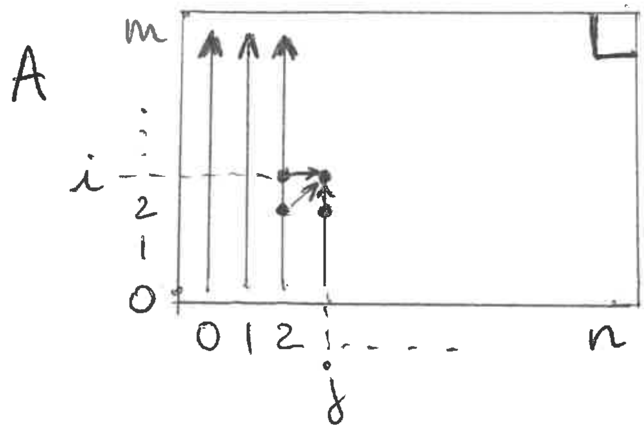
Sequence Alignment

- computing the sequence alignment of two sequences X and Y using the dynamic programming algorithm Alignment(X,Y) has $RT = O(m \cdot n)$ and space $= O(m \cdot n)$



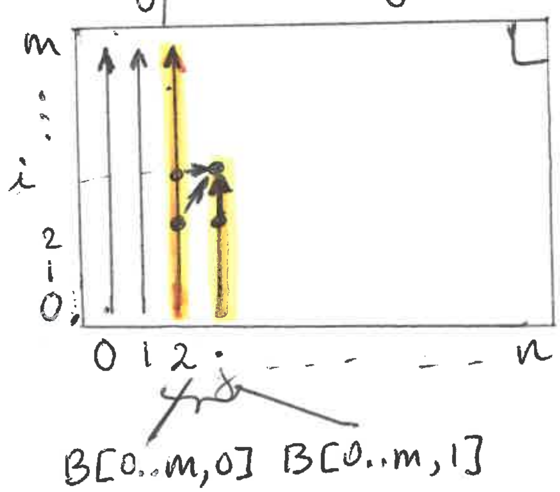
$X = x_1 x_2 \cdots x_m$

$Y = y_1 y_2 \cdots y_n$

$A[0..m, 0..n]$

$$A[i,j] = \min\{\alpha_{x_i y_j} + A[i-1,j-1], \delta + A[i-1,j], \delta + A[i,j-1]\}$$

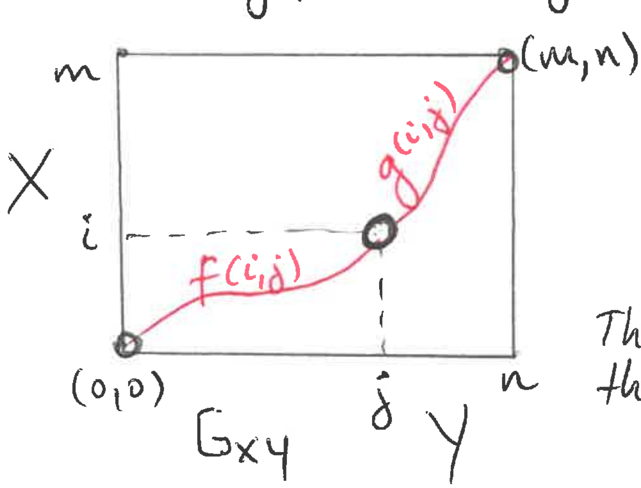
- Space-efficient alignment:



array $B[0..m, 0..1]$

B has size $(m+1) \times 2 = \Theta(m)$

- defining $f(i,j)$ and $g(i,j)$



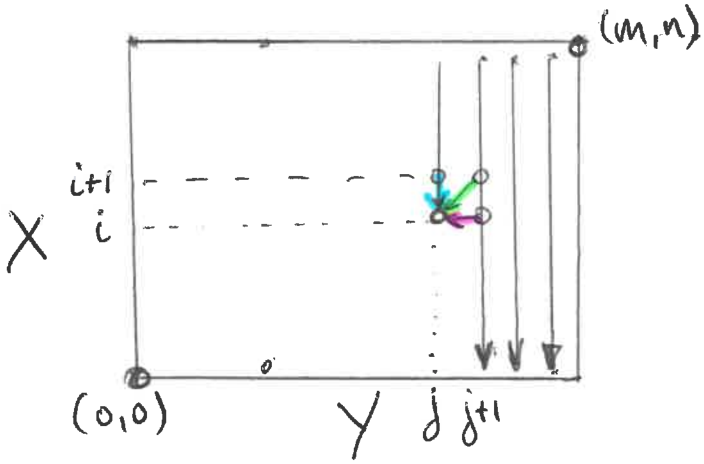
$f(i,j)$ – length of the shortest-path $(0,0)$ to (i,j)

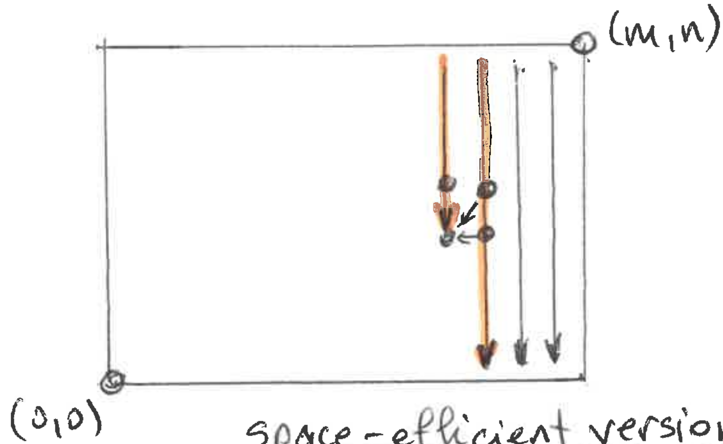
$f(i,j) = OPT(i,j)$

$g(i,j)$ – length of the shortest-path (i,j) to (m,n)

The length of the shortest-path $(0,0)$ to (m,n) that passes through (i,j) is $f(i,j) + g(i,j)$

- How can we compute $g(i,j)$?
  - use a backward formulation of the dynamic programming





space-efficient version

for $i<m$ and $j<n$, we have:

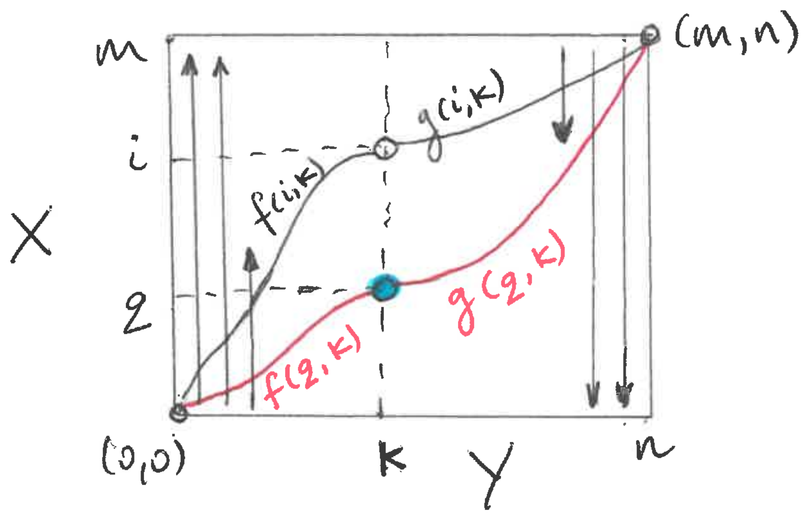
$$g(i,j) = \min\{\alpha_{x_{i+1}y_{j+1}} + g(i+1,j+1),\ \delta + g(i,j+1),\ \delta + g(i+1,j)\}$$

Backward-Space-Efficient-Alignment (X,Y) has

$RT = O(m \cdot n)$

$\text{space} = O(m)$

- Combining the forward and backward formulations



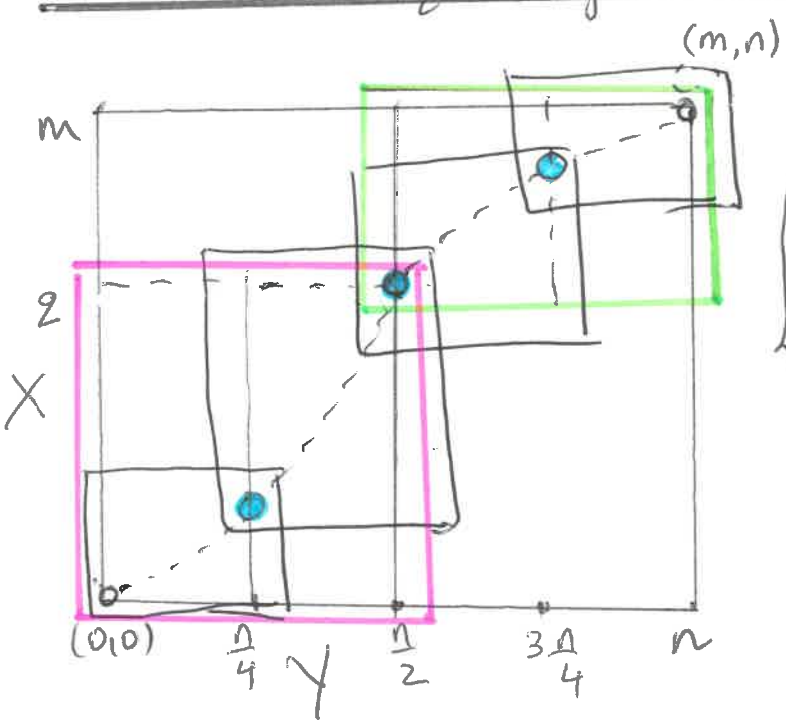
- Let k be any number in $\{0,1,\dots,n\}$

- compute $f(i,k)$ for each $i = 0..m$ using Space-Efficient-Alignment()
  $RT = O(m \cdot n)$, space $O(m)$
- compute $g(i,k)$ for each index $i = 0..m$ using Backward-Space-Efficient-Alignment()
  $RT = O(m \cdot n)$, space $O(m)$
- for each index $i = 0..m$ compute $f(i,k) + g(i,k)$
  $RT = O(m)$, space $O(m)$
- let q be the index for which $f(q,k) + g(q,k)$ has a minimum value

Then there is a shortest path (minimum length path) from $(0,0)$ to (m,n) which contains the node (q,k).

- the node (q,k) is part of an <u>optimal alignment</u>
- we computed (q,k) with $RT = O(m \cdot n)$, space $O(m)$

Divide-and-conquer Algorithm



- assume that n is a power of 2

$T(m,n)$ – running time

<u>Divide</u> takes $O(m \cdot n)$

<u>Conquer</u>

$T(q, \frac{n}{2}) + T(m-q, \frac{n}{2})$

<u>Combine</u>: nothing

$$T(m,n) = O(m \cdot n) + T\left(q, \frac{n}{2}\right) + T\left(m-q, \frac{n}{2}\right)$$

$$\boxed{\begin{array}{l} T(m,n) = O(m \cdot n) \\ \text{space } O(m+n) \end{array}}$$