CS325 - Project 2

Group #6

William Jernigan, Alexander Merrill, Sean Rettig

October 28, 2014

Correctness

Proof of Claim 3: Direct Proof

Claim 3: If $\{z_1, z_2, ..., z_t\}$ and $\{z'_1, z'_2, ..., z'_s\}$ are two visible sets of lines (each ordered by increasing slope), then the visible subset of $\{z_1, z_2, ..., z_t\} \cup \{z'_1, z'_2, ..., z'_s\}$ is $\{z_1, ..., z_i\} \cup \{z'_j, ..., z'_s\}$ for some $i \geq 1$ and $j \leq s$.

Let $A = \{a_1, a_2, ..., a_t\}$ be the set $\{z_1, z_2, ..., z_t\}$ and let $B = \{b_1, b_2, ..., b_s\}$ be the set $\{z'_1, z'_2, ..., z'_s\}$ for improved clarity.

Prove that $\{a_1, ..., a_i\}$ is visible:

Because all elements in A were defined to be visible with respect to each other, any covering line b_q must be from B.
Given that $m_{a_t} < m_{b_1}$, a covered line $a_p \in A$ has $m_{a_p} < m_{b_q}$.

Prove that for any covered line a_p, all lines to the right of p in A are also covered:

Because a_p is defined to be invisible, then by Claim 2 in the P1 Visible Lines Handout:
$y_{a_{p-1}}(x_{a_{p-1},b_q}) > y_{a_p}(x_{a_{p-1},b_q})$, where x_{a_{p-1},b_q} is the x coordinate where a_{p-1} and b_q intersect.
Because a_{p+1} is defined to not cover a_p, we know that $y_{a_p}(x_{a_{p-1},b_q}) > y_{a_{p+1}}(x_{a_{p-1},b_q})$.
For $x < x_{a_{p-1},a_p}, a_{p+1}$ is covered by a_p.
We need to show that a_{p+1} is covered for $x \geq x_{a_{p-1},a_p}$.
Because b_q is covering a_p, $y_{b_q}(x) > y_{a_q}(x)$ for $x \geq x_{a_{p-1},b_q} \geq x_{a_{p-1},a_p}$.
$\therefore a_{p+1}$ is invisible if a_p is invisible.
This follows for all p.
Then let $p = i + 1$.
$\therefore \{a_1, ..., a_i\}$ is visible and $\{a_p, ..., a_t\}$ is invisible.

Prove that $\{b_j, ..., b_s\}$ is visible

Because all elements in B were defined to be visible with respect to each other, any covering line a_o must be from A.
Given that $m_{a_t} < m_{b_1}$, a covered line $b_r \in B$ has $m_{a_o} < m_{b_r}$.

Prove that for any covered line $b_r, r = s - 1$, all lines to the left of r in B are also covered:

Because b_r is defined to be invisible, then by Claim 2 in the P1 Visible Lines Handout:
$y_{b_{r+1}}(x_{b_{r+1},a_o}) > y_{b_r}(x_{b_{r+1},a_o})$, where x_{b_{r+1},a_o} is the x coordinate where b_{r+1} and a_o intersect.

Because b_{r-1} is defined to not cover b_r, we know that $y_{b_r}(x_{b_{r+1},a_o}) > y_{b_{r-1}}(x_{b_{r+1},a_o})$.

For $x < x_{b_{r+1},b_r}$, b_{r-1} is covered by b_r.

We need to show that b_{r-1} is covered for $x \geq x_{b_{r+1},b_r}$.

Because a_o is covering b_r, $y_{a_o}(x) > y_{b_r}(x)$ for $x \geq x_{b_{r+1},a_o} \geq x_{b_{r+1},b_r}$.

$\therefore b_{r-1}$ is invisible if b_r is invisible.

This follows for all r.

Then let $r = j - 1$.

$\therefore \{b_j, ..., b_s\}$ is visible and $\{b_1, ..., b_r\}$ is invisible.

Proof of MergeVisible: Direct Proof

Prove that for each line that MergeVisible checks, it correctly determines the visibility of the line:

For each line at i that MergeVisible checks, it chooses j, k with $j < i < k$ and marks the line as invisible if $y_j(x_{j,k}) > y_i(x_{j,k})$.

Therefore, by the proof of Claims 2 and 3 in the "Visible Line Notes" handout, each line it checks and marks as invisible is actually invisible.

Additionally, for each line that MergeVisible checks the visibility of, it also checks to make sure that the previous line was not covered by the current line (using the same check), thereby exhausting all possible combinations of j, k.

Therefore, there is no i such that $j < i < k$ and $y_j(x_{j,k}) > y_i(x_{j,k})$.

Therefore, by the proof of Claims 2 and 3 in the "Visible Line Notes" handout, every line marked as visible is actually visible.

Prove that each line in the rest of each set is also invisible if an invisible line is found:

By the proof of Claim 3, we know that if a line in A with index p is invisible, then all lines in A with index $> p$ are also invisible.

By the proof of Claim 3, we know that if a line in B with index r is invisible, then all lines in B with index $< r$ are also invisible.

Proof of Algorithm 4: Proof by Induction

Let Y be a list of n lines sorted in ascending order by slope. We claim that Algorithm 4 will return the visible subset of Y.

Base case:

For any set of size < 2, we know that each line in the set is trivially visible because no other lines exist to cover the single line in the set.

Inductive hypothesis:

By our proof of MergeVisible, we know that if MergeVisible is passed two sets of visible lines, that it will correctly return the merged set of visible lines.

Applying the axiom of induction:

Because Algorithm 4 subdivides the given set of lines Y into sets of size 1 (which are visible according to the base case) before passing them to MergeVisible, we know that the result of each first MergeVisible call

is correct.

Because Algorithm 4 then only passes the results of correct previous MergeVisible calls to MergeVisible, each subsequent call of MergeVisible also returns a correct result.

Therefore, once all subsets of Y have been merged using MergeVisible, we know that the result of the final MergeVisible call is correct.
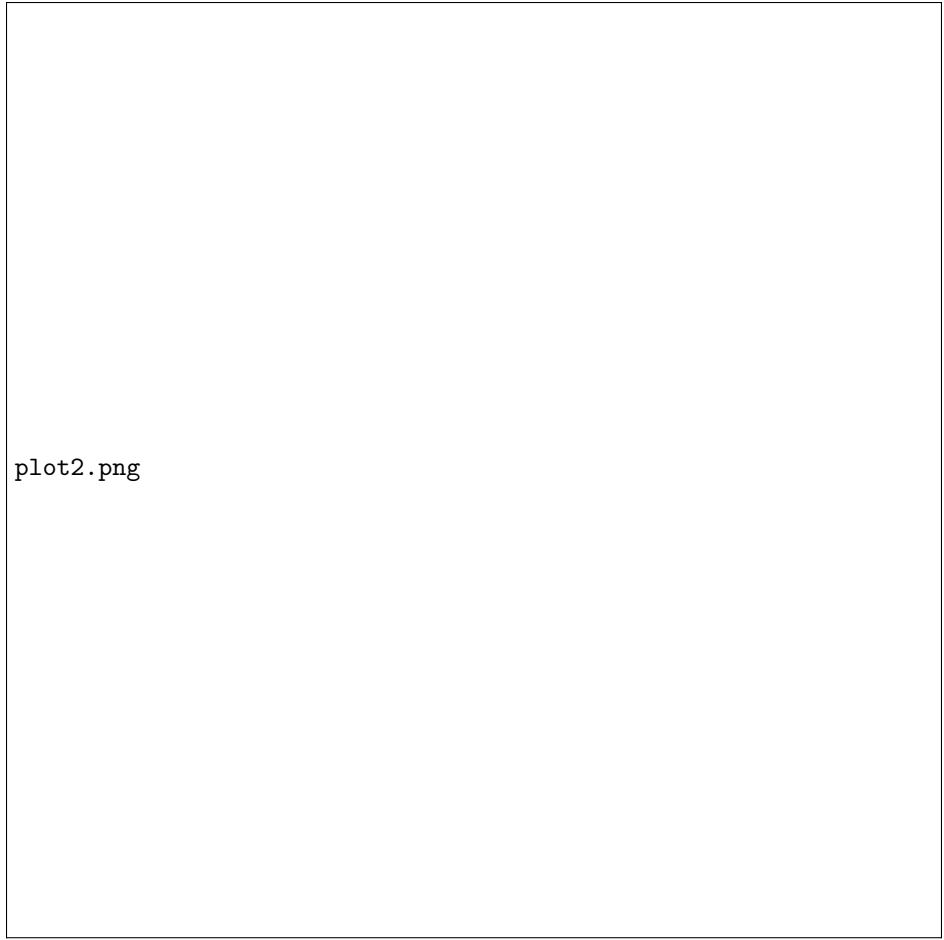
Therefore, because Algorithm 4 returns this result, we know that Algorithm 4 correctly returns the set of visible lines given any set of lines Y.

Experimental and Asymptotic Run Time Analysis

Experimental Run Time Data

Experimental Run Time Plots

plot1.png
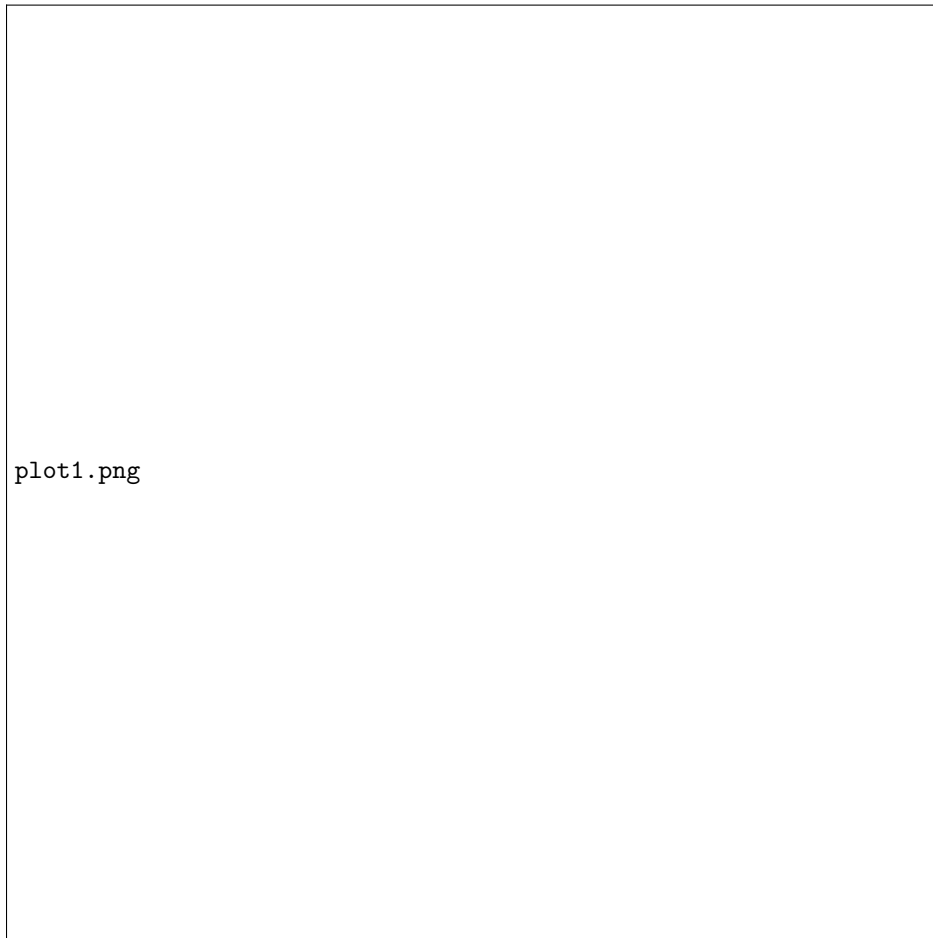
plot2.png

Experimental Run Time Analysis

Algorithm 1: $y = 8 \times 10^{-8} x^{3.0026}$
Algorithm 2: $y = 4 \times 10^{-8} x^{2.7892}$
Algorithm 3: $y = 1 \times 10^{-6} x^{1.0921}$
Algorithm 4: $y = 2 \times 10^{-6} x^{1.1231}$

Given these equations, we can use both the slopes and our code to analyze the run times of the algorithms. We can also determine the biggest instance that can be solved with each algorithm in an hour.

Asymptotic Run Time Analysis

Algorithm 1: $\Theta(n^3)$
Algorithm 2: $\Theta(n^3)$
Algorithm 3: $O(n^2), \Omega(n)$
Algorithm 4: $O(n \log n), \Omega(\log n)$

Extrapolation and Interpretation

Estimated Number of Lines Per Hour

Algorithm 1: $\sim 3,532$
Algorithm 2: $\sim 8,460$
Algorithm 3: $\sim 562,848,000$
Algorithm 4: $\sim 174,111,000$

Discrepancies

We note that the slopes from our experimentally-derived equations are within the asymptotic run-time range of $O(n^2)$, $\Omega(n)$ and $O(n \log n)$, $\Omega(\log n)$. However, our implementation of Algorithm 4 actually seems to run slightly slower compared to our Algorithm 3, despite Algorithm 4 having smaller minimum and maximum asymptotic runtimes. This discrepancy has many possible contributing factors, including a small sample size, differences in how the compiler and system the operations performed in each algorithm, the randomness of the input data sets. This randomness is due to both algorithms not performing a fixed number of operations, but rather changing what operations they run depending on the input lines given. It is possible that our tests are biased toward data sets that provide an advantage to Algorithm 3 (or a disadvantage to Algorithm 4). This randomness and compiler/system code handling are likely the primary causes of these differences.

Pseudocode