Randomized Methods in Computational Geometry

(Marco Pellegrini)

Lecture 7: Random Sampling

Randomized divide and conquer in 1D

Given a set S of n points on the real line, we select a random sample R of S of size r. We build the decomposition H(R), and we repeat this construction recursively in each interval on the points of the conflict list of that interval.

- How large is the expected size of a conflict list?
- is the conflict size O(n/r) with high probability?

Such a construction allows to do predecessor search in S in time $O(r \log_r n)$.

A Basic sampling theorem

Theorem:
Let S be a set of n points on the real line and R a random sample of S of size r. Let c>2 and s ≥r >2.
The conflict size of each interval in H(R) is at most

$$c\frac{n}{r}\log s$$

with probability $1 - O(\frac{1}{s^{c-2}})$

Proof.

Let Σ be a segment with endpoints in S (including semi-infinite ones).

For $\sigma \in \Sigma$, $T(\sigma)$ is the set of its endpoints, $L(\sigma)$ the points of S in its interior.

For a random sample R ⊆S, σ is in H(R) if and only if $T(\sigma) \subseteq R$ and $L(\sigma) \cap R = \emptyset$.

Set: $E_1(\sigma) = (L(\sigma) \cap R = \emptyset) \quad E_2(\sigma) = T(\sigma) \subseteq R$

By the law of conditional probabilities:

$$\Pr[E_1(\sigma) \wedge E_2(\sigma)] = \Pr[E_1(\sigma) \mid E_2(\sigma)] \Pr[E_2(\sigma)]$$

The probability over r-t(σ) draws to miss all point in the conflict list L(σ) is, setting $|L(\sigma)|=l(\sigma)$:

$$\prod_{j=1}^{r-t(\sigma)}\left(1-\frac{l(\sigma)}{n-t(\sigma)-j}\right) \le \prod_{j=1}^{r-t(\sigma)}\left(1-\frac{l(\sigma)}{n}\right) \le \exp\left(-\frac{l(\sigma)}{n}(r-t(\sigma))\right)$$

We call bad a segment with $l(\sigma) \ge c(n/(r-2))\log s$. The probability that σ is bad (conditional to existing in R) is:

$$\Pr[bad(\sigma) \mid E_2(\sigma)] \le \exp(-c\log s) = \frac{1}{s^c}$$

Using *Boole's Inequality* (known also as *Union Bound*), the probability of having one bad segment is at most:

$$\sum_{\sigma\in\Sigma} \Pr[bad(\sigma)] = \sum_{\sigma\in\Sigma} \Pr[bad(\sigma) \mid E_2(\sigma)] \Pr[E_2(\sigma)] \le \left(\frac{1}{s^c}\right) \sum_{\sigma\in\Sigma} \Pr[E_2(\sigma)]$$

But the summation is just the expected number of segmentns with endpoints in R, which is $O(r^2)$. Thus the probability of a bad segment in H(R) is at most $O(1/s^{c-2})$.

The converse, that is the probability of all elements of H(R) being good is at leat 1- $O(1/s^{c-2})$, which can be made arbitrarily close to 1, for c>2.

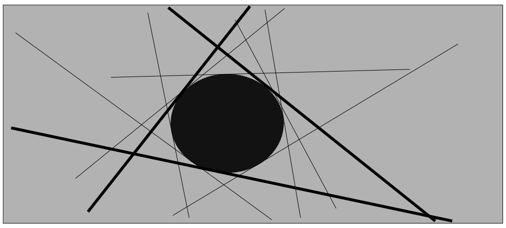
Cutting lemma

Given a set S of n points on the line, for 0< c < 1, a *c-cutting* is a partition of the line so that no region intersects more than cn point in S. The size of the cutting is the number of regions.

We nave seen that for a random sample R⊆S, H(R) is a (1/u)-cutting of size O(ulog(u+1)).

Given a set S of n lines on the plane, for c < 1, a *c-cutting* is a partition of the plane so that no region intersects more than cn lines in S. The size of the cutting is the number of regions.

Can we use a random sample of the set of lines to build a c-cutting for lines on the plane? Not directly (see counterexample)



C-cutting via trapezoidal maps

Theorem:
Let S be a set of n lines on the real plane and R a random sample of S of size r. Let c>4 and s $\geq$r >4. The conflict size of each trapezoid in the trapezoidal decomposition H(R) is at most:

$$c\frac{n}{r}\log s$$

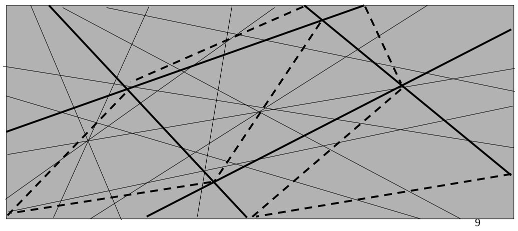
with probability

$$1-O(\frac{1}{s^{c-4}})$$

Proof. It follows the proof of the previous theorem, however a trapezoid σ is defined by two vertices, i.e. up to 4 lines: $|T(\sigma)|\leq 4$. Moreover, we have the bounded valence proeprty: 4 lines define a constant number of trapezoids.

Canonical triangulations

- Any Configuration space of bounded degree and valence built on a set of lines will do. For example the triangulation obtained by connecting to the bottom vertex.



Top-down construction of searching structures.

General building strategy:
(1) Choose a random subset R of S of size r large enough.
(2) Build H(R) and a search structure for H(R).
(3) Find for each s in H(R) the conflict list with S.
(4) Recursively build the data structure for each σ in H(R) on the conflict list of L(σ).

General searching strategy:

(1) Given a query q, solve q in H(R) finding σ_q.

(2) Ask the same query recursively in the recursive data structure for L(σ_q) of objects conflicting with s1.

(3) from the sequence of regions visited derive the answer for the set S.

Point Location in Arrangements of Lines

- Given a set S of lines in the plane build a data structure such that for a query point q, the cell of the arrangement A(S) containing p is found quickly.
- We build on the side the arrangement A(N) in time O(n^2).
- We take a random sample R of S, and we build the trapezoidal decomposition H(R). For each $\sigma\in$ H(R) we find the conflict list L(σ). Moreover we build A(S) $\cap$ σ and we set pointers from cells in H(R) to cells in A(S).
- If a conflict list is too long (bad). We repeat the sampling loop. This loop is executed an expected number of times O(1).
- If r is constant the time spent in these operations is O(n).
- When the size of the conflict list becomes a constant we just store all cells of the arrangement of these lines.

Query procedure and time analysis

- Given a query point q. We locate q among the trapezoids of H(R) finding σ. Then we recurse on the data structure for L(σ). At the bottom we find a cell in a A(S) and we follow the pointers.
- On each level of the structure the time spent is constant.
- So if Q(n) is the query time we have: Q(n) $\leq$ O(1) + Q(cn log r /r). That solves to Q(n)= O(log n).
- Let T(n) be the expected time to build the data structure. T(O(1))=O(1). T(n) = O(n^2) + O(r^2)T(cn log r /r), whose solution is O($n^{2+\epsilon}$) for any ϵ >0.
- The storage used S(n) has bound in the worst case O($n^{2+\epsilon}$).
- We choose r constant in dependence of the required value for ϵ.

Advanced c-cuts

We have proved:

Theorem: Given a set S of n lines in the plane, there is a (1/u)-cutting of size O($u^2 \log^2 (u+1)$).

Can we do better? There exists a (1/u)-cutting of size O(u^2)?

For a region σ in H(R) set $|L(\sigma)|=w(n/r)$. We know that with high probability w<c log r for all σ in H(R). Call t the "excess". How many regions in H(R) can have a fixed excess w? Better, what is the expected number of regions with excess w?

Lemma: Let S be a set of n lines in general position on the plane, p = r/n, for 1≤r ≤n/2, let R be a sample of S obtained with independent Bernoulli trials with success probability p, H(R) the trapezoidel decomposition of R, then:

$$E[|\{\sigma \in H(R) : |L(\sigma)| \geq w(n/r)\}|] \leq O(2^{-w} r^2)$$

Proof. Consider two Bernoulli processes.

The first process, with parameter p produces R.

The second process with parameter p'=p/t produces R'. Call:

$$H_{\geq w}(R) = \{\sigma \in H(R) : |L(\sigma)| \geq wn/r\}$$

Now we estimate E[|H(R')|) in terms of E[|$H_{>w}$(R)|]

$$E[|H(R')|] = \sum_{\sigma\in\Sigma} \Pr[\sigma \in H(R')] \geq \sum_{\sigma\in\Pi_i, i\geq wn/r} \Pr[\sigma \in H(R')] =$$

$$= \sum_{\sigma\in\Pi_i, i\geq wn/r} \Pr[\sigma \in H(R)] \frac{\Pr[\sigma \in H(R')]}{\Pr[\sigma \in H(R)]} \geq$$

$$\geq \min_{\sigma\in\Pi_i, i\geq wn/r} \left\{ \frac{\Pr[\sigma \in H(R')]}{\Pr[\sigma \in H(R)]} \right\} \sum_{\sigma\in\Pi_i, i\geq wn/r} \Pr[\sigma \in H(R)] =$$

$$\geq \min_{\sigma\in\Pi_i, i\geq wn/r} \left\{ \frac{\Pr[\sigma \in H(R')]}{\Pr[\sigma \in H(R)]} \right\} E[|H_{\geq w}(R)|]$$

Inverting:

$$E[|H_{\geq w}(R)|] \leq \left(\frac{1}{\min_{\sigma\in\Pi_i, i\geq wn/r} \left\{ \frac{\Pr[\sigma \in H(R')]}{\Pr[\sigma \in H(R)]} \right\}} \right) E[|H\ (R')|]$$

Now we upper bound E[|H(R')|] and lower bound the ratio of the two probabilities.

$$\frac{\Pr[\sigma \in H(R')]}{\Pr[\sigma \in H(R)]} = \frac{(p')^{t(\sigma)}(1-p')^{l(\sigma)}}{(p)^{t(\sigma)}(1-p)^{l(\sigma)}} \geq w^{-4} \left(\frac{1-p'}{1-p} \right)^{wn/r}$$

For every p: 1-p ≤exp(-p); for 0 ≤ p' ≤1/2: 1-p'≥exp(-2p')

$$\left(\frac{1-p/w}{1-p} \right)^{w/p} \geq \left(\frac{\exp(-2p/w)}{\exp(-p)} \right)^{w/p} = \exp(p(1-2/w)(w/p)) = \exp(w-2$$

Therefore:

$$E[|H_{\geq w}(R)|] \leq w^4 \exp(2-w) E[|H(R')|]$$

Now we estimate the expected size of the trapezoidal map of the R' sample. Clearly: $|H(R')| = O(|R'|^2)$

$$E[|R'|^2) = Var(|R'|) + E[|R'|]^2 = \sum_{i=1}^{n} Var(X_i) + (\sum_{i=1}^{n} E[X_i])^2 = np' + (np')^2$$

Where X_i's are the inependent Bernoulli trial. For p'=(r/nw) it becomes O(r^2/w^2).

Putting bits and pieces together:

$$E[|H_{\geq w}(R)|] \leq w^4 \exp(2-w) O(r^2/w^2) = O(2^{-w} r^2)$$

So we have proved that the number of trapezoid with excess w decays exponentially fast in w.

Optimal cuttings

Theorem: Given a set S of n lines in the plane, and parameter r, $1 < r < n$, there is a $(1/r)$-cutting of size $O(r^2)$.

Algorithm. Take R a Bernoulli sample of S with parameter $p=n/r$ and build the trapezoidal decomposition H(R). For every trapezoid $\sigma \in$ H(R) with excess $w>1$ build a $(1/w)$-cutting on $L(\sigma)$ and clip it to σ. Add these new trapezoids to teh output.

Proof. Clearly a trapezoid in H(R) either is cut by at most n/r lines or it is refined so that each refined trapezoid is cut by n/r lines.

19

Let us estimate the expected number of trapezoids produced:

$$E\left[\sum_{\sigma\in H(R)} 1 + C(w_\sigma \log w_\sigma)^2\right] \le E[|H(R)|] + CE\left[\sum_{\sigma\in H(R)} {w_\sigma}^4\right] \le$$

$$\le O(r^2) + CE\left[\sum_{i=0}^{\infty} \sum_{\sigma\in H(R), w_\sigma\in[2^i, 2^{i+1}]} {w_\sigma}^4\right] \le O(r^2) + C\sum_{i=0}^{\infty} E[|H_{\ge 2^i}(R)|] 2^{4(i+1)} \le$$

$$O(r^2) + C'r^2 \sum_{i=0}^{\infty} 2^{-2^i} 2^{4(i+1)} = O(r^2)$$

The summation converges to a constant, easy to prove via the criterio of the ratio: $\lim_{i\to\infty} \frac{a_{i+1}}{a_i} < 1$

20