CENG 465
Spring 2010-2011

Assignment #1
(Programming Assignment)
SOLUTIONS

Question 1: Describe your modification to the original NW algorithm.

Modified DP recurrence relations are as follows:

$F(0,0) = 0$

$$F(i,0) = \begin{cases} F(i-1,0) - 1 & \text{if flanking(seq1.i)} \\ F(i-1,0) - 3 & \text{if !flanking(seq1.i)} \end{cases}$$

$$F(0,j) = \begin{cases} F(0,j-1) - 1 & \text{if flanking(seq2.j)} \\ F(0,j-1) - 3 & \text{if !flanking(seq2.j)} \end{cases}$$

$$F(i,j) = \max \begin{cases} F(i,j-1) - 1 & \text{if (flanking(seq2.j))} \\ F(i,j-1) - 3 & \text{if (!flanking(seq2.j))} \\ F(i-1,j) - 1 & \text{if (flanking(seq1.i))} \\ F(i-1,j) - 3 & \text{if (!flanking(seq1.i))} \\ F(i-1,j-1) + s(seq1.i, seq2.j) \end{cases}$$

$$s(seq1.i, seq2.j) = \begin{cases} 7 & \text{if } (seq1.i == seq2.j \,\&\&\, \text{!flanking}(seq1.j) \,\&\&\, \text{!flanking}(seq2.j)) \\ 4 & \text{if } (seq1.i == seq2.j \,\&\&\, \text{flanking}(seq1.j) \,\&\&\, \text{flanking}(seq2.j)) \\ 2 & \text{if } (seq1.i == seq2.j \,\&\&\, \text{flanking}(seq1.j) != \text{flanking}(seq2.j)) \\ -5 & \text{if } (seq1.i != seq2.j \,\&\&\, \text{!flanking}(seq1.j) \,\&\&\, \text{!flanking}(seq2.j)) \\ -3 & \text{if } (seq1.i != seq2.j \,\&\&\, \text{flanking}(seq1.j) \,\&\&\, \text{flanking}(seq2.j)) \\ -8 & \text{if } (seq1.i != seq2.j \,\&\&\, \text{flanking}(seq1.j) != \text{flanking}(seq2.j)) \end{cases}$$

Question 2: Analyze the running time complexity of your program theoretically. Did the modifications to the original algorithm cause an increase in the running time? You may ignore constant time increases.

The modifications are constant time condition checks to check whether the current row or the current column is a flanking region when filling the table. So, the asymptotic time complexity in Big-Oh notation does not change. The solution is still O(mn). However, these additional "if" statements will make it a little bit slower compared to the original version.

Question 3: What are the alignment scores for the 5 test sequence pairs?

The alignment scores are as follows:

test1.seq → 373
test2.seq → 226
test3.seq → 509
test4.seq → 1780
test5.seq → 272

Question 4: What are the alignments for the 5 test sequence pairs?

The alignments are provided in the output files test?.out

Question 5: For test5, i.e., the fifth test sequence pair, if the length of the flanking region was different, could you get a better alignment score? In other words, ignore the length provided in the test file and then find the best length of the flanking region which results in the maximum alignment score. There are $\lceil L/2 \rceil - 1$ different lengths you should try if the length of the shorter sequence is L and any region has at least on nucleotide.

Yes, we could get a better score. For flanking length 3 and 4 an alignment of length 286 is possible.