Lecture 10
Dynamic Programming

View in slide-show mode
Introduction

• An algorithm design paradigm like divide-and-conquer
• “Programming”: A tabular method (not writing computer code)
  Older sense of planning or scheduling, typically by filling in a table
• Divide-and-Conquer (DAC): subproblems are independent
• Dynamic Programming (DP): subproblems are not independent
• Overlapping subproblems: subproblems share sub-subproblems
  – In solving problems with overlapping subproblems
    • A DAC algorithm does redundant work
      – Repeatedly solves common subproblems
    • A DP algorithm solves each problem just once
      – Saves its result in a table
Problem 1:
Fibonacci Numbers
Example: Fibonacci Numbers (Recursive Solution)

Reminder:
F(0) = 0 and F(1) = 1
F(n) = F(n-1) + F(n-2)

REC-FIBO(n)
if n < 2
  return n
else
  return REC-FIBO(n-1) + REC-FIBO(n-2)

Overlapping subproblems in different recursive calls. Repeated work!
Example: Fibonacci Numbers (Recursive Solution)

Recurrence:

\[ T(n) = T(n-1) + T(n-2) + 1 \]

\( \Rightarrow \) exponential runtime

Recursive algorithm inefficient because it recomputes the same \( F(i) \) repeatedly in different branches of the recursion tree.
Example: Fibonacci Numbers (Bottom-up Computation)

Reminder:
F(0) = 0 and F(1) = 1
F(n) = F(n-1) + F(n-2)

ITER-FIBO(n)
F[0] = 0
F[1] = 1
for i = 2 to n do
    F[i] = F[i-1] + F[i-2]
return F[n]

Runtime: \( \Theta(n) \)
Optimization Problems

- **DP** typically applied to optimization problems
- In an optimization problem
  - There are many possible solutions (feasible solutions)
  - Each solution has a value
  - Want to find an optimal solution to the problem
    - A solution with the optimal value (min or max value)
  - Wrong to say “the” optimal solution to the problem
    - There may be several solutions with the same optimal value
Development of a DP Algorithm

1. Characterize the structure of an optimal solution
2. Recursively define the value of an optimal solution
3. Compute the value of an optimal solution in a bottom-up fashion
4. Construct an optimal solution from the information computed in Step 3
Problem 2: Matric Chain Multiplication
Example: Matrix-chain Multiplication

- **Input**: a sequence (chain) \( \langle A_1, A_2, \ldots, A_n \rangle \) of \( n \) matrices
- **Aim**: compute the product \( A_1 \cdot A_2 \cdot \ldots \cdot A_n \)
- A product of matrices is fully parenthesized if
  - It is either a single matrix
  - Or, the product of two fully parenthesized matrix products surrounded by a pair of parentheses.

\[
\left( A_i (A_{i+1} A_{i+2} \ldots A_j) \right)
\]

\[
\left( (A_i A_{i+1} A_{i+2} \ldots A_{j-1}) A_j \right)
\]

\[
\left( (A_i A_{i+1} A_{i+2} \ldots A_k) (A_{k+1} A_{k+2} \ldots A_j) \right) \quad \text{for } i \leq k < j
\]

- All parenthesizations yield the same product; matrix product is associative
Matrix-chain Multiplication: An Example Parenthesization

- **Input**: \( \langle A_1, A_2, A_3, A_4 \rangle \)
- 5 distinct ways of full parenthesization
  
  \[
  (A_1(A_2(A_3A_4))) \\
  (A_1((A_2A_3)A_4)) \\
  ((A_1A_2)(A_3A_4)) \\
  ((A_1(A_2A_3))A_4) \\
  (((A_1A_2)A_3)A_4)
  \]
- The way we parenthesize a chain of matrices can have a dramatic effect on the cost of computing the product
Reminder: Matrix Multiplication

\textbf{MATRIX-MULTIPLY}(A, B)

\begin{algorithm}
\textbf{if} \text{cols}[A] \neq \text{rows}[B] \textbf{then}
\textbf{error}(\text{"incompatible dimensions")}
\textbf{for} i \leftarrow 1 \textbf{to} \text{rows}[A] \textbf{do}
\textbf{for} j \leftarrow 1 \textbf{to} \text{cols}[B] \textbf{do}
\quad C[i,j] \leftarrow 0
\textbf{for} k \leftarrow 1 \textbf{to} \text{cols}[A] \textbf{do}
\quad C[i,j] \leftarrow C[i,j] + A[i,k] \cdot B[k,j]
\textbf{return} C
\end{algorithm}

\[ A \times B = C \]

\begin{align*}
\text{rows}(A) &= p & \text{rows}(B) &= q \\
\text{cols}(A) &= q & \text{cols}(B) &= r \\
\text{rows}(C) &= p & \text{cols}(C) &= r
\end{align*}
Reminder: Matrix Multiplication

**MATRIX-MULTIPLY**(A, B)

if cols[A] ≠ rows[B] then error ("incompatible dimensions")

for i ← 1 to rows[A] do
  for j ← 1 to cols[B] do
    C[i,j] ← 0
    for k ← 1 to cols[A] do
      C[i,j] ← C[i,j] + A[i,k] · B[k,j]
  return C

A: p x q  B: q x r  C: p x r

# of mult-add ops = rows[A] x cols[B] x cols[A]

# of mult-add ops = p x q x r
Matrix Chain Multiplication: Example

\( A_1: 10 \times 100 \quad A_2: 100 \times 5 \quad A_3: 5 \times 50 \)

Which paranthesesization is better? \((A_1A_2)A_3\) or \(A_1(A_2A_3)\)?

\[
\begin{array}{c}
A_1: \\
10 \begin{bmatrix} 100 \\ 10 \end{bmatrix} \\
\mathbb{A}_1
\end{array} \times 
\begin{array}{c}
A_2: \\
100 \begin{bmatrix} 5 \\ 100 \end{bmatrix} \\
\mathbb{A}_2
\end{array} = 
\begin{array}{c}
A_1A_2: \\
5 \begin{bmatrix} 5 \\ 5 \end{bmatrix} \\
\mathbb{A}_1A_2
\end{array} \\
\text{# of ops: } 10 \cdot 100 \cdot 5 = 5000
\]

\[
\begin{array}{c}
A_1A_2: \\
10 \begin{bmatrix} 5 \\ 50 \end{bmatrix} \\
\mathbb{A}_1A_2
\end{array} \times 
\begin{array}{c}
A_3: \\
5 \begin{bmatrix} 50 \\ 5 \end{bmatrix} \\
\mathbb{A}_3
\end{array} = 
\begin{array}{c}
A_1A_2A_3: \\
50 \begin{bmatrix} 50 \\ 10 \end{bmatrix} \\
\mathbb{A}_1A_2A_3
\end{array} \\
\text{# of ops: } 10 \cdot 5 \cdot 50 = 2500
\]

Total # of ops: 7500
Matrix Chain Multiplication: Example

\[ A_1: 10 \times 100 \quad A_2: 100 \times 5 \quad A_3: 5 \times 50 \]

Which paranthesization is better? \((A_1A_2)A_3\) or \(A_1(A_2A_3)\)?

\[
\begin{align*}
100 \begin{bmatrix} 5 \\ A_2 \end{bmatrix} & \times 5 \begin{bmatrix} 50 \\ A_3 \end{bmatrix} = 100 \begin{bmatrix} 50 \\ A_2A_3 \end{bmatrix} & \text{# of ops: } 100 \cdot 5 \cdot 50 = 25000 \\
10 \begin{bmatrix} A_1 \end{bmatrix} & \times 100 \begin{bmatrix} 50 \\ A_2A_3 \end{bmatrix} = 10 \begin{bmatrix} 50 \\ A_1A_2A_3 \end{bmatrix} & \text{# of ops: } 10 \cdot 100 \cdot 50 = 50000
\end{align*}
\]

Total # of ops: 75000
Matrix Chain Multiplication: Example

\[ A_1: 10 \times 100 \quad A_2: 100 \times 5 \quad A_3: 5 \times 50 \]

Which paranthesization is better? \((A_1A_2)A_3\) or \(A_1(A_2A_3)\)?

In summary: \((A_1A_2)A_3 \Rightarrow \# \text{ of multiply-add ops: 7500}\)
\[ A_1(A_2A_3) \Rightarrow \# \text{ of multiple-add ops: 75000} \]

\(\Rightarrow\) First parenthesization yields 10x faster computation
Matrix-chain Multiplication Problem

**Input:** A chain \(\langle A_1, A_2, \ldots, A_n \rangle\) of \(n\) matrices, where \(A_i\) is a \(p_{i-1} \times p_i\) matrix

**Objective:** Fully parenthesize the product

\[ A_1 \cdot A_2 \cdot \ldots \cdot A_n \]

such that the number of scalar mult-adds is minimized.
Counting the Number of Parenthesizations

- **Brute force approach**: exhaustively check all parenthesizations
- \( P(n) \): # of parenthesizations of a sequence of \( n \) matrices
- We can split sequence between \( k^{\text{th}} \) and \( (k+1)^{\text{st}} \) matrices for any \( k=1, 2, \ldots, n-1 \), then parenthesize the two resulting sequences independently, i.e.,
  \[
  (A_1A_2A_3 \ldots A_k)(A_{k+1}A_{k+2} \ldots A_n)
  \]
- We obtain the recurrence
  \[
P(1) = 1 \quad \text{and} \quad P(n) = \sum_{k=1}^{n-1} P(k)P(n-k)
  \]
Number of Parenthesizations: \( \sum_{k=1}^{n-1} P(k) P(n-k) \)

- The recurrence generates the sequence of Catalan Numbers
- Solution is \( P(n) = C(n-1) \) where
  \[
  C(n) = \frac{1}{n+1} \binom{2n}{n} = \Omega\left(\frac{4^n}{n^{3/2}}\right)
  \]
- The number of solutions is exponential in \( n \)
- Therefore, brute force approach is a poor strategy
The Structure of Optimal Parenthesization

**Notation**: $A_{i..j}$: The matrix that results from evaluation of the product: $A_i \ A_{i+1} \ A_{i+2} \ ... \ A_j$

**Observation**: Consider the last multiplication operation in any parenthesization: $(A_1 \ A_2 \ ... \ A_k) \cdot (A_{k+1} \ A_{k+2} \ ... \ A_n)$

There is a $k$ value ($1 \leq k < n$) such that:

1. First, the product $A_{1..k}$ is computed
2. Then, the product $A_{k+1..n}$ is computed
3. Finally, the matrices $A_{1..k}$ and $A_{k+1..n}$ are multiplied
Step 1: Characterize the structure of an optimal solution

- An optimal parenthesization of product $A_1A_2…A_n$ will be:  
  $$(A_1 A_2 \ldots A_k) . (A_{k+1} A_{k+2} \ldots A_n)$$  
  for some $k$ value

- The cost of this optimal parenthesization will be:  
  Cost of computing $A_{1..k}$  
  + Cost of computing $A_{k+1..n}$  
  + Cost of multiplying $A_{1..k} \cdot A_{k+1..n}$
Step 1: Characterize the Structure of an Optimal Solution

- **Key observation**: Given optimal parenthesization

  \[(A_1A_2A_3 \ldots A_k) \cdot (A_{k+1}A_{k+2} \ldots A_n)\]

  - Parenthesization of the subchain \(A_1A_2A_3 \ldots A_k\)
  - Parenthesization of the subchain \(A_{k+1}A_{k+2} \ldots A_n\)

  should both be optimal

Thus, optimal solution to an instance of the problem contains optimal solutions to subproblem instances

i.e., **optimal substructure** within an optimal solution exists.
Step 2: A Recursive Solution

**Step 2**: Define the value of an optimal solution recursively in terms of optimal solutions to the subproblems

Assume we are trying to determine the min cost of computing $A_{i..j}$

$m_{i,j}$: min # of scalar multiply-add opns needed to compute $A_{i..j}$

*Note: The optimal cost of the original problem: $m_{1,n}$*

How to compute $m_{i,j}$ recursively?
Step 2: A recursive Solution

Base case: \( m_{i,i} = 0 \) (single matrix, no multiplication)

Let the size of matrix \( A_i \) be \((p_{i-1} \times p_i)\)
Consider an optimal parenthesization of chain \( A_i \ldots A_j \):

\[
(A_i \ldots A_k) \cdot (A_{k+1} \ldots A_j)
\]

The optimal cost:

\[
m_{i,j} = m_{i,k} + m_{k+1,j} + p_{i-1} \times p_k \times p_j
\]

where:

\( m_{i,k} \): Optimal cost of computing \( A_{i..k} \)
\( m_{k+1,j} \): Optimal cost of computing \( A_{k+1..j} \)
\( p_{i-1} \times p_k \times p_j \): Cost of multiplying \( A_{i..k} \) and \( A_{k+1..j} \)
Step 2: A Recursive Solution

In an optimal parenthesization:

\( k \) must be chosen to minimize \( m_{ij} \)

The recursive formulation for \( m_{ij} \):

\[
m_{ij} = \begin{cases} 
0 & \text{if } i=j \\
\text{MIN} \{ m_{ik} + m_{k+1,j} + p_{i-1}p_kp_j \} & \text{if } i < j 
\end{cases}
\]
Step 2: A Recursive Solution

• The $m_{ij}$ values give the costs of optimal solutions to subproblems

• In order to keep track of how to construct an optimal solution
  – Define $s_{ij}$ to be the value of $k$ which yields the optimal split of the subchain $A_{i..j}$
  That is, $s_{ij} = k$ such that

\[ m_{ij} = m_{ik} + m_{k+1,j} + p_{i-1}p_k p_j \]
holds
Direct Recursion: Inefficient!

Recursive matrix-chain order

\[ \text{RMC}(p, i, j) \]

\[
\begin{align*}
\text{if } i &= j \text{ then} \\
& \quad \text{return 0} \\
\text{m}[i, j] &\leftarrow \infty \\
\text{for } k &\leftarrow i \text{ to } j - 1 \text{ do} \\
q &\leftarrow \text{RMC}(p, i, k) + \text{RMC}(p, k+1, j) + p_{i-1} p_k p_j \\
&\quad \text{if } q < \text{m}[i, j] \text{ then} \\
\text{m}[i, j] &\leftarrow q \\
\text{return } m[i, j]
\end{align*}
\]
Direct Recursion: Inefficient!

Recursion tree for $\text{RMC}(p,1,4)$

Nodes are labeled with $i$ and $j$ values

Redundant calls are filled
Computing the Optimal Cost (Matrix-Chain Multiplication)

An important observation:

• We have relatively few subproblems
  – one problem for each choice of $i$ and $j$ satisfying $1 \leq i \leq j \leq n$
  – total $n + (n-1) + \ldots + 2 + 1 = \frac{1}{2} n(n+1) = \Theta(n^2)$ subproblems
• We can write a recursive algorithm based on recurrence.
• However, a recursive algorithm may encounter each subproblem many times in different branches of the recursion tree
• This property, overlapping subproblems, is the second important feature for applicability of dynamic programming
Computing the Optimal Cost (Matrix-Chain Multiplication)

Compute the value of an optimal solution in a bottom-up fashion
- matrix $A_i$ has dimensions $p_{i-1} \times p_i$ for $i = 1, 2, \ldots, n$
- the input is a sequence $\langle p_0, p_1, \ldots, p_n \rangle$ where $\text{length}[p] = n + 1$

Procedure uses the following auxiliary tables:
- $m[1\ldots n, 1\ldots n]$: for storing the $m[i, j]$ costs
- $s[1\ldots n, 1\ldots n]$: records which index of $k$ achieved the optimal cost in computing $m[i, j]$
Bottom-up computation

$$m_{ij} = \min_{i \leq k < j} \{ m_{ik} + m_{k+1,j} + p_i 1p_k p_j \}$$

How to choose the order in which we process $m_{ij}$ values?

Before computing $m_{ij}$, we have to make sure that the values for $m_{ik}$ and $m_{k+1,j}$ have been computed for all $k$. 
$m_{ij} = \min_{i < k < j} \left\{ m_{ik} + m_{k+1,j} + p_i \cdot 1_{p_k p_j} \right\}$

**Reminder:** $m_{ij}$ computed only for $j > i$

$m_{ij}$ must be processed after $m_{ik}$ and $m_{j,k+1}$
\[ m_{ij} = \min_{i \leq k < j} \{ m_{ik} + m_{k+1,j} + p_i \cdot 1_p k p_j \} \]

How to set up the iterations over \( i \) and \( j \) to compute \( m_{ij} \)?

\( m_{ij} \) must be processed after \( m_{ik} \) and \( m_{j,k+1} \)
\[ m_{ij} = \min_{i \leq k < j} \{ m_{ik} + m_{k+1,j} + p_{i-1}p_{k}p_{j} \} \]

If the entries \( m_{ij} \) are computed in the shown order, then \( m_{ik} \) and \( m_{k+1,j} \) values are guaranteed to be computed before \( m_{ij} \).
\[ m_{ij} = \min_{i \leq k < j} \{ m_{ik} + m_{k+1,j} + p_i - p_{k+1} \} \]
\[ m_{ij} = \min_{i \neq k < j} \{ m_{ik} + m_{k+1,j} + p_i \cdot p_k \cdot p_j \} \]

for \( \ell = 2 \) to \( n \)

for \( i = 1 \) to \( n-\ell+1 \)

\[ j = i + \ell - 1 \]

\[ m_{ij} = \ldots \]

\[ \ldots \]
Algorithm for Computing the Optimal Costs

MATRIX-CHAIN-ORDER($p$)

$n \leftarrow \text{length}[p] - 1$
for $i \leftarrow 1$ to $n$ do
    $m[i, i] \leftarrow 0$
for $\ell \leftarrow 2$ to $n$ do
    for $i \leftarrow 1$ to $n - \ell + 1$ do
        $j \leftarrow i + \ell - 1$
        $m[i, j] \leftarrow \infty$
        for $k \leftarrow i$ to $j - 1$ do
            $q \leftarrow m[i, k] + m[k+1, j] + p_{i-1}p_kp_j$
            if $q < m[i, j]$ then
                $m[i, j] \leftarrow q$
                $s[i, j] \leftarrow k$
return $m$ and $s$
Algorithm for Computing the Optimal Costs

- The algorithm first computes
  \[ m[i, i] \leftarrow 0 \text{ for } i = 1, 2, \ldots, n \]
  min costs for all chains of length 1

- Then, for \( \ell = 2, 3, \ldots, n \) computes
  \[ m[i, i+\ell-1] \text{ for } i = 1, \ldots, n-\ell+1 \]
  min costs for all chains of length \( \ell \)

- For each value of \( \ell = 2, 3, \ldots, n \),
  \[ m[i, i+\ell-1] \]
  depends only on table entries \( m[i, k] \) & \( m[k+1, i+\ell-1] \)
  for \( i \leq k < i+\ell-1 \), which are already computed
Algorithm for Computing the Optimal Costs

\( \ell = 2 \)
for \( i = 1 \) to \( n - 1 \)
\( m[i, i+1] = \infty \)
for \( k = i \) to \( i \) do
\[
. 
\]
\[
. 
\]
\( \ell = 3 \)
for \( i = 1 \) to \( n - 2 \)
\( m[i, i+2] = \infty \)
for \( k = i \) to \( i+1 \) do
\[
. 
\]
\[
. 
\]
\( \ell = 4 \)
for \( i = 1 \) to \( n - 3 \)
\( m[i, i+3] = \infty \)
for \( k = i \) to \( i+2 \) do
\[
. 
\]
\[
. 
\]
Table access pattern in computing $m[i, j]$s for $\ell = j - i + 1$

for $k \leftarrow i$ to $j - 1$ do

$q \leftarrow m[i, k] + m[k+1, j] + p_{i-1}p_kp_j$

- Table entries currently computed
- Table entries already computed
- Table entries referenced
Table access pattern in computing $m[i, j]$s for $\ell = j-i+1$

$$((A_i) (A_{i+1}A_{i+2} \ldots A_j))$$

for $k \leftarrow i$ to $j-1$ do

$$q \leftarrow m[i, k] + m[k+1, j] + p_{i-1}p_kp_j$$

- Table entries currently computed
- Table entries already computed
- Table entries referenced
Table access pattern in computing $m[i, j]$s for $\ell = j - i + 1$

$$((A_i A_{i+1}) (A_{i+2} \ldots A_j))$$

for $k \leftarrow i$ to $j-1$ do

$q \leftarrow m[i, k] + m[k+1, j] + p_{i-1} p_k p_j$

- Table entries **currently** computed
- Table entries **already** computed
- Table entries **referenced**
Table access pattern in computing $m[i, j]$s for $\ell = j - i + 1$

for $k \leftarrow i$ to $j - 1$ do
  \[ q \leftarrow m[i, k] + m[k+1, j] + p_{i-1}p_kp_j \]

- Table entries currently computed
- Table entries already computed
- Table entries referenced
Table access pattern in computing \( m[i, j]s \) for \( \ell = j - i + 1 \)

\[
\begin{align*}
((A_i A_{i+1} \ldots A_{j-1}) (A_j))
\end{align*}
\]

for \( k \leftarrow i \) to \( j-1 \) do

\[
q \leftarrow m[i, k] + m[k+1, j] + p_{i-1}p_kp_j
\]

- Table entries currently computed
- Table entries already computed
- Table entries referenced
Example

\[ m_{ij} = \min_{i \leq k < j} \{ m_{ik} + m_{k+1,j} + p_i \cdot p_k \cdot p_j \} \]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁: (30x35)</td>
<td>0</td>
<td>15750</td>
<td>7875</td>
<td>9375</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A₂: (35x15)</td>
<td>0</td>
<td>2625</td>
<td>4375</td>
<td>??</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>A₃: (15x5)</td>
<td>0</td>
<td>750</td>
<td>2500</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>A₄: (5x10)</td>
<td>0</td>
<td>1000</td>
<td>3500</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>A₅: (10x20)</td>
<td>0</td>
<td>5000</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>A₆: (20x25)</td>
<td>0</td>
<td>6</td>
<td></td>
<td></td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Compute \( m_{25} \)

\( k = 2 \)

\( (A₂) \cdot (A₃A₄A₅) \)

\[
\text{cost} = m_{22} + m_{35} + p_1p_2p_5
\]

\[
= 0 + 2500 + 35 \times 15 \times 20
\]

\[
= 13000
\]

Choose the \( k \) value that leads to min cost
### Example

The formula for computing the minimum cost is:

\[ m_{ij} = \min_{i \leq k < j} \{ m_{ik} + m_{k+1,j} + p_i \cdot p_k \cdot p_j \} \]

### Computation

#### A_1: (30x35)

\[ 0 \hspace{1cm} 15750 \hspace{1cm} 7875 \hspace{1cm} 9375 \hspace{1cm} \]  

#### A_2: (35x15)

\[ 0 \hspace{1cm} 2625 \hspace{1cm} 4375 \hspace{1cm} ??? \hspace{1cm} \]

#### A_3: (15x5)

\[ 0 \hspace{1cm} 750 \hspace{1cm} 2500 \hspace{1cm} \]

#### A_4: (5x10)

\[ 0 \hspace{1cm} 1000 \hspace{1cm} 3500 \hspace{1cm} \]

#### A_5: (10x20)

\[ 0 \hspace{1cm} 5000 \hspace{1cm} 6 \]

#### A_6: (20x25)

\[ \downarrow \]

*(A_2A_3) (A_4A_5)*

To compute \( m_{25} \) with \( k = 3 \):

\[ m_{25} = 2625 + 1000 + 35 \times 5 \times 20 \]

\[ = 2625 + 1000 + 35 \times 100 \]

\[ = 2625 + 1000 + 3500 \]

\[ = 7125 \]

Choose the \( k \) value that leads to min cost.
Example

$$m_{ij} = \min_{i \ k < j} \left\{ m_{ik} + m_{k+1,j} + p_i \cdot p_k \cdot p_j \right\}$$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁: (30x35)</td>
<td>0</td>
<td>15750</td>
<td>7875</td>
<td>9375</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>A₂: (35x15)</td>
<td>0</td>
<td>2625</td>
<td>4375</td>
<td>???</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>A₃: (15x5)</td>
<td>0</td>
<td>750</td>
<td>2500</td>
<td>1000</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>A₄: (5x10)</td>
<td>0</td>
<td>1000</td>
<td>3500</td>
<td>5000</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>A₅: (10x20)</td>
<td>0</td>
<td>5000</td>
<td>6</td>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>A₆: (20x25)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Compute $m_{25}$

Compute $m_{25}$

$$k = 4$$

$$\text{(A}_2\text{A}_3\text{A}_4) \ (\text{A}_5)$$

$$\text{cost} = m_{24} + m_{55} + p_1 p_4 p_5$$

$$= 4375 + 0 + 35 \times 10 \times 20$$

$$= 11375$$

Choose the $k$ value that leads to min cost
Example

\[ m_{ij} = \min_{i, k < j} \{ m_{ik} + m_{k+1,j} + p_i p_{k+1} p_j \} \]

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A_1: (30x35)</td>
<td>0</td>
<td>15750</td>
<td>7875</td>
<td>9375</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>A_2: (35x15)</td>
<td></td>
<td>0</td>
<td>2625</td>
<td>4375</td>
<td>7125</td>
<td>2</td>
</tr>
<tr>
<td>A_3: (15x5)</td>
<td></td>
<td></td>
<td>0</td>
<td>750</td>
<td>2500</td>
<td>3</td>
</tr>
<tr>
<td>A_4: (5x10)</td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>1000</td>
<td>3500</td>
</tr>
<tr>
<td>A_5: (10x20)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>5000</td>
</tr>
<tr>
<td>A_6: (20x25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Compute \( m_{25} \)

Choose \( k = 3 \)

\( m_{25} = 7125 \)

\( s_{25} = 3 \)
Constructing an Optimal Solution

- **MATRIX-CHAIN-ORDER** determines the optimal # of scalar mults/adds
  - needed to compute a matrix-chain product
  - it does not directly show how to multiply the matrices

- That is,
  - it determines the cost of the optimal solution(s)
  - it does not show how to obtain an optimal solution

- Each entry $s[i,j]$ records the value of $k$ such that
  optimal parenthesization of $A_i \ldots A_j$ splits the product between $A_k \& A_{k+1}$

- We know that the final matrix multiplication in computing $A_{1\ldots n}$ optimally
  is $A_{1\ldots s[1,n]} \times A_{s[1,n]+1\ldots n}$
Example: Constructing an Optimal Solution

Reminder: $s_{ij}$ is the optimal top-level split of $A_i \ldots A_j$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>1</th>
<th>3</th>
<th>3</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

What is the optimal top-level split for:

$A_1 A_2 A_3 A_4 A_5 A_6$

$s_{16} = 3$
Example: Constructing an Optimal Solution

**Reminder:** $s_{ij}$ is the optimal top-level split of $A_i \ldots A_j$

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

$k = 3$

$(A_1A_2A_3) (A_4A_5A_6)$

What is the optimal split for $A_1 \ldots A_3$? $s_{13} = 1$

What is the optimal split for $A_4 \ldots A_6$? $s_{46} = 5$
Example: Constructing an Optimal Solution

**Reminder:** $s_{ij}$ is the optimal top-level split of $A_i \ldots A_j$

$$
\begin{array}{cccccc}
1 & 1 & 3 & 3 & 3 & 1 \\
2 & 3 & 3 & 3 & & 2 \\
3 & 3 & 3 & & & 3 \\
4 & & & & 5 & 5 \\
\end{array}
$$

What is the optimal split for $A_1 \ldots A_3$?
$s_{13} = 1$

What is the optimal split for $A_4 \ldots A_6$?
$s_{46} = 5$
Example: Constructing an Optimal Solution

**Reminder:** $s_{ij}$ is the optimal top-level split of $A_i \ldots A_j$

$$(((A_1) (A_2A_3)) \,( (A_4A_5) \,(A_6))$$

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

What is the optimal split for $A_2A_3$? $s_{23} = 2$

What is the optimal split for $A_4A_5$? $s_{45} = 4$
Example: Constructing an Optimal Solution

**Reminder:** \( s_{ij} \) is the optimal top-level split of \( A_i \ldots A_j \)

\[
((A_1) ((A_2) (A_3))) (((A_4) (A_5)) (A_6))
\]

What is the optimal split for \( A_2A_3 \)? \( s_{23} = 2 \)

What is the optimal split for \( A_4A_5 \)? \( s_{45} = 4 \)
Constructing an Optimal Solution

Earlier optimal matrix multiplications can be computed recursively.

Given:
- the chain of matrices \( A = \langle A_1, A_2, \ldots A_n \rangle \)
- the \( s \) table computed by `MATRIX-CHAIN-ORDER`

The following recursive procedure computes the matrix-chain product \( A_{i...j} \)

\[
\text{MATRICE-CHAIN-MULTIPLY}(A, s, i, j) \\
\text{if } j > i \text{ then} \\
\quad X \leftarrow \text{MATRICE-CHAIN-MULTIPLY}(A, s, i, s[i, j]) \\
\quad Y \leftarrow \text{MATRICE-CHAIN-MULTIPLY}(A, s, s[i, j]+1, j) \\
\quad \text{return } \text{MATRICE-MULTIPLY}(X, Y) \\
\text{else} \\
\quad \text{return } A_i
\]

Invocation: `MATRICE-CHAIN-MULTIPLY(A, s, 1, n)`
Example: Recursive Construction of an Optimal Solution

\[ s[1...6, 1...6] \]

\[
\begin{array}{cccccc}
1 & 1 & 1 & 3 & 3 & 3 \\
2 & 2 & 3 & 4 & 3 & 3 \\
3 & 3 & 3 & 3 & 3 & 3 \\
4 & 4 & 5 & & & \\
5 & 5 & & & & \\
\end{array}
\]

\[
MCM(1,6) \quad X \leftarrow MCM(1,3) = (A_1A_2A_3) \quad \quad \text{return } A_1
\]

\[
Y \leftarrow MCM(4,6) = (A_4A_5A_6) \quad X \leftarrow MCM(1,1) = A_1 \\
\text{return (??)} \quad Y \leftarrow MCM(2,3) = (A_2A_3) \quad \text{return (??)}
\]
Example: Recursive Construction of an Optimal Solution

\[
\text{MCM}(1,6) \quad X \leftarrow \text{MCM}(1,3) = (A_1(A_2A_3)) \quad \text{MCM}(1,3) \quad \text{return } A_1
\]

\[
Y \leftarrow \text{MCM}(4,6) = (A_4A_5A_6) \quad \text{return } (A_1(A_2A_3))
\]

\[
\text{MCM}(2,3) \quad X \leftarrow \text{MCM}(2,2) = A_2 \quad \text{return } A_2
\]

\[
Y \leftarrow \text{MCM}(3,3) = A_3 \quad \text{return } (A_2A_3)
\]
Example: Recursive Construction of an Optimal Solution

\[
\begin{align*}
\text{MCM}(1,6) & \quad X \leftarrow \text{MCM}(1,3) = (A_1(A_2A_3)) \\
& \quad Y \leftarrow \text{MCM}(4,6) = ((A_4A_5)A_6) \\
& \quad \text{return } (A_1(A_2A_3))((A_4A_5)A_6)
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(1,3) & \quad X \leftarrow \text{MCM}(1,1) = A_1 \\
& \quad Y \leftarrow \text{MCM}(2,3) = (A_2A_3) \\
& \quad \text{return } (A_1(A_2A_3)) \\
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(2,3) & \quad X \leftarrow \text{MCM}(2,2) = A_2 \\
& \quad Y \leftarrow \text{MCM}(3,3) = A_3 \\
& \quad \text{return } (A_2A_3) \\
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(4,6) & \quad X \leftarrow \text{MCM}(4,5) = (A_4A_5) \\
& \quad Y \leftarrow \text{MCM}(6,6) = A_6 \\
& \quad \text{return } ((A_4A_5)A_6)
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(4,5) & \quad X \leftarrow \text{MCM}(4,4) = A_4 \\
& \quad Y \leftarrow \text{MCM}(5,5) = A_5 \\
& \quad \text{return } (A_4A_5)
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(5,5) & \quad \text{return } A_5
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(4,4) & \quad \text{return } A_4
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(6,6) & \quad \text{return } A_6
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(3,3) & \quad \text{return } A_3
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(2,2) & \quad \text{return } A_2
\end{align*}
\]

\[
\begin{align*}
\text{MCM}(1,1) & \quad \text{return } A_1
\end{align*}
\]
Table reference pattern for $m[i, j]$ ($1 \leq i \leq j \leq n$)

$m[i, j]$ is referenced for the computation of

- $m[i, r]$ for $j < r \leq n$  \((n - j)\) times
- $m[r, j]$ for $1 \leq r < i$  \((i - 1)\) times
Table reference pattern for $m[i, j]$ ($1 \leq i \leq j \leq n$)

$R(i, j) = \#$ of times that $m[i, j]$ is referenced in computing other entries

$R(i, j) = (n-j) + (i-1)$

$= (n-1) - (j-i)$

The total # of references for the entire table is

$$R(i, j) = \sum_{i=1}^{n} \sum_{j=i}^{n} \frac{n^3}{3}$$
Summary

1. Identification of the optimal substructure property

2. Recursive formulation to compute the cost of the optimal solution

3. Bottom-up computation of the table entries

4. Constructing the optimal solution by backtracing the table entries
Elements of Dynamic Programming

• When should we look for a DP solution to an optimization problem?
• Two key ingredients for the problem
  – Optimal substructure
  – Overlapping subproblems
DP Hallmark #1

Optimal Substructure

• A problem exhibits optimal substructure
  – if an optimal solution to a problem contains within it optimal solutions to subproblems

• Example: matrix-chain-multiplication

  Optimal parenthesization of $A_1 A_2 \ldots A_n$ that splits the product between $A_k$ and $A_{k+1}$, contains within it optimal soln’s to the problems of parenthesizing $A_1 A_2 \ldots A_k$ and $A_{k+1} A_{k+2} \ldots A_n$
Optimal Substructure

Finding a suitable space of subproblems

• Iterate on subproblem instances

• Example: matrix-chain-multiplication
  – Iterate and look at the structure of optimal soln’s to subproblems, sub-subproblems, and so forth
  – Discover that all subproblems consists of subchains of \( \langle A_1, A_2, \ldots, A_n \rangle \)
  – Thus, the set of chains of the form \( \langle A_i, A_{i+1}, \ldots, A_j \rangle \) for \( 1 \leq i \leq j \leq n \)
  – Makes a natural and reasonable space of subproblems
DP Hallmark #2

Overlapping Subproblems

• Total number of distinct subproblems should be polynomial in the input size
• When a recursive algorithm revisits the same problem over and over again we say that the optimization problem has overlapping subproblems
Overlapping Subproblems

• DP algorithms typically take advantage of overlapping subproblems
  – by solving each problem once
  – then storing the solutions in a table
    where it can be looked up when needed
  – using constant time per lookup
Overlapping Subproblems

Recursive matrix-chain order

\textbf{RMC}(p, i, j)

\begin{itemize}
  \item \textbf{if} \ i = j \ \textbf{then}
  \item \quad return 0
  \item \quad \textbf{m}[i, j] \leftarrow \infty
  \item \textbf{for} \ k \leftarrow i \textbf{ to } j - 1 \ \textbf{do}
  \item \quad q \leftarrow \text{RMC}(p, i, k) + \text{RMC}(p, k+1, j) + p_{i-1}p_kp_j
  \item \quad \textbf{if} \ q < \text{m}[i, j] \ \textbf{then}
  \item \quad \quad \text{m}[i, j] \leftarrow q
  \item \quad \quad return \text{m}[i, j]
\end{itemize}
Recursive Matrix-chain Order

Recursion tree for $\text{RMC}(p,1,4)$

Nodes are labeled with $i$ and $j$ values

Redundant calls are filled
Running Time of RMC

\[ T(1) \geq 1 \]

\[ T(n) \geq 1 + \sum_{k=1}^{n-1} (T(k) + T(n-k) + 1) \text{ for } n > 1 \]

- For \( i = 1, 2, \ldots, n \) each term \( T(i) \) appears twice
  - Once as \( T(k) \), and once as \( T(n-k) \)
- Collect \( n-1 \) 1’s in the summation together with the front 1

\[ T(n) \geq 2 \sum_{i=1}^{n-1} T(i) + n \]

- Prove that \( T(n) = \Omega(2^n) \) using the substitution method
Running Time of RMC: Prove that $T(n) = \Omega(2^n)$

- Try to show that $T(n) \geq 2^{n-1}$ (by substitution)

Base case: $T(1) \geq 1 = 2^0 = 2^{1-1}$ for $n = 1$

IH: $T(i) \geq 2^{i-1}$ for all $i = 1, 2, \ldots, n-1$ and $n \geq 2$

$$T(n) \geq 2 \sum_{i=1}^{n-1} 2^{i-1} + n$$

$$= 2 \sum_{i=0}^{n-2} 2^i + n = 2(2^{n-1} - 1) + n$$

$$= 2^{n-1} + (2^{n-1} - 2 + n)$$

$\Rightarrow T(n) \geq 2^{n-1}$

Q.E.D.
Running Time of RMC: $T(n) \geq 2^{n-1}$

Whenever

- a recursion tree for the natural recursive solution to a problem contains the same subproblem repeatedly
- the total number of different subproblems is small

it is a good idea to see if DP can be applied
Memoization

• Offers the efficiency of the usual DP approach while maintaining top-down strategy
• Idea is to memoize the natural, but inefficient, recursive algorithm
Memoized Recursive Algorithm

- Maintains an entry in a table for the solution to each subproblem
- Each table entry contains a special value to indicate that the entry has yet to be filled in
- When the subproblem is first encountered its solution is computed and then stored in the table
- Each subsequent time that the subproblem encountered the value stored in the table is simply looked up and returned
Memoized Recursive Matrix-chain Order

**LookupC**\((p, i, j)\)

\[
\text{if } m[i, j] = \infty \text{ then }
\]

\[
\text{if } i = j \text{ then }
        m[i, j] \leftarrow 0
\]

\[
\text{else }
\]

\[
\text{for } k \leftarrow i \text{ to } j - 1 \text{ do }
\]

\[
q \leftarrow \text{LookupC}(p, i, k) + \text{LookupC}(p, k+1, j) + p_{i-1} p_k p_j
\]

\[
\text{if } q < m[i, j] \text{ then }
        m[i, j] \leftarrow q
\]

\[
\text{return } m[i, j]
\]

**MemoizedMatrixChain**\((p)\)

\[
n \leftarrow \text{length}[p] - 1
\]

\[
\text{for } i \leftarrow 1 \text{ to } n \text{ do }
\]

\[
\quad \text{for } j \leftarrow 1 \text{ to } n \text{ do }
\]

\[
\quad m[i, j] \leftarrow \infty
\]

\[
\text{return } \text{LookupC}(p, 1, n)
\]

\(\triangleright\) Shaded subtrees are looked-up rather than recomputing
Memoized Recursive Algorithm

• The approach assumes that
  – The set of all possible subproblem parameters are known
  – The relation between the table positions and subproblems is established

• Another approach is to memoize
  – by using hashing with subproblem parameters as key

Memoization-based solutions will NOT BE ACCEPTED in the exams!
Dynamic Programming vs Memoization

Summary

• Matrix-chain multiplication can be solved in $O(n^3)$ time
  – by either a top-down memoized recursive algorithm
  – or a bottom-up dynamic programming algorithm

• Both methods exploit the overlapping subproblems property
  – There are only $\Theta(n^2)$ different subproblems in total
  – Both methods compute the soln to each problem once

• Without memoization the natural recursive algorithm runs in exponential time since subproblems are solved repeatedly
Dynamic Programming vs Memoization Summary

In general practice

• If all subproblems must be solved at once
  – a bottom-up DP algorithm always outperforms a top-down memoized algorithm by a constant factor

because, bottom-up DP algorithm
  • Has no overhead for recursion
  • Less overhead for maintaining the table

• DP: Regular pattern of table accesses can be exploited to reduce the time and/or space requirements even further

• Memoized: If some problems need not be solved at all, it has the advantage of avoiding solutions to those subproblems
Problem 3:
Longest Common Subsequence
Definitions

- A subsequence of a given sequence is just the given sequence with some elements (possibly none) left out.

Example:

\[ X = < A, B, C, B, D, A, B > \]
\[ Z = < B, C, D, B > \]

\( \Rightarrow \) Z is a subsequence of X.
Definitions

**Formal definition:** Given a sequence \( X = \langle x_1, x_2, \ldots, x_m \rangle \),

sequence \( Z = \langle z_1, z_2, \ldots, z_k \rangle \) is a subsequence of \( X \)

if \( \exists \) a strictly increasing sequence \( \langle i_1, i_2, \ldots, i_k \rangle \) of indices of \( X \)

such that \( x_i = z_j \) for all \( j = 1, 2, \ldots, k \), where \( 1 \leq k \leq m \)

**Example:** \( Z = \langle B, C, D, B \rangle \) is a subsequence of \( X = \langle A, B, C, B, D, A, B \rangle \)

with the index sequence \( \langle i_1, i_2, i_3, i_4 \rangle = \langle 2, 3, 5, 7 \rangle \)
Definitions

If $Z$ is a subsequence of both $X$ and $Y$, we denote $Z$ as a *common subsequence* of $X$ and $Y$.

*Example*: $X = \langle A, B, C, B, D, A, B \rangle$ and $Y = \langle B, D, C, A, B, A \rangle$

$Z_1 = \langle B, C, A \rangle$ is a common subsequence (of length 3) of $X$ and $Y$.

Two longest common subsequence (LCSs) of $X$ and $Y$?

$Z_2 = \langle B, C, B, A \rangle$ of length 4

$Z_3 = \langle B, D, A, B \rangle$ of length 4

The optimal solution value = 4
Longest Common Subsequence (LCS) Problem

- **LCS problem**: Given two sequences \( X = \langle x_1, x_2, \ldots, x_m \rangle \) and \( Y = \langle y_1, y_2, \ldots, y_n \rangle \), find the LCS of \( X \) & \( Y \)

- **Brute force approach**:
  - Enumerate all subsequences of \( X \)
  - Check if each subsequence is also a subsequence of \( Y \)
  - Keep track of the LCS
  - What is the complexity?
    - There are \( 2^m \) subsequences of \( X \)
    - Exponential runtime
Notation

**Notation**: Let $X_i$ denote the $i^{th}$ prefix of $X$

i.e. $X_i = \langle x_1, x_2, \ldots, x_i \rangle$

**Example**: $X = \langle A, B, C, B, D, A, B \rangle$

$X_4 = \langle A, B, C, B \rangle$, $X_0 = \langle \rangle$
Optimal Substructure of an LCS

Let $X = \langle x_1, x_2, \ldots, x_m \rangle$ and $Y = \langle y_1, y_2, \ldots, y_n \rangle$ are given

Let $Z = \langle z_1, z_2, \ldots, z_k \rangle$ be an LCS of $X$ and $Y$

**Question 1**: If $x_m = y_n$, how to define the optimal substructure?

We must have $z_k = x_m = y_n$ and $Z_{k-1} = \text{LCS}(X_{m-1}, Y_{n-1})$
Optimal Substructure of an LCS

Let $X = \langle x_1, x_2, \ldots, x_m \rangle$ and $Y = \langle y_1, y_2, \ldots, y_n \rangle$ are given.

Let $Z = \langle z_1, z_2, \ldots, z_k \rangle$ be an LCS of $X$ and $Y$.

**Question 2**: If $x_m \neq y_n$ and $z_k \neq x_m$, how to define the optimal substructure?

We must have $Z = \text{LCS}(X_{m-1}, Y)$.
Optimal Substructure of an LCS

Let \( X = <x_1, x_2, \ldots, x_m> \) and \( Y = <y_1, y_2, \ldots, y_n> \) are given

Let \( Z = <z_1, z_2, \ldots, z_k> \) be an LCS of \( X \) and \( Y \)

**Question 3**: If \( x_m \neq y_n \) and \( z_k \neq y_n \), how to define the optimal substructure?

We must have \( Z = \text{LCS}(X, Y_{n-1}) \)
Theorem: Optimal Substructure of an LCS

Let \( X = \langle x_1, x_2, \ldots, x_m \rangle \) and \( Y = \langle y_1, y_2, \ldots, y_n \rangle \) are given.

Let \( Z = \langle z_1, z_2, \ldots, z_k \rangle \) be an LCS of \( X \) and \( Y \).

**Theorem:** Optimal substructure of an LCS:

1. **If** \( x_m = y_n \)
   
   then \( z_k = x_m = y_n \) and \( Z_{k-1} \) is an LCS of \( X_{m-1} \) and \( Y_{n-1} \).

2. **If** \( x_m \neq y_n \) and \( z_k \neq x_m \)
   
   then \( Z \) is an LCS of \( X_{m-1} \) and \( Y \).

3. **If** \( x_m \neq y_n \) and \( z_k \neq y_n \)
   
   then \( Z \) is an LCS of \( X \) and \( Y_{n-1} \).
Optimal Substructure Theorem (case 1)

If $x_m = y_n$ then $z_k = x_m = y_n$ and $Z_{k-1}$ is an LCS of $X_{m-1}$ and $Y_{n-1}$
Optimal Substructure Theorem (case 2)

If $x_m \neq y_n$ and $z_k \neq x_m$ then $Z$ is an LCS of $X_{m-1}$ and $Y$
Optimal Substructure Theorem (case 3)

If \( x_m \neq y_n \) and \( z_k \neq y_n \) then \( Z \) is an LCS of \( X \) and \( Y_{n-1} \)

\[
X = \begin{array}{c|c|c}
1 & 2 & m \\
\end{array} \\
Y = \begin{array}{c|c|c|c}
1 & 2 & n \\
D \\
\end{array} \\
Z = \begin{array}{c|c|c|c}
1 & 2 & k \\
\end{array}
\]
Proof of Optimal Substructure Theorem (case 1)

If \( x_m = y_n \) then \( z_k = x_m = y_n \) and \( Z_{k-1} \) is an LCS of \( X_{m-1} \) and \( Y_{n-1} \)

Proof: If \( z_k \neq x_m = y_n \) then

we can append \( x_m = y_n \) to \( Z \) to obtain a common subsequence of length \( k+1 \) \( \Rightarrow \) contradiction

Thus, we must have \( z_k = x_m = y_n \)

Hence, the prefix \( Z_{k-1} \) is a length-(\( k-1 \)) CS of \( X_{m-1} \) and \( Y_{n-1} \)

We have to show that \( Z_{k-1} \) is in fact an LCS of \( X_{m-1} \) and \( Y_{n-1} \)

Proof by contradiction:

Assume that \( \exists \) a CS \( W \) of \( X_{m-1} \) and \( Y_{n-1} \) with \( |W| = k \)

Then appending \( x_m = y_n \) to \( W \) produces a CS of length \( k+1 \)
Proof of Optimal Substructure Theorem (case 2)

If \( x_m \neq y_n \) and \( z_k \neq x_m \) then \( Z \) is an LCS of \( X_{m-1} \) and \( Y \)

Proof: If \( z_k \neq x_m \) then \( Z \) is a CS of \( X_{m-1} \) and \( Y_n \)

We have to show that \( Z \) is in fact an LCS of \( X_{m-1} \) and \( Y_n \) (Proof by contradiction)

Assume that \( \exists \) a CS \( W \) of \( X_{m-1} \) and \( Y_n \) with \( |W| > k \)

Then \( W \) would also be a CS of \( X \) and \( Y \)

Contradiction to the assumption that

\( Z \) is an LCS of \( X \) and \( Y \) with \( |Z| = k \)

Case 3: Dual of the proof for (case 2)
A Recursive Solution to Subproblems

Theorem implies that there are one or two subproblems to examine
if $x_m = y_n$ then
  we must solve the subproblem of finding an LCS of $X_{m-1} \& Y_{n-1}$
  appending $x_m = y_n$ to this LCS yields an LCS of $X \& Y$
else
  we must solve two subproblems
  – finding an LCS of $X_{m-1} \& Y$
  – finding an LCS of $X \& Y_{n-1}$
  longer of these two LCSs is an LCS of $X \& Y$
endif
Recursive Algorithm (Inefficient!!!)

\[
\text{LCS}(X, Y) \\
\begin{align*}
  m &\leftarrow \text{length}[X] \\
  n &\leftarrow \text{length}[Y] \\
  \text{if } x_m = y_n \text{ then} & \quad \text{solve one subproblem} \\
  & \quad Z \leftarrow \text{LCS}(X_{m-1}, Y_{n-1}) \\
  & \quad \text{return } \langle Z, x_m = y_n \rangle \\
  \text{else} & \quad \text{solve two subproblems} \\
  & \quad Z' \leftarrow \text{LCS}(X_{m-1}, Y) \\
  & \quad Z'' \leftarrow \text{LCS}(X, Y_{n-1}) \\
  & \quad \text{return longer of } Z' \text{ and } Z''
\end{align*}
\]
A Recursive Solution

c[i, j]: length of an LCS of $X_i$ and $Y_j$

\[
c[i, j] = \begin{cases} 
0 & \text{if } i = 0 \text{ or } j = 0 \\
 c[i-1, j-1] + 1 & \text{if } i, j > 0 \text{ and } x_i = y_j \\
 \max\{c[i, j-1], c[i-1, j]\} & \text{if } i, j > 0 \text{ and } x_i \neq y_j 
\end{cases}
\]
Computing the Length of an LCS

- We can easily write an exponential-time recursive algorithm based on the given recurrence. ➔ Inefficient!

- How many distinct subproblems to solve?
  \[ \Theta(mn) \]

- Overlapping subproblems property: Many subproblems share the same sub-subproblems.
  e.g. Finding an LCS to \( X_{m-1} \) & \( Y \) and an LCS to \( X \) & \( Y_{n-1} \) has the sub-subproblem of finding an LCS to \( X_{m-1} \) & \( Y_{n-1} \)

- Therefore, we can use dynamic programming.
Data Structures

Let:

\[ c[i, j] \]: length of an LCS of \( X_i \) and \( Y_j \)

\[ b[i, j] \]: direction towards the table entry corresponding to the optimal subproblem solution chosen when computing \( c[i, j] \). Used to simplify the construction of an optimal solution at the end.

Maintain the following tables:

\[ c[0\ldots m, 0\ldots n] \]

\[ b[1\ldots m, 1\ldots n] \]
Bottom-up Computation

**Reminder:**

\[
c[i, j] = \begin{cases} 
0 & \text{if } i = 0 \text{ or } j = 0 \\
c[i-1, j-1] + 1 & \text{if } i, j > 0 \text{ and } x_i = y_j \\
\max\{ c[i, j-1], c[i-1, j] \} & \text{if } i, j > 0 \text{ and } x_i \neq y_j 
\end{cases}
\]

How to choose the order in which we process \( c[i, j] \) values?

The values for \( c[i-1, j-1], c[i, j-1], \) and \( c[i-1,j] \) must be computed before computing \( c[i, j] \).
The equation for the cost matrix \( c[i, j] \) is given by:

\[
c[i, j] = \begin{cases} 
0 & \text{if } i = 0 \text{ or } j = 0 \\
\max\{ c[i-1, j], c[i, j-1] \} & \text{if } i, j > 0 \text{ and } x_i = y_j \\
\max\{ c[i-1, j-1], c[i-1, j], c[i, j-1] \} & \text{otherwise}
\end{cases}
\]

Need to process:
- \( c[i, j] \) after computing:
  - \( c[i-1, j-1] \),
  - \( c[i, j-1] \),
  - \( c[i-1, j] \)
\[ c[i, j] = \begin{cases} 
0 & \text{if } i = 0 \text{ or } j = 0 \\
\ c[i-1, j-1] + 1 & \text{if } i, j > 0 \text{ and } x_i = y_j \\
\max\{c[i, j-1], c[i-1, j]\} & \text{if } i, j > 0 \text{ and } x_i \neq y_j 
\end{cases} \]

for \( i \leftarrow 1 \) to \( m \)

for \( j \leftarrow 1 \) to \( n \)

....

....

\[ c[i, j] = \]
Computing the Length of an LCS

LCS-LENGTH(X,Y)

\[ m \leftarrow \text{length}[X]; \quad n \leftarrow \text{length}[Y] \]

for \( i \leftarrow 0 \) to \( m \) do \( c[i, 0] \leftarrow 0 \)

for \( j \leftarrow 0 \) to \( n \) do \( c[0, j] \leftarrow 0 \)

for \( i \leftarrow 1 \) to \( m \) do

\[ \text{for } j \leftarrow 1 \text{ to } n \text{ do} \]

if \( x_i = y_j \) then

\[ c[i, j] \leftarrow c[i-1, j-1] + 1 \]

\[ b[i, j] \leftarrow "\leftarrow" \]

else if \( c[i - 1, j] \geq c[i, j-1] \)

\[ c[i, j] \leftarrow c[i-1, j] \]

\[ b[i, j] \leftarrow "\uparrow" \]

else

\[ c[i, j] \leftarrow c[i, j-1] \]

\[ b[i, j] \leftarrow "\leftarrow" \]

\[ \text{Total runtime} = \Theta(mn) \]

\[ \text{Total space} = \Theta(mn) \]
Computing the Length of an LCS

Operation of **LCS-LENGTH** on the sequences

\[ X = \langle A, B, C, B, D, A, B \rangle \]

\[ Y = \langle B, D, C, A, B, A \rangle \]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Computing the Length of an LCS

Operation of **LCS-LENGTH** on the sequences

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>B</td>
<td>D</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Y</td>
<td>B</td>
<td>D</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td></td>
</tr>
</tbody>
</table>

### Table:

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>↑</td>
<td>↑</td>
<td>↑</td>
<td>←</td>
<td>←</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Computing the Length of an LCS

**Operation of LCS-LENGTH**

on the sequences

$X = \langle A, B, C, B, D, A, B \rangle$

$Y = \langle B, D, C, A, B, A \rangle$

<table>
<thead>
<tr>
<th>$i$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$j$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1 A</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>2 B</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3 C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4 B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5 D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6 A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7 B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Computing the Length of an LCS

Operation of **LCS-LENGTH** on the sequences

\[ X = \langle A, B, C, B, D, A, B \rangle \]

\[ Y = \langle B, D, C, A, B, A \rangle \]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>x_i</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Computing the Length of an LCS

Operation of **LCS-LENGTH**
on the sequences

\[
\begin{align*}
X &= \langle A, B, C, B, D, A, B \rangle \\
Y &= \langle B, D, C, A, B, A \rangle
\end{align*}
\]
Computing the Length of an LCS

Operation of **LCS-LENGTH** on the sequences

\[ X = \langle A, B, C, B, D, A, B \rangle \]

\[ Y = \langle B, D, C, A, B, A \rangle \]
Computing the Length of an LCS

Operation of **LCS-LENGTH** on the sequences

\[ X = \langle A, B, C, B, D, A, B \rangle \]

\[ Y = \langle B, D, C, A, B, A \rangle \]

<table>
<thead>
<tr>
<th>( j )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i )</td>
<td>0</td>
<td>A</td>
<td>B</td>
<td>D</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
### Computing the Length of an LCS

**Operation of LCS-LENGTH** on the sequences

- **X = \langle A, B, C, B, D, A, B\rangle**
- **Y = \langle B, D, C, A, B, A\rangle**

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>←1</td>
<td>←1</td>
</tr>
<tr>
<td>2</td>
<td>←1</td>
<td>←1</td>
<td>←1</td>
<td>←1</td>
<td>←2</td>
<td>←2</td>
<td>←2</td>
</tr>
<tr>
<td>3</td>
<td>←2</td>
<td>←2</td>
<td>←2</td>
<td>←2</td>
<td>←2</td>
<td>←2</td>
<td>←2</td>
</tr>
<tr>
<td>4</td>
<td>←3</td>
<td>←3</td>
<td>←3</td>
<td>←3</td>
<td>←3</td>
<td>←3</td>
<td>←3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

X: A, B, C, B, D, A, B

Y: B, D, C, A, B, A
Computing the Length of an LCS

Operation of **LCS-LENGTH** on the sequences

\[ X = \langle A, B, C, B, D, A, B \rangle \]

\[ Y = \langle B, D, C, A, B, A \rangle \]
Computing the Length of an LCS

Operation of **LCS-LENGTH**
on the sequences

\[ X = \langle A, B, C, B, D, A, B \rangle \]
\[ Y = \langle B, D, C, A, B, A \rangle \]

\[ 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \]

<table>
<thead>
<tr>
<th>i</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Computing the Length of an LCS

Operation of **LCS-LENGTH** on the sequences

\[
X = \langle A, B, C, B, D, A, B \rangle
\]
\[
Y = \langle B, D, C, A, B, A \rangle
\]
Computing the Length of an LCS

Operation of **LCS-LENGTH** on the sequences

\[
\begin{array}{ccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 \\
\end{array}
\]

\[
\begin{array}{ccccccc}
X = <A, B, C, B, D, A, B> \\
\end{array}
\]

\[
\begin{array}{ccccccc}
Y = <B, D, C, A, B, A> \\
\end{array}
\]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>y_j</strong></td>
<td>B</td>
<td>D</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td><strong>x_i</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{array}{ccccccc}
\begin{array}{cccccccc}
0 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\
\end{array}
\end{array}
\]

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>y_j</strong></td>
<td>B</td>
<td>D</td>
<td>C</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td></td>
</tr>
<tr>
<td><strong>x_i</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{array}{ccccccc}
\begin{array}{ccccccc}
0 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \\
\end{array}
\end{array}
\]
Computing the Length of an LCS

Operation of \textbf{LCS-LENGTH} on the sequences

\begin{align*}
X &= \langle A, B, C, B, D, A, B \rangle \\
Y &= \langle B, D, C, A, B, A \rangle
\end{align*}

Running-time = $O(mn)$ since each table entry takes $O(1)$ time to compute
Computing the Length of an LCS

Operation of \textbf{LCS-LENGTH} on the sequences

\begin{align*}
X &= \langle A, B, C, B, D, A, B \rangle \\
Y &= \langle B, D, C, A, B, A \rangle
\end{align*}

Running-time = $O(mn)$ since each table entry takes $O(1)$ time to compute

LCS of $X$ & $Y$ = $\langle B, C, B, A \rangle$
Constructing an LCS

The $b$ table returned by \texttt{LCS-LENGTH} can be used to quickly construct an LCS of $X$ & $Y$

Begin at $b[m, n]$ and trace through the table following arrows

Whenever you encounter a “\\(\wedge\)” in entry $b[i, j]$

it implies that $x_i = y_j$ is an element of LCS

The elements of LCS are encountered in reverse order
Constructing an LCS

PRINT-LCS\((b, X, i, j)\)

if \(i = 0\) or \(j = 0\) then
    return

if \(b[i, j]\) = “↖” then
    PRINT-LCS\((b, X, i-1, j-1)\)
    print \(x_i\)
else if \(b[i, j]\) = “↑” then
    PRINT-LCS\((b, X, i-1, j)\)
else
    PRINT-LCS\((b, X, i, j-1)\)

The initial invocation: PRINT-LCS\((b, X, \text{length}[X], \text{length}[Y])\)

The recursive procedure PRINT-LCS prints out LCS in proper order

This procedure takes \(O(m+n)\) time

since at least one of \(i\) and \(j\) is decremented in each stage of the recursion
Do we really need the b table (back-pointers)?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ø</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

**Question:** From which neighbor did we expand to the highlighted cell?

**Answer:** Upper-left neighbor, because $X[i] = Y[j]$. 
Do we really need the b table (back-pointers)?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>A</th>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

**Question:** From which neighbor did we expand to the highlighted cell?

**Answer:** Left neighbor, because \( X[i] \neq Y[j] \) and \( LCS[i, j-1] > LCS[i-1, j] \).
Do we really need the b table (back-pointers)?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>φ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

**Question**: From which neighbor did we expand to the highlighted cell?

**Answer**: Upper neighbor, because $X[i] \neq Y[j]$ and $LCS[i, j-1] = LCS[i-1, j]$.

*(See pseudo-code to see how ties are handled.)*
Improving the Space Requirements

We can eliminate the $b$ table altogether

- each $c[i, j]$ entry depends only on 3 other $c$ table entries: $c[i-1, j-1]$, $c[i-1, j]$ and $c[i, j-1]$

Given the value of $c[i, j]$:

- We can determine in $O(1)$ time which of these 3 values was used to compute $c[i, j]$ without inspecting table $b$
- We save $\Theta(mn)$ space by this method
- However, space requirement is still $\Theta(mn)$ since we need $\Theta(mn)$ space for the $c$ table anyway
What if we store the last 2 rows only?

To compute $c[i, j]$, we only need $c[i-1, j-1], c[i-1, j], \text{and } c[i-1, j-1]$

So, we can store only the last two rows.
What if we store the last 2 rows only?

To compute $c[i, j]$, we only need $c[i-1, j-1]$, $c[i-1, j]$, and $c[i-1, j-1]$. So, we can store only the last two rows.
What if we store the last 2 rows only?

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>φ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>φ</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To compute $c[i, j]$, we only need $c[i-1, j-1]$, $c[i-1, j]$, and $c[i-1, j-1]$. So, we can store only the last two rows.

This reduces space complexity from $\Theta(mn)$ to $\Theta(n)$.

Is there a problem with this approach?
What if we store the last 2 rows only?

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>D</th>
<th>C</th>
<th>A</th>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>∅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>∅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Is there a problem with this approach?

We cannot construct the optimal solution because we cannot backtrace anymore.

This approach works if we only need the length of an LCS, not the actual LCS.
Problem 4

Optimal Binary Search Tree
Reminder: Binary Search Tree (BST)

All keys in the left subtree less than 8

All keys in the right subtree greater than 8

This property holds for all nodes.
Binary Search Tree Example

*Example*: English-to-French translation

Organize (English, French) word pairs in a BST

- **Keyword**: English word
- **Satellite data**: French word

We can search for an English word (node key) efficiently, and return the corresponding French word (satellite data).
Binary Search Tree Example

Suppose we know the frequency of each keyword in texts:

\[
\begin{align*}
\text{begin} & \quad \text{do} & \quad \text{else} & \quad \text{end} & \quad \text{if} & \quad \text{then} & \quad \text{while} \\
5\% & \quad 40\% & \quad 8\% & \quad 4\% & \quad 10\% & \quad 10\% & \quad 23\%
\end{align*}
\]
Cost of a Binary Search Tree

Example: If we search for keyword “while”, we need to access 3 nodes. So, 23% of the queries will have cost of 3.

Total cost = \( (\text{depth}(i)+1) \times \text{freq}(i) \)

\[
= 1 \times 0.04 + 2 \times 0.4 + 2 \times 0.1 + 3 \times 0.05 + 3 \times 0.08 + 3 \times 0.1 + 3 \times 0.23 \\
= 2.42
\]
A different binary search tree (BST) leads to a different total cost:

\[
\text{Total cost} = 1 \times 0.4 + 2 \times 0.05 + 2 \times 0.23 + 3 \times 0.1 + 4 \times 0.08 + 4 \times 0.1 + 5 \times 0.04
\]

\[
= 2.18
\]

This is in fact an optimal BST.
Optimal Binary Search Tree Problem

Given:

A collection of $n$ keys $K_1 < K_2 < \ldots K_n$ to be stored in a BST.
The corresponding $p_i$ values for $1 \leq i \leq n$

$p_i$: probability of searching for key $K_i$

Find:

An optimal BST with minimum total cost:

$$\text{Total cost} = \sum_{i=1}^{n} (\text{depth}(i)+1) \times \text{freq}(i)$$

**Note:** The BST will be static. Only search operations will be performed. No insert, no delete, etc.
Cost of a Binary Search Tree

**Lemma 1**: Let $T_{ij}$ be a BST containing keys $K_i < K_{i+1} < \ldots < K_j$. Let $T_L$ and $T_R$ be the left and right subtrees of $T$. Then we have:

$$\text{cost}(T_{ij}) = \text{cost}(T_L) + \text{cost}(T_R) + \sum_{h=i}^{j} p_h$$

**Intuition**: When we add the root node, the depth of each node in $T_L$ and $T_R$ increases by 1. So, the cost of node $h$ increases by $p_h$. In addition, the cost of root node $r$ is $p_r$. That’s why, we have the last term at the end of the formula above.
Lemma 2: Optimal substructure property

Consider an optimal BST $T_{ij}$ for keys $K_i < K_{i+1} < \ldots < K_j$

Let $K_m$ be the key at the root of $T_{ij}$

Then:

$T_{i,m-1}$ is an optimal BST for subproblem containing keys: $K_i < \ldots < K_{m-1}$

$T_{m+1,j}$ is an optimal BST for subproblem containing keys: $K_{m+1} < \ldots < K_j$

$$\text{cost}(T_{ij}) = \text{cost}(T_{i,m-1}) + \text{cost}(T_{m+1,j}) + \sum_{h=i}^{j} p_h$$
Recursive Formulation

Note: We don’t know which root vertex leads to the minimum total cost. So, we need to try each vertex \( m \), and choose the one with minimum total cost.

\( c[i, j] \): cost of an optimal BST \( T_{ij} \) for the subproblem \( K_i < \ldots < K_j \)

\[
c[i, j] = \begin{cases} 
0 & \text{if } i > j \\
\min_{i \leq r \leq j} \{ c[i, r-1] + c[r+1, j] + P_{ij} \} & \text{otherwise}
\end{cases}
\]

where \( P_{ij} = \begin{cases} 
p_h & \text{if } h = i 
\end{cases} \)
Bottom-up computation

\[ c[i, j] = \begin{cases} 0 & \text{if } i > j \\ \min_{i < r < j} \{c[i, r - 1] + c[r + 1, j] + P_{ij}\} & \text{otherwise} \end{cases} \]

How to choose the order in which we process \( c[i, j] \) values?

Before computing \( c[i, j] \), we have to make sure that the values for \( c[i, r-1] \) and \( c[r+1, j] \) have been computed for all \( r \).
\[ c[i, j] = \begin{cases} 0 & \text{if } i > j \\ \min_{i \leq r \leq j} \{ c[i, r-1] + c[r+1, j] + P_{ij} \} & \text{otherwise} \end{cases} \]
If the entries $c[i,j]$ are computed in the shown order, then $c[i,r-1]$ and $c[r+1,j]$ values are guaranteed to be computed before $c[i,j]$. 

$$c[i,j] = \begin{cases} 0 & \text{if } i > j \\ \min_{i \leq r \leq j} \{c[i,r-1] + c[r+1,j] + P_{ij}\} & \text{otherwise} \end{cases}$$
Computing the Optimal BST Cost

OPTIMAL-BST-COST (p, n)

for i ← 1 to n do
    c[i, i-1] ← 0
    c[i, i] ← p[i]
    R[i, j] ← i


for i ← 2 to n do  // compute the prefix sum
    PS[i] ← p[i] + PS[i-1]

for d ← 1 to n-1 do  // BSTs with d+1 consecutive keys
    for i ← 1 to n - d do
        j ← i + d
        c[i, j] ← ∞

        for r ← i to j do
            q ← min{c[i, r-1] + c[r+1, j]} + PS[j] - PS[i-1]

            if q < c[i, j] then
                c[i, j] ← q
                R[i, j] ← r

return c[1, n], R
Note on Prefix Sum

- We need $P_{ij}$ values for each $i, j$ ($1 \leq i \leq n$ and $1 \leq j \leq n$),

\[
\sum_{h=i}^{j} P_{h} = P_{ij}
\]

where:

- If we compute the summation directly for every $(i, j)$ pair, the runtime would be $\Theta(n^3)$.

- Instead, we spend $O(n)$ time in preprocessing to compute the prefix sum array $PS$. Then we can compute each $P_{ij}$ in $O(1)$ time using $PS$. 
Note on Prefix Sum

In preprocessing, compute for each \( i \):

\[
PS[i]: \text{the sum of } p[j] \text{ values for } 1 \leq j \leq i
\]

Then, we can compute \( P_{ij} \) in \( O(1) \) time as follows:

\[
P_{ij} = PS[i] - PS[j-1]
\]

**Example:**

\[
\begin{array}{cccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
\text{p:} & 0.05 & 0.02 & 0.06 & 0.07 & 0.20 & 0.05 & 0.08 & 0.02 \\
\text{PS:} & 0.05 & 0.07 & 0.13 & 0.20 & 0.40 & 0.45 & 0.53 & 0.55 \\
\end{array}
\]

\[
P_{27} = PS[7] - PS[1] = 0.53 - 0.05 = 0.48
\]

\[
P_{36} = PS[6] - PS[2] = 0.45 - 0.07 = 0.38
\]