Weekplan: Range Reporting

Philip Bille

References and Reading

[1] Scribe notes from MIT.

[2] Computational Geometry: Algorithms and Applications, M. de Berg, O. Cheong, M. van Kreveld and M. Overmars,

[3] Fractional cascading: I. A data structuring technique, B. Chazelle and L. Guibas, Algoritmica, 1986

[4] Analysis of range searches in quad trees, J. L. Bentley and D. F. Stanat, Inf. Process. Lett., 1975

We recommend reading [1] in detail. [3] and [4] provide background on range trees and kD trees.

Exercises

1 Preprocessing Times Let $P \subseteq \mathscr{R}^2$ be a set of n points. Solve the following exercises.

1.1 Show how to construct a 2D range tree for P quickly.

1.2 Generalize your solution to 2D range trees with fractional cascading.

1.3 Show how to construct a kD tree for P quickly.

2 kD Tree Analysis Let T be a kD tree for a set of n points P. Consider a query for a range R. We want to bound the number of regions in T intersected by R to get a bound the query time for R. The number of regions intersected by any rectangle is at most 4 times the number of regions intersected by any vertical or horizontal line (why?) leading to our upper bound. Solve the following exercises.

2.1 Let $Q(n)$ denote the number of regions intersected by a vertical line in a kD tree for n points. Assume that the first split in kD tree is on the x-axis. Show that $Q(n)$ satisfies the following recurrence.

$$Q(n) = \begin{cases} 2Q(n/4) + O(1) & n > 1 \\ O(1) & n = 0 \end{cases}$$

2.2 Show that $Q(n) = O(\sqrt{n})$. *Hint:* draw recursion tree.

2.3 Argue that the query time for a kD tree is $O(\sqrt{n} + \text{occ})$.

2.4 Show that for some points set P of size n and some range R, the regions of the kD tree intersects with R in $\Omega(\sqrt{n})$ regions. Conclude that the upper bound analysis is tight up to constant factors.

3 Interval Trees Let $I = [l_1, r_1], \ldots, [l_n, r_n]$ be a set n of intervals. Give an efficient data structure that supports the following operation.

- intersect(x): return the set of intervals that contain the point x.

4 Quad Trees Let P be a set of n points in the plane. A *quad tree* Q is a rooted tree obtained recursively a follows. If P consists of < 2 points the quadtree Q is a single leaf. Otherwise, divide the plane into four (equally sized) quadrants NW, NE, SW, SE and recursively build a quadtree for each quadrant Q_{NW}, Q_{NE}, Q_{SW}, Q_{SE}. The quadtree Q consist of single node connected to the roots of Q_{NW}, Q_{NE}, Q_{SW}, Q_{SE}. Solve the following exercises.

4.1 Explain how quadtree can be used for range reporting queries.

4.2 Explain how quadtrees compare to other range reporting data structures. What time and space bounds can you give for quadtrees?

4.3 The quadtree use superlinear space. Show how to modify them to use only linear space. *Hint:* compress chains.

5 Fractional Cascading for General Arrays Let A_1 and A_2 be two sorted arrays. Solve the following exercises.

5.1 A fellow student wants to compactly store A_1 and A_2 to support efficient range reporting queries on both arrays using a single binary search. He suggest using fractional cascading (as described in the lecture). Explain why this will not work.

5.2 [$*$] Can you modify the data structure to make it work? *Hint:* Add more elements to A_1. This is where the name *fractional cascading* comes from.

6 [$*$] **Fast 1D Range Reporting** Give a data structure for a set of integers $S \subseteq U = \{0, \dots, u-1\}$ of n values that supports the following operation:

- report(x, y): return all values in S between x and y, that is, the set of values $\{z \mid z \in S, x \leq z \leq y\}$.

The data structure should use $O(n \log u)$ space and report queries should take $O(1 + \text{occ})$ time. *Hint:* x-fast tries and nearest common ancestors on complete binary trees.