

Problem 1

DNA sequences: (1) ATCAGAGTA, (2) TTCAGTA

Scoring function: Match = +2, Mismatch = -1, Gap = -1

(a) Refer to the ``A1_Q1_Aaryan_2020004.py`` script file for the global alignment code using dynamic programming. Open terminal in the working directory and run

→ `python .\A1_Q1_Aaryan_2020004.py`

→ You will be able to see the desired output on the terminal itself

Substitution Matrix

	A	T	C	A	G	A	G	T	A
0.0	-1.0	-2.0	-3.0	-4.0	-5.0	-6.0	-7.0	-8.0	-9.0
T	-1.0	-1.0	1.0	0.0	-1.0	-2.0	-3.0	-4.0	-5.0
T	-2.0	-2.0	1.0	0.0	-1.0	-2.0	-3.0	-4.0	-2.0
C	-3.0	-3.0	0.0	3.0	2.0	1.0	0.0	-1.0	-2.0
A	-4.0	-1.0	-1.0	2.0	5.0	4.0	3.0	2.0	1.0
G	-5.0	-2.0	-2.0	1.0	4.0	7.0	6.0	5.0	4.0
T	-6.0	-3.0	0.0	0.0	3.0	6.0	6.0	5.0	7.0
A	-7.0	-4.0	-1.0	-1.0	2.0	5.0	8.0	7.0	6.0
									9.0

(b) **Yes**, there is more than one possible optimal alignment of the given sequences. This is because when we trace the substitution matrix for finding optimal alignment, we get some points where we can align two strings in more than one way. We can have multiple decisions at a particular point which will generate multiple alignments. There are two such points in the above substitution matrix.

One such point is (5, 7) where we can go either go to (4, 6) (i.e. aligning 4th character of seq2 with 6th character of seq1) or we can go to (5, 6) (i.e. aligning 5th character of seq2 with a gap). As a result, we will get multiple optimal alignments.

(c) Yes, all optimal global alignments with their scores are as follow:

Optimal Alignments with Scores

ATCAGAGTA

 | | | | |
TTC -- AGTA

Score: 9.0

ATCAGAGTA

 | | | | |
TTCA -- GTA

Score: 9.0

ATCAGAGTA

 | | | | |
TTCAG -- TA

Score: 9.0

Problem 2

(a) Refer to the `A1_Q2_Aaryan_2020004.py` script file for the local alignment code using dynamic programming. Open terminal in the working directory and run

- `python .\A1_Q2_Aaryan_2020004.py`
- You will be able to see the desired output on the terminal itself

Substitution Matrix

		A	T	C	A	G	A	G	T	A
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T	0.0	0.0	2.0	1.0	0.0	0.0	0.0	0.0	2.0	1.0
T	0.0	0.0	2.0	1.0	0.0	0.0	0.0	0.0	2.0	1.0
C	0.0	0.0	1.0	4.0	3.0	2.0	1.0	0.0	1.0	1.0
A	0.0	2.0	1.0	3.0	6.0	5.0	4.0	3.0	2.0	3.0
G	0.0	1.0	1.0	2.0	5.0	8.0	7.0	6.0	5.0	4.0
T	0.0	0.0	3.0	2.0	4.0	7.0	7.0	6.0	8.0	7.0
A	0.0	2.0	2.0	2.0	4.0	6.0	9.0	8.0	7.0	10.0

(b)

Optimal Alignments with Scores

TCAGAGTA
 || ||||
 TC--AGTA

Score: 10.0

TCAGAGTA
 ||| |||
 TCA--GTA

Score: 10.0

TCAGAGTA
 |||| ||
 TCAG--TA

Score: 10.0

Problem 3

Changes done in the program to perform local rather than global pairwise alignment are as follows:

- ☐ Initialize the **first row** and the **first column** with **zero** (Note that we want to ignore the gaps in the front and end of the alignment).
- ☐ The **lowest allowable value** inside a cell is **zero** (Note that we only 'local' alignment so we can negative score in order to maximize the score of final alignment).
- ☐ For finding the optimal alignment, **start** from the **highest value** (i, j) and extend using the same algorithm as in the case of global alignment.
- ☐ The alignment will **go** till the **first zero** value only.
- ☐ The **score** of the optimal alignment is the **highest value** in the substitution matrix.

Problem 4

New Scoring Scheme: Match = +2, Mismatch = -1, Gap = -2

By changing the scoring scheme, the results obtained in questions 1 and 2 will change.

New results:

For global alignment

Substitution Matrix										
	A	T	C	A	G	A	G	T	A	
	0.0	-2.0	-4.0	-6.0	-8.0	-10.0	-12.0	-14.0	-16.0	-18.0
T	-2.0	-1.0	0.0	-2.0	-4.0	-6.0	-8.0	-10.0	-12.0	-14.0
T	-4.0	-3.0	1.0	-1.0	-3.0	-5.0	-7.0	-9.0	-8.0	-10.0
C	-6.0	-5.0	-1.0	3.0	1.0	-1.0	-3.0	-5.0	-7.0	-9.0
A	-8.0	-4.0	-3.0	1.0	5.0	3.0	1.0	-1.0	-3.0	-5.0
G	-10.0	-6.0	-5.0	-1.0	3.0	7.0	5.0	3.0	1.0	-1.0
T	-12.0	-8.0	-4.0	-3.0	1.0	5.0	6.0	4.0	5.0	3.0
A	-14.0	-10.0	-6.0	-5.0	-1.0	3.0	7.0	5.0	3.0	7.0

Optimal Alignments with Scores

ATCAGAGTA

| | | | |

TTC -- AGTA

Score: 7.0

ATCAGAGTA

| | | | |

TTCA -- GTA

Score: 7.0

ATCAGAGTA

| | | | |

TTCAG -- TA

Score: 7.0

New results:

For local alignment

Substitution Matrix

		A	T	C	A	G	A	G	T	A
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
T	0.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0
T	0.0	0.0	2.0	1.0	0.0	0.0	0.0	0.0	2.0	1.0
C	0.0	0.0	0.0	4.0	2.0	0.0	0.0	0.0	0.0	1.0
A	0.0	2.0	0.0	2.0	6.0	4.0	2.0	0.0	0.0	2.0
G	0.0	0.0	1.0	0.0	4.0	8.0	6.0	4.0	2.0	0.0
T	0.0	0.0	2.0	0.0	2.0	6.0	7.0	5.0	6.0	4.0
A	0.0	2.0	0.0	1.0	2.0	4.0	8.0	6.0	4.0	8.0

Optimal Alignments with Scores

TCAG

| | | |

TCAG

Score: 8.0

TCAG - A

| | | | |

TCAGAA

Score: 8.0

AGTA

| | | |

AGTA

Score: 8.0

TCAGAGTA

| | | | | |

TCA -- GTA

Score: 8.0

TCAGAGTA

| | | | | |

TCAG -- TA

Score: 8.0

Running Code:

Refer to the `**A1_Q3(1)_Aaryan_2020004.py**` and `**A1_Q3(2)_Aaryan_2020004.py**` script files for the global and local alignment code using dynamic programming respectively with the new scoring scheme. Open terminal in the working directory and run

- `python .\A1_Q3(1)_Aaryan_2020004.py`
- `python .\A1_Q3(2)_Aaryan_2020004.py`
- You will be able to see the desired output on the terminal itself

Justification:

- ☐ If a higher gap penalty (more negative) is given then it's most likely that the score will get lower depending on dissimilarities between the sequences. The higher the dissimilarities, the lower is the score of optimal alignment. This is because optimal alignment cannot afford to have as many gaps as there were when the gap penalty was less.
- ☐ The same principle applies to the MISMATCH penalty.
- ☐ But in the case of MATCH reward, it is just the opposite. The higher it is (more positive), the higher is the score of the optimal alignment.
- ☐ Note that for filling each entry of the substitution matrix we refer to the scoring scheme, therefore, the optimal alignment or result (which is constructed using the substitution matrix) is directly related to the scoring scheme.
- ☐ Hence, the choice of GAP penalty and substitution matrix (with MATCH reward and MISMATCH penalty) is critically important.