Original Lecture #17: 5 March 1992
Topics: Randomized Divide and Conquer
Scribe: David Karger

Randomized Divide and Conquer

We have introduced the divide and conquer paradigm in a previous lecture. Recall that given a problem, we split it up into smaller, less complex subproblems which we can solve more easily. We then merge the solutions to obtain a solution for the entire problem. One of the key questions is how to perform the division step so that no subproblem has too great a complexity. The division has to be done in a 'balanced' manner — in other words, we'd like the subproblems to be approximately equal in size. This guarantees that the depth of divide and conquer 'tree' is logarithmic in size.

In the geometric setting, where problems involve collections of objects in a geometric domain, an obvious approach to the division is to divide the ambient space itself into regions so that no region contains too many objects. This is the problem we study in this lecture.

The first problem we consider is the following. We are given n lines in the plane and we are required to partition the plane into trapezoids, so that no trapezoid is cut by more than n/r of the lines, for some r. As we define below, such a partition is called an *r-cutting*.

Definition 1. *Given an arrangement of n lines in the plane, an r-*cutting* is a partition of the plane into trapezoids such that no trapezoid is cut by more than n/r of the lines in the arrangement.*

Constructing such a cutting is a first step in many divide and conquer geometric algorithms. A combinatorial version of this problem, that of dividing the lines into groups of r lines, is of course trivial; we can just take the first r lines, then the next r, and so on. Here, however, we are looking for more structure than that. In particular, if we just group the lines then each line will be in only one group; but if we partition the plane then a single line can cut more than one region.

How many trapezoids are needed to create an r-cutting? There is an obvious lower bound. Any trapezoid which has at most n/r lines passing through it can contain at most $\binom{n/r}{2}$ vertices (one for each pair of lines). Thus, since there can be up to $\binom{n}{2}$ vertices in the whole arrangement, there must be $\Omega(r^2)$ trapezoids in order to contain all the vertices. It turns out that this bound can always be achieved, but that the deterministic algorithm to do so is very complicated. We therefore turn to an extremely simple randomized algorithm which gets within a $\log n$ factor of the optimal bound.

The algorithm is based on a very simple intuition from the 1-dimensional world. Suppose we have a set of points on the line and we want to divide them into intervals each containing at most n/r points. If we choose r points at random from the collection, and use them as the endpoints of the intervals, then the probability is very high that no more than $O(n\log n/r)$ points will occur between any pair of selected points. The 2 dimensional algorithm is exactly analogous. To form the trapezoidal partition of the plane, we choose r lines uniformly at random from the arrangement and construct the trapezoidation of this sub-arrangement. Since there r lines, the arrangement will contain $O(r^2)$ vertices and thus the trapezoidation will contain $O(r^2)$ trapezoids. How likely is it that a trapezoid in our construction will have a lot of lines crossing it? We present an argument similar to that used in homework 2:

Let the *weight* of a trapezoid T, $w(T)$, be the number of lines from our arrangement which cut it. Consider some trapezoid T which might arise because some set of r lines was selected. Observe that any such trapezoid is defined by at most 4 lines from the arrangement: the two lines which form the top and bottom edges of the trapezoid, and the two lines which are responsible for the side threads of the trapezoid. Thus there are at most $\binom{n}{4}$ such trapezoids to consider. Under what circumstances does T actually turn up as a trapezoid in the cutting? It must certainly be the case that the four lines which define the trapezoid are chosen and that none of the $w(T)$ lines which cut the trapezoid are chosen. This happens with probability,

$$\begin{aligned} \text{Prob}[\Delta \text{ arises as a trapezoid}] &= \frac{\binom{n-w(T)-4}{r-4}}{\binom{n}{r}} \\ &\leq e^{-jr/n}, \end{aligned}$$

where $j = w(T)$.

Some trapezoids in our trapezoidation may also be defined two or three lines instead of four (consider what happens when the arrangement contains a triangle as a face). In this case, the probability reduces to,

$$\begin{aligned} \text{Prob}[\Delta \text{ arises as a trapezoid}] &= \frac{\binom{n-w(T)-2}{r-2}}{\binom{n}{r}} \\ &\leq e^{-jr/n}, \end{aligned}$$

where $j = w(T)$.

If we take $j \geq cn\log n/r$, then T has probability at most $1/n^c$ of arising. There are at most n^4 possible trapezoids may ever arise in a cutting (to see this, recall that a trapezoid is completely defined by at most 4 lines). Therefore, the chance that any trapezoid of weight exceeding $5n\log n/r$ turns up is at most $1/n$.

We can also show a related result, namely that the expected weight of a trapezoid in the cutting is n/r. This has to be made more precise. Consider some point p in the plane, and let T_p be the random variable denoting the trapezoid which contains p in the cutting we randomly construct. Then

$$
\begin{aligned}
E[w(T_p)] &= \sum_{T \supset p} w(T)\,\Pr[T = T_p] \\
&= \sum_T w(T) \binom{n - w(T) - 4}{r-4} \Big/ \binom{n}{r} \\
&= \sum_T w(T) \frac{n - w(T) - r + 1}{r-4} \binom{n - w(T) - 4}{r-5} \Big/ \binom{n}{r} \\
&\leq \frac{n}{r-4} \sum_T w(T) \binom{n - w(T) - 4}{r-5} \Big/ \binom{n}{r} \\
&\leq \frac{4n}{r-4} \\
&= O(n/r).
\end{aligned}
$$

The last transformation here is based on the fact that the final sum has value at most 4.
To summarize the analysis,

- With very high probability, none of the funnels is cut by more than $\Theta(n \log r/r)$ lines. The constant factor involved is quite small (around 5).
- The expected number of lines cutting any particular funnel is $\Theta(n/r)$.
- The expected total number of 'crossings' of funnels by lines is $\Theta(nr)$.

Application

The divide and conquer method introduced in this lecture can be applied in solving a lot of problems. One example is the "problem of many faces" which is stated as follows.

Given any n lines and m points on the plane, determine a boundary description of all the faces containing any of the m points. We show that the total combinatorial complexity of those cells is $k(m,n) = \Theta(m^{2/3}n^{2/3} + n)$. We give the analysis here very briefly.

We first introduce a variation of *Canham's lemma* that was presented in Homework 1. Recall that Canham's lemma states that given any m faces in an arrangement of n lines, their total complexity is at most $n + 4\binom{m}{2}$. A variation of this lemma states that their complexity is, in fact, at most $\Theta(m\sqrt{n} + n)$. This is a better bound if $m \geqslant \sqrt{n}$.

Choose r of the n lines at random. Decompose the arrangement of these r lines into trapezoids using the usual trapezoidal decomposition. For each trapezoid in this decomposition, we now have a sub-problem in which the participants are the lines that intersect the trapezoid.

It is not sufficient, however to find the solutions to all the subproblems. We need to 'merge' the results for adjacent trapezoids. This adds an additional complexity to our algorithm that is proportional to the total size of the 'boundaries' of the trapezoids. But this is nothing but the zone of the threads that define the trapezoids, which we know is linear.

Hence, we have the following recurrence (summing over all trapezoids),

$$k(m,n) = \sum_{\Delta} [\Theta(m_\Delta \sqrt{n_\Delta} + n_\Delta) + \Theta(n_\Delta)].$$

But we have,

$$\begin{aligned}
\sum_{\Delta} m_\Delta &= m \\
n_\Delta &= \Theta(n/r) \\
\Rightarrow k(m,n) &= \Theta(m\sqrt{n/r} + (n/r)r^2 + (n/r)r^2) \\
&= \Theta(m\sqrt{n/r} + nr)
\end{aligned}$$

Choose $r = m^{2/3}/n^{1/3}$ to 'balance' the two terms. We get,

$$k(m,n) = \Theta(m^{2/3} n^{2/3})$$

But we're assuming $r > 1$, that is, $m^2 > n$. But if that is not the case, we can use Canham's lemma to obtain a tight bound.

Therefore,

$$\mathcal{E}(k) = m^{2/3} n^{2/3} + n.$$

But notice that this also gives us a worst-case total complexity of the sum of the sizes of the faces. If the expected (or average) size is the given quantity, it is of course true that there is *some* choice of the lines such that the complexity is atleast the average. Hence, the given bound is a worst-case bound as well.