Problem Set 1

Due 11:59pm, Friday, September 17th

Collaboration is allowed on this homework. You may discuss the problems with your colleagues, but each student must prepare and submit a separate assignment. Please list the names of the people you worked with:

The goal of this assignment is to help you to understand the details of how pairwise alignment algorithms work. To achieve this, you should work out the alignments by hand. Do not use an alignment program to calculate the alignments on this problem set. Please provide answers to all questions. In order to obtain full credit, explain your reasoning on eqch question and show all intermediate steps leading to your solution.

You may submit your assignments in any of the following ways:

- Download the assignment in pdf format and print it out. Write your solutions in the space provided. Scan your handwritten solution and upload it to Canvas. Attach additional pages, if needed.
- Download the assignment in pdf format. Use the commenting features in adobe or a similar tool to enter your solution directly on the pdf. Upload the completed assignment, annotated with your solutions.
- Download the assignment in latex format. Enter your solutions in latex format, compile the assignment, and turn in the resulting pdf.

Problems 1 and 2 ask you to align a pair of sequences. An alignment template, consisting of a grid with cells for the alignment score and the traceback arrows, is provided for these problems. Submit the alignment matrices with your assignment in one of the following ways:

- Download alignmentTemplate.pdf. Using a pad or tablet, fill in the template with a stylus. Save the annotated pdf and upload it to Canvas with your assignment.
- Download and print out alignmentTemplate.pdf. Enter the alignment scores and traceback arrows in the appropriate cells with a pen or pencil. Scan your handwritten alignment matrix and upload it to Canvas.
- Download alignmentTemplate.xlsx. Using excel as a tool for recording a table, type the alignment scores and traceback arrows in the appropriate cells of the template. Save the result as a pdf and upload it to Canvas.

1. **Global pairwise alignment:** This assignment asks you to align s_1 = `HOUSTON` and s_2 = `TUCSON`, using a scoring scheme that assigns a value of 0 to matches, a value of 1 to mismatches, and a value of 2 to gaps.

   (a) Is this a distance score or a similarity score? How do you know?

   (b) Calculate the global alignment of s_1 and s_2 using this scoring scheme. Hand in your alignment matrix with scores and traceback on the alignment template provided.

   (c) How many different optimal alignments are there? Show them.

2. **Local alignment:**

   (a) Compute the local alignment of $s_1 =$ `TOPOLOGY` and $s_2 =$ `POLYTOPE`, using the following scoring system: matches = 2, mismatches = -1, indels = -2. Hand in your alignment matrix with scores and traceback on the alignment template provided.

   (b) What is the optimal local alignment score? Show all optimal local alignment(s).

3. **Scoring functions for local alignments:** The following local alignment was constructed with the similarity function, $M = 4$, $m = -1$, $g = -3$.

		0	P	H	O	T	O	T	Y	P	E
0	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	4	1	4	1	0	0
Y	0	0	0	0	0	1	3	1	8	5	2
P	0	4	1	0	0	0	0	2	5	12	9
H	0	1	8	5	2	0	0	0	2	9	11
O	0	0	5	12	9	6	3	0	6	8	
O	0	0	2	9	11	13	10	7	4	5	
N	0	0	0	6	8	10	12	9	6	3	

(a) What four properties must a scoring function satisfy to be appropriate for local alignment? Demonstrate that this scoring function satisfies these properties.

(b) Let α_1 be the optimal local alignment in the alignment matrix above

```
PHOTO
PHO_O
```

and let α_2 be the highest scoring suboptimal alignment that *does not overlap with this optimal local alignment*:

```
TYP
TYP
```

What are the scores of α_1 and α_2?

(c) It is possible to specify a scoring function (i.e., to assign values to M, m, and g) for which α_2 has a *higher* score than the alignment α_1. Give an example of such a scoring function and rescore the above alignments with this function. Your scoring function should still satisfy the criteria in (a).

(d) In fact, there are an infinite number of scoring functions for which the alignment α_2 has a *higher* score than the alignment α_1. Give an equation or inequality in terms of M, m, and g that specifies this set of scoring functions. It is not necessary to restate the criteria in (a); simply give the additional constraints required. Explain your reasoning. (Note you do not need to work another alignment matrix to answer this question.)