12.1 Computing alignments in only linear space

about local alignment

20160405

Chanyoung Song

Basic Idea for Computing Similarity in O(m) space

- S_1 = cacd
- S_2 = cadb

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

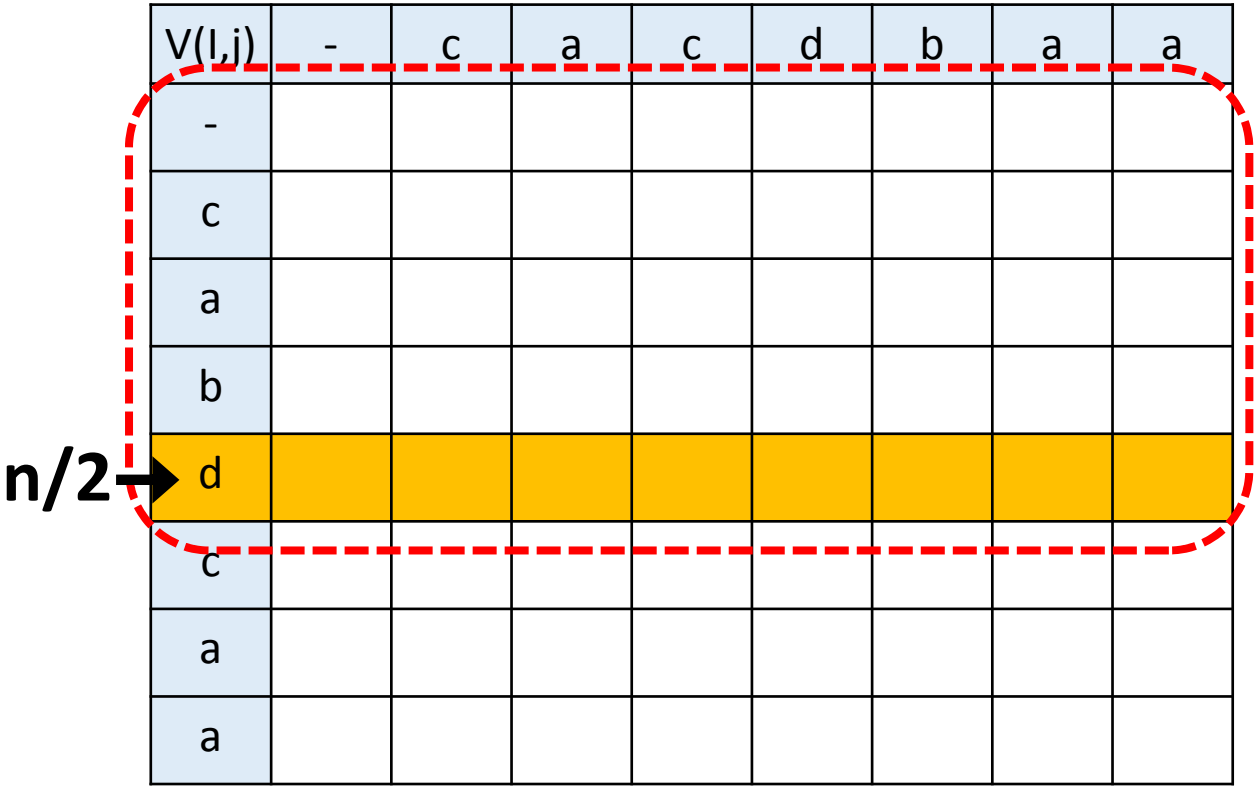
V(i,j)	-	c	a	c	d
-	0	-2	-3	-5	-6
c	-2	0	-1	-3	-4
a					
d					
b					

i-th row

ONLY NEED i-1-th ROW
FOR COMPUTING i-th ROW

V(i,j) = max[V(i-1, j-1) +s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)_s(-,S_2(j))]

1) Find a position k* in row n/2

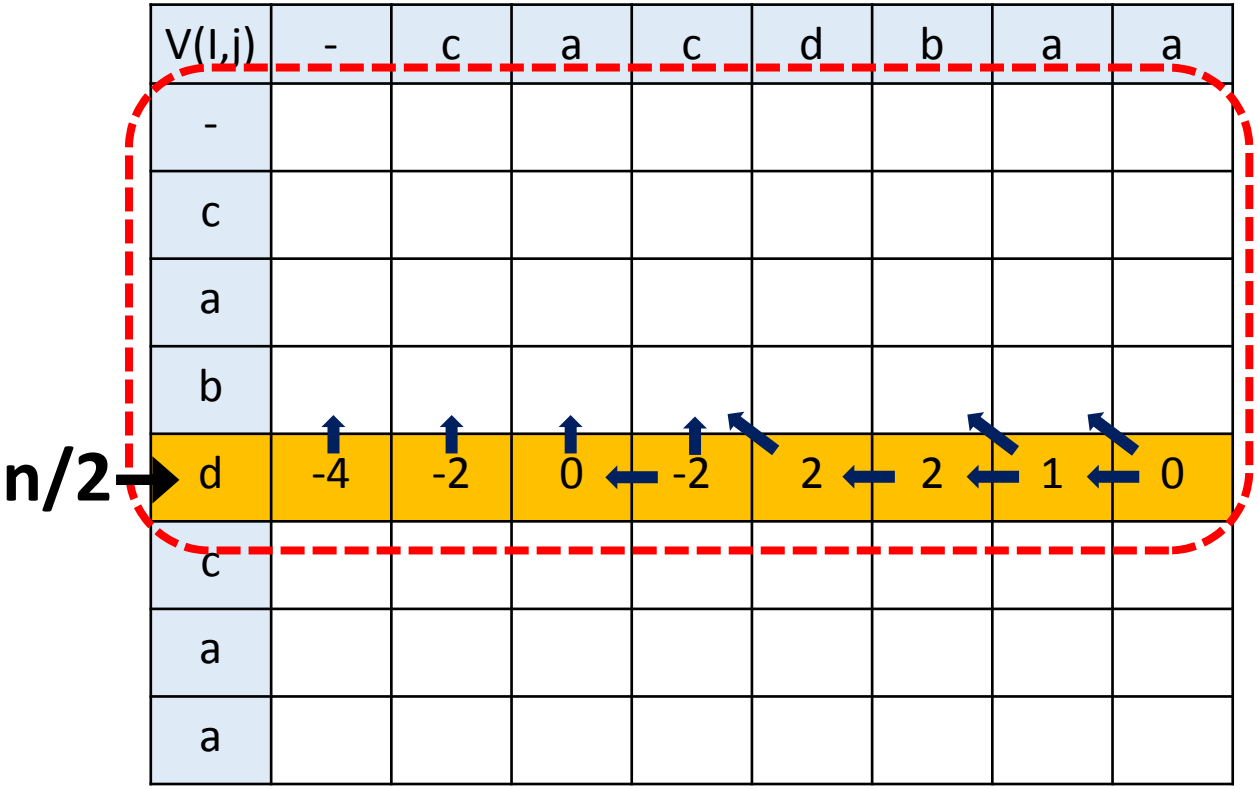


1) For V(n/2, k), $0 \leq k \leq m$
   - Time : O(nm/2)
   - Space : O(m)
   - Pointer set at row n/2

- Let k^* be a position k

  that maximizes [V(n/2, k) + V^r(n/2, m-k)]

1) Find a position k* in row n/2

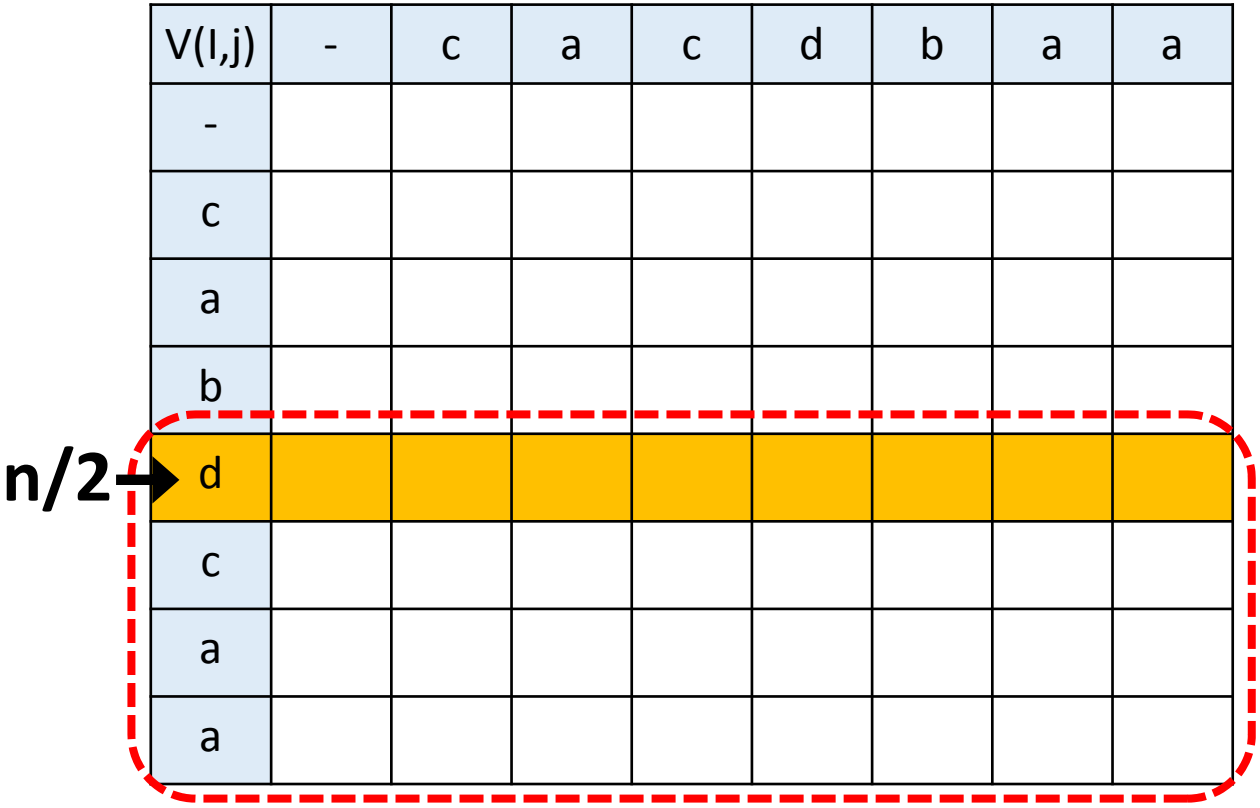


1) For V(n/2, k), $0 \leq k \leq m$
- Time : O(nm/2)
- Space : O(m)
- Pointer set at row n/2

- Let k^* be a position k

  that maximizes [V(n/2, k) + V^r(n/2, m-k)]

1) Find a position k* in row n/2



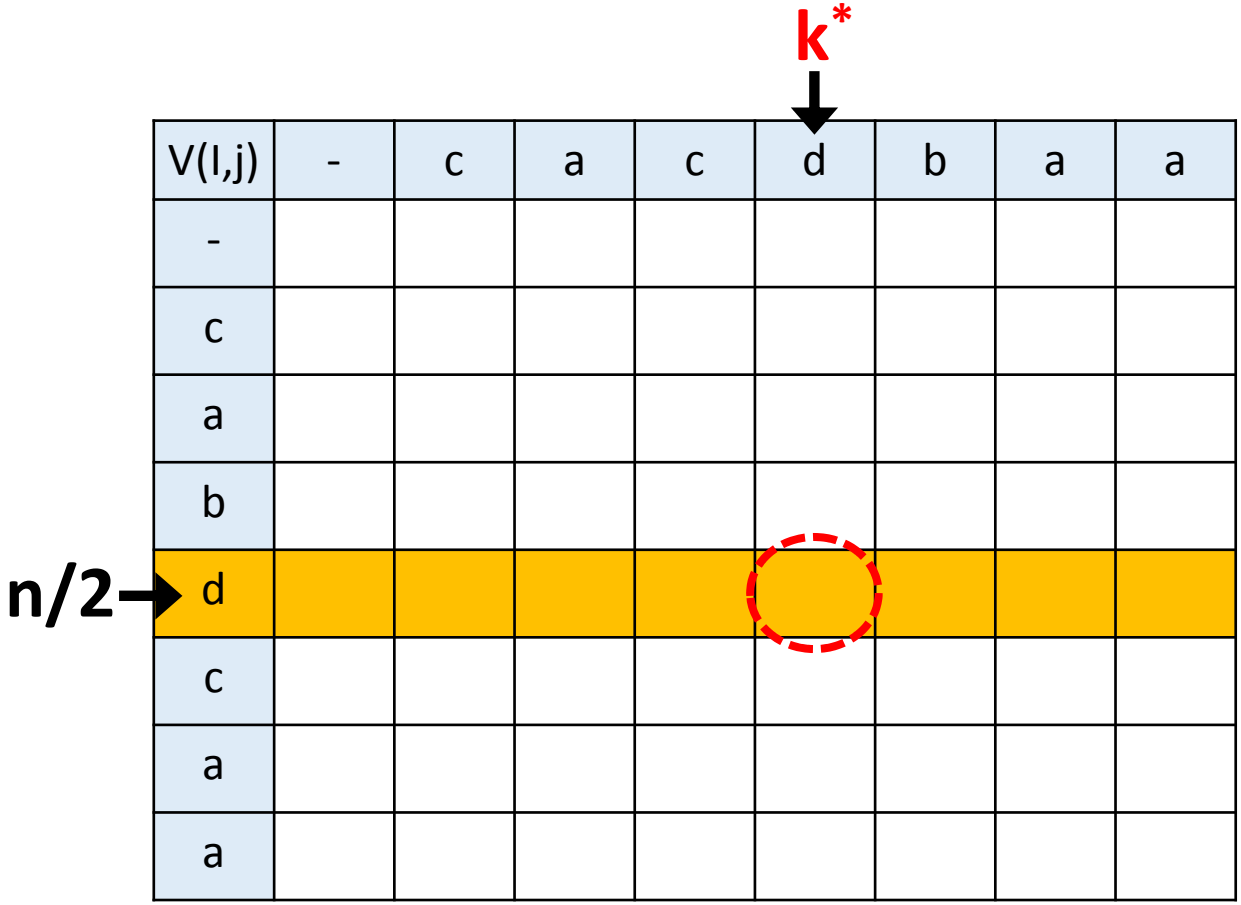
1) For V(n/2, k), $0 \leq k \leq m$
   - Time : O(nm/2)
   - Space : O(m)
   - Pointer set at row n/2

2) For $V^r(n/2, m-k)$, $0 \leq k \leq m$
   - Time : O(nm/2)
   - Space : O(m)
   - Pointer set at row n/2

- Let k^* be a position k

  that maximizes [$V(n/2, k) + V^r(n/2, m-k)$]

1) Find a position k* in row n/2



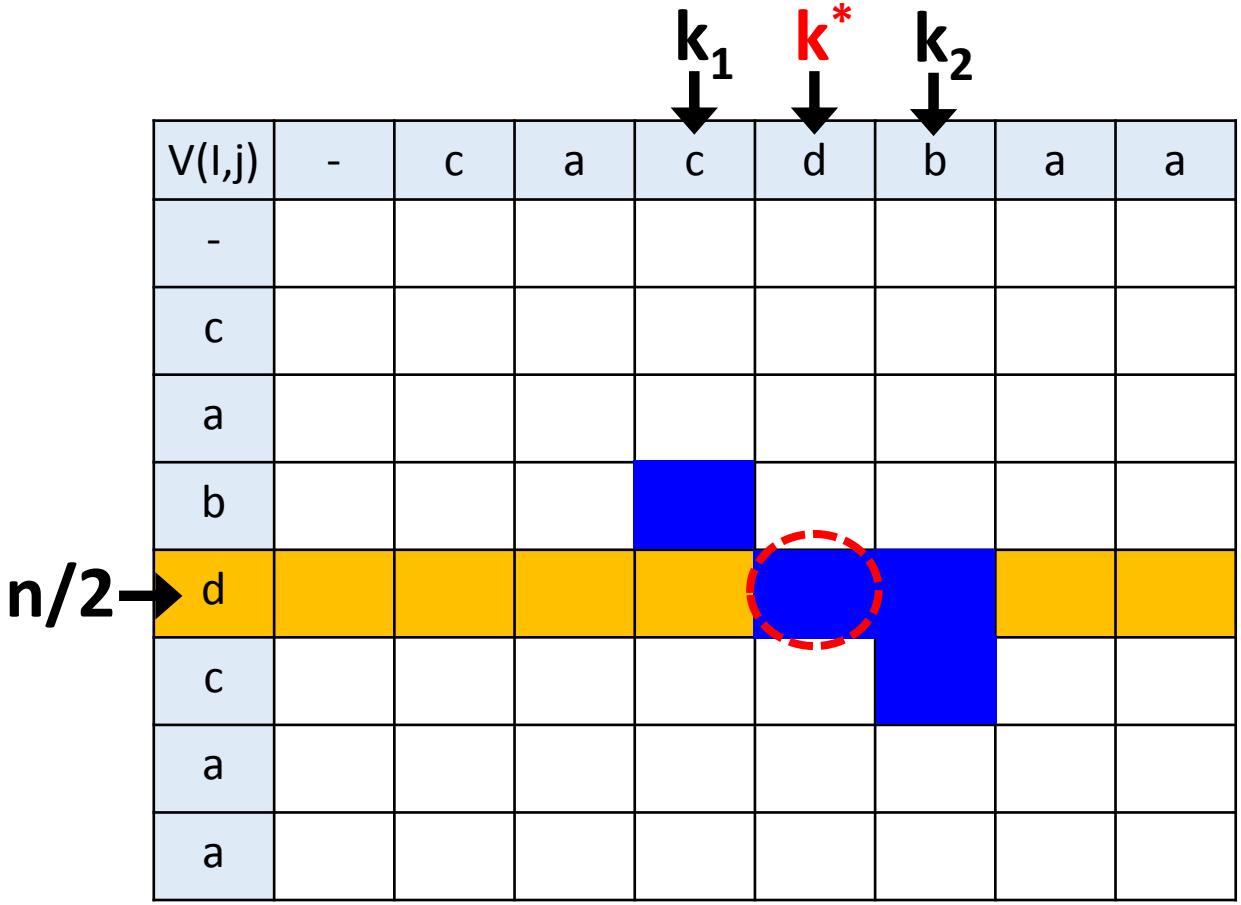
- Let k^* be a position k that maximizes [$V(n/2, k) + V^r(n/2, m-k)$]

1) For $V(n/2, k)$, $0 \leq k \leq m$
   - Time : O(nm/2)
   - Space : O(m)
   - Pointer set at row n/2

2) For $V^r(n/2, m-k)$, $0 \leq k \leq m$
   - Time : O(nm/2)
   - Space : O(m)
   - Pointer set at row n/2

A position k^* in row n/2 can be found in O(nm) time and O(m) space.

2) Find a Subpath $L_{n/2}$



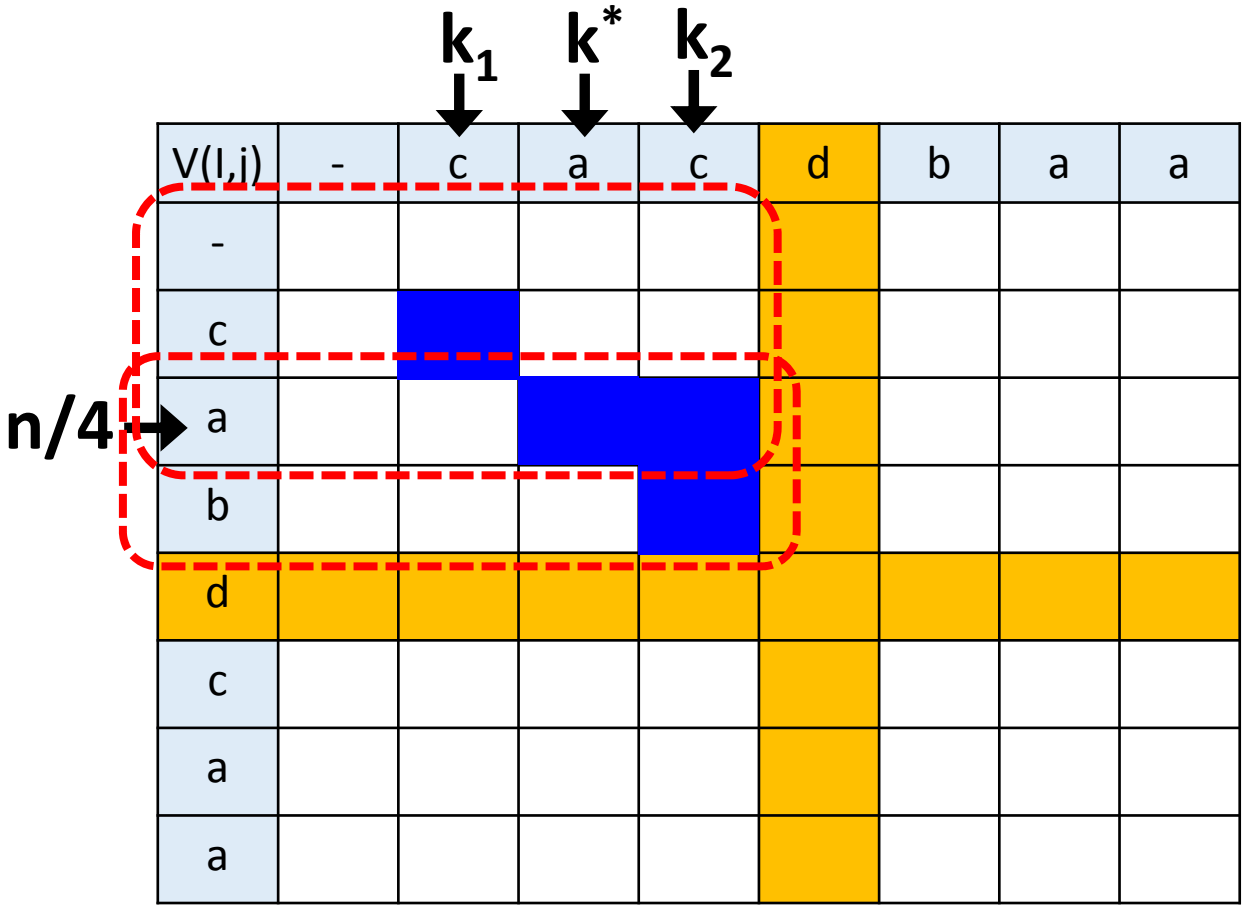
- $L_{n/2}$ (subpath)
- k^*, k_1, k_2
- Time : cnm

  = c(n/2)m×2
- Space : O(m)

k_1 : Col. Idx of 1st cell in n/2 - 1 row from cell (n/2, k^*)

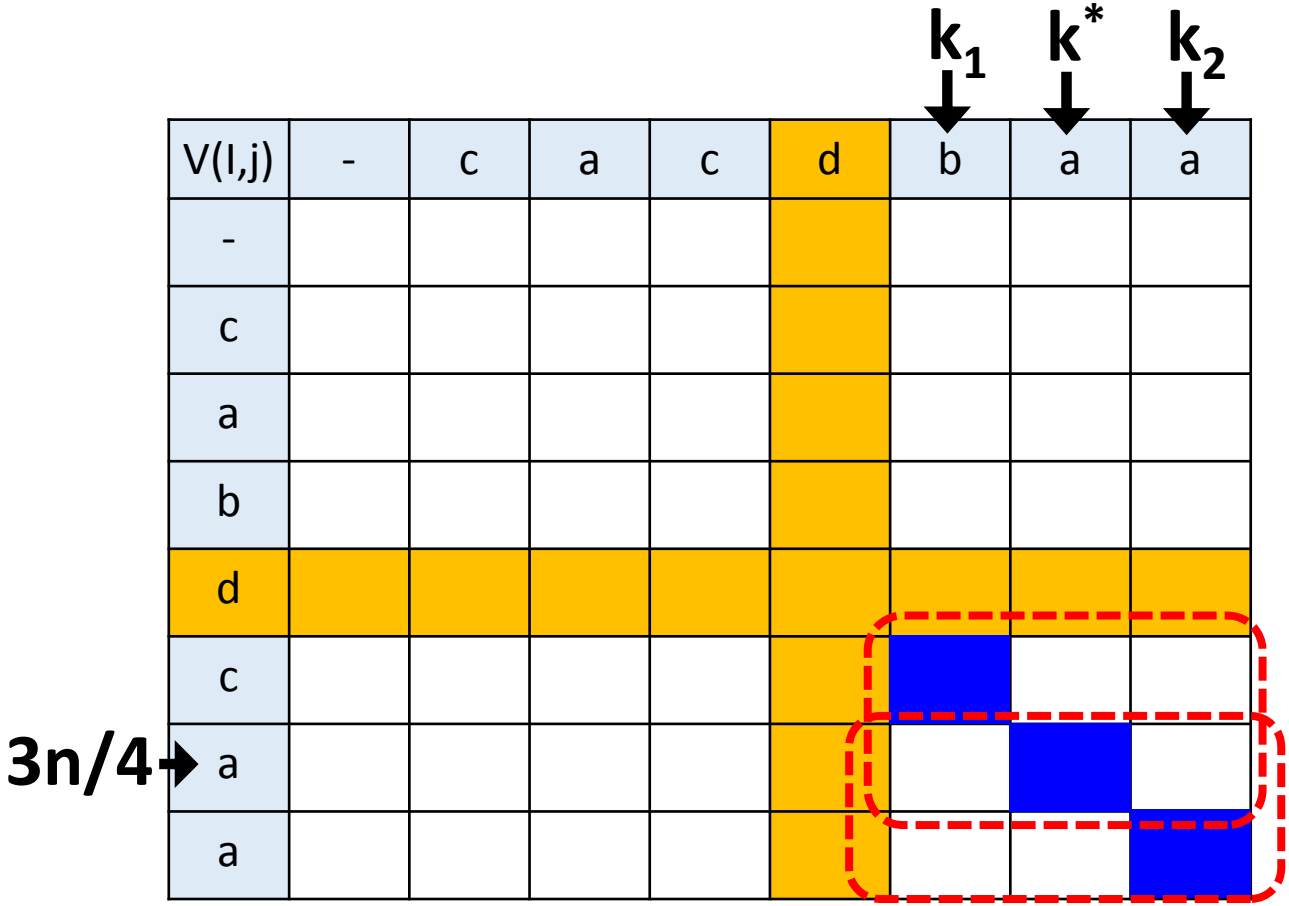
k_2 : Col. Idx of 1st cell in n/2 + 1 row from cell (n/2, k^*)

Full Idea : Use Recursion



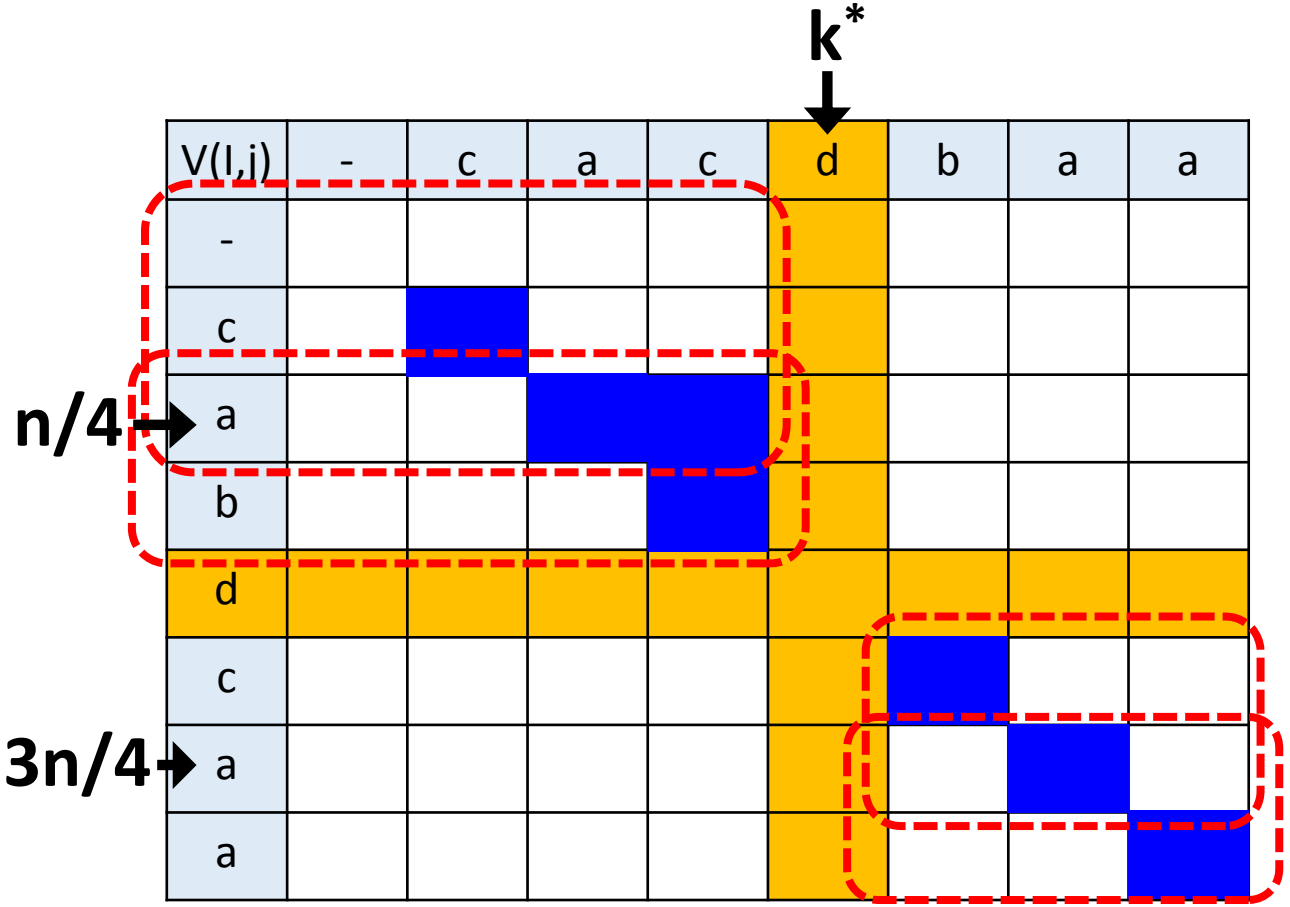
- $L_{n/4}$ (subpath)
- k^*, k_1, k_2

Full Idea : Use Recursion



- $L_{3n/4}$ (subpath)
- k^*, k_1, k_2

Full Idea : Use Recursion



- $L_{n/4}$, $L_{3n/4}$ (subpath)
- k …

- Time : (1/2)cnm

  = $c(n/4)(k^*)\times 2$

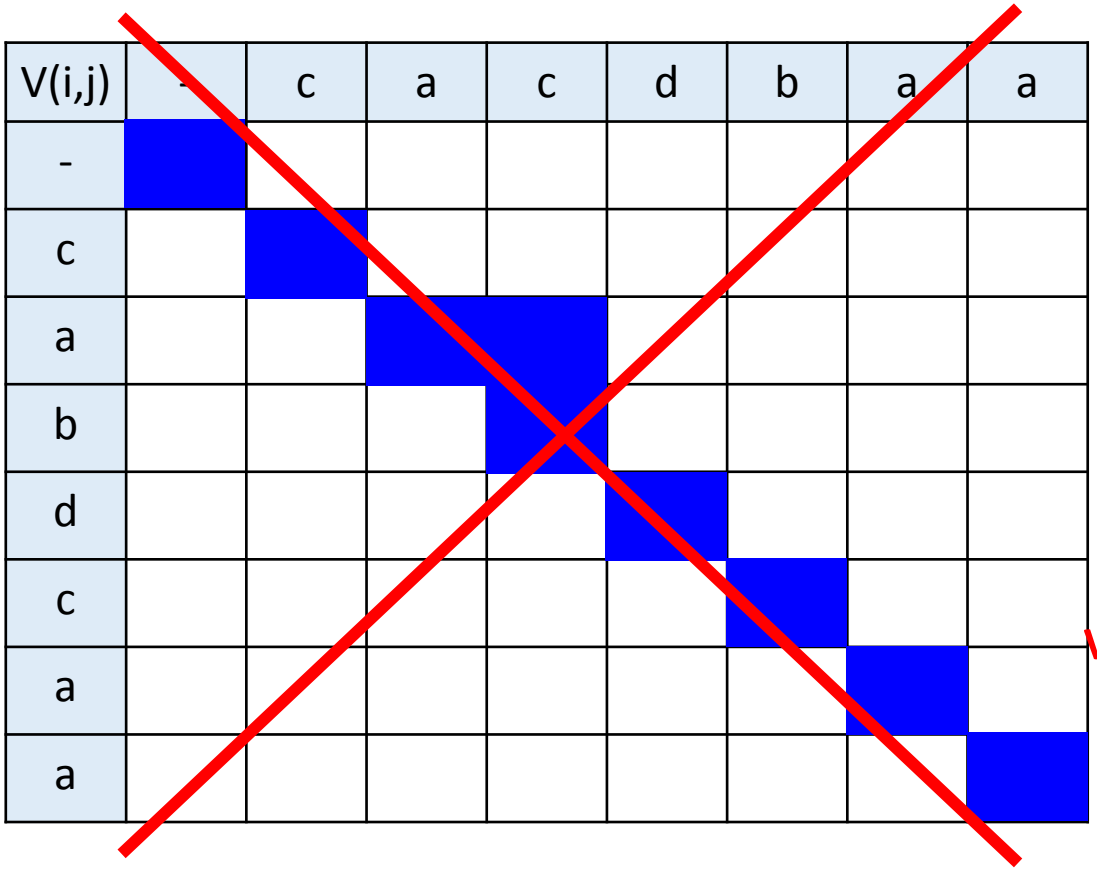
  + $c(n/4)(m - k^*) \times 2$

- Space : O(m)

Theorem 12.1.1

- Using the procedure OPTA, an optimal alignment of two strings of length n and m can be found in $\sum_{i=1}^{logn} cnm/2^{i-1} \leq 2cnm$ time and O(m) space.

Extension to Local Alignment



V(i,j)	-	c	a	c	d	b	a	a
-								
c								
a								
b								
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

WE HAVE TO FIND A PATH FROM CELL (I^*,J^*) OF MAX V(I^*,J^*)

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0,\ V(i-1,\ j-1) + s(S_1(i),\ S_2(j)),\ V(i-1,j) + s(S_1(i),\ -),\ V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0,\ j > 0)$

Extension to Local Alignment

1) Find substrings α, β from S_1, S_2
2) Do O(m) space alignment for α, β

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-	0	0	0	0	0	0	0	0
c	0	0	0	0	0	0	0	0
a								
b								
d								
c								
a								
a								

STORE MAX V(I,J) INDEX

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX V(i^*, j^*)

V(i,0) = V(0,j) = 0

V(i,j) = max[0, V(i-1, j-1) +s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)_s(-,S_2(j))]

(i > 0, j > 0)

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c	0	0	0	0	0	0	0	0
a								
b								
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c	0	0	0	0	0	0	0	0
a	0	0						
b								
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c	0	0	0	0	0	0	0	0
a	0	0	1					
b								
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c	0	0	0	0	0	0	0	0
a	0	0	1	0				
b								
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\, 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j)) \,]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c								
a	0	0	1	0	0	0	1	1
b	0	0	1	0	0	3	2	1
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX V(i^*, j^*)

V(i,0) = V(0,j) = 0

V(i,j) = max[0, V(i-1, j-1) +s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)_s(-,S_2(j))]

(i > 0, j > 0)

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c								
a								
b								
d								
c								
a	0	0	1	0	1	1	2	2
a	0	0	1	0	0	0	2	3

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX V(i^*, j^*)

WHERE IS THE START CELL?

V(i,0) = V(0,j) = 0

V(i,j) = max[0, V(i-1, j-1) +s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)_s(-,S_2(j))]

(i > 0, j > 0)

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c	0	0	0	0	0	0	0	0
a	0	0	1					
b								
d								
c								
a								
a								

STORE START CELL INDEX BY EACH CELL

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

2) Find Start Cell
: $V(i', j') = 0$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c	0	0	0	0	0	0	0	0
a	0	0	1					
b								
d								
c								
a								
a								

STORE START CELL INDEX OF MAX V(I,J)

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

2) Find Start Cell
: $V(i', j') = 0$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c	0	0	0	0	0	0	0	0
a	0	0	1	0	0	0	1	1
b								
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

2) Find Start Cell
: $V(i', j') = 0$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\, 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j)) \,]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c								
a	0	0	1	0	0	0	1	1
b	0	0						
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

2) Find Start Cell
: $V(i', j') = 0$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2

V(I,j)	-	c	a	c	d	b	a	a
-								
c								
a	0	0	1	0	0	0	1	1
b	0	0	1					
d								
c								
a								
a								

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

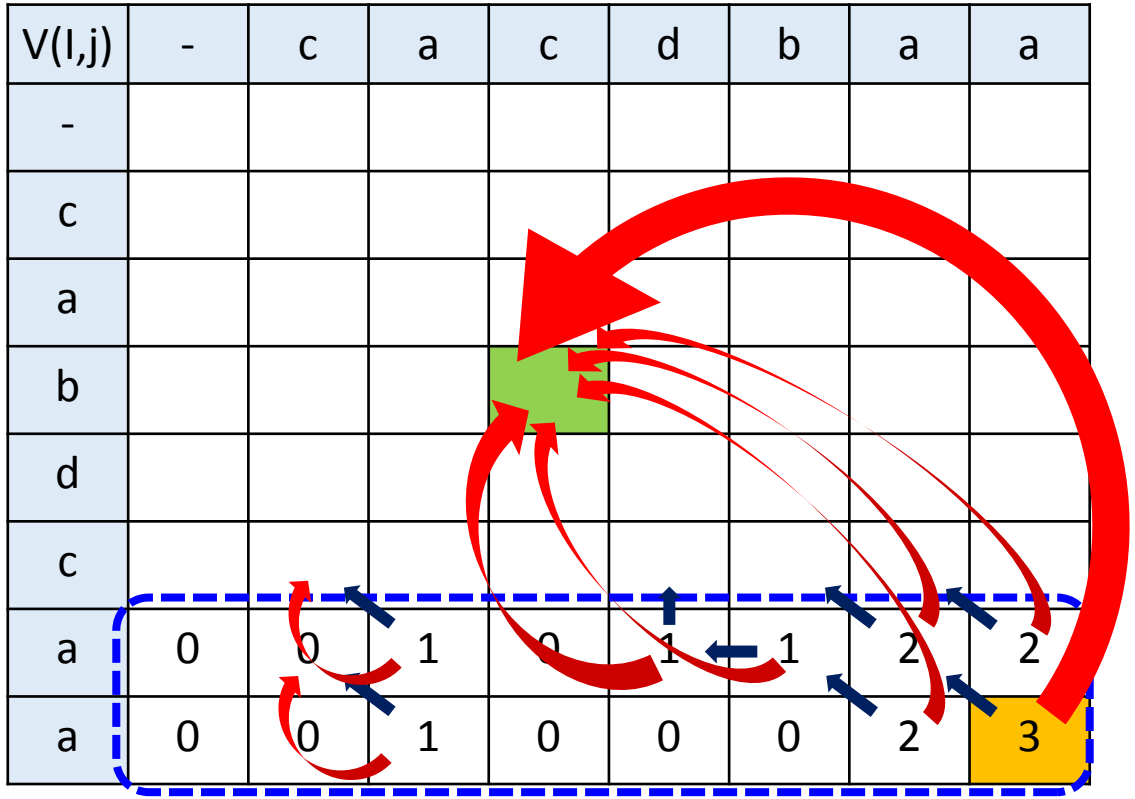
2) Find Start Cell
: $V(i', j') = 0$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0, j > 0)$

1) Find substrings α, β from S_1, S_2



V(I,j)	-	c	a	c	d	b	a	a
-								
c								
a								
b								
d								
c								
a	0	0	1	0	1	1	2	2
a	0	0	1	0	0	0	2	3

s	a	b	c	d	-
a	1	-3	-2	0	-1
b		3	-2	-1	0
c			0	-4	-2
d				3	-1
-					0

1) Find End Cell
: MAX $V(i^*, j^*)$

2) Find Start Cell
: $V(i', j') = 0$

$V(i,0) = V(0,j) = 0$

$V(i,j) = \max[\ 0, V(i-1, j-1) + s(S_1(i), S_2(j)), V(i-1,j) + s(S_1(i), -), V(i,j-1)\_s(-,S_2(j))\ ]$

$(i > 0, j > 0)$

Extension to Local Alignment

1) Find substrings α, β from S_1, S_2

2) Do O(m) space alignment for α, β

O(m) space
O(nm) time