Analyzing the greedy algorithm for interval (activity) scheduling.

Problem: Given a set $A = \{a_1, a_2, \cdots, a_n\}$ of n intervals with start and finish times (s_i, f_i), $1 \leq i \leq n$, find a maximal set of mutually non-overlapping intervals.

Solution: sort in increasing order of finish times, and pick greedily the interval with the smallest finish time, eliminate all overlapping intervals, and repeat.

The intuition of the solution is that we want the resource to become free as soon as possible.

Why does this work?

First, the selected intervals are all compatible (non-overlapping). That's trivial. But, is the number of selected interval optimal? This is the question.

So, let O be an optimal set of intervals for A. Let G be the set of greedily selected intervals. Ideally one would want to show that G = O, but, that's too much to ask—there may be several optimal solutions, and we are happy if G is one of them. That is, we are happy if $|G| = |O|$, that is, G contains the same number of intervals as O.

The idea of the proof is to take an optimal solution O, and transform it into another solution that uses the greedy choice. We'll do this in a step-by-step fashion, so eventually the entire optimal solution can be transformed into a solution formed entirely of the greedy choices.

Notation: Assume the original set of intervals A i ssorted increasingly by finish time. let $i_1, i_2, ..., i_k$ be the set of intervals on G in the order in which they were adde to G (so $i_1 = 1$). Thus, $|G| = k$. Similarly, let the set of intervals in O be denoted by $j_1, j_2, ...j_m$. Assume that the intervals in G and O are ordered from left to right (note for a set of intervals that's non-overlapping, the order by start time is the same as the order by finish time). Our goal is to prove that $k = m$.

Note that the greedy choice guarantees that $f_{i_1} < f_{j_1}$, that is, the first interval in G finishes before the first interval in O. We will prove that this is true for the r-th interval as well, by induction:

Proposition 1 *For all $r \leq k$ we have that $f_{i_r} \leq f_{j_r}$.*

Proof 1 *By induction on r. For $r = 1$ this is true, by the greedy choice (see above).*

Now let $r > 1$, and assume the claim is true for $r - 1$: that $f_{i_{r-1}} \leq f_{j_{r-1}}$. We'll prove it for r.

Assume that the r^{th} interval of G does NOT finish earlier than the r^{th} interval of O. But this cannot happen, and we argue as follows: at the moment when Greedy was making its choice for the next interval to add to G,interval j_r MUST have been in the set of considered intervals, because, by induction hypothesis, $f_{i_{r-1}} \leq f_{j_{r-1}}$ and $f_{j_{r-1}} \leq s_{j_r}$. Thus, the greedy algorithm would have selected the interval with minimum finish time, thus, since it selected interval i_r it means that $f_{i_r} < f_{j_r}$. Contradiction.

Thus, the greedy algorithm "stays ahead" of any optimal solution. This implies optimality, as follows:

Proposition 2 *The greedy algorithm returns an optimal set G.*

Proof 2 *Assume, by contradiction, that the optimal solution O has $m > k$. Apply the proposition above, for $r = k$: we get that $f_{i_k} \leq f_{j_k}$. Since $m > k$, there is an interval j_{k+1} in O. This interval starts after interval j_k ends, thus after interval i_k ends. So, after deleting all intervals that are not compatible with i_k, the greedy algorithms still remains with at leats one interval, j_{k+1}. But the Greedy algorithm only stops when it has no more intervals. Contradiction.*