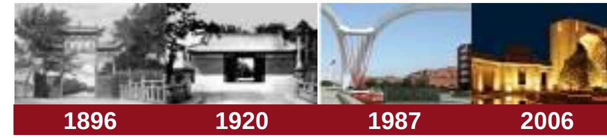
上海交通大學
SHANGHAI JIAO TONG UNIVERSITY



Chapter 3: Dynamic Programming

Chaochun Wei

Spring 2019

Contents

- **Reading materials**
- **Introduction**
  - **Dynamic programming**
  - **Greedy algorithm**

Reading

Cormen book:

Thomas, H. ,Cormen, Charles, E., Leiserson, and Ronald, L., Rivest . Introduction to Algorithms, The MIT Press.

(read Chapter 16 and 17, page 299-355).

Elements of dynamic programming

- **Two elements are required**
  1. **Optimal substructure**
     - An optimal solution contains within it optimal solutions to the subproblems
  2. **Overlapping subproblems**
     - Recursive formula exists

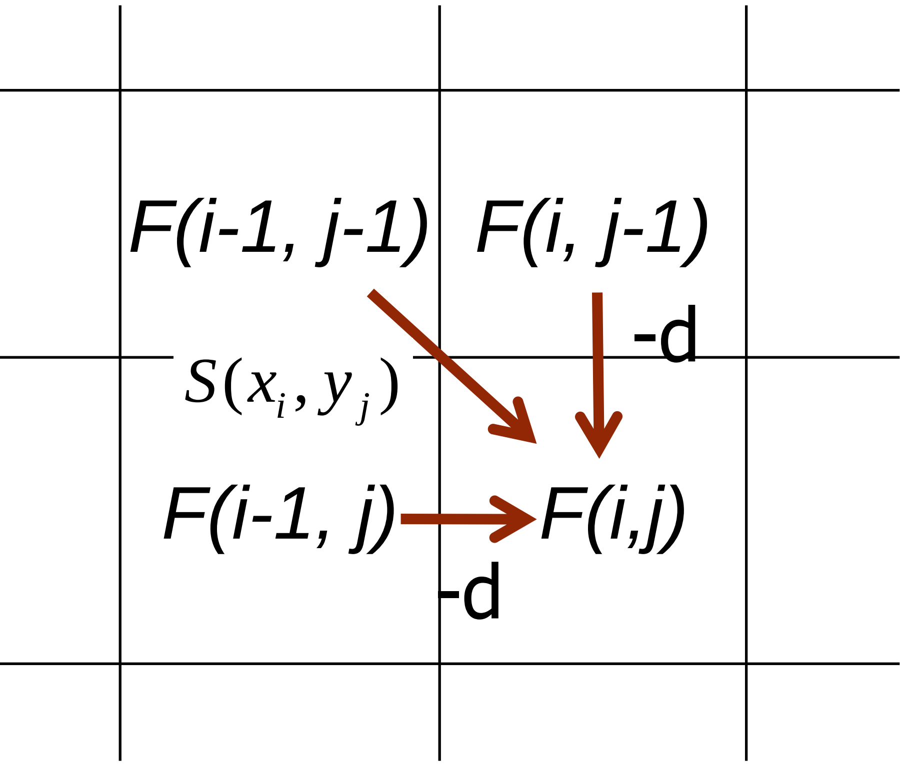
Needleman/Wunsch global alignment (1970)

- Two sequences $X = x_1 \ldots x_n$ and $Y = y_1 \ldots y_m$
- Let $F(i, j)$ be the optimal alignment score *of* $X_{1\ldots i}$ of X up to x_i and $Y_{1\ldots j}$ of Y up to Y_j $(0 \le i \le n,\ 0 \le j \le m)$, *then we have*

$$F(0,0) = 0$$

$$F(i,j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - d \\ F(i, j-1) - d \end{cases}$$

Needleman/Wunsch global alignment (1970)



$$F(0,0) = 0$$

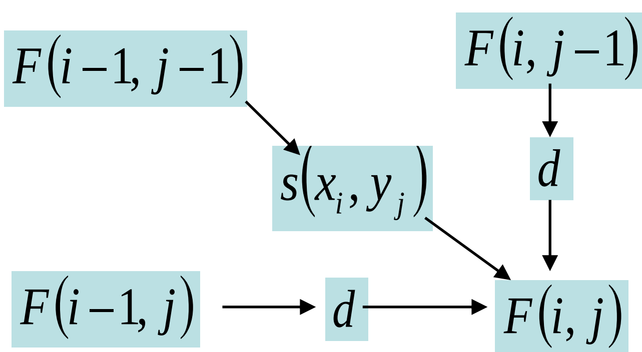
$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - d \\ F(i, j-1) - d \end{cases}$$

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal alignment of AAG and AGC.
Use a gap penalty of d=-5.

		A	A	G
A				
G				
C				

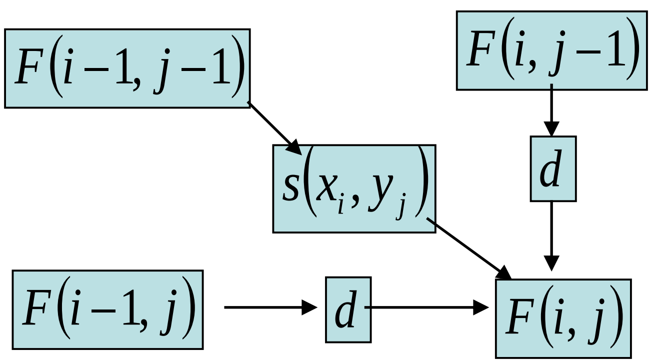


A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal alignment of AAG and A
Use a gap penalty of d=-5.

		A	A	G
	0			
A				
G				
C				

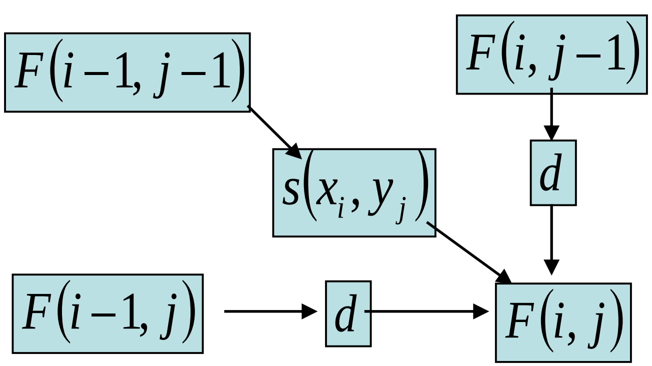


A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal alignment of AAG and AC
Use a gap penalty of d=-5.

		A	A	G
	0 →	-5 →	-10 →	-15
A	↓ -5			
G	↓ -10			
C	↓ -15			

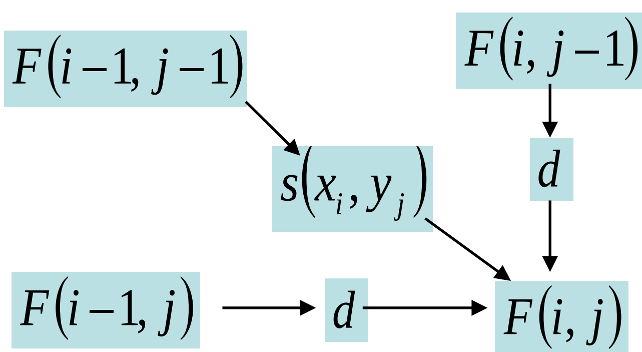


A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal alignment of AAG and A
Use a gap penalty of d=-5.

		A	A	G
	0	-5	-10	-15
A	-5	2	-3	-8
G	-10	-3	-3	-1
C	-15	-8	-8	-6



Traceback

- **Start from the lower right corner and trace back to the upper left.**
- **Each arrow introduces one character at the end of each aligned sequence.**
- **A horizontal move puts a gap in the left sequence.**
- **A vertical move puts a gap in the top sequence.**
- **A diagonal move uses one character from each sequence.**

A simple example

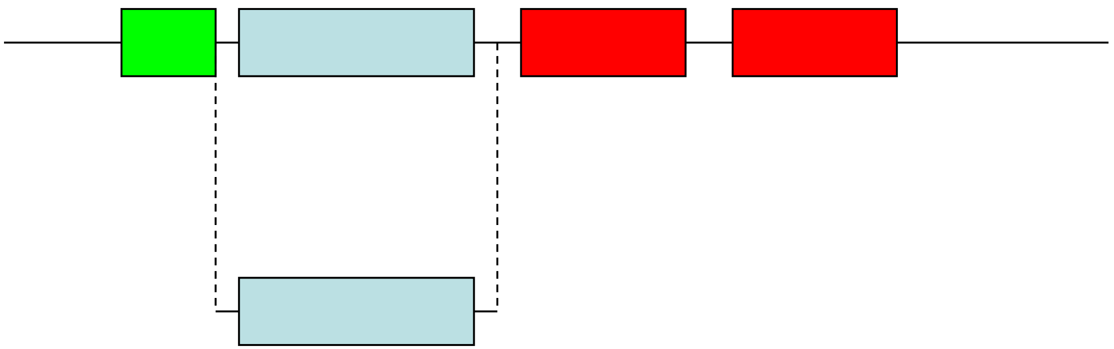
- Start from the lower right corner and trace back to the upper left.
- Each arrow introduces one character at the end of each aligned sequence.
- A horizontal move puts a gap in the left sequence.
- A vertical move puts a gap in the top sequence.
- A diagonal move uses one character from each sequence.

Find the optimal alignment of AAG and AC
Use a gap penalty of d=-5.

		A	A	G
	0	-5		
A		2	-3	
G				-1
C				**-6**

```
AAG-      AAG-
-AGC      A-GC
```

Local alignment



- A single-domain protein may be homologous to a region within a multi-domain protein.
- Usually, an alignment that spans the complete length of both sequences is not required.

Smith/Waterman local alignment (1981)

- Two sequences $X = x_1 \ldots x_n$ and $Y = y_1 \ldots y_m$
- Let $F(i, j)$ be the optimal alignment score *of* $X_{1\ldots i}$ of X up to x_i and $Y_{1\ldots j}$ of Y up to Y_j $(0 \le i \le n, 0 \le j \le m)$, *then we have*

$$F(0,0) = 0$$

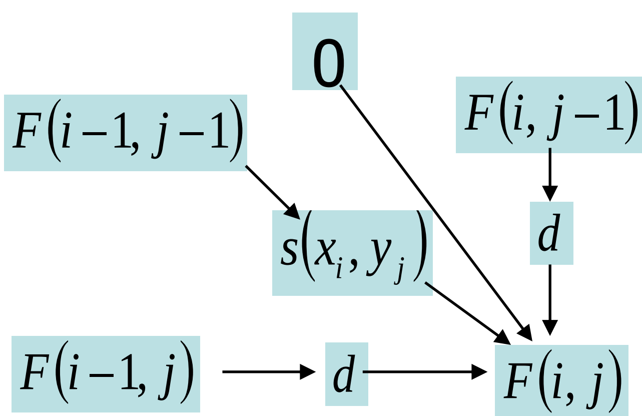
$$F(i,j) = \max \begin{cases} 0 \\ F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - d \\ F(i, j-1) - d \end{cases}$$

Local alignment

- Two differences with respect to global alignment:
  - No score is negative.
  - Traceback begins at the highest score in the matrix and continues until you reach 0.
- Global alignment algorithm: *Needleman-Wunsch*.
- Local alignment algorithm: *Smith-Waterman*.

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2



Find the optimal local alignment of AAG and AGC.
Use a gap penalty of d=-5.

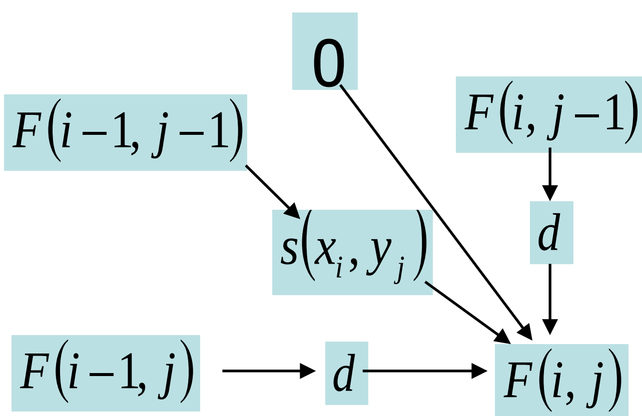
		A	A	G
A				
G				
C				

A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal local alignment of AAG and AGC.
Use a gap penalty of d=-5.

		A	A	G
	0	0	0	0
A	0			
G	0			
C	0			

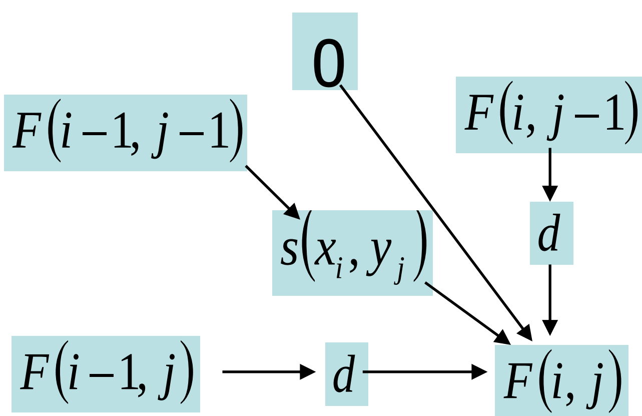


A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal local alignment of AAG and AGC.
Use a gap penalty of d=-5.

		A	A	G
	0	0	0	0
A	0	2	2	0
G	0	0	0	4
C	0	0	0	0

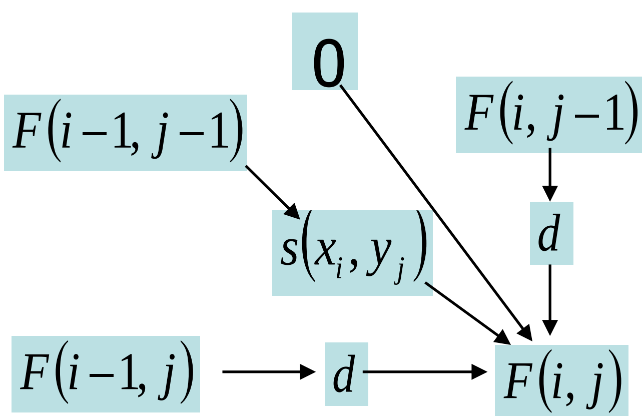


A simple example

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal local alignment of AAG and AGC.
Use a gap penalty of d=-5.

		A	A	G
	0	0	0	0
A	0	2	2	0
G	0	0	0	4
C	0	0	0	0



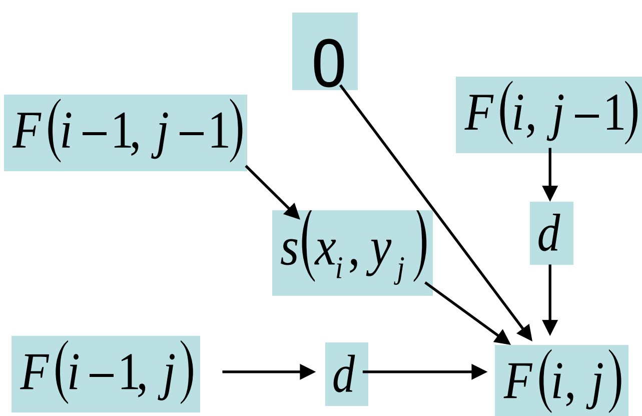
AG
AG

Local alignment

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal local alignment of AAG and GAAGGC.
Use a gap penalty of d=-5.

		A	A	G
	0	0	0	0
G	0			
A	0			
A	0			
G	0			
G	0			
C	0			

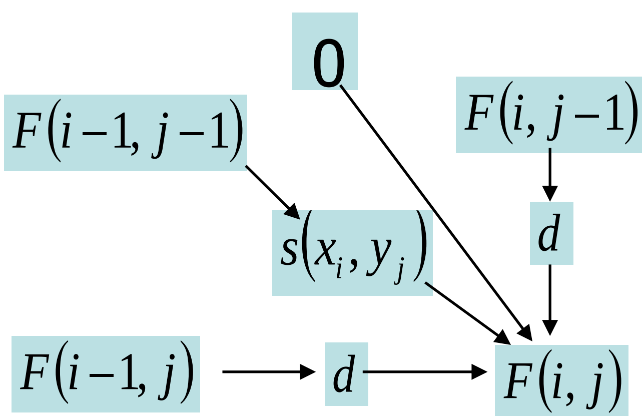


Local alignment

	A	C	G	T
A	2	-7	-5	-7
C	-7	2	-7	-5
G	-5	-7	2	-7
T	-7	-5	-7	2

Find the optimal local alignment of AAG and GAAGGC.
Use a gap penalty of d=-5.

		A	A	G
	0	0	0	0
G	0	0	0	2
A	0	2	2	0
A	0	2	4	0
G	0	0	0	6
G	0	0	0	2
C	0	0	0	0



Greedy algorithm:
Choose the best at the moment

- **Not always produce the optimal result**
- **Two elements are required to find an optimal solution by greedy algorithm**

1. **Greedy-choice property**
   - Global optimal can be reached by local optimal (greedy)
2. **Optimal substructure**
   - An optimal solution contains within it optimal solutions to the subproblems

Acknowledgement

PPTs for examples in dynamic programming are kindly provided by Dr. Qi Liu.