COMP 5970/6970-004
Computational Biology: Genomics and Transcriptomics
Homework 1 solutions

Haynes Heaton

Spring, 2022

Question 2: What is the property that makes problems amenable to solutions using dynamic programming?

Answer: Given a base case solution and a solution at point i, there is an easy way to find the optimal solution at point $i + 1$. This is the inductive property. Partial credit given to any answer about combining sub problems as that is also applicable to divide and conquer.

Question 3: True or False, in global alignment with scoring system match: +1, mismatch: -1, indel: -1, the score can go up when coming from a horizontal or vertical prior cell in the scoring matrix.

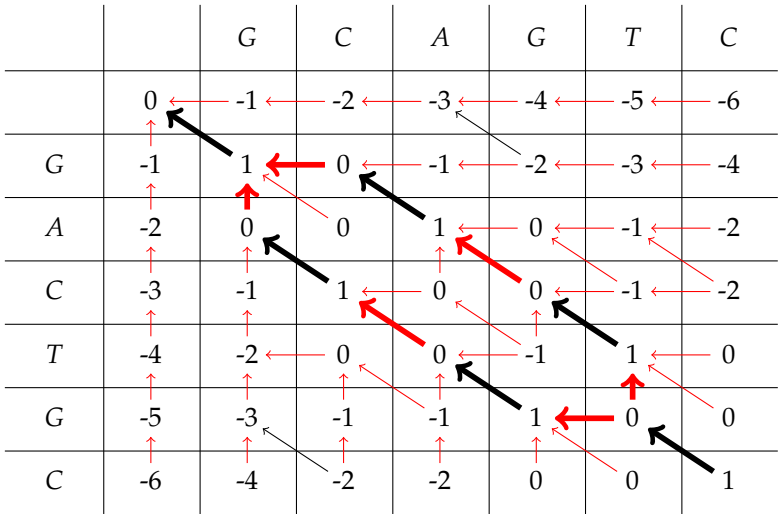
Answer: False, all moves vertical or horizontal are indels and thus reduce the alignment score.

Question 4: What is the formal name of the backtrace algorithm?

Answer: Viterbi algorithm

Question 5: Provide two different optimal solutions for the global alignment of GCAGTC and GACTGC in the form of two lines of strings with dashes for indels. Switching to courier new font will make it look nicer (monospaced fonts will make the two lines match character for character)

Answer:



Giving us two optimal alignments of
(upper path)
`GCAGT-C`
`G-ACTGC`
(and lower path)
`G-CAGTC`
`GACTG-C`

Question 6: What is the final score (number in lower right cell in the scoring matrix) of question 5?

Answer: 1

Question 7: Describe a dynamic programming approach to finding the largest checkerboard pattern in a matrix of zeros and ones where zero represents white and one represents black (checkerboard pattern is just alternating 0s and 1s, it does not matter which color starts or ends the pattern). Checkerboard must be square. Don't spend over an hour on this. If you find it difficult, come to my office hours after class and we can work through it together. Starting with 7a may help. And coming up with rules and testing them out may lead to the correct rules so use the example matrix below.

Answer: Generally in dynamic programming you will have a table that keeps track of the optimal solution(s) up to this point. We need to decide what our base case solution is. In this case, the simplest base cases to use would be a $0x0$ size checkerboard or a $1x1$ checkerboard. For this example, I will use $1x1$. So a single cell of 1 or 0 are both $1x1$ checkerboards. Now we need to decide what the values in the DP table represent. Similar to our global alignment problem, we could have a DP table of the same size as the input matrix. In the global alignment problem, the value was the score of the alignment up until now. Here it is natural to make the DP table keep track of the largest checkerboard cell i, j is the bottom right corner of. If that is the case, the left column and top row are all the base case of a $1x1$ checkerboard. We will denote the input matrix I and the DP table S for score. So the DP table S is initialized as

1	1	1	1	1	1
1					
1					
1					
1					

Now at each $i = 1..n$ and $j = 1..m$, we can check the cell above and to the left of it in the input matrix and make sure they are different and the one diagonal to the upper left and make sure they are they same. If that is the case, this square is the bottom right of a checkerboard that is $min(S[i-1][j-1], S[i][j-1], S[i-1][j]) + 1$. If this isn't the case, this is only the base case checkerboard of $1x1$. So our formal recurrence relation is

$$S[i,j] = \begin{cases} min(S[i-1,j-1], S[i-1][j], S[i][j-1]) + 1 \\ \quad \text{if } I[i][j] = I[i-1][j-1] \\ \quad\quad \text{and } I[i][j] \neq I[i-1][j] \\ \quad\quad \text{and } I[i][j] \neq I[i][j-1] \text{ (continuation of a checkerboard)} \\ 1 \text{ otherwise} \end{cases}$$

So filling the DP table out, we have

1	1	1	1	1	1
1	2	1	1	2	2
1	2	2	2	2	3
1	2	3	3	3	3
1	1	1	2	1	1

So we have several ties for different locations of a $3x3$ checkerboard in the given input matrix.

Question 7a: What is the base case for this algorithm?

Answer: 1x1 checkerboard of either 0 or 1. Also 0x0 would be acceptable. And since I didn't give a formal definition of a checkerboard, I also accepted 2x2.