MAX-PLANCK-INSTITUT FÜR INFORMATIK

Tail Estimates for the

Space Complexity of Randomized

Incremental Algorithms

Kurt Mehlhorn Micha Sharir Emo Welzl

Trait Month Mana Shari Dino Well

August 1991

MPI-I-91-113



Im Stadtwald 66123 Saarbrücken Germany

Tail Estimates for the Space Complexity of Randomized Incremental Algorithms *

K. Mehlhorn Max-Planck-Institut für Informatik Saarbrücken M. Sharir School of Mathematical Sciences Tel-Aviv University

Courant Institute of Mathematical Sciences New York University

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

27 May 1991

Abstract

We give tail estimates for the space complexity of randomized incremental algorithms for line segment intersection in the plane. For n the number of segments, m is the number of intersections, and $m \geq n \ln n \ln^{10} n$, there is a constant c such that the probability that the total space cost exceeds c times the expected space cost is $e^{-\Omega m/(n \ln n)}$.

1 Introduction

Randomized incremental algorithms have received considerable attention recently; $\mathbf{c}(S)$, $\mathbf{l}(S)$ and $\mathbf{l}(S)$. They solve a large number of geometric problems, including the construction of Voronoi diagrams and convex hulls and intersection of line segments, in optimal expected time and space. Mehlhorn |S| has given a tail estimate for the space complexity of some of these algorithms. In particular, he has shown that, for the construction of planar Voronoi diagrams and Delaunay triangulations (using the algorithm of |S|), the probability that the space cost exceeds \mathbf{c} times the expected value is at most $|\mathbf{c}(\mathbf{c})| = \mathbf{f}(\mathbf{c})$ for every constant $\mathbf{c} \geq 1$. In this paper we obtain a considerably sharper bound for line segment intersection which is exponential in the size of the problem

This paper is structured as follows. In Section 2 we prove a simple probabilistic lemma which we then apply in Section 3 to derive the tail estimate mentioned above.

2 A Probalistic Lemma

For functions $M: \mathcal{N} \to \mathcal{N}$ and $d: \mathcal{N} \to \mathcal{N}$ and integers n and r with $n \ge r \ge 0$, call a rooted tree T an (n,r)-tree respecting M and d iff either r=0 and T consists of a

^{*}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. 3075(project ALCOM).

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 $d_i \le d(n)$ for $1 \le i \le n$ and $\sum_{1 \le i \le n} d_i \le M(n)$.

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

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

Prob
$$(X > B) < e^{-tB + \sum_{0 \le i \le r-1} \frac{M(n-i)}{(n-i)d(n-i)}} (e^{td(n-i)} - 1)}$$
.

Remark: The proof is an adaption of the standard proof for Hoeffding's inequality, cf. [8]. The case r = n and d(i) = M(i) for all i was previously treated in [9].

Proof: For $0 \le i \le r-1$, let X_i be the weight of the (i+1)-st edge on the path π . Then $X = \sum_{0 \le i \le r-1} X_i$ and

$$\mathbf{Prob}(X \geq B) = e^{-tB}e^{tB}\mathbf{Prob}(e^{tX} \geq e^{tB}) \leq e^{-tB}\mathbf{E}(e^{tX}) \leq e^{-tB}\mathbf{E}(\prod_{0 \leq i \leq r-1} e^{tX_i}).$$

Put $Z_i = e^{tX_i}$ and $Q_i = Z_0 Z_1 \cdots Z_i$. Then, for $i \ge 1$,

$$\mathbf{E}(Q_i) = \mathbf{E}(Q_{i-1}Z_i) = \mathbf{E}(Q_{i-1}\mathbf{E}(Z_i|Q_{i-1}))$$
.

But

$$\mathbf{E}(Z_i|Q_{i-1}) = \frac{1}{n-i} (e^{td_1} + \cdots + e^{td_{n-i}}),$$

where $d_1, d_2, \cdots, d_{n-i}$ are the weights of the edges emanating from the node corresponding to Q_{i-1} . Since $d_j \leq d(n-i)$ for $1 \leq j \leq n-i$ and $\sum_{1 \leq j \leq n-i} d_j \leq M(n-i)$, the last expression is maximized when $\lfloor M(n-i)/d(n-i) \rfloor$ weights d_j are equal to d(n-i), one weight is equal to $M(n-i) - \lfloor M(n-i)/d(n-i) \rfloor d(n-i)$ and the remaining weights are equal to x(n-i) follows that

$$\begin{split} \mathbf{E}(Z_i|Q_{i-1}) &\leq \frac{\frac{M(n-i)}{d(n-i)}}{e^{td}(n-i)} + n - i - \frac{M(n-i)}{d(n-i)} \\ &= 1 + \frac{M(n-i)}{(n-i)} \left(e^{td(n-i)} - 1\right). \end{split}$$

the case i = 0 is treated similarly, except that no conditional probability is used. Hence,

$$\mathbf{E}(\prod_{0 \leq i \leq r-1} e^{tX_i}) \leq e^{\sum_{0 \leq i \leq r-1} \frac{M(n-i)}{(n-i)d(n-i)} (e^{td(n-i)}-1)} \;,$$

since $\prod_{0 \le i \le r-1} (1+y_i) \le \prod_{0 \le i \le r-1} e^{y_i} = e^{\sum_{0 \le i \le r-1} y_i}$ for all reals y_0, \dots, y_{r-1} .

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

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

This simplifies Lemma 1 to

Prob
$$(X \ge B) \le e^{-tB + ((e^{td(n)} - 1)/d(n)) \cdot \sum_{0 \le i \le r-1} \frac{M(n-i)}{n-i}}$$
.

In what follows we will use this revised bound.

3 The Space Complexity of Randomized Incremental Constructions

Randomized incremental constructions take a random permutation of some set S of n objects, e.g., points in the plane, and construct a set of so-called regions, e.g., regions of the Voronoi diagram, in an incremental fashion. The algorithm maintains the collection of regions for the current subset of objects, and updates it after each insertion of a new object. The space complexity of the algorithm is defined to be the overall number of regions created during the random insertion process. This notation is appropriate because one of the main techniques for randomized incremental constructions maintains all regions ever constructed "on too of each other" without erasing old regions; see [1, 2, 5, 12, 11].

Whenever an object is added, an incremental space cost occurs. In most applications this space cost depends only on the object x added and the $x \in R$ of objects previously added (i.e. it is independent of the order in which the elements in R have been inserted,) for $R \subseteq S$ and $x \in S - R$ let c(R, x) denote the incremental space cost when object x is added to set R of objects. For the analysis of the total space cost, we define a space cost tree T_0 as follows. In tree T_0 the nodes of depth i, or space i is a transverse i in a natural way. The nodes of depth is correspond to the subsets of S of size n-i in a natural way. The root corresponds to S and S is an S of S of size S in the tended of the object S of the S is the object S of the S of the object S of S of the object S of the object S of the object S of S of the object S of the object S of S of S of the object S of S of

Let

$$M(r) = \max\{\sum_{x \in R} c(R - \{x\}, x); R \subseteq S, |R| = r\}$$

and

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

Then T_0 is an (n, n)-tree respecting M and d and hence Lemma 1 gives a tail estimate for the total space cost. For many randomized incremental constructions, e.g., planar Voronoi diagrams and Delaunay triangulations, trapezoidal decompositions, and convex hulls in three dimensions, one has M(r) = d(r) = ar for some constant a. Then, as shown in [9], if we choose $t = \frac{1}{n}$ in c and r = n in Lemma 1, we obtain

$$\mathbf{Prob}(X \geq c \cdot an) \leq \frac{1}{e} \left(\frac{e}{c}\right)^{c}$$

for all $c \ge 1$. We will now discuss an application where d is much smaller than M.

Line Segment Intersections. Randomized incremental algorithms for line segment intersection are described in [3] and [10]. These algorithms have expected running time $O(n\log n + m)$ and expected space cost O(n + m), where n is the number of segments and m is the number of intersections. Both algorithms maintain the trapezoidal decomposition T(R) of the plane with respect to the set R of segments. This decomposition is defined as the planar map obtained by drawing, from each endpoint and intersection point of the segments in R, a vertical segment up and down and extend it until it meets other segments

of R. The incremental space cost c(R, s) of adding a segment $s \in S - R$ to the current set R of segments is the number of trapezoids in $T(R \cup \{s\})$ incident to s. Since each trapezoid is incident to at most four segments and since the number of trapezoids in T(R)is at most 1 + 3|R| + 3m(R), where m(R) is the number of pairs of intersecting segments in R, we have

$$M(R) := \sum_{s \in R} c(R - \{s\}, s) \le 4(1 + 3|R| + 3m(R)).$$

Clearly, $0 \le m(R) \le |R|^2/2$. We can therefore apply Lemma 1 with $d(r) = M(r) = O(r + \min(r^2, m))$ to derive a tail estimate for the total space cost, albeit with a fairly weak bound of the form:

$$\operatorname{Prob}(X \ge c(n+m)\log n) \le c^{-\Omega(c)}$$

for all constants c. We leave it to the reader to verify this bound, which we now proceed to improve. There are two weak spots in the argument just offered. First, the expected value of m(R) is only $O(r+m-\frac{r}{\pi^2})$, so the upper bound M(r) is a gross overestimation for most values of M(R). Second, as will be shown momentarily, d(r) is always much smaller than the bound eiven above.

To obtain a sharper tail estimate we thus proceed as follows. We first show that $d(r) = O(r\alpha(r))$, where α is the functional inverse of Ackermann's function (cf. [7, page 156] for a definition). We next prove a tail estimate for (a quantity related to) m(R), and then finally use this tail estimate to apply Lemma 1 with a function M(r) smaller than $O(r + \min(r^2, m))$.

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

Proof: Let l be the line supporting s. Then $c(R, s) \le 6 + c(R, l)$ since every trapezoid in $T(R \cup \{s\})$ incident to s but not incident to an endpoint of s is also incident to l in $T(R \cup \{s\})$, and each endpoint of s is incident to only three trapezoids in $T(R \cup \{s\})$. In order to bound the number of trapezoids incident to l and above l, perform a scan along l and consider the list of top segments of these trapezoids. This list is an (r, 3) Davenport-Schinzel sequence $(c, \lceil r, page 15 \rceil]$ for a definition) as there can be no five indices $i_1 < i_2 < i_3 < i_4 < i_5$ in this list such that the trapezoids at places i_1, i_3, i_5 have top segment a and the trapezoids at places i_2, i_4 , have top segment $b, for b \neq a$. Since the length of an (r, 3) Davenport-Schinzel sequence is bounded by O(ra(r)) (cf. $\lceil r, page 15 \rceil$), the lemma follows.

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, $D(R) \ge 2m(R)$; 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.

Lemma 3 Let $a_1 \geq 4e^2$, $a_2 > 0$, $x \geq 1$, and $1 \leq r \leq n$. Then

$$\mathbf{Prob}(D(R) \geq (\frac{a_1r}{n} + \frac{a_2}{x}) \ m) \ \leq \ \min \left\{ e^{-a_1mr/n^2}, \left(\frac{4exr}{a_2n}\right)^{a_2m/(nx)} \right\}$$

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

Proof: For $r \geq \frac{2n}{a_1}$ there is nothing to prove since $D(R) \leq 2m$ always holds. For $r < \frac{2n}{a_1} \leq \frac{n}{2}$, we use Lemma 1, as follows. Consider the following (n,r)-tree T. The nodes

of T of depth $i, 0 \le i \le r$, correspond to subsets of S of cardinality i; the correspondence is many to one I mode v of T corresponds to $R' \subseteq S$ then the n - |R'| didden 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 such 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. Thus

$$\begin{array}{lcl} \mathbf{Prob}(D(R) \geq B) & \leq & e^{-tB + \sum_{0 \leq i \leq r-1} \frac{2m}{(n-i)n}(e^{tn} - 1)} \\ & \leq & e^{-tB + \frac{4mr}{n^2}(e^{tn} - 1)} \end{array}$$

for all $B \ge 0$ and $t \ge 0$. Put $B = m(\frac{a_1r}{n} + \frac{a_2}{x})$ and $t = \frac{1}{n}\ln(\frac{Bn}{4mr})$. Then

$$\begin{array}{lcl} \mathbf{Prob}(D(R) \geq B) & \leq & e^{\frac{R}{n}(1-\ln(\frac{Bn}{4mr}))} \\ & \leq & e^{\frac{a_1mr}{n^2}(-\ln(\frac{a_1}{4}))} \\ & \leq & e^{-\frac{a_1mr}{n^2}} \,. \end{array}$$

Similarly, we have

$$\begin{aligned} \mathbf{Prob}(D(R) \geq B) & \leq & e^{\frac{a_2 m}{n x} \left(1 - \ln\left(\frac{a_2 n}{4 r x}\right)\right)} \\ & = & e^{\frac{a_2 m}{n x} \ln\left(\frac{a_2 n}{2 r n}\right)}. \end{aligned}$$

ı

Theorem 1 There are absolute constants $C, \gamma > 0$ such that $\mathbf{Prob}(X \ge Cm) \le e^{-\gamma(\frac{m}{\ln \ln n})}$ for $m > n \ln n \ln \ln n$.

Proof: Put $z = \ln n$. Let T_0 be the space cost tree. Let $a_1 = 4e^2$, $a_2 = 16e$, and let $M(r) = 8(1 + 3r + \frac{3}{2}m(\frac{a_1r}{r} + \frac{a_2}{r}))$.

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

Let T_1 be the following (n, n)-tree respecting M and d, where $d(r) = \beta r \alpha(n)$ 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 $M(R) \leq M(|R|)$, then the labels of the edges emanating from w in T_1 are identical to the labels of the edges emanating from w in T_0 are identical to 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 .

Claim 1 $\operatorname{Prob}(X > B) < \operatorname{Prob}(X_1 > B) + \operatorname{Prob}(Y = 1)$ for any B > 0.

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

$$\mathbf{Prob}(X \ge B) \le \mathbf{Prob}(Y = 1) + \mathbf{Prob}(X \ge B) \text{ and } Y = 0$$

$$= \mathbf{Prob}(Y = 1) + \mathbf{Prob}(X_1 \ge B) \text{ and } Y = 0$$

$$< \mathbf{Prob}(Y = 1) + \mathbf{Prob}(X_1 \ge B)$$

We next bound Prob(Y = 1), proceeding in two steps. First we use a standard "power-of-two-trick" to show that we may essentially concentrate on Y_r for r being a power of two and then use Lemma 3 to bound the probability that Y_r is 1.

Claim 2 Let π be a path in T_0 . If $Y(\pi) = 1$ then there is an $r = 2^l$ for some l such that M(R) > M(r)/2 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 $1 \le r' \le n$, such that 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 $r \le n$, since M(R') < M(r') for all $r' \ge \frac{n}{2}$. Let R be the set corresponding to the node of depth n - r on path π . Then $M(R) \ge M(R') > M(r') \ge \frac{M(r)}{2}$.

Claim 3 Let $R \subseteq S$, |R| = r, and $M(R) \ge M(r)/2$. Then $D(R) > m(a_1r/n + a_2/x)$.

Proof: We have $M(R) \le 4(1+3|R|+3m(R))$ and $M(r) = 8(1+3r+\frac{3}{2}m(a_1r/n+a_2/x))$. Thus $M(R) \ge M(r)/2$ implies $m(R) \ge \frac{1}{2}m(a_1r/n+a_2/x)$. Since $D(R) \ge 2m(R)$, the claim follows.

Claim 4 Let $m \ge nx$. Then

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

Proof: Let

$$f(r) = \left\{ \begin{array}{ll} e^{-a_1 m r/n^2} & \text{if} \quad r \geq n/z \\ \left(\frac{4e z r}{a_2 n}\right)^{a_2 m/(n z)} & \text{if} \quad r < n/z \end{array} \right. .$$

Then

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

according to Lemma 3, and Claims 2 and 3. Next observe that this sum can be split into two subsums, the first is

$$\sum_{\substack{r \in [n/n,n/2] \\ r \equiv 1 \text{ for some } i}} f(r) \leq e^{-(a_1 m/n^2) \cdot (n/a)} \cdot \sum_{i=0}^{\lfloor \log(x/2) \rfloor} e^{-(a_1 m/n^2) 2^i} \\ \leq e^{-a_1 m/(na)} \cdot \log x$$

and the second is

$$\begin{split} \sum_{l=0}^{\lfloor \log(n/a)\rfloor} f(2^l) &= \left(\frac{4ex}{a_2n}\right)^{a_2m/(na)} \cdot \sum_{l=0}^{\lfloor \log(n/a)\rfloor} (2^{a_2m/(na)})^l \\ &\leq \left(\frac{4ex}{a_2n}\right)^{a_2m/(na)} 2^{a_2m/(na)} (1+\lfloor \log(n/a)\rfloor) \\ &\leq \left(\frac{4ex}{a_2n} \cdot \frac{2n}{a}\right)^{a_2m/(na)} \leq \left(\frac{1}{2}\right)^{16em/(na)} \\ &\leq e^{-4e^2m/(na)} \end{aligned}$$

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

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

as claimed.

Claim 5 Let $x = \ln n$ and assume that $m \ge n$. Then, for all $B \ge 0$,

$$\mathbf{Prob}(X_1 \geq B) \leq \left(\frac{B}{3a_3em}\right)^{-B/(\beta n\alpha(n))}$$

where $a_3 = 12a_2 = 192e$.

Proof: The tree T_1 respects M and d. Since $m \ge n$, it is easily verified that T_1 also respects \overline{M} and d where $\overline{M}(r) = a_3 m(r/n+1/x)$. Thus, Lemma 1 and the remark following it imply, for any B > 0,

$$\begin{aligned} \mathbf{Prob}(X_1 \geq B) &\leq e^{-tB + \sum_{1 \leq r \leq n} \frac{s_{\min}(r, s_1 + j_1)}{s_1}} e^{\theta r \sin(n) - 1} \\ &\leq e^{-tB + \theta^{\theta r \sin(n)}} \frac{s_n}{s_n} \sum_{1 \leq r \leq n} \left(\frac{1}{n} + \frac{1}{n^r}\right) \\ &\leq e^{-tB + \theta^{\theta r \sin(n)}} \frac{s_n}{s_n} \left(2 + \frac{1}{n^r}\right) \\ &\leq e^{-tB + \theta^{\theta r \sin(n)}} \frac{s_n}{s_n} \\ &\leq e^{-tB + \theta^{\theta r \sin(n)}} \frac{s_n}{s_n} \end{aligned}$$

Put $t = 1/(\beta n\alpha(n)) \ln(B/(3a_3m))$. Then

$$\mathbf{Prob}(X_1 \geq B) \leq e^{-\frac{B}{\beta n \alpha(n)} \ln \left(\frac{B}{3 \epsilon a_3 m}\right)}$$

We can now complete the proof of Theorem 1 by combining Claims 1, 4 and 5. Since $m \ge nz \ln nz$, we have $-4e^2m/(nz) + \ln \log(2z) \le -m/(nz)$, so that, by Claim 4, **Prob** $(Y = 1) \le e^{-m/(nz)}$. Claim 5 with $B > 576e^2m$ implies $\mathbf{Prob}(X_1 \ge B) \le e^{-4(m/n\alpha/n)}$. The Theorem now follows from Claim 1.

Remark: The constants C, γ in Theorem 1 that can be derived from our analysis are probably much too large and much too small, respectively. It would be interesting to obtain finer calibration of these constants.

References

- J.D. Boissonnat, O. Devillers, R. Schott, M. Teillaud and M. Yvinec, Applications of random sampling to on-line algorithms in computational geometry, Tech. rept., INRIA, 1990.
- [2] B. Chazelle, H. Edelsbrunner, L. Guibas, M. Sharir and J. Snoeyink, Computing a single face in an arrangement of line segments, Proc. 2nd ACM-SIAM Symp. on Discrete Algorithms, 1991, pp. 441-448.
- [3] K. L. Clarkson and P. W. Shor, Applications of random sampling in computational geometry, II, Discrete and Computational Geometry 4 (1989), 387-421.

- [4] H. Edelsbrunner and L. Guibas and M. Sharir, The complexity and construction of many faces in arrangements of lines and of segments, Discrete and Computational Geometry 5 (1990), 161-196.
- [5] L. Guibas, D. Knuth and M. Sharir, Randomized incremental construction of Voronoi and Delaunay diagrams, Proc. 17th Int. Colloq. on Automata, Languages, and Programming, 1990, pp. 414-431.
- [6] L. Guibas, M. Sharir and S. Sifrony, On the general motion planning problem with two degrees of freedom, Discrete and Computational Geometry 4 (1989), 491-521.
- [7] S. Hart and M. Sharir, Nonlinearity of Davenport-Schinzel sequences and of generalized path compression schemes, Combinatorica, 6 (1986), 151-177.
- [8] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Am. Stat. Ass, 58 (1963), 13 - 30.
- [9] K. Mehlhorn, A tail estimate for the space complexity of some randomized incremental constructions, unpublished manuscript.
- [10] K. Mulmuley, A fast planar partition algorithm, I, J. Symbolic Comput., 10 (1990), 253-280.
- [11] O. Schwarzkopf, Dynamic maintainance of geometric structures made easy, Proc. 32nd IEEE Symp. on Foundations of Computer Science, 1991, to appear.
- [12] R. Seidel, A simple and fast incremental randomized algorithm for computing trapezoidal decompositions and for triangulating simple polygons, Comput. Geom. Theory and Appl. 1 (1991), to appear.