4n} \cdot \frac{n}{x}\right)^{a_2 m/(nx)} \leq \left(\frac{1}{2}\right)^{16em/(nx)} \\ &\leq e^{-4e^2 m/(nx)}\end{aligned}$$

(because $\left(\frac{1}{2}\right)^4 < e^{-e}$). Hence

$$\mathbf{Prob}(Y = 1) \leq \log(2x) \cdot e^{-4e^2 m/(nx)}.$$

Since $m \geq nx \ln \ln x$, we have $-4e^2 m/(nx) + \ln \log(2x) \leq -m/(nx)$, so that $\mathbf{Prob}(Y = 1) \leq e^{-m/(nx)}$. This completes the proof of Claim 2 and hence the proof of Theorem 3. ∎

We now discuss the other randomized line segment intersection algorithms. Clarkson and Shor [CS89] describe two algorithms. Both algorithms maintain the trapezoidal decomposition $\mathcal{D}(R)$ defined by the segments in R. When a segment $s \in S \setminus R$ is to be added, the two algorithms use different methods to find the trapezoids intersected by s. The first algorithm maintains for each $s \in S \setminus R$ the set of trapezoids of $\mathcal{D}(R)$ intersected by s and for each trapezoid the set of segments intersecting it (the so-called *conflict graph*). Our methods do not seem to imply anything for this algorithm. The second algorithm maintains for each trapezoid the set of segment endpoints contained in it. When a trapezoid is split during execution of the algorithm this list of points is scanned and the points are distributed among the resulting trapezoids. When a segment $s \in S \setminus R$ is to be added to $\mathcal{D}(R)$ the set of trapezoids intersected by s is determined by a walk through $\mathcal{D}(R)$ as described above for $\mathcal{T}(R)$. The walk through $\mathcal{D}(R)$ takes no longer than the walk through $\mathcal{T}(R)$ since $\mathcal{T}(R)$ is a refinement of $\mathcal{D}(R)$. We still need to estimate the time needed to maintain the conflict information.

Lemma 4 *Let $p \in \mathbb{R}^2$ be arbitrary and let X be the number of times p changes trapezoids during the incremental construction of $\mathcal{D}(S)$. Then* $\mathbf{Prob}(X \geq 6cH_n) \leq (e/c)^{cH_n}$ *for all $c \geq 0$.*

Proof: We use Theorem 1 with $M(i) = 6$ and $d(i) = 1$ for all i. Consider the tree T representing the backwards execution of the algorithm. Label the edge connecting vertices associated with sets R and $R \setminus \{x\}$ by 1 if the segment x is incident to the trapezoid of $\mathcal{D}(R)$ containing p and by 0 otherwise. Then at most six edges incident to any vertex are labeled 1 and hence T is an (n, n)–tree respecting M and d. Also X is the sum of the edge labels along a random path of T and $\sum_{1 \leq i \leq n} M(i)/i = 6H_n$. Thus $\mathbf{Prob}(X \geq 6cH_n) \leq (e/c)^{cH_n}$. ∎

Let $\overline{X}$ be the time needed to maintain the conflict information. Then $\overline{X} = O(X_1 + \ldots X_n)$ where X_i is the number of times the endpoints of the i–th segment in S changes trapezoids. Thus $\mathbf{Prob}(\overline{X} \geq c\gamma n H_n) \leq (e/c)^{cH_n}$ for a suitable constant γ and hence $\mathbf{Prob}(\overline{X} \geq \gamma m) \leq (enH_n/m)^{m/n} \leq e^{-m/n}$ for $m \geq e^2 n H_n$. Thus Theorem 3 holds for the second algorithm in [CS89].

The algorithm in [BDS+92] maintains $\mathcal{D}(R)$ and the so–called history of the construction. It determines the set of trapezoids intersected by s by determining all trapezoids in the history intersected by s. Our results do not seem to yield a tail estimate for the running of this algorithms. The expected space complexity of this algorithm is $O(n + m)$. It is easy to see that the incremental space cost after adding a segment $s \in S \setminus R$ to $\mathcal{D}(R)$ is no larger than the time needed to add s to $\mathcal{T}(R)$. Thus Theorem 3 holds for the space complexity of this algorithm.

Finally, Mulmuley's first algorithm [Mul88] maintains a subdiagram of $\mathcal{T}(R)$ in which vertical rays only emanate from the endpoints of the segments but not from the intersection points. Our results do not give a tail estimate for this algorithm.

References

[ASE91] N. Alon, J. Spencer, and P. Erdös. *The Probabilistic Method*. Wiley, 1991.

[BDS+92] J.D. Boissonnat, O. Devillers, R. Schott, M. Teillaud, and M. Yvinec. Applications of random sampling to on-line algorithms in computational geometry. *Discrete Comput. Geom.*, 8:51–71, 1992.

[CMS92] Kenneth L. Clarkson, Kurt Mehlhorn, and Raimund Seidel. Four results on randomized incremental constructions. In *Symposium on Theorectical Aspects of Computer Science, LNCS*, volume 577, pages 463–472, 1992. full version available as MPI-report MPI-I-92-112.

[CS89] K. L. Clarkson and P. W. Shor. Applications of random sampling in computational geometry, II. *Discrete Comput. Geom.*, 4:387–421, 1989.

[GSS89] L. Guibas, M. Sharir, and S. Sifrony. On the general motion planning problem with two degrees of freedom. *Discrete Comput. Geom.*, 4:491–521, 1989.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Ass.*, 58:13–30, 1963.

[MS92] J. Matoušek and R. Seidel. A tail estimate for mulmuley's segment intersection algorithm. *Proc. 19th Int. Colloq. Automata, Languages and Programming*, 1992, 427–438.

[Mul88] K. Mulmuley. A fast planar partition algorithm, I. *Proc. 29th IEEE Symp. on Foundations of Computer Science*, 1988, 580–589.

[Sei91] R. Seidel. A simple and fast incremental randomized algorithm for computing trapezoidal decompositions and for triangulating polygons. *Computational Geometry: Theory and Applications*, 1:51 – 64, 1991.