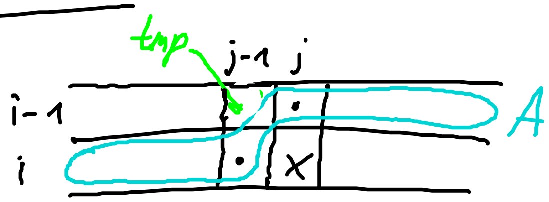
Exercise 3

Problem 7



procedure ScoresLine(S, T):

  for $j = 1, \dots, n$

    $A[j] = j \cdot g$

  for $i = 1, \dots, m$

    $tmp = (i-1) \cdot g$

    for $j = 1, \dots, n$

$$new = \begin{cases} tmp + p(S_i, T_j) \\ A[j] + g \\ g + \begin{cases} i \cdot g, & j = 1 \\ A[j-1], & j > 1 \end{cases} \end{cases}$$

      $tmp = A[j]$

      $A[j] = new$

  return A

procedure Score(S, T)

  $A := \text{ScoresLine}(S, T)$

  return $A[n]$

Space: array $A \rightsquigarrow O(n)$ entries

$+$ $O(1)$ temporary variables

Runtime: · initialisation of A: $O(n)$

· "filling out" the matrix $M \rightsquigarrow O(m \cdot n)$

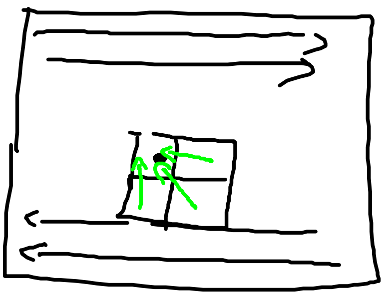
7b)

Observation 1: To align S, T, $m = |S|$, $n = |T|$

$\rightsquigarrow$ a matrix of size $n \times m$ is enough

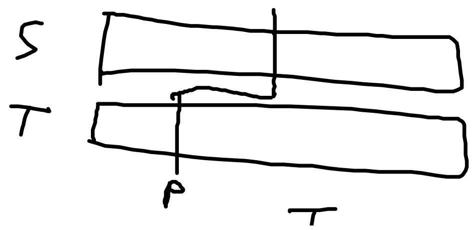
(eliminate first row and column)

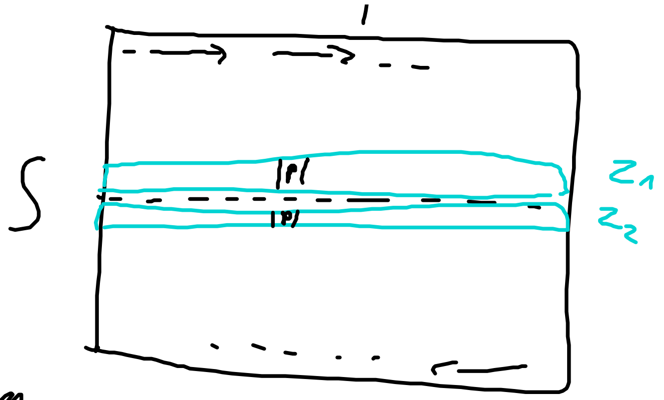
Observation 2:



We can fill out the alignment matrix from bottom to top.

Idea:





- Fill out upper half from top to bottom and
- lower half from bottom to top

$Z_1[p]$ will contain the scoring of an optimal alignment of $S_{1,\frac{m}{2}}$ and $T_{1,p}$

$Z_2[p]$ will contain the scoring of an optimal alignment of $S_{\frac{m}{2}+1,m}$ and $T_{p+1,n}$

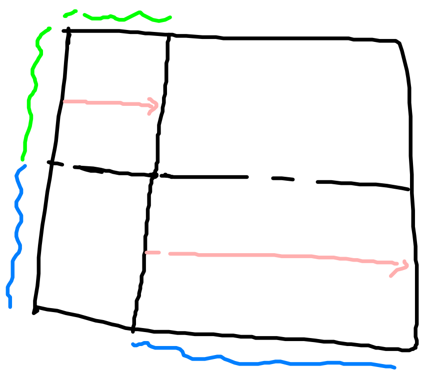
$\rightsquigarrow$ minimize $Z_1[p] + Z_2[p]$ by one iteration over $1,\dots,n$

- Find recursively optimal alignments of $S_{1,\frac{m}{2}}$ and $T_{1,p}$ and $S_{\frac{m}{2}+1,m}$ and $T_{p+1,n}$
- Stop recursion if $|S| = 1$
  - $|T| = 0$ $\rightsquigarrow$ deletion

- $|T| > 0 \rightsquigarrow$ iterate over $i = 1, \dots, |T|$, calculate $\delta(S_1, T_i)$, find minimum m
  - If $m > 2g \rightsquigarrow$ 1 deletion, $|T|$ insertions
  - If $m \leq 2g \rightsquigarrow$ replace that one character, insert the $|T|-1$ others
- output alignment immediately
  $\rightsquigarrow$ works if we do the "left" recursion first

Running time

- First level of recursion
  $\rightsquigarrow$ $m \times n$ Matrix, fill out in $m \cdot n$ constant time steps
  calculate p in $n+1$ steps
- Second Level of recursion



$$\rightsquigarrow \frac{m}{2} \cdot p + \frac{m}{2} \cdot (n-p)$$

$$= \frac{m \cdot n}{2} \text{ cells to fill out}$$

Calculate p in $(p+1) + (n-p+1) = n+2$ steps

$\vdots$

$ld(m)$ recursion levels

$$\Rightarrow T = \sum_{i=0}^{ld(m)} \left[\frac{m \cdot n}{2^i} + n + 2^i \right] \leq m \cdot n + n[ld(m)+1] + 2m - 1$$

$$= O(m \cdot n)$$

Space complexity: • Array (see a)): $\mathcal{O}(n)$

• recursion depth $ld(n) \rightsquigarrow \mathcal{O}(\log(n))$

$\Rightarrow \mathcal{O}(n + \log(n))$

Problem 8

DSSP $\leq_p$ MASP

• NP: a) guessing a solution $\rightsquigarrow \mathcal{O}(l \cdot k)$ where $l = \sum_{i=1}^{k} |S_i|$

b) compute the score $\rightsquigarrow \mathcal{O}(l \cdot k^2)$

• $S_1, \dots, S_k \in \{0,1\}^*$, $m \in \mathbb{N}$

$$p(0,0) = p(1,1) = p(-,-) = 1$$
$$p(0,1) = p(1,0) = m\binom{k}{2} + 1$$
$$g = 1$$
$$d = m\binom{k}{2}$$

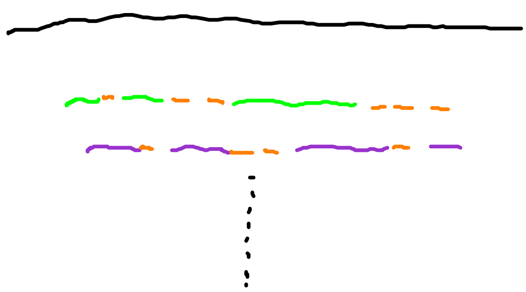
I_{DSSP} yes-instance for DSSP $\Leftrightarrow f(I_{DSSP})$ yes-instance for MASP

$\Leftarrow$: MA with score $\leq d$ exists

$\rightsquigarrow$ every column contributes $\binom{k}{2}$ to the total score since there can't be any (real) substitution

$\rightsquigarrow$ construct unique consensus from the MA (every column contains only $\{-,0\}$ or $\{-,1\}$) which is also a supersequence for $S_1,\dots,S_k$

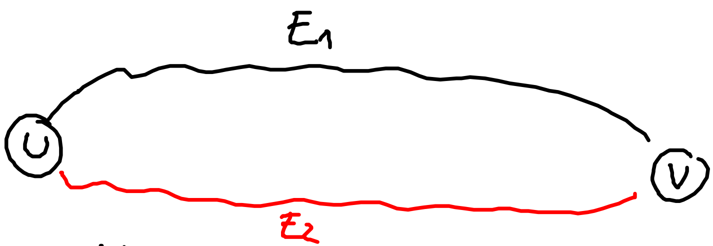
$\Longrightarrow$ we have supersequence for $S_1,\dots,S_k$ of length $\leq n$,



This sequence allows immediately the construction of a MA of $S_1,\dots,S_k$ without mismatches.

g) $\left(g^*(u,v) = g(u,v) - (\xi(u) - \xi(v))\right)$

Let $E_1 \subseteq E$, $E_1 = \{(v_1,v_2),\dots,(v_{k-1},v_k)\}$ be arbitrary shortest path in G w.r.t g^*



Suppose that path $E_2 \neq E_1$ is shorter w.r.t g

$$E_2 = \{ (v_1', v_2'), \dots, (v_{n-1}', v_n') \}$$

$$\Rightarrow \sum_{(v_i, v_{i+1}) \in E_1} g(v_i, v_{i+1}) > \sum_{(v_i', v_{i+1}') \in E_2} g(v_i', v_{i+1}')$$

$$\Rightarrow \sum \left(g^*(v_i, v_{i+1}) + \left(\xi(v_i) - \xi(v_{i+1}) \right) \right)$$

$$> \sum \left(g^*(v_i', v_{i+1}') + \left(\xi(v_i') - \xi(v_{i+1}') \right) \right)$$

telescopic
$\Rightarrow$
sum

$$\sum \left(g^*(v_i, v_{i+1}) \right) + \xi(\overset{v_1}{\overset{\shortparallel}{u}}) - \xi(\overset{v_k}{\overset{\shortparallel}{v}})$$

$$> \sum_{E_2} \left(g^*(v_i', v_{i+1}') \right) + \cancel{\xi(u)} - \cancel{\xi(v)}$$

$$\Rightarrow \sum_{E_1} g^*(v_i, v_{i+1}) > \sum_{E_2} g^*(v_i', v_{i+1}')$$

E_1 is shortest path wrt g^*