

Homework 2

CS 4364/5364
Spring 2021

Due: 25 February 2021

Because of the reliance of the particular assignments in this class on mathematical notation, and the fact that all assignments will be submitted electronically, students are encouraged to use \LaTeX to formalize their responses. **For those enrolled in the graduate section the use of latex is *required*.** This assignment (like all others) will be posted on the course `github`¹ as source code as well as in PDF form on the course website. Please submit your assignment to the professor via email, either as a link to your assignment online (i.e. overleaf or github) or as an attachment. Graduate students will need to include the `.tex` files as well as a PDF, this is optional but encouraged for undergraduates.

1. **(20 points)** Given two sequences S and T (not necessarily the same length), let G , L , and H be the scores of an optimal global alignment, an optimal local alignment, and an optimal global alignment without counting the indels at the beginning of S and the end of T , respectively.
 - (a) Give an example of S and T so that all 3 scores, G , L , and H , are different.
 - (b) Prove or disprove the statement $L \geq H \geq G$.

Note: for the proof, this can be a sketch, you just need to convince me that you're correct.

2. **(20 points)** Given two strings $S[1..n]$ and $T[1..m]$, we would like to find the two non-overlapping alignments $(S[i_1..i_2], T[j_1..j_2])$ and $(S[i_3..i_4], T[j_3..j_4])$ such that $i_2 < i_3$ and $j_2 < j_3$ and the total alignment score is maximized in running time $O(mn)$. Hints: remember that the local alignment between two sequences is the alignment of a pair of substrings from S and T such that the alignment score is maximized, and that the question does not say anything about the relationship between i_1 and i_2 nor i_3 and i_4 that if $i_3 > i_4$, $S[i_3..i_4]$ is the empty string.

¹github.com/deblasiolab/CS4364-documents