It dissapoints me but after writing this I realized there is a (small) bug in my code, but It was too late for me to fix it unfortunately.

**Using seq1 and seq2, can you reproduce the alignments from figure 5.9, 5.11, 5.12?**

**5.9**

$ python assignment1\_joris.py -seqA THISLINE -seqB ISALIGNED -gap\_penalty 8 -end\_gap\_penalty 8

**Result:**

ISALIGNED

||

THISLINE-

Identity: 22.2222

**5.11**

**$** python assignment1\_joris.py -seqA THISLINE -seqB ISALIGNED -gap\_penalty 4 -end\_gap\_penalty 4

**Result:**

ISALIGNED

|| || ||

IS-LI-NE-

Identity: 66.6667

**5.12**

**$** python assignment1\_joris.py -seqA THISLINE -seqB ISALIGNED -gap\_penalty 4 -end\_gap\_penalty 4

does not give the same result

**Describe your traceback strategy in words. In case of equal scores from multiple cells, which direction is preferred?**

The parent\_matrix object is filled positions of the “parent column”. The traceback starts at the extreme bottom right of the matrix and traverses back using the indexes of the parents until the topleft position is reached. Diagonal is always preferred because it makes more sense to do an actual alignment instead of a gap. Next left is preferred over top, but I don’t necessarily have good reasons for that decision.

**When you align seq1 and seq2 using different linear gap penalties, ranging from 1 to 20 (and no separate end gap penalty), how many different alignments do you get? List the different alignments. Explain the differences, given the settings.**

*When the gap cost increases past 4 the most optimal/true alignment can no longer be made.*

**Sizes 1,2,3,4**

ISALIGNED

|| || ||

IS-LI-NE-

**Sizes 5,6**

IS-ALIGNED

|| ||

THISLI-NE-

**Sizes 7 t/m 20**

ISALIGNED

||

THISLINE-

**Align two related proteins, seq3 and seq4, and report the alignment score (right bottom in the matrix) using 2 different settings:**

- linear gap penalty = 5 & end gap = 1

score: 1505

Identity: 74.346

- linear gap penalty = 5 & end gap = 10

score: 1505

identity: 74.346

**Can you reproduce your results using the EMBOSS needle program?**

EMBOSS needle on seq3 and seq4 with:

- gapopen 5

- gap extend: 5

- endopen 5

- endextend 5

Alignment score: 1505

Identity: 74.3%

My implementation

Alignment score: 1505

Identity: 74,3%

However this is cherry picked, my code has a bug somewhere. If the endopen/extend is 1 for example the results are no longer the same.

**What is the criterion that the algorithm optimizes?**

The optimization step would be the traceback, so imagining that the matrix is filled with possible alignments, the tracebacks looks for the best possible one. The criterion that is optimized is therefore the similarity score.

**Is the algorithm exact or approximate?**

The algorithm seems exact because it looks for a global optimum instead of a local one.

**What are the time and space complexity of the algorithm you implemented?**

Using the unix time command:

|  |  |  |
| --- | --- | --- |
| Sum of both sequence lengths | Real time (seconds) | Memory usage (kb) |
| 760 | 0.688 | 25.784 |
| 1520 | 2.633 | 87.044 |
| 3040 | 11.043 | 340.012 |

**Describe how you could decrease running time by limiting the number of gaps allowed in the alignment.**

When introducing gaps, the number of steps needed for the traceback will be higher, en therefore running time will be slower. Raising the cost for gaps should make the running time faster because less gaps will be incorporated. (I did test this and it seems very slightly faster with a high gap penalty)