Closest Pair Report

Sigurt Dinesen, Gustav Røder & Lars Yndal Sørensen

September 7, 2015

Results

Addison–Wesley 2005. Our implementation produces the correct result, for all of the provided input/output examples (`closest-pair-out.txt`). The of calculated results, and given example outputs, is made when the program is run without parameters.

Implementation

The implementation follows the pseudo code from page 230 in Kleinberg and Tardos, *Algorithms Design* closely.

For input sizes $n \leq 3$, we run a naive, pairwise $O(n^2)$ comparison algorithm. For larger inputs, we recursively divide the plane into two parts, each with $n/2 \pm 1$ points and find the closest point-pair in the left half, the right half and the pairs who's connecting edges cross the partition – the result of course, being The recursive function has a cutoff to the $O(n^2)$ solution, as mentioned above.

Performance

The running time is $n \log(n)$. This is illustrated in the following, by use of the `master theorem`.

The running time is an instance of the equation

$$T(n) = aT\left(\frac{n}{b}\right) + f(n)$$

Which means we can analyse it using the `master theorem`: As we recursively divide the input into two parts of $O(n/2)$ size, we have $a = b = 2$ in the above equation, which gives us

$$k = \log_b(a) = \log_2(2) = 1$$

At each level of recursion:

- The four lists (named `Q_x, Q_y, R_x, and R_y` in both the book and our implementation) containing the points of each partition, in increasing order by x and y coordinate respectively, are constructed at $O(n)$ cost.
- The set of points within δ distance of the partitioning line is found at $O(n)$ cost.

- Each point in that set is compared to the 15 following points, at $O(n * 15) \in O(n)$ cost.

In summary: $f(n) \in O(n)$

Returning to the `master theorem`, we observe that as $n = n^1 = n^k$:

$$f(n) \in O\left(n^k \log^0(n)\right)$$

which, according to the `master theorem`, means that the recursive algorithm runs in

$$O\left(n \log(n)\right)$$

Before the recursion, the input is ordered by both x and y coordinates, giving $O(n \log(n))$ preprocessing cost, which leaves the total running time as

$$O\left(n \log(n)\right)$$