Solutions Exercise set 4

1. Assume the existence of a linear time algorithm that finds whether a sequence S is a subsequence of T. Give a linear time algorithm to check whether a sequence S is a circular shift of another sequence T. For example, ciseexer is a circular shift of exercise. What happens if one wants to test whether S is a subsequence of a circular shift of T?

Solution:

It is enough to test whether S is a subsequence of string TT and it has the same length as T. For the second part, it is enough to test if S is a subsequence of TT.

2. We apply the Sanger sequencing technique to the single stranded DNA sequence ``CATCAGTGCAGTGATC''. What single stranded DNA sequences will you find after the reaction in each of the four test tubes and what will be the picture obtained by running those test tubes through gel electrophoresis? What do you read from that picture?

Solution:

5’ G T A G T C A C G T C A C T A G 3’

3’ C A T C A G T G C A G T G A T C 5’

Tube A

GTA

GTAGTCA

GTAGTCACGTCA

GTAGTCACGTCACTA

GTAGTCACGTCACTAG

Tube T

GT

GTAGT

GTAGTCACGT

GTAGTCACGTCACT

GTAGTCACGTCACTAG

Tube G

G

GTAG

GTAGTCACG

GTAGTCACGTCACTAG

GTAGTCACGTCACTAG

Tube C

GTAGTC

GTAGTCAC

GTAGTCACGTC

GTAGTCACGTCAC

GTAGTCACGTCACTAG

\- +

A			I				I					I			I	
C						I		I			I		I			
G	I			I					I							I
T		I			I					I				I		
	G	T	A	G	T	C	A	C	G	T	C	A	C	T	A	G

5’ 3’

3. The following two DNA sequences contain the same 4 genes in different arrangements (in different order and on different strands). Find those arrangements and give one way to transform one arrangement into the other one using reversals only, that is, in each step of your strategy, you can only choose a certain substring and invert it. Take into account the orientation (strand) of each gene.

GGTCA AATTAC AGTCTGG CACTAG **CCAGACT TGACC CACTAG GTAATT**

CCAGT TTAATG TCAGACC GTGATC **GGTCTGA ACTGG GTGATC CATTAA**

Solution:

Solution: Here are the first 4 genes

1	2	3	4
GGTCA	AATTAC	AGTCTGG	CACTAG
CCAGT	TTAATG	TCAGACC	GTGATC

Sequences GGTCA AATTAC AGTCTGG CACTAG CCAGACT TGACC CACTAG GTAATT

CCAGT TTAATG TCAGACC GTGATC GGTCTGA ACTGG GTGATC CATTAA

The first sequence is the gene permutation 1234 and the second one is -3 -1 4 -2. The second permutation should be now sorted as in the lecture notes.

Sorting signed permutation -3 -1 4 -2

Step 1:

p = (0 -3 -1 4 -2 5)	oriented pairs: (0,-1),(-3,4)

Pairs	Permutation	Obtain oriented pairs	score
(0,-1)	(0,1,3,4,-2,5)	(1,-2), (3,-2)	2
(-3,4)	(0,1,3,4,-2,5)	(1,-2),(3,-2)	2

selecting (0,-1).

Step 2:

p = (0,1,2,-4,-3 5)	oriented pairs: (2,-3),(-4,5)

Pairs	Permutation	Obtain oriented pairs	score
(2,-3)	(0 1 2 3 -4 5)	(-4,5)	1
(-4,5)	(0 1 2 3 -4 5)	(-4,5)	1

selecting (-4,5)

Step 3:

p = (0 1 2 3 -4 5)	oriented pairs: (-4,5)

Pairs	Permutation	Obtain oriented pairs	score
(-4,5)	(0 1 2 3 4 5)	-None-	0

The solution is obtain in 3 steps.

4. Find a solution for a partial digest experiment giving the following results: 3, 4, 6, 10, 12, 15, 16, 20, 22, 25, 28, 31, 32, 35. Note that some of these lengths may have occurred in the experiment more than once, although they are only reported by your bio-lab only once in the list above.

<u>Solution</u>:

D(X) = {3, 4, 6, 10, 12, 15, 16, 20, 22, 25, 28, 31, 32, 35}

L = D(X).

First point in the set x is x0 = 0. X = {0}.

<u>35 ϵ X:</u>

The second point in the set is X = {0,35}. L = {3,4, 6, 10, 12, 15, 16, 20, 22, 25, 28, 31,32}

<u>32 ϵ X:</u>

Either 32 ϵ X and 3 ϵ X.

Choose 3 ϵ X.

X = {0,3,35} and L = {4, 6, 10, 12, 15, 16, 20, 22, 25, 28, 31}

<u>31 ϵ X:</u>

Either 31 ϵ X or 4 ϵ X.

4 ϵ X gives contradiction as |4-3| = 1 ∉ L. Considering 31 ϵ X.

X = {0,3,31,35} and L = { 6, 10, 12, 15, 16, 20, 22, 25}

<u>25 ϵ X:</u>

Either 25 ϵ X or 10 ϵ X. Considering 25 ϵ X. X = {0,3,25,31,35} and L = { 12, 15, 16, 20}

<u>20 ϵ X:</u>

Either 20 ϵ X or 15 ϵ X.

Considering 15 ϵ X. X = {0,3,15,25,31,35} Removing the distances from L. This solution is possible if 10 occurs 2 times in L.

	0	3	15	25	31	35
0		3	15	25	31	35
3			12	22	28	32
15				10	16	20
25					6	10
31						4
35						

The solution is X = {0,3,15,25,31,35} for L = {3,4,6,10,10,12,15,16,20,22,25,31,32,35}.

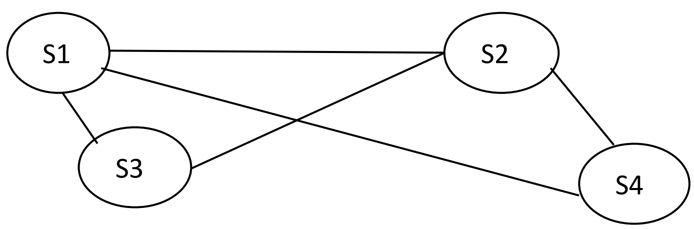
5. For the following matrix, find a permutation of its columns that either transforms it into a matrix with consecutive 1s in each row, or into one having at most two gaps in its blocks of ones (at most two gaps in the whole matrix).

0	1	1	0	1	1	1	0
1	1	0	0	1	0	1	0
0	1	1	0	1	1	0	1
1	0	0	1	0	0	1	0

Solution:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	0	1	1	0	1	1	1	0
2	1	1	0	0	1	0	1	0
3	0	1	1	0	1	1	0	1
4	1	0	0	1	0	0	1	0

S1 = {2,3,5,6,7}	S1 ∩S3 = {2,3,5,6} S1∩S4 = {7} , S1∩S2 = {2,5,7}
S2 = {1,2,5,7}	S2∩S3 = {2,5} , S2∩S4 = {1,7}
S3 = {2,3,5,6,8}	
S4 = {1,4,7}	



First we place the columns in S4:

	……	{1,4,7}	{1,4,7}	{1,4,7}	……	……	……
S4		1	1	1			

Then we handle the columns in S2. This leads to the decision of having {1,7} next to each other, with columns 4 and 5 on either side of them. We consider only one of the options, as the other is symmetric:

	{4}	{1,7}	{1,7}	{2,5}	{2,5}	……	……
S4	1	1	1	0	0		
S2	0	1	1	1	1		

Then we handle the columns in S1={2,3,5,6,7}. It means that column 7 must be adjacent to {2,5} on their left, and columns {3,6} adjacent to their right:

	{4}	1	7	{2,5}	{2,5}	{3,6}	{3,6}
S4	1	1	1	0	0	0	0

S2	0	1	1	1	1	0	0
S1	0	0	1	1	1	1	1

Now we handle the columns in S3={2,3,5,6,8}. It means that column 8 has to be placed to the right of {3,6}:

	{4}	{1}	{7}	{2,5}	{2,5}	{3,6}	{3,6}	{8}
S4	1	1	1	0	0	0	0	0
S2	0	1	1	1	1	0	0	0
S1	0	0	1	1	1	1	1	0
S3	0	0	0	1	1	1	1	1

So we obtained four column permutations that transform the matrix with one with C1P:

- 41725268
- 41725638
- 41752368
- 41752638

To these we are adding the symmetric solutions:

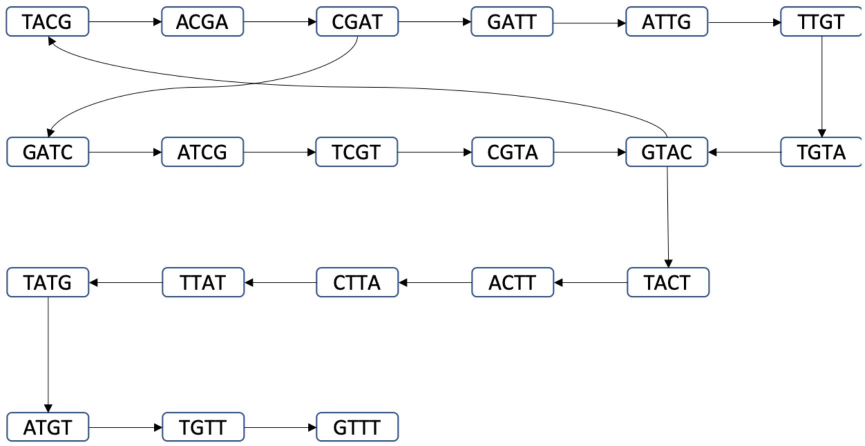
- 8625714
- 8362714
- 86325714
- 83625714

6. In an experiment on sequencing by hybridization, the following error-free results were return: ACTTA, ACGAT, ATCGT, ATTGT, ATGTT, CGATC, CGATT, CGTAC, CTTAT, GATCG, GATTG, GTACG, GTACT, TACGA, TACTT, TATGT, TCGTA, TGTAC, TGTTT, TTGTA, TTATG. Find at least one sequence that may give this result.

<u>**Solution:**</u>

Building graph containing sequences of length *l = 4.*

We will try to find Eulerian path in this graph.



Following the Eulerian path from CGAT to GTTT, the following sequence is found :

CGAT→GATT→ATTG→TTGT→TGTA→GTAC→TACG→ACGA→CGAT→GATC→ATCG→TCGT→CGTA→GTAC→TACT→ACTT→CTTA→TTAT→TATG→ATGT→TGTT→GTTT

The solution is CGATTGTACGATCGTACTTATGTTT.

Other Eulerian paths (and hence solutions to our problem) also exist. All of them are equally good.

7. Find a shortest common superstring of the following strings: TCAGT, AGTGC, GCA, CATCAG, GTG. Prove that no shorter solutions exist.

<u>**Solution**</u>:

Here is a common superstring of length 10: CATCAGTGCA.

To get a shorter superstring, i.e. of length 9 or less, we should have an overlap of strings CATCAG (of length 6) and AGTGC (of length 5) of at least two letters. Their longest possible overlap is indeed of two letters: CATCAG

AGTGC

yielding their superstring CATCAGTGC of length 9. All the other strings in the input should be now substrings of CATCAGTGC, but this is not the case for GCA. The conclusion is that there is no common superstring of length 9 or less.

8. Sort the following permutations using reversals only:

a. 2 7 3 8 1 5 4 6
b. -8 2 -5 6 -1 7 -4 3
c. 3 8 2 7 1 4 6 5
d. 2 -7 1 8 -4 3 5 -6

<u>**Solution:**</u>

<u>**(a) 2 7 3 8 1 5 4 6**</u>

Introducing break points at the place where the two adjacent numbers are not consecutive. Start with L and end with R.

L .2.7.3.8.1.5 4.6.R

Using the algorithm for sorting unsigned permutation. The strips having only one element are increasing or decreasing.

Start with the smallest number in the decreasing strip. It is 1.

k= 1.

Applying reversal which cuts after 1 and after L.

L 1.8.3.7.2.5 4.6.R

Next is 2. Applying reversal which cuts after 2 and after 1.

L 1 2.7.3.8.5 4.6.R

Next is 3. Applying reversal which cuts after 3 and after 2.

L 1 2 3.7 8.5 4.6.R

Next smallest in decreasing strip is 4. After reversal we obtain L 1 2 3 4 5 . 8 7 6. R

Next smallest in decreasing strip is 6. After reversal we obtain L 1 2 3 4 5 6 7 8 R

(b). -8 2 -5 6 -1 7 -4 3

Step 1:

p =(0 -8 2 -5 6 -1 7 -4 3 9)	Oriented pairs : (-8,9),(-8,7), (2,-1), (-5,6), (-4,3),(0,-1)

Pairs (pi,pj)	pi-pj	Reversal	Permulation	Score
(0,-1)	-1	(i+1,j)	(0 -8 2 -5 6 -1 7 -4 3 9) → (0 1 -6 5 -2 8 7 -4 3 9) oriented pairs: (1,-2),(-6,5),(-6,7),(5,-4),(-2,3),(-4,3)	6
(-8,9)	1	(i,j-1)	(0 -8 2 -5 6 -1 7 -4 3 9) → (0 -3 4 -7 1 -6 5 -2 8 9) oriented pairs: (-3,4),(-7,8),(1-2),(-6,5)	4
(2,-1)	1	(i,j-1)	(0 -8 2 -5 6 -1 7 -4 3 9) → (0 -8 -6 5 -2 -1 7 -4 3 9) oriented pairs : (0,-1),(-8,9),(-6,5),(-6,7),(5,-4),(-2,3), (-8,7),(-4,3)	8
(-5,6)	1	(i,j-1)	(0 -8 2 -5 6 -1 7 -4 3 9) → (0 -8 2 5 6 -1 7 -4 3 9) oriented pairs: (0,-1),(-8,9),(5,-4),(-8,7),(2,-1),(-4,3)	6
(-8,7)	-1	(i+1,j)	(0 -8 2 -5 6 -1 7 -4 3 9)→ (0 -8 -7 1 -6 5 -2 -4 3 9) oriented pairs : (-8,9),(-6,5),(1,-2),(5,-4),(-4,3),(-2,3)	6

Choose (2,-1)

Step 2:

p = (0 -8 -6 5 -2 -1 7 -4 3 9)	Oriented pairs : (0,-1),(-8,9),(-6,5),(-6,7),(5,-4),(-2,3), (-8,7),(-4,3)

Pairs (pi,pj)	pi-pj	Reversal	Permulation	Score

(0,-1)	-1	(i+1,j)	(0 -8 -6 5 -2 -1 7 -4 3 9) →(0 1 2 -5 6 8 7 -4 3 9) Oriented pairs : (-5,6), (-4,3)	2
(-8,9)	1	(i,j-1)	(0 -8 -6 5 -2 -1 7 -4 3 9) →(0 -3 4 -7 1 2 -5 6 8 9) Oriented pairs : (-3,4), (-3,2),(-7,6),(-7,8),(4,-5),(-5,6)	6
(-6,7)	1	(i,j-1)	(0 -8 -6 5 -2 -1 7 -4 3 9) →(0 -8 -6 -5 -2 -1 7 -4 3 9) Oriented pairs : (-8,9),(-5,6),(-8,7),(-4,3)	4
(5,-4)	1	(i,j-1)	(0 -8 -6 5 -2 -1 7 -4 3 9) →(0 -8 -6 -7 1 2 -5 -4 3 9) Oriented pairs (-8,9),(-4,3)	2
(-2,3)	1	(i,j-1)	(0 -8 -6 5 -2 -1 7 -4 3 9) →(0 -8 -6 5 4 -7 1 2 3 9) Oriented pairs : (-8,9),(-6,5)	2
(-8,7)	-1	(i+1,j)	(0 -8 -6 5 -2 -1 7 -4 3 9) →(0 -8 -7 1 2 -5 6 -4 3 9) Oriented pairs : (-8,9), (-7,6),(-5,6),(-4,3)	4
(-4,3)	-1	(i+1,j)	(0 -8 -6 5 -2 -1 7 -4 3 9) →(0 -8 -6 5 -2 -1 7 -4 -3 9) Oriented pairs: (0,-1),(-8,7),(-6,7),(5,-4),(-8,9),(-6,5)	6

Choosing (-8,9)

Step 3:

p = →(0 -3 4 -7 1 2 -5 6 8 9)	Oriented pairs : (-3,4), (-3,2),(-7,6),(-7,8),(4,-5),(-5,6)

Performing calculations similarly as shows in previous steps. Here in step 3 we choose (-7,8) who has maximum score 4 and obtained permulation is (0 -3 4 -6 5 -2 -1 7 8 9).

Step 4:

p = (0 -3 4 -6 5 -2 -1 7 8 9).	Oriented pairs : (0,-1),(-3,4),(-6,5),(-6,7)

After performing reversal in each pair, (-6,7) is chosen and it has maximum score 4. Obtained permutation is (0,-3,4,1,2,-5,6,7,8,9)

Step 5:

p = (0 -3 4 1 2 -5 6 7 8 9)	Oriented pairs : (-3,4),(-3,2),(4,-5),(-5,6)

After performing reversal in each pair, (-3,2) is chosen and it has maximum score 2. Obtained permutation is (0 -3 -2 -1 -4 -5 6 7 8 9)

Step 6:

p = (0 -3 -2 -1 -4 -5 6 7 8 9)	Oriented pairs : (0,-1),(-5,6)

Chosen pair is (-5,6). I t has score 2 and obtained permutation is (0,-3,-2,-1,-4, 5,6,7,8,9)

p = (0,-3,-2,-1,-4, 5,6,7,8,9)	Oriented pairs : (-4,5),(0,-1)

Step 7:

Chosen pair is (-4,5). I t has score 2 and obtained permutation is (0,-3,-2,-1,4, 5,6,7,8,9). The oriented pairs (0,-1),(-3,4)

Step 8:

For pair (0,-1) : (0,-3,-2,-1,4, 5,6,7,8,9) → (0 1 2 3 4 5 6 7 8 9)

For pair (-3,4) : (0,-3,-2,-1,4 , 5,6,7,8,9)→ (0,1,2,3,4,5,6,7,8,9)

Both have score 0.

The solution is obtained in 8 steps.

Alternate way: Reducing signed permutation into unsigned permutation (Ref: Lecture10, slide 25).

Given P: (-8 2 -5 6 -1 7 -4 3)

Replacing +x with (2x-1) (2x) and –x with (2x)(2x-1) will give us following unsigned permutation. This can be sorted by applying the same method used in (a)

P: (16 15 3 4 10 9 11 12 2 1 13 14 8 7 5 6) .

(c) 3 8 2 7 1 4 6 5

L.3.8.2.7.1.4.6 5.R	1 is the smallest in decreasing sequence.
L 1.7.2.8.3 4. 6 5.R	2 is the next smallest
L 1 2. 7 8.3 4. 6 5. R	5 is the next smallest
L 1 2.7 8. 3 4 5 6.R	No decreasing strip found. Applying 2-approximation algorithm for unsigned reversal. If there is no decreasing strip then reverse any increasing strip to create a decreasing strip.
L 1 2.7 8. 6 5 4 3. R	3 is the next smallest
L 1 2 3 4 5 6 . 8 7. R	7 is the smallest
L 1 2 3 4 5 6 7 8 R	Sorted permutation.

(d) (2 -7 1 8 -4 3 5 -6)

step 1:

p =(0 2 -7 1 8 -4 3 5 -6 9)	Oriented pairs : (-7,8),(-4,3),(-4,5),(5,-6)

Maximum score gain is 6 and corresponding pair is (-7,8) .obtained permutation is (0 2 -1 7 8 -4 3 5 -6 9)

Step 2:

p = (0 2 -1 7 8 -4 3 5 -6 9)	Oriented pairs : (0,-1),(7,-6),(-4,5),(2,-1),(-4,3),(5,-6)

Maximum score gain is 6 and choosing pair (2,-1) .obtained permutation is (0 -2 -1 7 8 -4 3 5 -6 9)

Step 3:

p = (0 -2 -1 7 8 -4 3 5 -6 9)	Oriented pairs : (0,-1),(-2,3),(7,-6),(-4,3),(-4,5),(5,-6)

Maximum score gain is 4 and choosing pair (5,-6) .obtained permutation is (0 -2 -1 7 8 -4 3 5 6 9)

Step 4:

p = (0 -2 -1 7 8 -4 3 5 6 9)	Oriented pairs : (0,-1),(-4,5),(-4,3),(-2,3)

Maximum score gain is 2 and choosing pair (-2,3). Obtained permutation is (0 4 -8 -7 1 2 3 5 6 9)

Step 5:

p = (0 4 -8 -7 1 2 3 5 6 9)	Oriented pairs : (-7,6), (-8,9)

Maximum score gain is 4 and choosing pair (-8,9). Obtained permutation is (0,4,-6,-5,-3,-2,-1,7,8,9)

Step 6:

p = (0,4,-6,-5,-3,-2,-1,7,8,9)	Oriented pairs : (0,-1),(-6,7),(4,-3),(4,-5)

Maximum score gain is 2 and choosing pair (0,-1), Obtained permutation is (0 1 2 3 5 6 -4 7 8 ,9)

Step 7:

p = (0 1 2 3 5 6 -4 7 8 ,9)	Oriented pairs : (5,-4)(3,-4)

Maximum score gain is 2 and choosing pair (3,-4), Obtained permutation is (0 1 2 3 4 -5 -6 7 8 ,9)

Step 7:

p = (0 1 2 3 4 -5 -6 7 8 ,9)	Oriented pairs : (-6,7), (4,-5)

Both oriented pairs yields same permutation (0 1 2 3 4 5 6 7 8 9)

The solution is obtained in 7 steps.