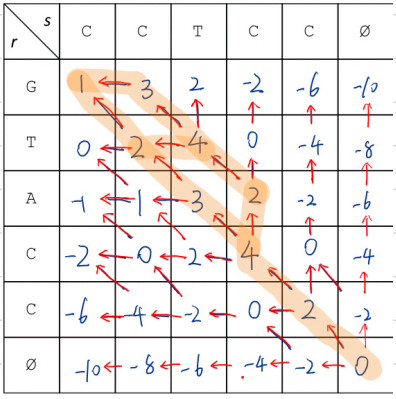
CSCI 3220.

Assignment 1.

1155077072 . ZHANG, Chongzhi .

1.



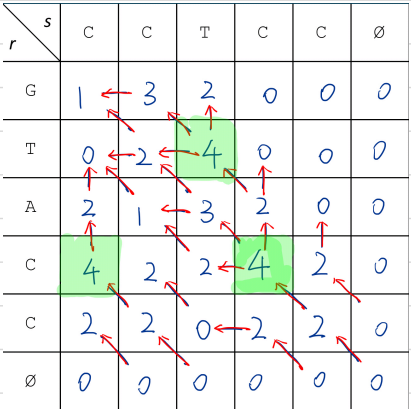
r \ s	C	C	T	C	C	Ø
G	1	3	2	-2	-6	-10
T	0	2	4	0	-4	-8
A	-1	1	3	2	-2	-6
C	-2	0	2	4	0	-4
C	-6	-4	-2	0	2	-2
Ø	-10	-8	-6	-4	-2	0

Answer:

Best alignment Score = 1

1) r = _ GTACC
   s = C CT_ CC
2). r = G_ TACC
   s = CCT_ CC
3). r = GTACC
   s = CCTCC

2.



r \ s	C	C	T	C	C	Ø
G	1	3	2	0	0	0
T	0	2	4	0	0	0
A	2	1	3	2	0	0
C	4	2	2	4	2	0
C	2	2	0	2	2	0
Ø	0	0	0	0	0	0

Answer:

Best Local alignment Score = 4.

1). r = GTA CC
   s = CCT CC
2). r = GTA CC
   CC TCC
3). r = G TACC
   s = CC T_CC

3. (Programming Part).

4. (a).

It obvious to know that the worst case of the optimal global alignment score is -5, if all the characters in r mismatchs with s. Thus if through one entry, the best situation of it is lower than -5, it will clearly useless.

r \ s	s_1	s_2	s_3	s_4	s_5	ϕ
r_1				X	X	X
r_2					X	X
r_3						X
r_4	X					
r_5	X	X				
ϕ	X	X	X			

Set $S_{(N)}$ means that the highest global alignment score of parth certain entry N:

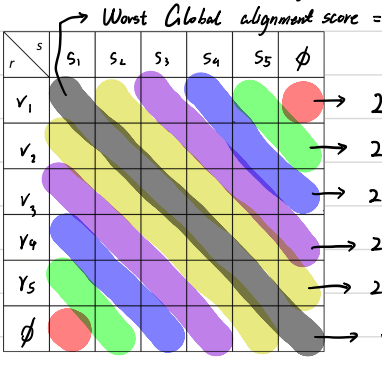
Calculate and we get $S_{blue} = (3 \times (-2)) \times 2 + 2 \times 2 = -8$

$S_{green} = (4 \times (-2)) \times 2 + 1 \times 2 = -14$.

$S_{red} = (5 \times (-2)) \times 2 + 0 \times 2 = -20$.

Thus, all the X marks above can neve been involved.

(b). All the $S_{(N)}$ in diagnal is the same.



$2 \times (5 \times S_{indel})$

$2 \times (4 \times S_{indel}) + 1 \times S_{match}$

$2 \times (3 \times S_{indel}) + 2 \times S_{match}$.

$2 \times (2 \times S_{indel}) + 3 \times S_{match}$.

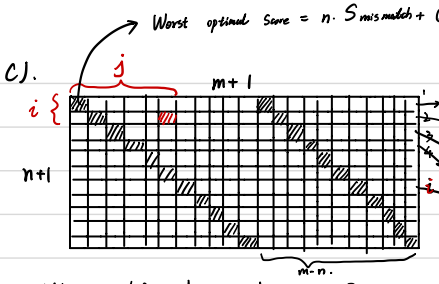
$2 \times (1 \times S_{indel}) + 4 \times S_{match}$.

$5 \times S_{match}$.

} Best case.

If the Best case of a entry is lower than the worst global alignment Score, Then it is never been involved.

c). Worst optimal Score = $n \cdot S_{mismatch} + (m-n) \times S_{indel}$



We splite the graph into 3 part. $j < i$. $i \le j \le i+m-n$, $j > i+m-n$.

Using similar method in (b).

- Case 1 if $i \le j \le i+m-n$.

  $S(i,j) = n \cdot S_{match} + (m-n) \times S_{indel} > n \cdot S_{mismatch} + (m-n) \times S_{indel}$.

  Thus this part will always been involved.

- Case 2 if $j > i+m-n$,

  The diagnal of S is equal. Thus $S(i,j) = S(i-j+m+1, m+1)$

  $= 2\times\{[n-(i-j+m)]\times S_{indel}\} + (i-j+m)\times S_{match} + (m-n) S_{indel}$.

  And let it compare with the worst optimal score, if lower, then it is not been involved.

- Case 3 if $j < i$.

  Using $S(i,j) = S(j, i+m-n)$ and transfer it to Case 2.

5. (a). Y = ACGCGTCA (b). S = CCGTGCAC

AC	1
CA	7
CG	2, 4
GC	3
GT	5
TC	6

v \ s	C	C	G	T	G	C	A	C	∅
A							1		
C		2						6	
G			3			4			
C		4							
G			5						
T				5					
C						6			
A							7		
∅									

Maximal Diagonal run:

1. CGTC

c). Persudo code.

```
for i = 1 to (length of S) - k + 1.
    lookup ( S[i, i+k-1] ), get a list of number.
     for each result c,
          push S[i, i+k-1] to vertex X.
            add ← 0
            for j = 1 to k + add
               if ( C+j ∈ lookup ( S[i+j, i+j+k-1] ) )
               push S[i, i+j+k-1] to vertex X.
               add ← add + j.
```

Find the longest element in X.