

**Problem 1:**

DNA sequence: GCATTGC and GATTAGC

Match: +2

Mismatch: -1

Gap penalty: 2(each gap column contributes -2)

Dynamic programming matrix:

		G	C	A	T	T	G	C
	0	-2	-4	-6	-8	-10	-12	-14
G	-2	2	0	-2	-4	-6	-8	-10
A	-4	0	1	2	0	-2	-4	-6
T	-6	-2	-1	0	4	2	0	-2
T	-8	-4	-3	-2	2	6	4	2
A	-10	-6	-5	-1	0	4	5	3
G	-12	-8	-7	-3	-2	2	6	4
C	-14	-10	-9	-5	-4	0	2	8

The optimal alignment score will be 8.

Optimal: GCATT-GC

G-ATTAGC

**Problem 2:**

Using  $2 \times k$  matrix ( $k \geq m, n$ )

Sequences: TCTGC**C**TCTGC and ACTGAC**C**ACTGAC

V = TCTGC**C**TCTGC,  $n = 11$

W = ACTGAC**C**ACTGAC,  $m = 13$

Match: +2

Mismatch: -1

Gap: -2

Global: requiring end-to-end alignment

V	T	C	T	G	-	C	C	T	C	T	G	-	C
W	A	C	T	G	A	C	C	A	C	T	G	A	C
	-1	+2	+2	+2	-2	+2	+2	-1	+2	+2	+2	-2	+2

The optimal score will be 12.

**Problem 3:****(A) Sequences: ATC and ATATCTC.**

The scoring function is:

MATCH: 5.

MISMATCH: -10.

GAP: -5.

		A	T	A	T	C	T	C
	0	-5	-10	-15	-20	-25	-30	-35
A	-5	5	0	-5	-10	-15	-20	-25
T	-10	0	10	5	0	-5	-10	-15
C	-15	-5	5	0	-5	5	0	-5

**The optimal score is -5**

Optimal alignment:

AT--C--

ATATCTC

**(B) optimal local alignment:**

The scoring function is:

MATCH: 5.

MISMATCH:-10.

GAP: -5.

		A	T	A	T	C	T	C
	0	0	0	0	0	0	0	0
A	0	5	0	5	0	0	0	0
T	0	0	10	5	10	5	5	0
C	0	0	5	0	5	15	10	10

**The optimal alignment score will be 15.**

Optimal alignment:

-- ATC--  
  ↓ ↓ ↓  
ATATCTC

(C) **(B) will have the better performance** in optimal alignment for these two sequences because it has the high score.

**Problem 4:**

1. Write down a dynamic programming algorithm for global pairwise alignment with the affine gap penalty. You can assume that the match score and mismatch score are also given.

$$g(n) = g_{open} + (n - 1) \times g_{extend}$$
$$OPT(i, j) = \max \begin{cases} score(x_i, y_j) + OPT(i - 1, j - 1) \\ OPT(k, j) + g(i - k) \\ OPT(i, k) + g(j - k) \end{cases}$$

2.  $O(nm)$ , it is optimal.