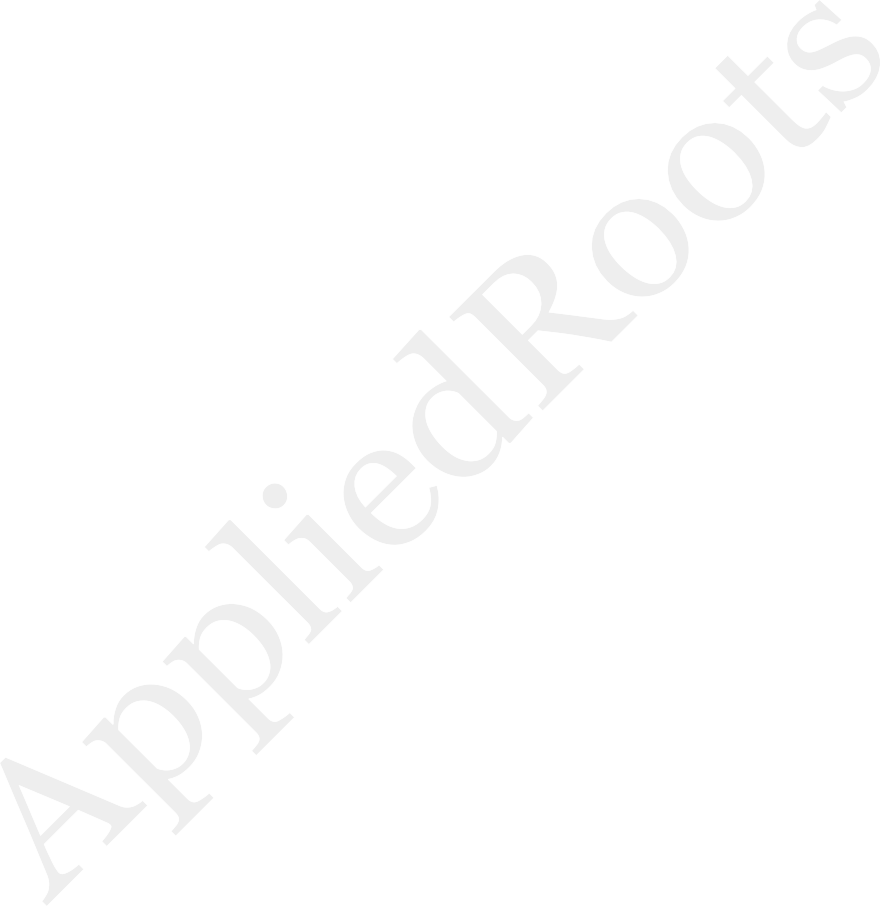
# Longest Common Subsequence

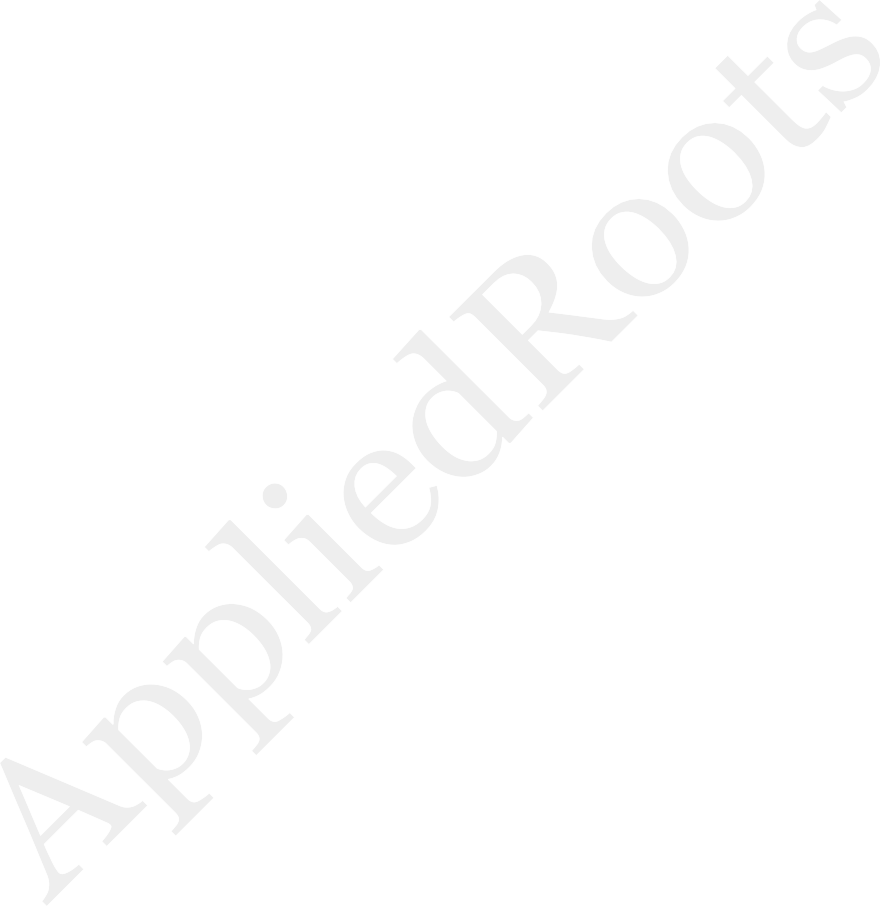
* given two sequences x[1..m] and y[1..n], find a longest subsequence LCS(x,y) common to both:

*x*: A B C B D A B

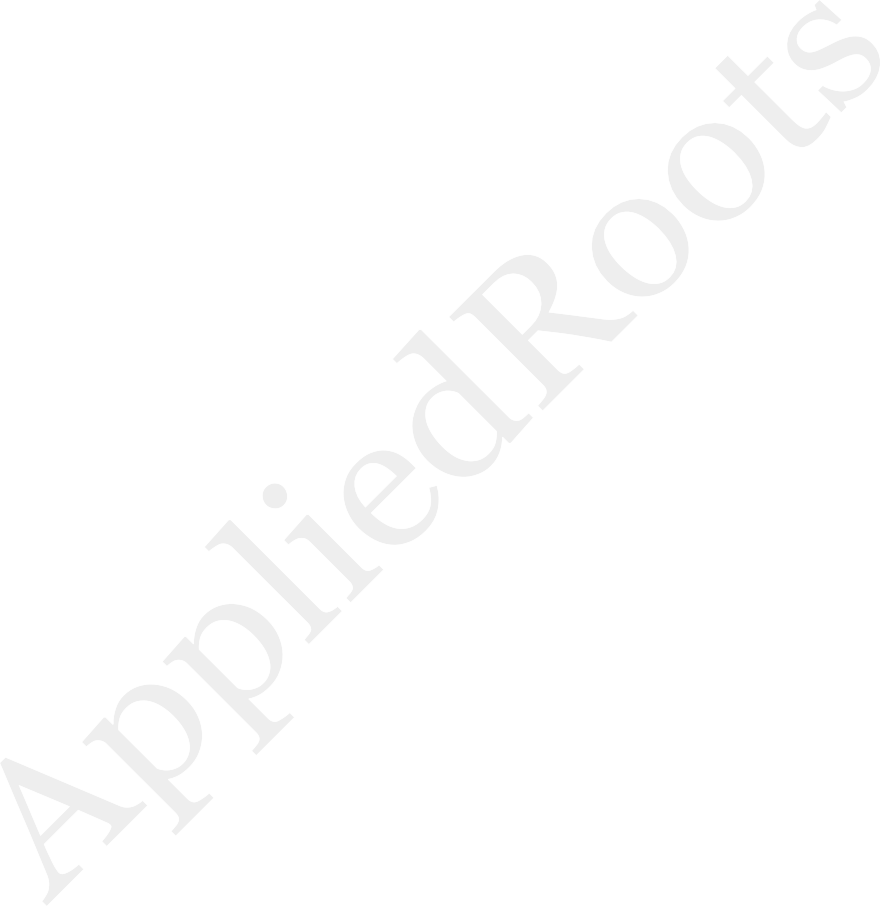
*y*: B D C A B A

* denote the length of a sequence *s* by **|***s***|**
* let us first try to get **|**LCS(x,y)**|**

# Brute force solution

* For every subsequence of x[1..m] , check if it is a subsequence of y[1..n]
* Analysis
  + *2m* subsequences of *x*
  + each check takes *(n)* time ...
  + worst case running time is *(n2m)*
* **Pretty bad (brute!)**

# Using prefixes

* consider prefixes of x and y
  + x[1..i] ith prefix of x[1..m]
  + y[1..j] jth prefix of y[1..n]
* define c[*i,j*] = **|**LCS(x[1..*i*],y[1..*j*])**|**
* so c[*m,n*] = **|**LCS(x,y)**|**
* recurrence?

# x[1..i] and y[1..j] end with xi=*y*j

**xi**

**x1 x2 … xi-1**

**yj=xi**

**y1 y2 … yj-1**

**zk =yj=xi**

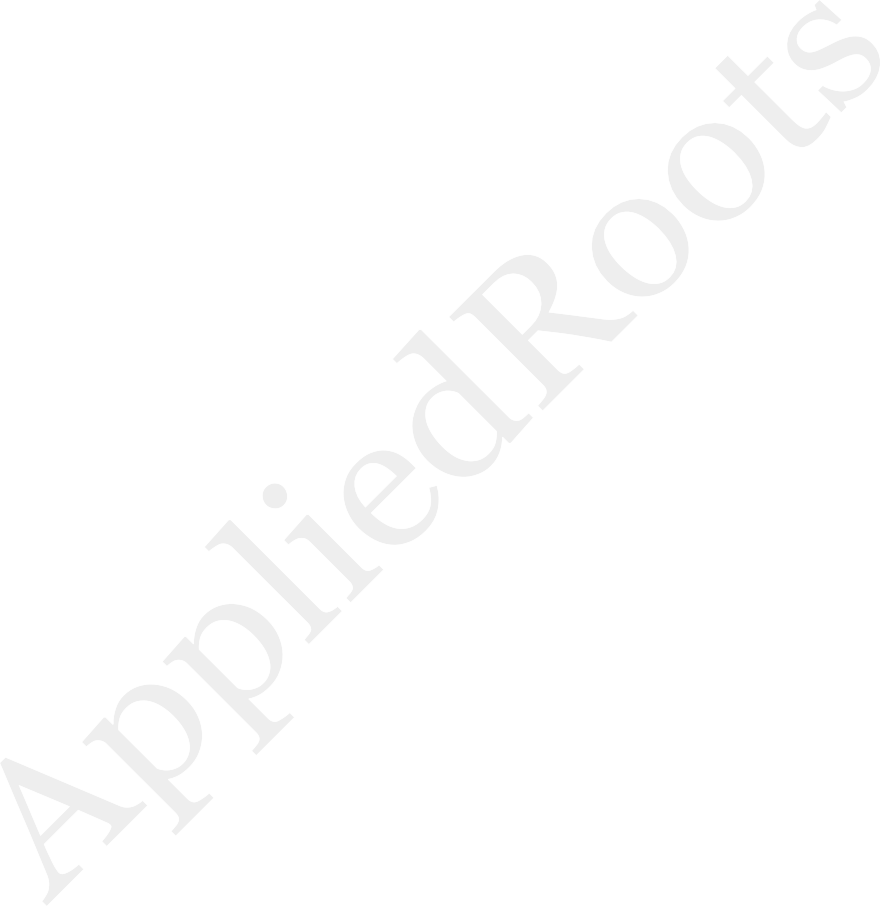
**z1 z2…zk-1**

*Z*k is *Z*k -1 followed by zk = *y*j = *x*i where

*Z*k-1 is an LCS of x[1..i-1] and y[1..j-1]

***c*(*i*, *j*) = c(*i*-1, *j*-1)+1**

# Example - use of property 1

*x*: B A N

A

N

A

*y*: A T

A

N

A

by inspection LCS of B A N and A T is A so LCS(x,y) is A A N A

# x[1..i] and y[1..j] end with *x*i  yj

**x1 x2 … xi-1 xi**

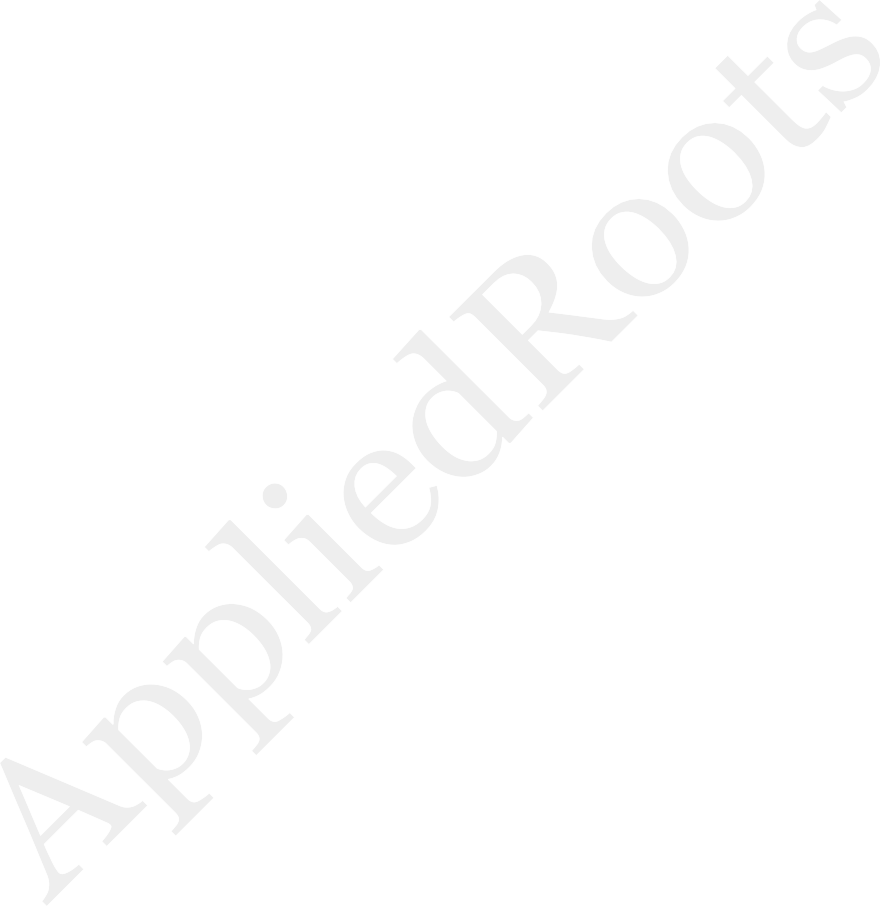
**x i**

**x1 x2 … xi-1**

**yj**

**y1 y2 … yj-1**

**yj y1 y2 …yj-1 yj**

***Z*k *Z*k**

**zk**  **xi**

**z1 z2…zk-1**

**zk** **yj**

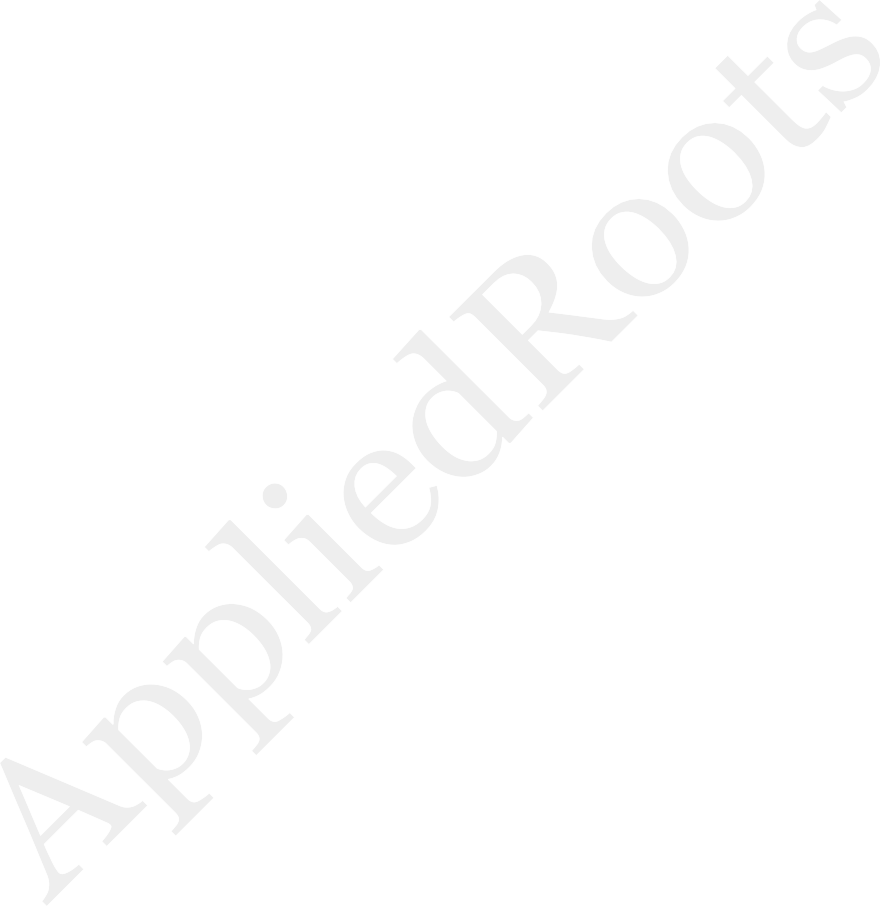
**z1 z2…zk-1**

**LCS of x[1..i] and y[1..j-1] LCS of x[1..i-1] and y[1..j]**

***c*(*i*, *j*)=max{*c*(*i*, *j*-1), *c*(*i*-1, *j*)}**

# Example: use of property 2

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *x*: | A | B | C | D | E | F | G |
| *y*: | B | C | D | G | K |  |  |

The last character of the LCS(x,y) either:

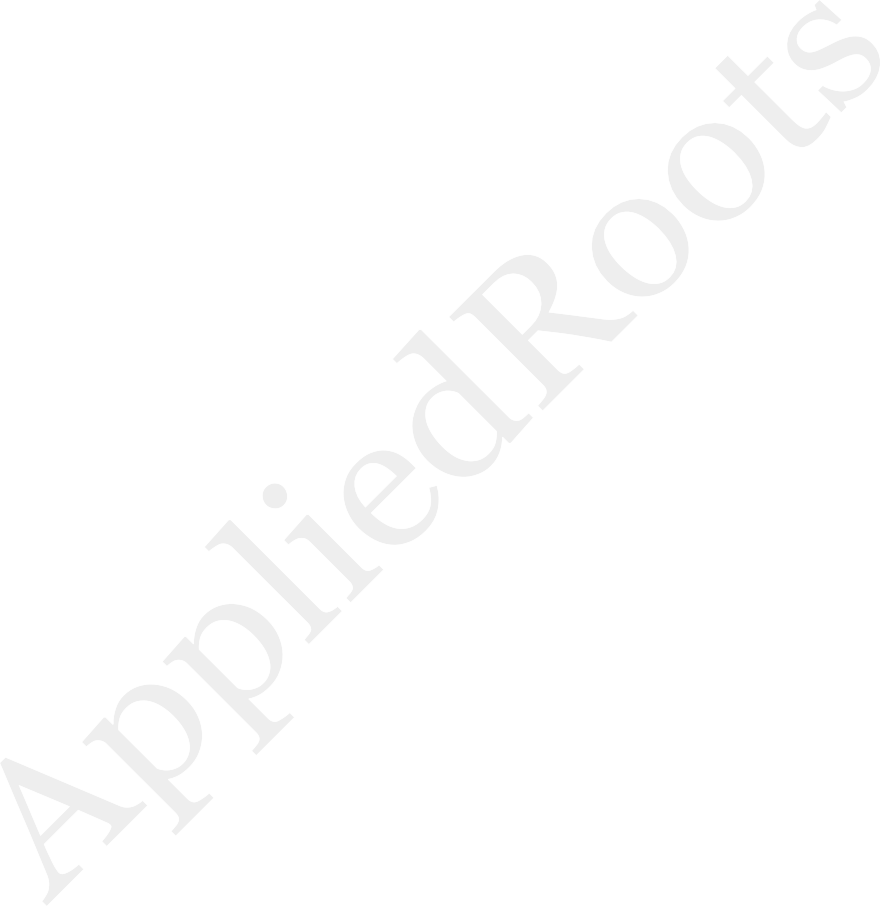
* ends with a G => can’t end with a K

• => can remove K from *y*

* doesn’t end with a G

• => can remove G from *x*

# A recurrence, summary

* consider prefixes of x and y
  + x[1..i] ith prefix of x[1..m]
  + y[1..j] jth prefix of y[1..n]
* define c[*i,j*] = **|**LCS(x[1..*i*],y[1..*j*])**|**
  + so c[*m,n*] = **|**LCS(x,y)**|**
* recurrence:

*c*[*i*,*j*]  

*c*[*i* 1, *j* 1] 1

if *xi* = *y j*

max{*c*[*i* 1, *j*],*c*[*i*, *j* 1]} otherwise

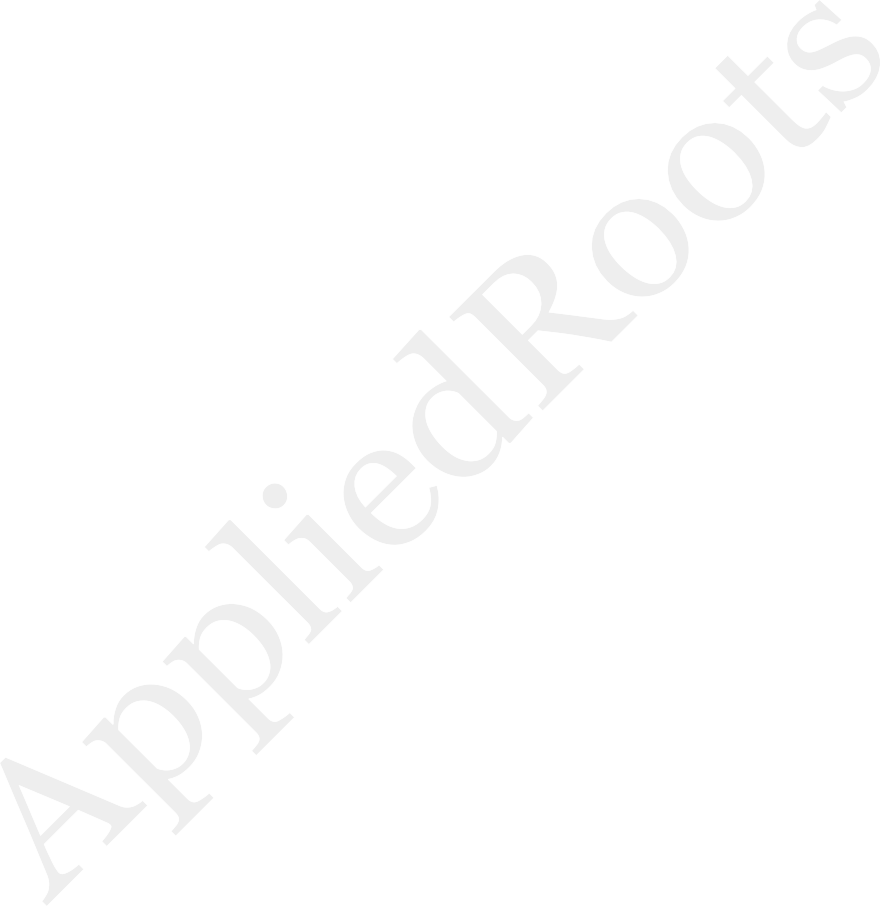
# Solving LCS with DP

*c*[*i*,*j*]  

*c*[*i* 1, *j* 1] 1

if *xi* = *y j*

max{*c*[*i* 1, *j*],*c*[*i*, *j* 1]} otherwise

* running time is ....
  + *O(n×m)*