Algorithm Design Techniques

Greedy Algorithms
Dynamic Programming

Longest Common Subsequence

- The Longest Common Subsequence (LCS) is as follows…

- We are given two strings, S of length n and T of length m
- Produce their longest common subsequence
- The longest subsequence of characters that appear left-to-right, but not necessarily in a contiguous block, in both strings

Longest Common Subsequence – Example

- S = ABAZDC
- T = BACBAD

- What is the longest common subsequence?

Longest Common Subsequence – Example

- S = **A**BAZDC
- T = B**A**CBAD

- What is the longest common subsequence?
  - A

Longest Common Subsequence – Example

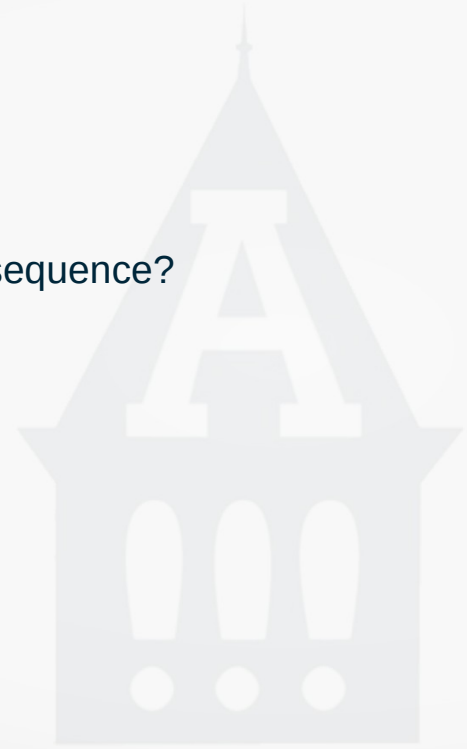
- S = **AB**AZDC
- T = B**A**C**B**AD

- What is the longest common subsequence?
  - AB

Longest Common Subsequence – Example

- S = **ABA**ZDC
- T = B**A**C**BA**D

- What is the longest common subsequence?
  - ABA



Longest Common Subsequence – Example

- S = **ABA**Z**D**C
- T = B**A**C**BAD**

- What is the longest common subsequence?
  - ABAD

Longest Common Subsequence – Example

- S = **ABA**Z**D**C
- T = B**A**C**BAD**

- How can we know we have the LCS?

Longest Common Subsequence – Example

- S = **ABA**Z**D**C
- T = B**A**C**BAD**

- How can we know we have the LCS?
  - Exhaustive search (you know I won't like that, it isn't efficient)

Longest Common Subsequence – Example

- S = **ABA**Z**D**C
- T = B**A**C**BAD**

- How can we know we have the LCS?
  - Exhaustive search (you know I won't like that, it isn't efficient)
  - Dynamic Programming!

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left

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

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [G, A]
  - no match
  - biggest: 0

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

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [G, G]
  - match: diagonal + 1 = 1

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1				
A	0						
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [G, C]
  - no match
  - biggest: 1
  -

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1			
A	0						
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [G, A]
  - no match
  - biggest: 1
  -

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1		
A	0						
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [G, T]
  - no match
  - biggest: 1
  -

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	
A	0						
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [G, G]
  - match: diagonal + 1 = 1
  - Note: comparing G with AGCATG

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0						
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [A, A]
  - match: diagonal + 1 = 1

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1					
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [A, G]
  - no match
  - biggest: 1

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1				
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [A, C]
  - no match
  - biggest: 1

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1			
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [A, A]
  - match: diagonal + 1 = 2

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2		
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [A, T]2
  - no match
  - biggest: 2

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left
- Cell [A, T]2
  - no match
  - biggest: 2

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0						
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0						
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0	1	2	2	2	2	3
T	0						
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0	1	2	2	2	2	3
T	0	1	2	2	2	3	3
A	0						

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
- An entry in the table is either
  - If the characters match, the diagonal element +1
  - The biggest of the element directly above or to the left

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0	1	2	2	2	2	3
T	0	1	2	2	2	3	3
A	0	1	2	2	3	3	3

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
  - 3
  - but which three?

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0	1	2	2	2	2	3
T	0	1	2	2	2	3	3
A	0	1	2	2	3	3	3

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
  - 3
  - but which three?
    - Either end with A, T, G

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0	1	2	2	2	2	3
T	0	1	2	2	2	3	3
A	0	1	2	2	3	3	3

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
  - 3
  - but which three?
    - Either end with A, T, G
    - If A, then next is G, C

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0	1	2	2	2	2	3
T	0	1	2	2	2	3	3
A	0	1	2	2	3	3	3

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
  - 3
  - but which three?
    - Either end with A, T, G
    - If A, then next is G, C
    - If G, then next is A

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0	1	2	2	2	2	3
T	0	1	2	2	2	3	3
A	0	1	2	2	3	3	3

Longest Common Subsequence

- Find LCS of GACGTA and AGCATG
  - 3
  - but which three?
    - Either end with A, T, G
    - If A, then next is G, C
    - If G, then next is A
    - AGA

		A	G	C	A	T	G
	0	0	0	0	0	0	0
G	0	0	1	1	1	1	1
A	0	1	1	1	2	2	2
C	0	1	1	2	2	2	2
G	0	1	2	2	2	2	3
T	0	1	2	2	2	3	3
A	0	1	2	2	3	3	3

Longest Common Subsequence

- Is this a hard problem?

Longest Common Subsequence

- Is this a hard problem?
- Can be done in n^2 time; just have to build the table