COMS4761: FALL 2007

ASSIGNMENT 2.1 SOLUTIONS

Problem 1: Alignment Of ezLife molecules

You are asked to align two ezGene sequences: X=0010010111 and Y=0110101100, using a fixed gap penalty g=2 and the scoring matrix

	0	1
0	2	-1
1	-1	2

.

A. (15 points) Compute optimal global alignments using the Needleman-Wunsch algorithm.

(i) Show the alignment matrix, score values and backtracking direction arrows as the algorithm completes:

(a) initialization,

		0	0	1	0	0	1	0	1	1	1
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18	-20
0	-2										
1	-4										
1	-6										
0	-8										
1	-10										
0	-12										
1	-14										
1	-16										
0	-18										
0	-20										

(b) filling the left-upper 4x4 sub-matrix;

		0	0	1	0	0	1	0	1	1	1
	0	-2	-4	-6	-8	-10	-12	-14	-16	-18	-20
0	-2	2	0	-2	-4						
1	-4	0	1	2	0						
1	-6	-2	-1	3	1						
0	-8	-4	0	1	5						
1	-10										
0	-12										
1	-14										
1	-16										
0	-18										
0	-20										

and (c) filling the entire alignment matrix.

		0	0	1	0	0	1	0	1	1	1
	0	**-2**	-4	-6	-8	-10	-12	-14	-16	-18	-20
0	-2	**2**	**0**	-2	-4	-6	-8	-10	-12	-14	-16
1	-4	0	**1**	**2**	0	-2	-4	-6	-8	-10	-12
1	-6	-2	-1	**3**	**1**	-1	0	-2	-4	-6	-8
0	-8	-4	0	1	**5**	**3**	1	2	0	-2	-4
1	-10	-6	-2	2	3	4	**5**	3	4	2	0
0	-12	-8	-4	0	4	5	3	**7**	5	3	1
1	-14	-10	-6	-2	2	3	7	5	**9**	7	5
1	-16	-12	-8	-4	0	1	5	6	7	**11**	9
0	-18	-14	-10	-6	-2	2	3	7	5	**9**	**10**
0	-20	-16	-12	-8	-4	0	1	5	6	7	**8**

Notes: arrows not leading along optimal paths were removed to simplify the picture. Entries providing optimal paths are depicted using red boldface.

(ii) Show the various optimal paths for backtracking. Compute 3 different optimal alignments.

		0	0	1	0	0	1	0	1	1	1
	0	**-2**	-4	-6	-8	-10	-12	-14	-16	-18	-20
0	-2	**2**	**0**	-2	-4	-6	-8	-10	-12	-14	-16
1	-4	0	**1**	**2**	0	-2	-4	-6	-8	-10	-12
1	-6	-2	-1	**3**	**1**	-1	0	-2	-4	-6	-8
0	-8	-4	0	1	**5**	**3**	1	2	0	-2	-4
1	-10	-6	-2	2	3	4	**5**	3	4	2	0
0	-12	-8	-4	0	4	5	3	**7**	5	3	1
1	-14	-10	-6	-2	2	3	7	2	**9**	7	5
1	-16	-12	-8	-4	0	1	5	6	4	**11**	9
0	-18	-14	-10	-6	-2	2	3	7	5	**9**	**10**
0	-20	-16	-12	-8	-4	0	1	5	6	7	**8**

The optimal paths can be obtained by backtracking the arrows of part (i)c. Here we consider three such paths --colored blue, green and red – and the respective alignments.

```
0010010111-     001001011-1     0010010111-
-0110101100     011-0101100     0110-101100
```

B. (15 points) Compute optimal local alignment using the Smith-Waterman algorithm. Show the computational steps. How many optimal local alignments are there? What are they?.

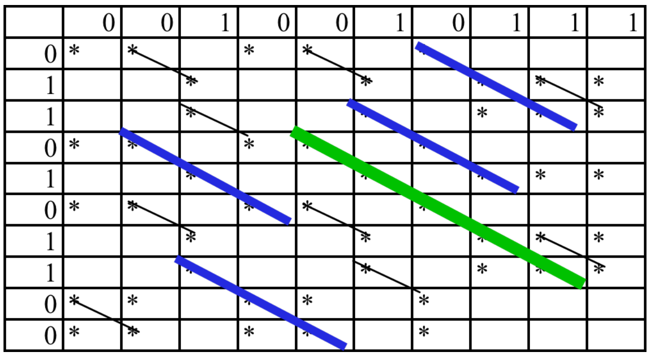
	0	0	1	0	0	1	0	1	1	1
0	2	**2**	0	2	2	0	2	0	0	0
1	0	1	**4**	2	1	4	2	4	2	2
1	0	0	3	**3**	1	3	3	4	6	4
0	2	2	1	5	**5**	3	5	3	3	5
1	0	1	4	3	4	**7**	5	7	5	5
0	2	2	2	6	5	5	**9**	7	6	4
1	0	1	4	4	5	7	7	**11**	9	8
1	0	0	3	3	3	7	6	9	**13**	11
0	2	2	1	5	5	5	9	7	11	12
0	2	4	2	3	7	5	7	8	9	10

There is one optimal local alignment described by the diagonal above:

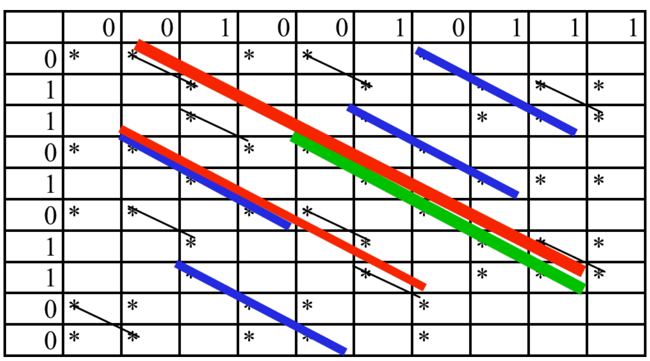
0010010111
-0110101100

C. (10 points) Compute local alignment using dot-matrix heuristic as follows.

(i) Show a dot matrix and draw all exact-match-diagonals of length ≥ 2.



(ii) Derive optimal non-gapped local alignment by extending exact-match-diagonals to maximize the score. Compare these local alignments with the results of B.



The red thick diagonal corresponds to the optimal local alignment of (B); it is obtained by extending the thick green diagonal to max its score. The thin red diagonal corresponds to a suboptimal local alignment below. This alignment scores 9, as can be seen from the Smith Waterman matrix of part (B); the optimal local alignment scores 13.

--0010010111
0110101100

(iii) Use the DP algorithm for a band near the exact-match-diagonals to compute optimal, possibly gapped, local alignment.

This is already accomplished by part (ii).

(iv) How will the results of (ii) change if the scoring matrix penalizes mismatch with a score of -10 instead of -1?

The thick green diagonal could not be extended and thus will represent the optimum non-gapped local alignment. However, with the gap penalty now lower than the mismatch penalty, another optimal gapped local-alignment is feasible. This one obtains by joining smaller diagonals to the thick green diagonal as depicted below. This gapped local alignment uses 2 gaps and thus pays a penalty of -4; which happens to be equal to the gain in score provided by the two small diagonals. So the score of this gapped alignment is the same as the score of the non-gapped alignment of the thick green diagonal.