

2 Pairwise alignment - space consumption

Generelt

Ensartethed i sekvensen har høj sandsynlighed for at have ensartet funktion eller struktur.

Edit Distance

Number of operations needed to transform A into B.

$a \rightarrow -$ deletion

$a \rightarrow b$ substitution

$- \rightarrow b$ insertion

In bioinformatics $a \rightarrow g$ might be more alike than $a \rightarrow c$. Therefore a scorematrix is defined along with a gapcost.

Linear gapcost

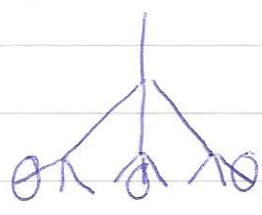
Given A, B, S , gapcost find $OPT(A, B)$

Trying all possible alignments too slow.

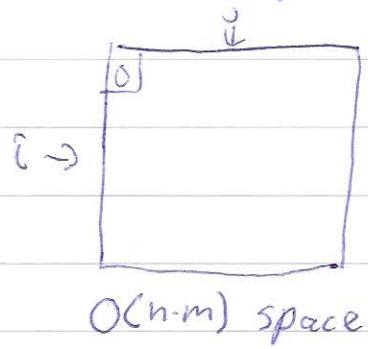
Define $cost(i, j) \equiv cost$ of $OPT(A[1...i], B[1...j])$

$$cost(i, j) = \max \begin{cases} 0 & \begin{array}{l} \text{---} A[i] \\ \text{---} B[j] \end{array} \\ cost(i-1, j-1) + S(A[i], B[j]) & \begin{array}{l} \text{---} A[i] \\ \text{---} B[j] \end{array} \\ cost(i-1, j) + gapcost & \begin{array}{l} \text{---} A[i] \\ \text{---} \end{array} \\ cost(i, j-1) + gapcost & \begin{array}{l} \text{---} \\ \text{---} B[j] \end{array} \end{cases}$$

recursion will calculate same entries multiple times

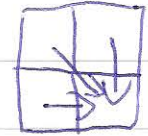


Solution: use dynamic table instead:



$(n+1) \times (m+1)$

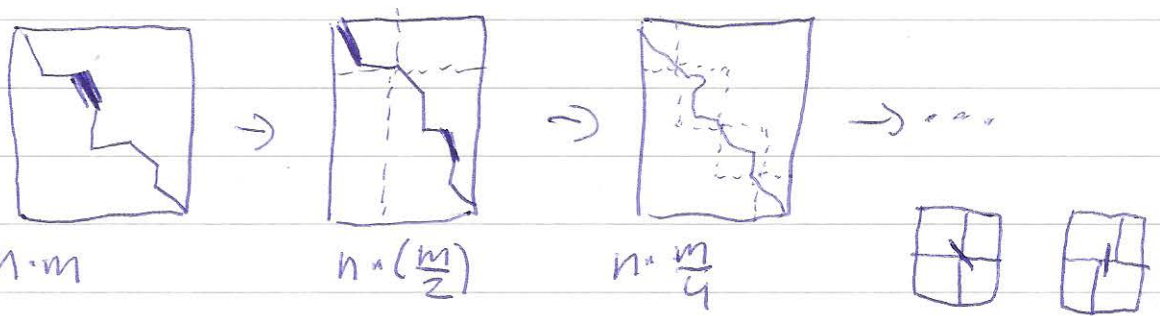
only need to store two rows in memory:



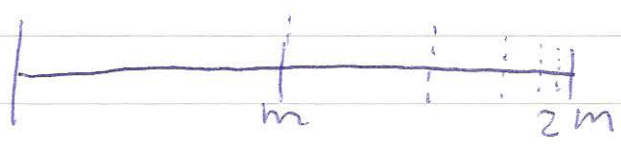
$O(n)$ space

Hirschberg

Find middle edge in $O(n \cdot m)$ time and $O(n)$ space.



$$n \cdot (m + \frac{m}{2} + \dots + 1) \leq n \cdot 2m = O(n \cdot m)$$



$$\text{OPT}(A, B) = \text{OPT}(A^R, B^R)$$

