Solution Set 1

Due: 2 pm Friday, September 14

Collaboration is allowed on this homework. You must hand in homework assignments individually. List the names of the people you worked with:

The goal of this assignment is to help you understand the details of how pairwise alignment algorithms work. To achieve this, you should work out the alignments by hand. You may not use an alignment program to calculate the alignments on this problem set. Turn in your handwritten answers on the alignment template that is available on the course website.

Homework must be submitted by 2 pm in MI646.

1. Global alignment

(a) Compute the global alignment of CRECHE with SCORCH with the scoring function $M = 4$, $m = -4$, and $g = -6$. Show your alignment matrix with scores and traceback on the alignment template provided.
(b) What is the score of the optimal pairwise alignment? You should get several optimal alignments. Show all optimal alignments here.

There are two optimal alignments with score: -8.
2. Scoring functions

(a) For the scoring function used in Problem 1, suppose you replaced the values of $M$, $m$ and $g$ with $M + d$, $m + d$ and $g + d$, where $d = -2$.

i. Write down the values of $M$, $m$, and $g$ for this new scoring function.

\[
\begin{align*}
M &= 2 \\
m &= -6 \\
g &= -8
\end{align*}
\]

ii. Is it valid? Why or why not?

Yes:

\[
M > 0 > m > 2g
\]

iii. If the scoring function is valid, rescore the optimal alignments you obtained in Problem 1. Show the rescored alignments. (Note that you do not need to recalculate the alignment matrix.) Do these alignments still have the same score? If not, which alignments have lower scores? Why?

Alignment 1, Score -24

\[
\text{-C-RECHE} \\
\text{SCOR-CH-}
\]

Alignment 2, Score -22

\[
\text{-CRECHE} \\
\text{SCORCH-}
\]

The alignment 1 score is $4M + 4g$. The alignment 2 score is $3M + 2m + 2g$. Alignment 1 is preferred when $M + 2g > 2m$. Alignment 2 is preferred when $2m > M + 2g$, which is the case here.
(b) Suppose you replaced the values of $M$, $m$ and $g$ with $M + d$, $m + d$ and $g + d$, where $d = 2$.

i. Write down the values of $M$, $m$, and $g$ for this new scoring function.

\begin{align*}
M &= 6 \\
m &= -2 \\
g &= -4
\end{align*}

ii. Is it valid? Why or why not?

Yes:

\begin{align*}
M > 0 > m > 2g
\end{align*}

iii. If the scoring function is valid, rescore the optimal alignments you obtained in Problem 1. Show the rescored alignments. (Note that you do not need to recalculate the alignment matrix.) Do these alignments still have the same score? If not, which alignments have lower scores? Why?

Alignment 1, Score 8

- C-RECHE
  SCOR-CH-

Alignment 2, Score 6

- CRECHE
  SCORCH-

In this case, $M + 2g > 2m$, so alignment 1 is preferred.
(c) Suppose you replaced the values of $M$, $m$, and $g$ with $M \cdot d$, $m \cdot d$ and $g \cdot d$, where $d = -2$.

i. Write down the values of $M$, $m$, and $g$ for this new scoring function.

\begin{align*}
M &= -8 \\
m &= 8 \\
g &= 12
\end{align*}

ii. Is it valid? Why or why not?

\textit{No, because }\text{M} < \text{m and } m < 2g.

iii. If the scoring function is valid, rescore the optimal alignments you obtained in Problem 1. Show the rescored alignments. (Note that you do not need to recalculate the alignment matrix.) Do these alignments still have the same score? If not, which alignments have lower scores? Why?

\textit{Not applicable.}
(d) Suppose you replaced the values of $M$, $m$ and $g$ with $M \cdot d$, $m \cdot d$ and $g \cdot d$, where $d = 2$.

i. Write down the values of $M$, $m$, and $g$ for this new scoring function.

\[
\begin{align*}
M &= 8 \\
m &= -8 \\
g &= -12
\end{align*}
\]

ii. Is it valid? Why or why not?

Yes: $M > 0 > m > 2g$

iii. If the scoring function is valid, rescore the optimal alignments you obtained in Problem 1. Show the rescored alignments. (Note that you do not need to recalculate the alignment matrix.) Do these alignments still have the same score? If not, which alignments have lower scores? Why?

Alignment 1, Score -16

\[
-C-RECHE \\
SCOR-CH-
\]

Alignment 2, Score -16

\[
-CRECHE \\
SCORCH-
\]

In part (a), $M + 2g = 2m$, so both alignments have the same score. Multiplying $M$, $m$, and $g$ by the same factor does not change their relative values, so both alignments have the same score here, as well.
3. Pairwise alignment

(a) Compute the local alignment of **HAIKU**, a form of Japanese poetry, with **KUWAIT**, a country located at the northern tip of the Persian Gulf, using the following scoring system: $M = 2$, $m = -3$, $g = -2$.

Hand in your alignment matrix with scores on the attached alignment template. Show the traceback on the matrix by highlighting or circling the cells in the matrix that correspond to the optimal solution.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>K</th>
<th>U</th>
<th>W</th>
<th>A</th>
<th>I</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>--</td>
</tr>
<tr>
<td>K</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>U</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>--</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

(b) What is the optimal local alignment score? How many different optimal alignments are there? Show them.

*There are two optimal local alignments with a score of 4:*

- **KU**
- **UK**

*and*

- **AI**
- **IA**
4. Semi-global alignment

(a) Compute the semi-global alignment of HAIKU with KUWAIT, with a scoring function of $M = 2$, $m = -3$, $g = -2$. Set up your semi-global calculation to allow the alignment to start late in HAIKU and end early in KUWAIT.

Hand in your alignment matrix with scores on the attached alignment template. Show the traceback on the matrix by highlighting or circling the cells in the matrix that correspond to the optimal solution.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>K</th>
<th>U</th>
<th>W</th>
<th>A</th>
<th>I</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>-2</td>
<td>-4</td>
<td>-6</td>
<td>-8</td>
<td>-10</td>
<td>-12</td>
</tr>
<tr>
<td>H</td>
<td>-2</td>
<td>0</td>
<td>-4</td>
<td>-6</td>
<td>-8</td>
<td>-12</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>-4</td>
<td>-4</td>
<td>0</td>
<td>-6</td>
<td>-8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>-6</td>
<td>-6</td>
<td>-8</td>
<td>0</td>
<td>-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>-10</td>
<td>-12</td>
<td>-5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U</td>
<td>-12</td>
<td>-4</td>
<td>-4</td>
<td>-2</td>
<td>-2</td>
<td>-4</td>
<td></td>
</tr>
</tbody>
</table>

(b) What is the score of the optimal semi-global alignment? How many different optimal alignments are there? Show the optimal semi-global alignment(s) you obtained. Do not put indels under leading and lagging symbols, where gaps were not penalized.

*There is one alignment with an optimal score of four.*

```
HAIKU
KUWAIT
```