1. Fill in the dynamic programming table and find all maximal *local* alignments for the strings $X = CAGGCT$ and $Y = TTAGCA$, using a scoring function $\sigma(x, y) = 3$ if $x = y$, $\sigma(x, y) = -1$ if $x \neq y$, and $\sigma(x, -) = \sigma(-, y) = -2$ for a gap.

2. [modified from Gusfield, Section 11.9, Problem 35] The sequence of a gene is made up of *introns* and *exons*. During transcription, the gene sequence is copied into RNA, but the introns are spliced out, leaving only the exons. Exome sequencing is becoming more and more common as a way of cheaply looking for variants in coding sequences. Describe a dynamic programming algorithm that would align an RNA transcript (just the exons) T to a reference genome G, in such a way that there are no gaps in the reference (potentially unrealistic) and no penalties for unaligned sequence at the beginning and end of the reference. Most introns start with the dinucleotide GT and end with AC. Modify your algorithm to enforce this constraint.

3. [BWT revisited] Fill in the table below for $\pi^{\text{sorted}}(S)$, for $S = ccabcbcabc$.

i	F L = BWT	occ(a,i) occ(b,i) occ(c,i)
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		

What are $M[a], M[b], M[c]$? Describe how to (exactly) match the pattern $P = abc$ to S using the recursions from lecture. Why might it be difficult to extend BWT to handle indels and mismatches?