Longest Prefix Suffix Matching

Shusen Wang

Prefix

• If string X can be written as X = PS for a nonempty string S, then P is a prefix of X.

```
Example string: X = "algorithm"
```

```
•prefixes = {"algorith", "algorit", "algori",
  "algor", "algo", "alg", "a"}.
```

Suffix

• If string X can be written as X = PS for a nonempty string P, then S is a suffix of X.

```
Example string: X = "algorithm"
```

```
• suffixes = {"lgorithm", "gorithm", "orithm",
   "rithm", "ithm", "thm", "hm", "m"}.
```

```
Example string: X = \text{``aba''}
```

```
•prefixes = { "ab", "a" }.
•suffixes = { "ba", "a" }.
```

```
Example string: X = \text{``aba''}
```

```
prefixes = { "ab", "a" }.
suffixes = { "ba", "a" }.
Their intersection: { "a" }.
The longest matching: "a".
```

```
Example string: X = "ababa"
```

```
•prefixes = { "abab", "aba", "ab", "a" }.
• suffixes = { "baba", "aba", "ba", "a" }.
```

Their intersection: { "aba", "a" }.

```
Example string: X = "ababa"

• prefixes = { "abab", "aba", "ab", "a" }.

• suffixes = { "baba", "aba", "ba", "a" }.
```

```
Example string: X = "ababa"
```

```
prefixes = { "abab", "aba", "ab", "a" }.
suffixes = { "baba", "aba", "ba", "a" }.
Their intersection: { "aba", "a" }.
```

• The longest matching: "aba".

| String: | X | = | a | b | a | b | a | b | С | a | |
|----------|---|---|---|---|---|---|---|---|---|---|--|
| Lengths: | L | = | ? | ? | ? | ? | ? | ? | ? | ? | |

```
String: X = a b a b a b c a

Lengths: L = ?
```

```
prefixes = { }.
suffixes = { }.
Their intersection: { }.
The longest matching: empty string. (Length = 0)
```

```
String: X = a b a b a b c a

Lengths: L = 0
```

```
• prefixes = { }.
• suffixes = { }.
• Their intersection: { }.
• The longest matching: empty string. (Length = 0)
```

```
String: X = a b a b a b c a

Lengths: L = 0 ?
```

```
prefixes = {"a"}.
suffixes = {"b"}.
Their intersection: { }.
The longest matching: empty string. (Length = 0)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0
```

```
prefixes = {"a"}.
suffixes = {"b"}.
Their intersection: { }.
The longest matching: empty string. (Length = 0)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0 ?
```

```
prefixes = {"ab", "a"}.
suffixes = {"ba", "a"}.
Their intersection: {"a"}.
The longest matching: "a". (Length = 1)
```

```
String: X = a b a b c a

Lengths: L = 0 0 1
```

```
prefixes = {"ab", "a"}.
suffixes = {"ba", "a"}.
Their intersection: {"a"}.
The longest matching: "a". (Length = 1)
```

```
String: X = a b a b a b a Lengths: L = 0 0 1 ?
```

```
prefixes = {"aba", "ab", "a"}.
suffixes = {"bab", "ab", "a"}.
Their intersection: {"ab", "a"}.
The longest matching: "ab". (Length = 2)
```

```
String: X = a b a b c a

Lengths: L = 0 0 1 2
```

```
prefixes = {"aba", "ab", "a"}.
suffixes = {"bab", "ab", "a"}.
Their intersection: {"ab", "a"}.
The longest matching: "ab". (Length = 2)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0 1 2 ?
```

```
prefixes = {"abab", "aba", "ab", "a"}.
suffixes = {"baba", "aba", "ba", "a"}.
Their intersection: {"aba", "a"}.
The longest matching: "aba". (Length = 3)
```

```
String: X = a b a b a b a Lengths: L = 0 0 1 2 3
```

```
prefixes = {"abab", "aba", "ab", "a"}.
suffixes = {"baba", "aba", "ba", "a"}.
Their intersection: {"aba", "a"}.
The longest matching: "aba". (Length = 3)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0 1 2 3 ?
```

```
prefixes = {"ababa", "abab", "aba", "ab", "a"}.
suffixes = {"babab", "abab", "bab", "ab", "b"}.
Their intersection: {"abab", "ab"}.
The longest matching: "abab". (Length = 4)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0 1 2 3 4
```

```
prefixes = {"ababa", "abab", "aba", "ab", "a"}.
suffixes = {"babab", "abab", "bab", "ab", "b"}.
Their intersection: {"abab", "ab"}.
The longest matching: "abab". (Length = 4)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0 1 2 3 4 ?
```

```
prefixes = {"ababab", "ababa", "abab", "aba", "aba", "ab", "a"}.
suffixes = {"bababc", "ababc", "babc", "abc", "bc", "c"}.
Their intersection: { }.
The longest matching: empty string. (Length = 0)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0 1 2 3 4 0
```

```
prefixes = {"ababab", "ababa", "abab", "aba", "aba", "ab", "a"}.
suffixes = {"bababc", "ababc", "babc", "abc", "bc", "c"}.
Their intersection: { }.
The longest matching: empty string. (Length = 0)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0 1 2 3 4 0 ?
```

```
prefixes = {"abababc", "ababab", "ababa", "abab", "aba", "ab", "a"}.
suffixes = {"bababca", "ababca", "babca", "abca", "bca", "ca", "a"}.
Their intersection: { "a" }.
The longest matching: "a". (Length = 1)
```

```
String: X = a b a b a b c a

Lengths: L = 0 0 1 2 3 4 0 1
```

```
prefixes = {"abababc", "ababab", "ababa", "abab", "aba", "ab", "a"}.
suffixes = {"bababca", "ababca", "babca", "abca", "bca", "ca", "a"}.
Their intersection: { "a" }.
The longest matching: "a". (Length = 1)
```

| String: | X | = | a | b | a | b | a | b | С | a |
|----------|---|---|---|---|---|---|---|---|---|---|
| Lengths: | L | = | 0 | 0 | 1 | 2 | 3 | 4 | 0 | 1 |

What does the number mean?

String: X = a b a b a b c a

Lengths: L = 0 0 1 2 3 4 0 1

What does the number mean?

Application

Why is the longest prefix suffix array interesting?

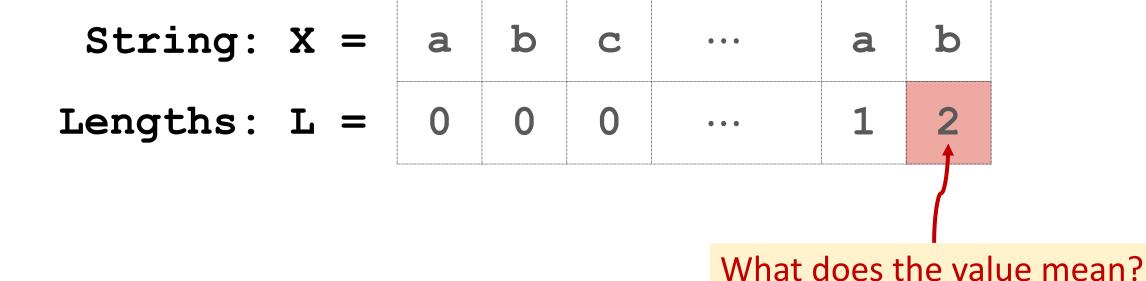
- The array is used by the Knuth–Morris–Pratt (KMP) algorithm [1].
- KMP algorithm solves the string matching problem.

Reference:

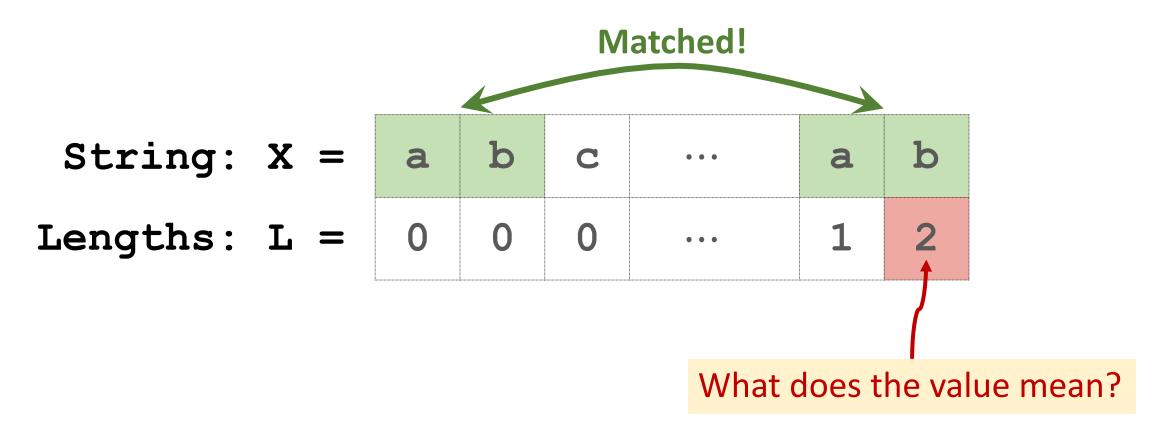
1. Knuth, Morris, & Pratt. Fast pattern matching in strings. SIAM Journal on Computing, 6 (2): 323–350, 1977.

Linear-Time Algorithm

| String: | X = | a | b | С | • • • | a | b |
|----------|-----|---|---|---|-------|---|---|
| Lengths: | L = | 0 | 0 | 0 | • • • | 1 | 2 |



| | | | | | ····· | | · | | |
|----------|-----|---|---|---|-------|--------|--------|--------|---------|
| String: | x = | a | b | С | | • • • | a | b | |
| Lengths: | L = | 0 | 0 | 0 | | • • • | 1 | 2 | |
| | | | | | | | | | |
| | | | | | | What o | does t | he val | ue mean |



What is the next element in the array?

| String: | x = | a | b | C | • • • | a | b |
|----------|-----|---|---|---|-------|---|---|
| Lengths: | L = | 0 | 0 | 0 | • • • | 1 | 2 |

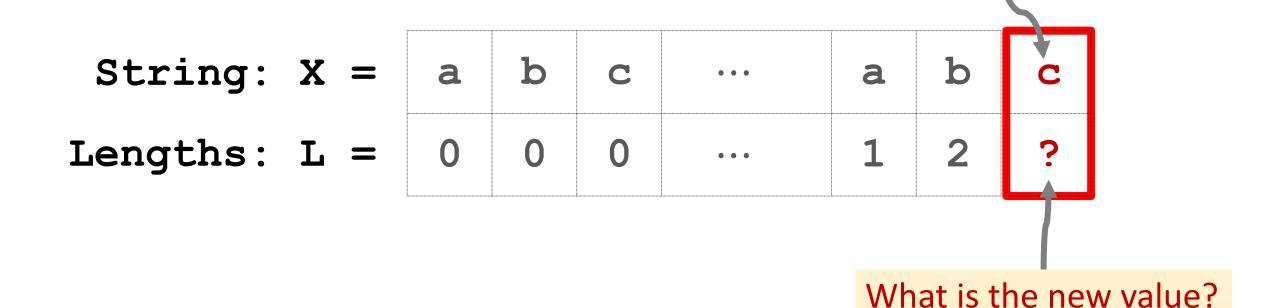
What is the next element in the array?

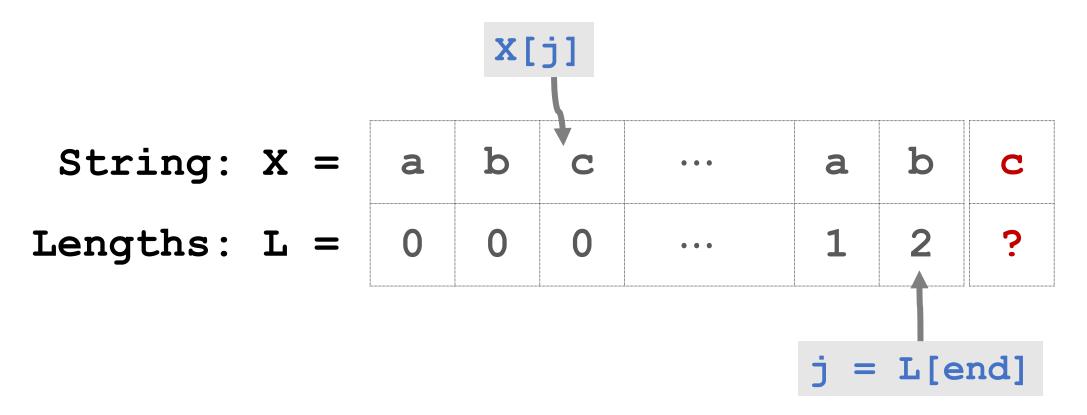
Append a new character to X

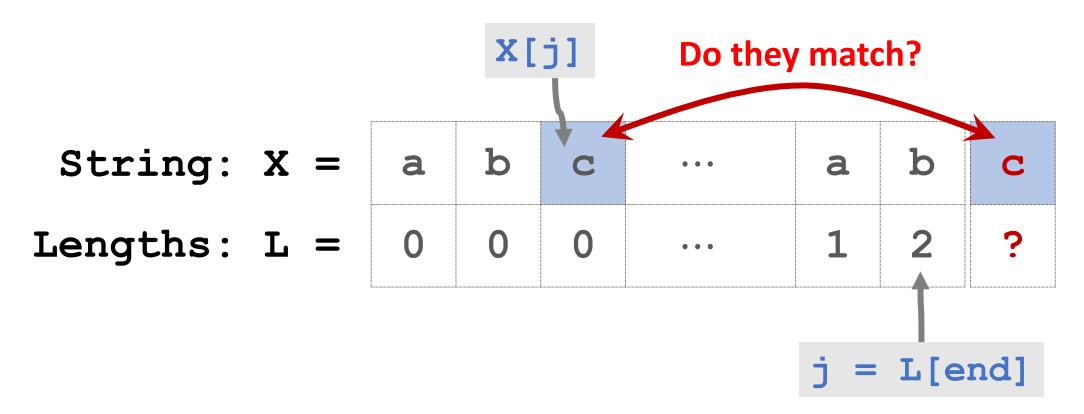
| String: | x = | a | b | С | • • • | a | b | C |
|----------|-----|---|---|---|-------|---|---|---|
| Lengths: | L = | 0 | 0 | 0 | • • • | 1 | 2 | |

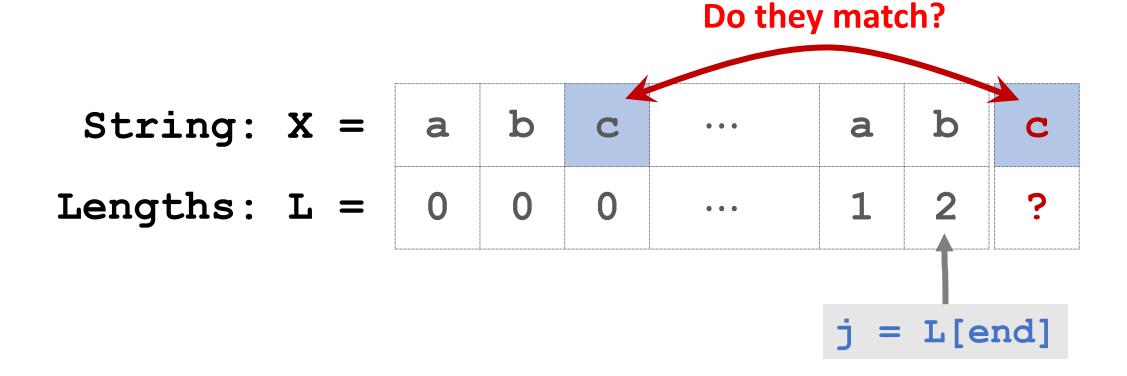
What is the next element in the array?

Append a new character to X

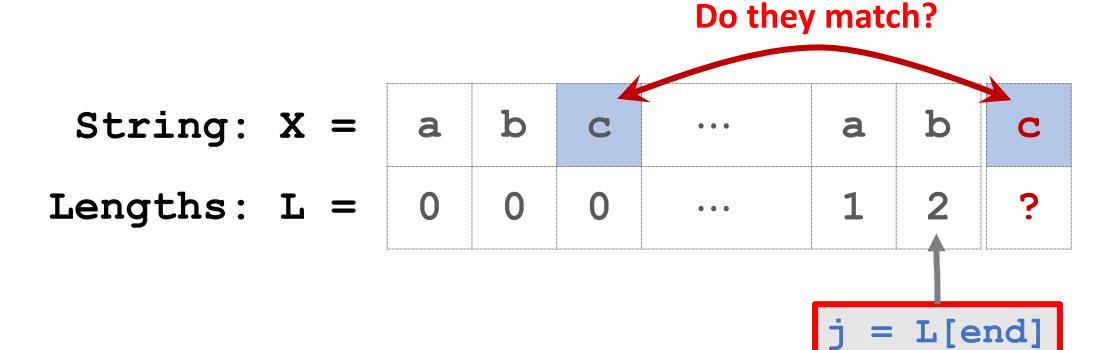




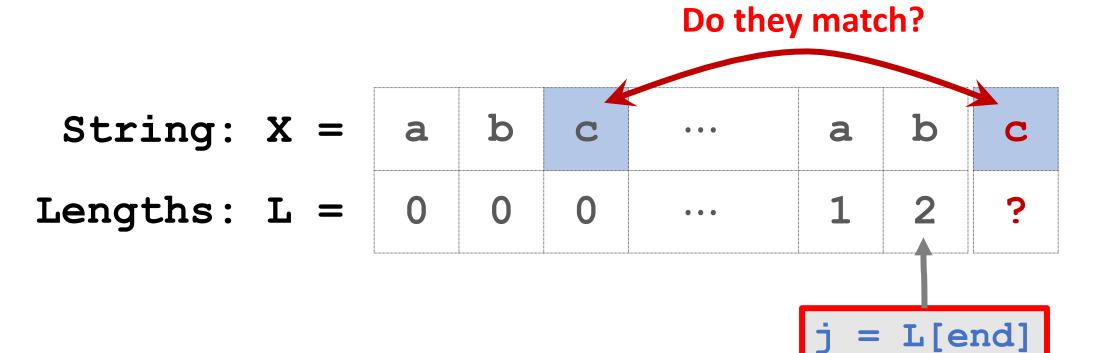


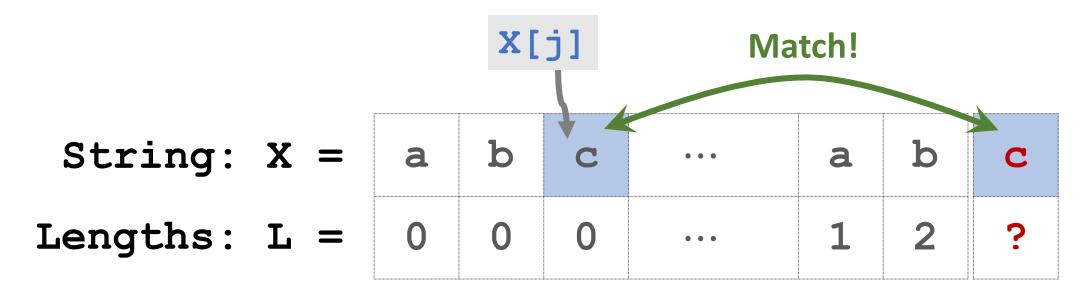


• Match ==> Case 1;



- Match ==> Case 1;
- Mismatch ==>





• Case 1: the new char is equal to X[j].

```
String: X = a b c ... a b c

Lengths: L = 0 0 0 ... 1 2 ?

j = L[end]
```

• Case 1 ==> Let the new value in L be j+1.

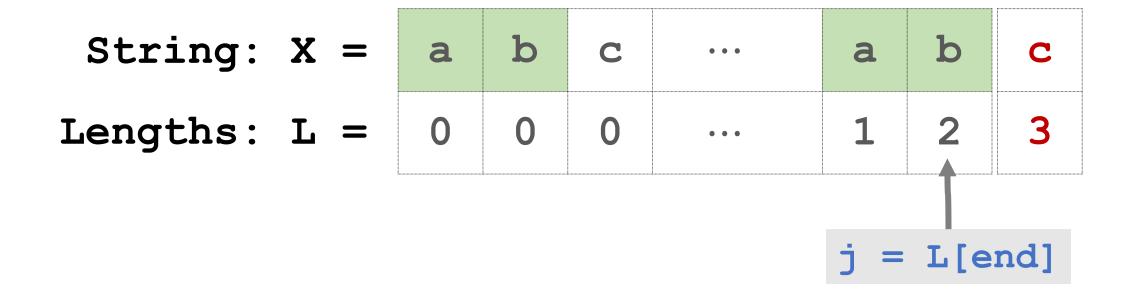
```
String: X = a b c ... a b c

Lengths: L = 0 0 0 ... 1 2 3

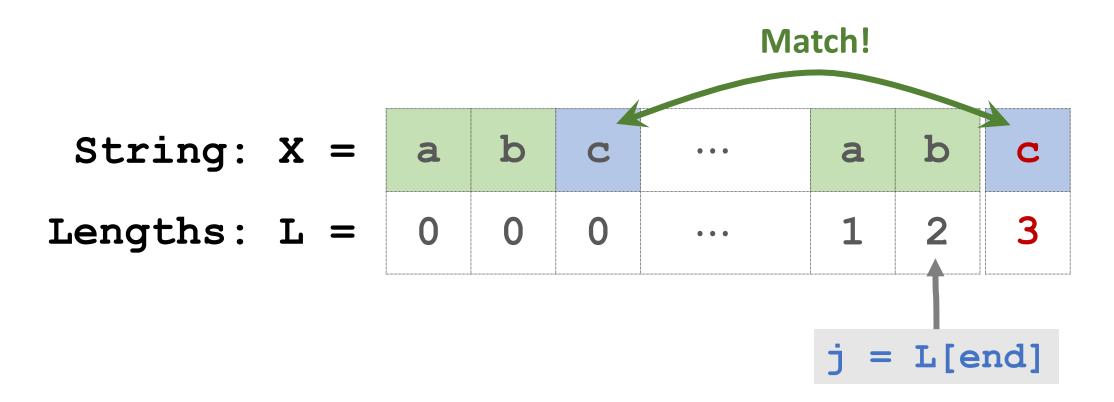
j = L[end]
```

• Case 1 ==> Let the new value in L be j+1.

Question: Why is the new value equal to j+1?



Question: Why is the new value equal to j+1?



Question: Why is the new value equal to j+1?

Another Example:

f

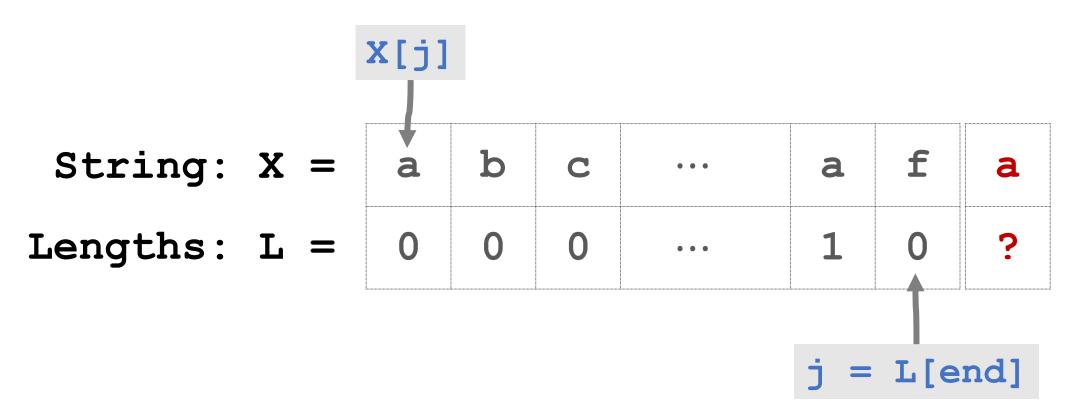
a

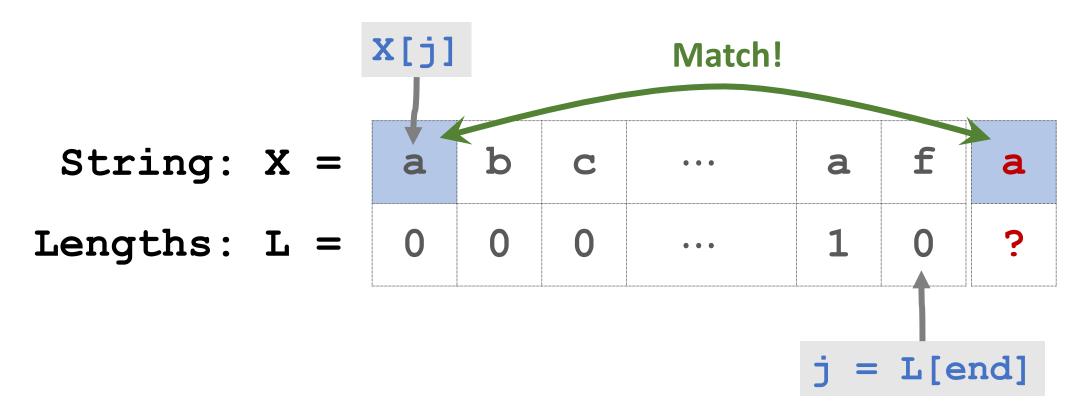
String: X = a b c ···

Lengths: $L = \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$

Append a new character to X

| String: | x = | a | b | С | • • • | a | f | a |
|----------|-----|---|---|---|-------|---|---|---|
| Lengths: | L = | 0 | 0 | 0 | • • • | 1 | 0 | ? |





• Case 1: the new char is equal to X[j].

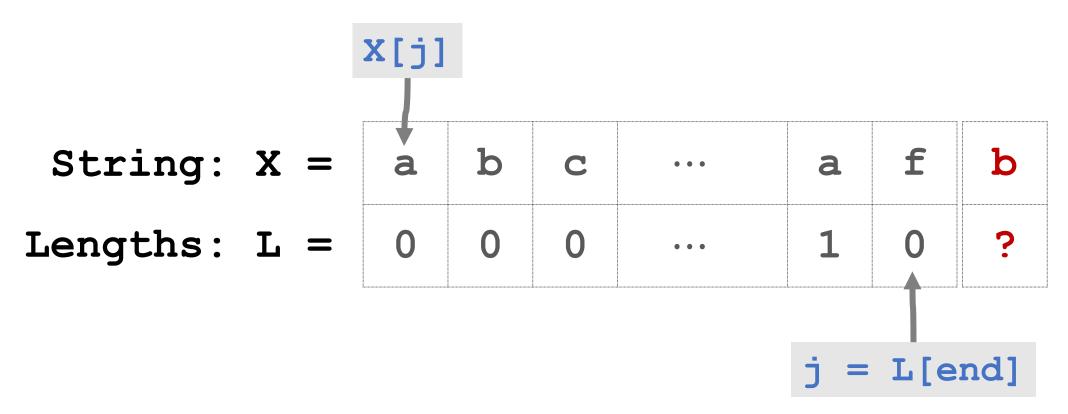
```
String: X = a b c ... a f a
Lengths: L = 0 0 0 ... 1 0 1
j = L[end]
```

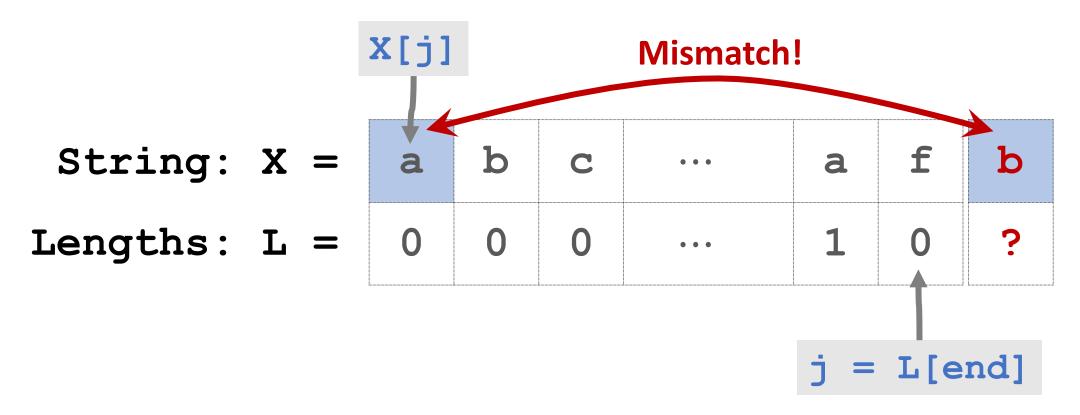
• Case 1 ==> Let the new value of L be j+1.

| String: | X | = | a | b | С | • • • | a | f | |
|----------|---|---|---|---|---|-------|---|---|--|
| Lengths: | L | = | 0 | 0 | 0 | • • • | 1 | 0 | |

Append a new character to X

| String: | x = | a | b | С | • • • | a | f | b | |
|----------|-----|---|---|---|-------|---|---|---|--|
| Lengths: | L = | 0 | 0 | 0 | • • • | 1 | 0 | ÿ | |





• Mismatch ==>
$$\begin{cases} j=0 & ==> \text{ Case 2;} \\ j\neq 0 & ==> \text{ Case 3.} \end{cases}$$

```
String: X = a b c ... a f b
Lengths: L = 0 0 0 ... 1 0 ?
j = L[end]
```

• Mismatch ==>
$$\begin{cases} j = 0 & ==> \text{ Case 2}; \\ j \neq 0 & ==> \text{ Case 3}. \end{cases}$$

String: X = a b c ... a f b

Lengths: L = 0 0 0 ... 1 0 ?

• Case 2 ==> Let the new value be 0.

String: $X = \begin{bmatrix} a & b & c & \cdots & a & f & b \end{bmatrix}$ Lengths: $L = \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 & 0 & 0 \end{bmatrix}$

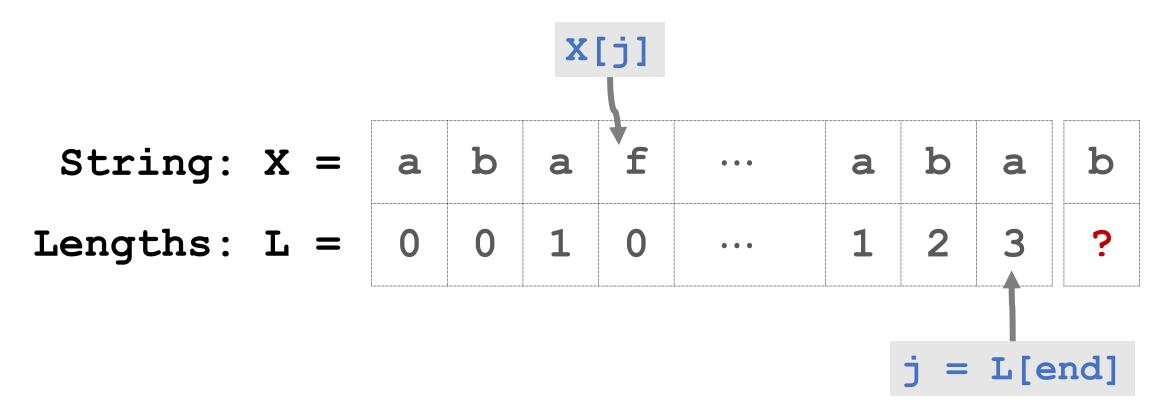
• Case 2 ==> Let the new value be 0.

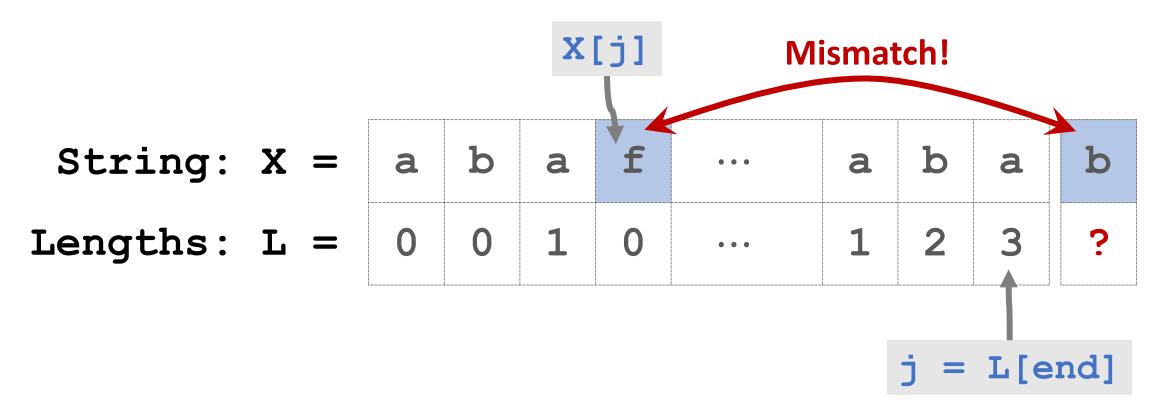
String: X = a b a f ... a b a

Lengths: L = 0 0 1 0 ... 1 2 3

Append a new character to X

| String: | X | = | a | b | a | f | • • • | a | b | a | b |
|----------|---|---|---|---|---|---|-------|---|---|---|---|
| Lengths: | L | = | 0 | 0 | 1 | 0 | • • • | 1 | 2 | 3 | 3 |





• Mismatch ==>
$$\begin{cases} j=0 & ==> \text{ Case 2}; \\ j\neq 0 & ==> \text{ Case 3}. \end{cases}$$

```
String: X = a b a f ... a b a b

Lengths: L = 0 0 1 0 ... 1 2 3 ?

j = L[end]
```

• Mismatch ==>
$$\begin{cases} j = 0 ==> \text{Case 2}; \\ j \neq 0 ==> \text{Case 3}. \end{cases}$$

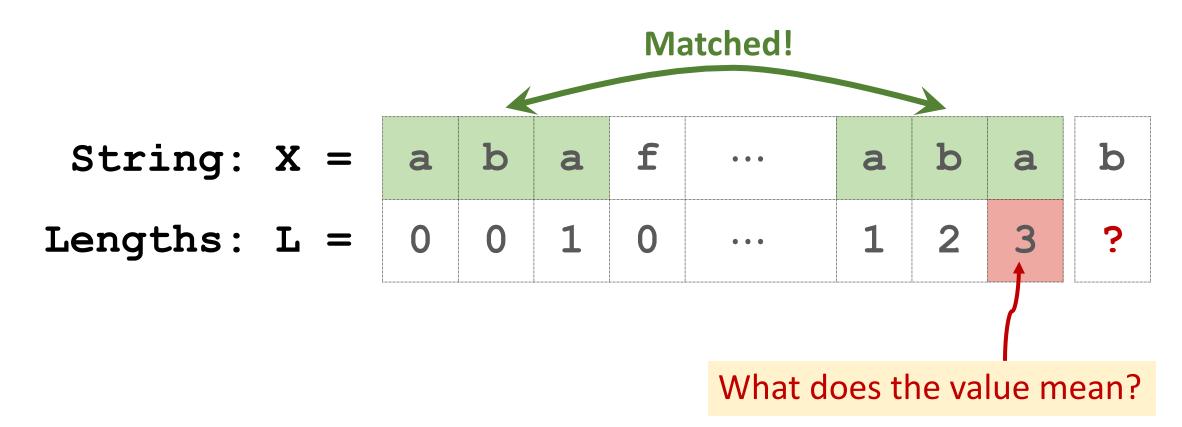
```
String: X = a b a f ... a b a b
Lengths: L = 0 0 1 0 ... 1 2 3 ?
j = L[end]
```

• Case 3 ==> Reduce the big problem to a smaller one.

String: X = a b a f ... a b a b

Lengths: L = 0 0 1 0 ... 1 2 3 ?

What does the value mean?



String: X = a b a f ... a b a b

Lengths: L = 0 0 1 0 ... 1 2 3 ?

String: X' =

Lengths: L' =

| String: | X | = | a | b | a | f | • • • | a | b | a | b |
|----------|---|---|---|---|---|---|-------|---|---|---|---|
| Lengths: | L | = | 0 | 0 | 1 | 0 | • • • | 1 | 2 | 3 | ? |

| String: | X' | | a | b | a |
|----------|----|---|---|---|---|
| Lengths: | L' | = | 0 | 0 | 1 |

String: X = a b a f ... a b a b

Lengths: L = 0 0 1 0 ... 1 2 3 ?

String: X' = a b a b

Lengths: L' = 0 0 1

| String: | x = | a | b | a | f | • • • | a | b | a | b |
|----------|------|---|---|---|---|-------|---|------|--|---|
| Lengths: | L = | 0 | 0 | 1 | 0 | • • • | 1 | 2 | 3 | ? |
| | | L | | | | | | | <u> </u> | |
| String: | x' = | a | b | a | b | | | Equa | | , |
| Lengths: | L' = | 0 | 0 | 1 | ? | | | | | |

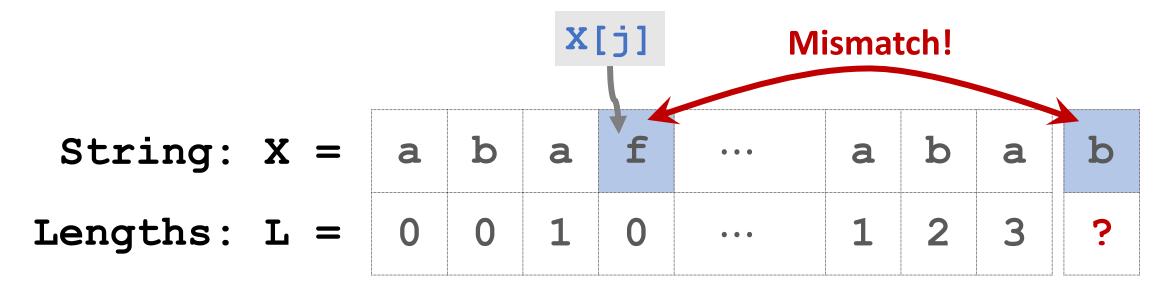
| String: | X = | = | a | b | a | f | • • • | a | b | a | b |
|----------|-----|----|---|----------|---|-----|-------|---|------|---|---|
| Lengths: | L = | = | 0 | 0 | 1 | 0 | • • • | 1 | 2 | 3 | ? |
| | | I. | | <u> </u> | | | | | | | |
| String: | X' | = | a | b | a | b | | | Equa | | |
| Lengths: | L' | = | 0 | 0 | 1 | 2 - | | | | | |

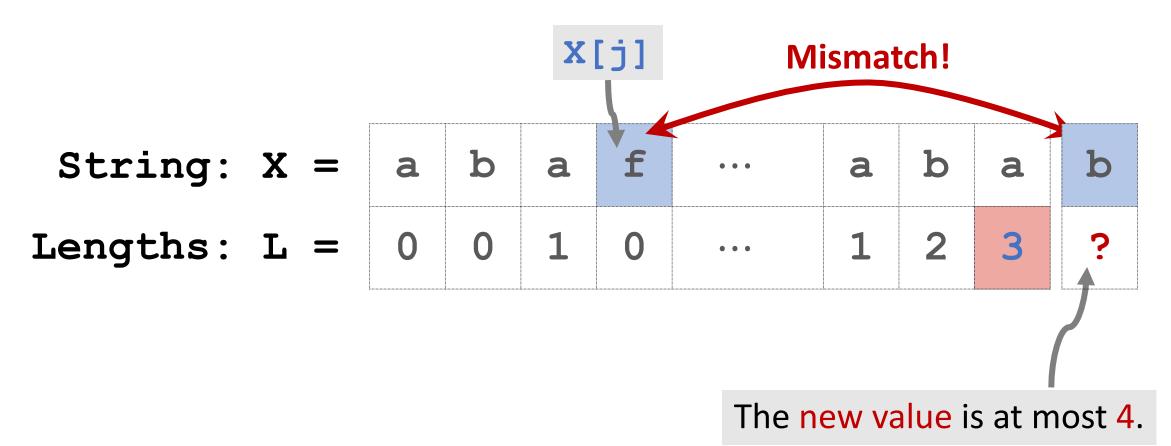
| String: | X : | - | a | b | a | f | • • • | a | b | a | b |
|----------|-----|----------|---|----------|---|-----|-------|---|------|---|---|
| Lengths: | L : | = | 0 | 0 | 1 | 0 | • • • | 1 | 2 | 3 | 2 |
| | | <u></u> | | <u> </u> | | | | | | | |
| String: | X' | | a | b | a | b | | | Equa | | |
| Lengths: | L' | | 0 | 0 | 1 | 2 - | | | | | |

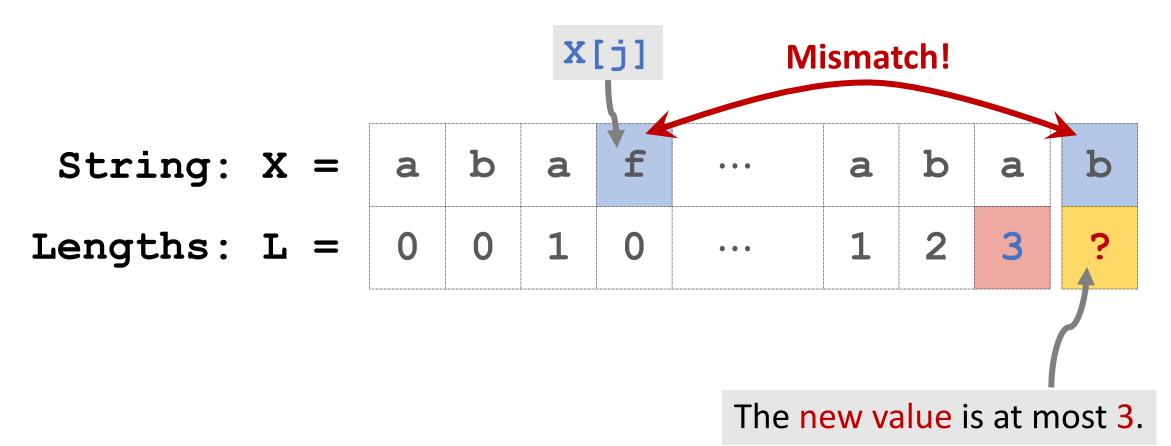
String: X = a b a f ... a b a b

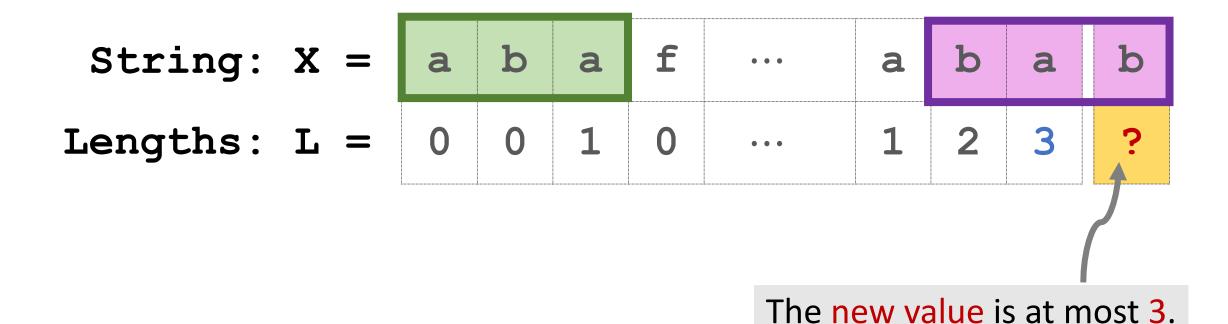
Lengths: L = 0 0 1 0 ... 1 2 3 ?

Question: Why can Case 3 be solved in this way?









| String: X | X = | a | b | a | f | • • • | a | b | a | b |
|------------|------------------|---|---|---|---|-------|---|---|---|---|
| Lengths: I | , = ['] | 0 | 0 | 1 | 0 | • • • | 1 | 2 | 3 | ? |

Question: Why are the yellow entry and blue entry equal?

String:
$$X = \begin{bmatrix} a & b & a & f & \cdots & a & b & a \\ Lengths: L = \begin{bmatrix} 0 & 0 & 1 & 0 & \cdots & 1 & 2 & 3 \\ \end{bmatrix}$$

String: $X' = \begin{bmatrix} a & b & a & b \\ \end{bmatrix}$

Equal

Lengths: L' =

Question: Why are the yellow entry and blue entry equal?

X and X' have the same prefixes.

String:
$$X' = \begin{bmatrix} a & b & a \end{bmatrix}$$
 b

Lengths: $L' = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$?

Question: Why are the yellow entry and blue entry equal?

X and X' have the same prefixes.

Question: Why are the yellow entry and blue entry equal?

X and X' have the same prefixes.

String:
$$X' = \begin{bmatrix} a & b & a & b \end{bmatrix}$$
Lengths: $L' = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}$?

Question: Why are the yellow entry and blue entry equal?

X and X' have the same suffixes.

String:
$$X' = a$$
 b a b

Lengths: $L' = 0$ 0 1

Question: Why are the yellow entry and blue entry equal?

X and X' have the same suffixes.

String:
$$X' = a b a b$$

Lengths: $L' = 0 0 1$

Question: Why are the yellow entry and blue entry equal?

→ String: X = a b a f ··· a b a b

Lengths: L = 0 0 1 0 ··· 1 2 3 2

X and X' have the same suffixes.

String: X' = a b a bLengths: L' = 0 0 1

Question: Why are the yellow entry and blue entry equal?

Equal

String: X' = a b a b

Lengths: L' = 0 0 1

Summary

Longest Prefixes Suffixes Matching

```
Example string: X = \text{``ababa''}
```

```
•prefixes = { "abab", "aba", "ab", "a" }.
• suffixes = { "baba", "aba", "ba", "a" }.
```

Longest Prefixes Suffixes Matching

```
Example string: X = "ababa"
```

```
•prefixes = { "abab", "aba", "ab", "a" }.
• suffixes = { "baba", "aba", "ba", "a" }.
• Their intersection: { "aba", "a" }.
```

Longest Prefixes Suffixes Matching

```
Example string: X = "ababa"
```

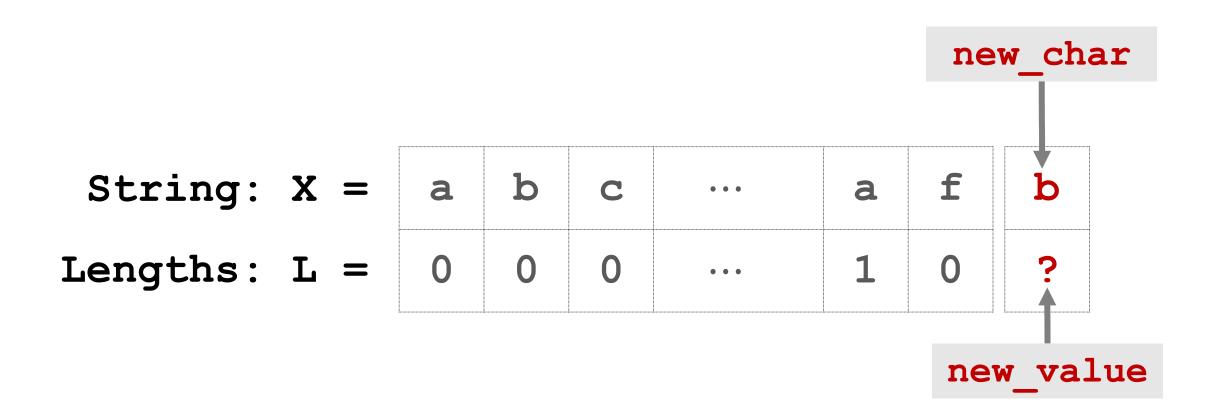
```
prefixes = { "abab", "aba", "ab", "a" }.
suffixes = { "baba", "aba", "ba", "a" }.
Their intersection: { "aba", "a" }.
```

• The longest matching: "aba".

Longest Prefix Suffix Array

| String: | X | = | a | b | С | • • • | a | f | |
|----------|---|---|---|---|---|-------|---|---|--|
| Lengths: | L | = | 0 | 0 | 0 | • • • | 1 | 0 | |

Longest Prefix Suffix Array



```
Function: new_value = f(X, L, new_char)
```



```
Function: new_value = f(X, L, new_char)
```

```
new char
 String: X =
                a
Lengths: L =
                                      new value
```

Time complexity: O(1) time (amortized) for running the function once.

```
Function: new_value = f(X, L, new_char)
```

```
Step 1: Decide the 3 cases
```

• Let j = L[end].

```
Function: new_value = f(X, L, new_char)
```

```
Step 1: Decide the 3 cases
```

- Let j = L[end].
- Match (i.e., new char == X[j]) or mismatch?

```
Function: new_value = f(X, L, new_char)
```

```
Step 1: Decide the 3 cases

• Let j = L[end].

• Match (i.e., new_char == X[j]) or mismatch?

• If Match ==> Case 1;
```

Function: new_value = f(X, L, new_char)

```
Step 1: Decide the 3 cases
• Let j = L[end].
• Match (i.e., new char == X[j]) or mismatch?
If Match ==> Case 1;
• If Mismatch ==> |j = 0| ==> Case 2;
j \neq 0 ==> Case 3.
```

```
Function: value = f(X, L, new_char)
```

Step 2: Different solutions to the 3 cases • Case 1 ==> Return j + 1. • Case 2 ==> Return 0. • Case 3 ==> Return f(X', L', new char).

X' = X[0:j] and L' = L[0:j]

Questions

Build Partial Matching Table

Question 1:

| x = | | | | | | | | | | | |
|-----|---|---|---|---|----|---|---|---|---|---|---|
| L = | 0 | 0 | 1 | ? | ·· | ? | 1 | 2 | 3 | 4 | ? |

Question 2:

| x = | a | a | a | a | b | • • • | a | a | a | a | a |
|-----|---|---|---|---|---|-------|---|---|---|---|---|
| L = | 0 | 1 | 2 | 3 | ? | • • • | 1 | 2 | 3 | 4 | ? |

Build Partial Matching Table

Question 3:

Question 4:

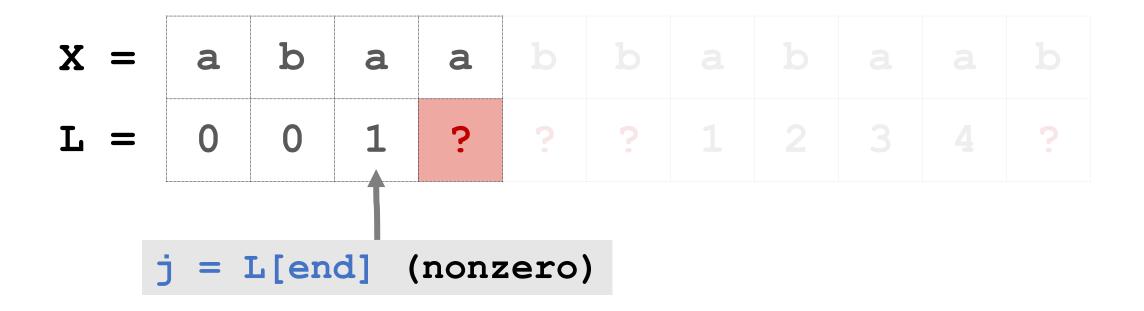
| x = | b | a | b | a | a | • • • | a | b | a | b | a | С | a |
|-----|---|---|---|---|---|-------|---|---|---|---|---|---|---|
| L = | 0 | 0 | 1 | 2 | ? | • • • | 2 | 3 | ? | ? | ? | 0 | ? |

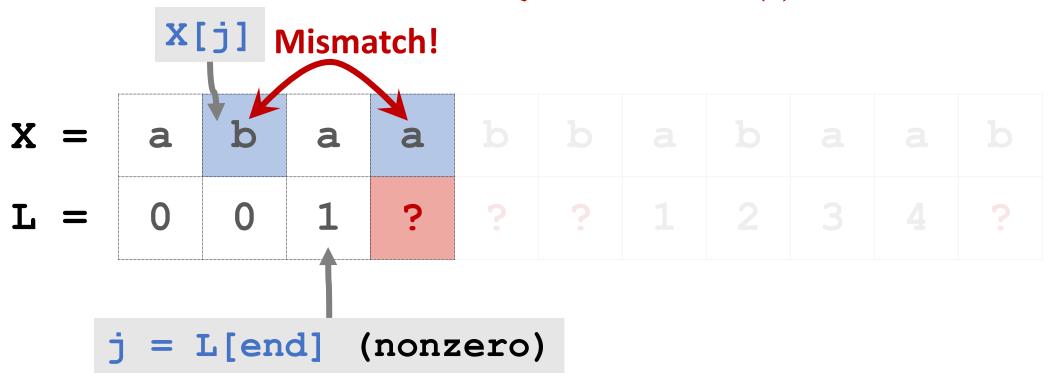
Thank You!

Solution to Question 1

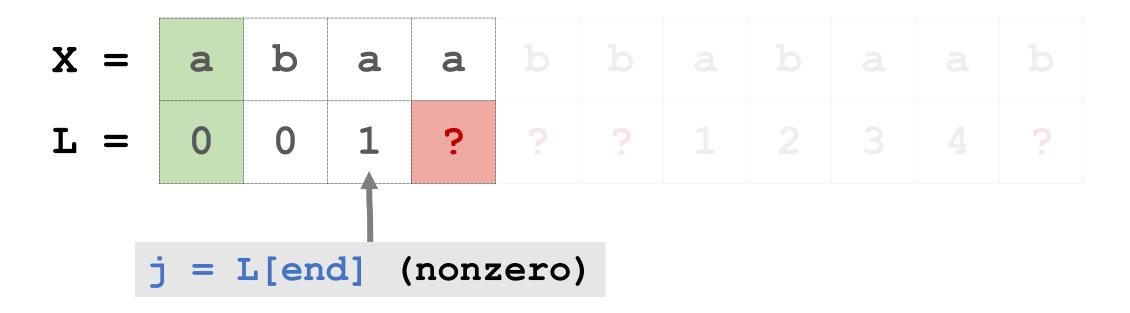
Solution to Question 1

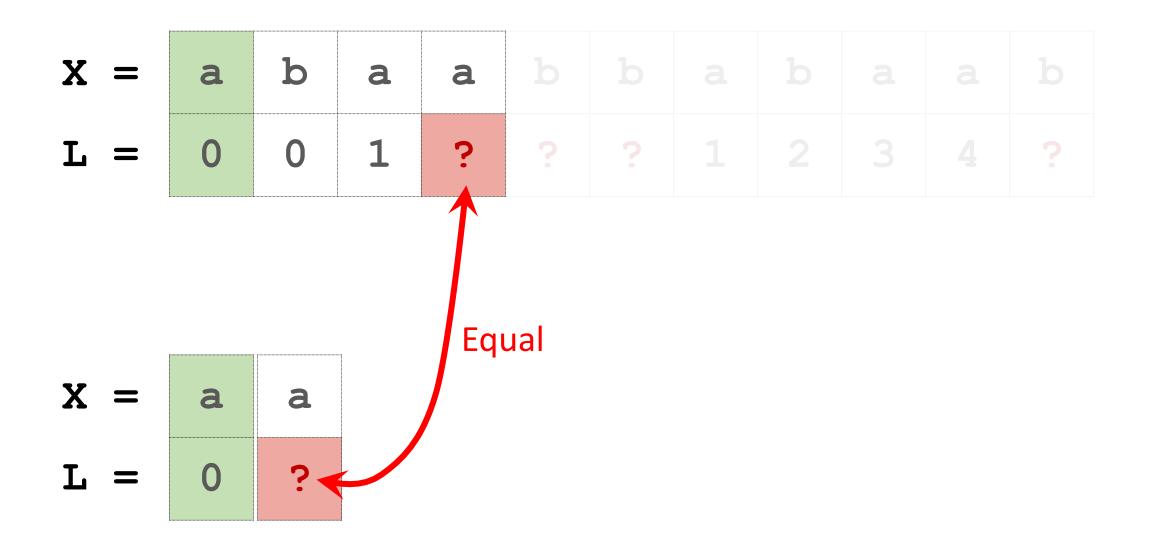
| x = | a | b | a | a | b | b | a | b | a | a | b |
|-----|---|---|---|----------|---|----|---|---|---|---|---|
| L = | 0 | 0 | 1 | . | ; | ÷. | 1 | 2 | 3 | 4 | ? |

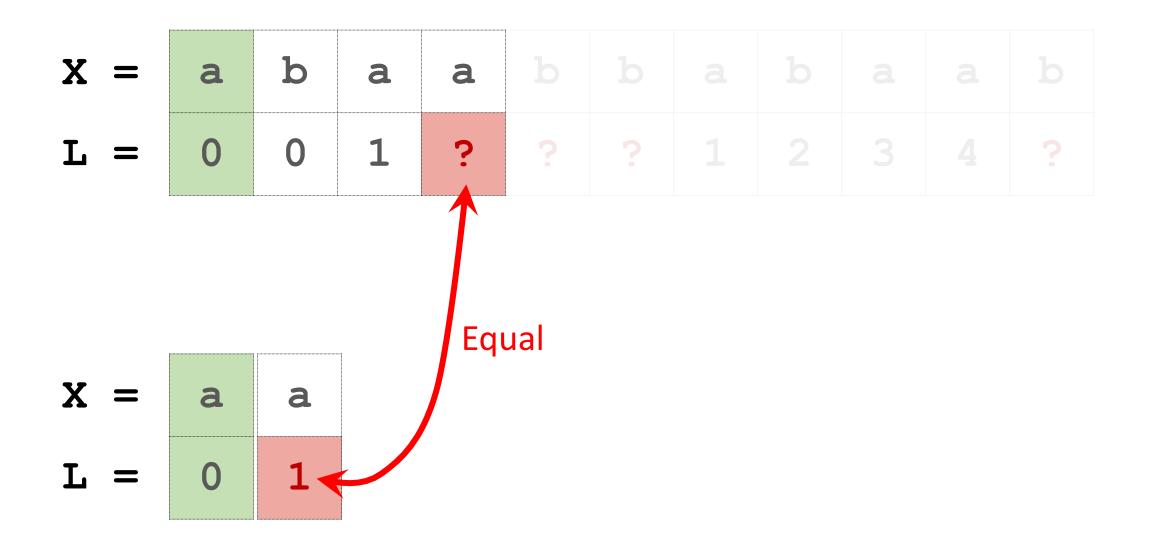


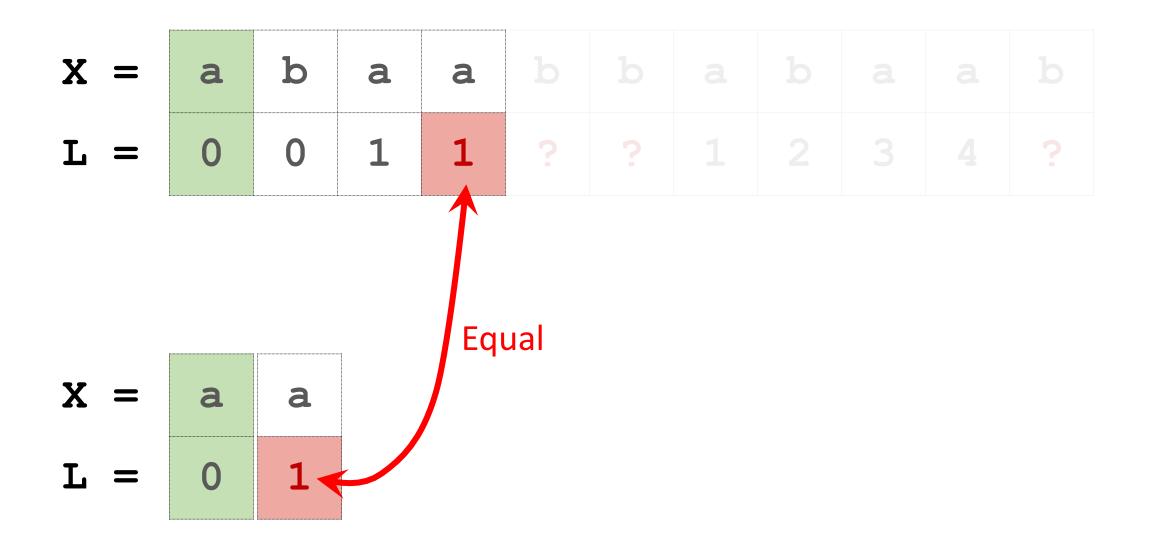


- This is Case 3.
- Reduce the problem to a smaller problem.

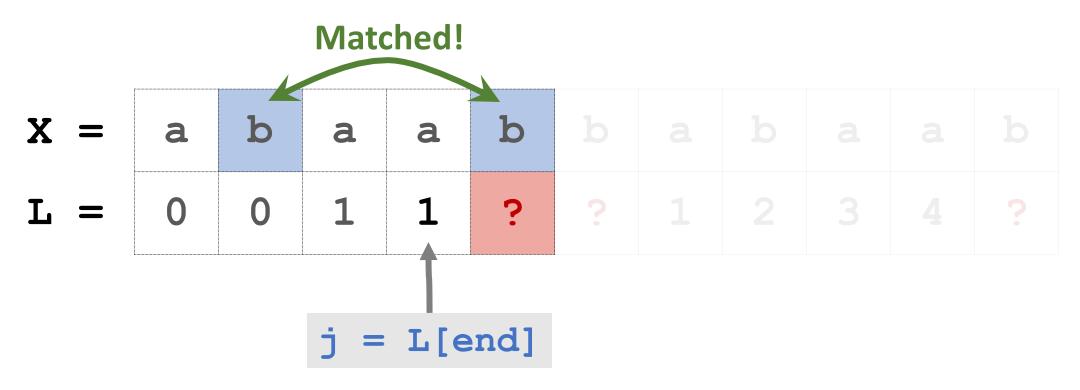




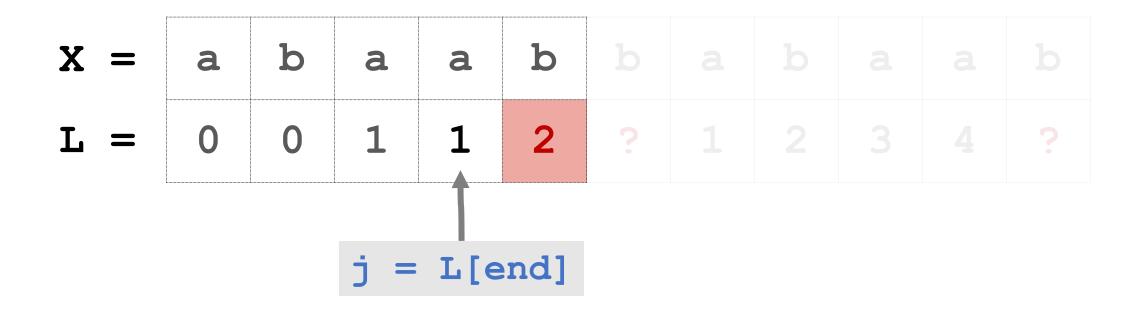




| x = | | | | | | | | | |
|-----|---|---|---|---|---|---|--|--|--|
| L = | 0 | 0 | 1 | 1 | ? | ? | | | |

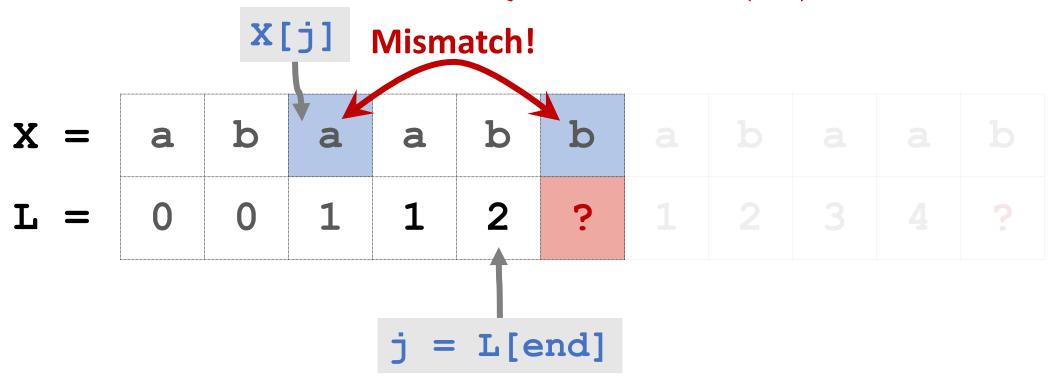


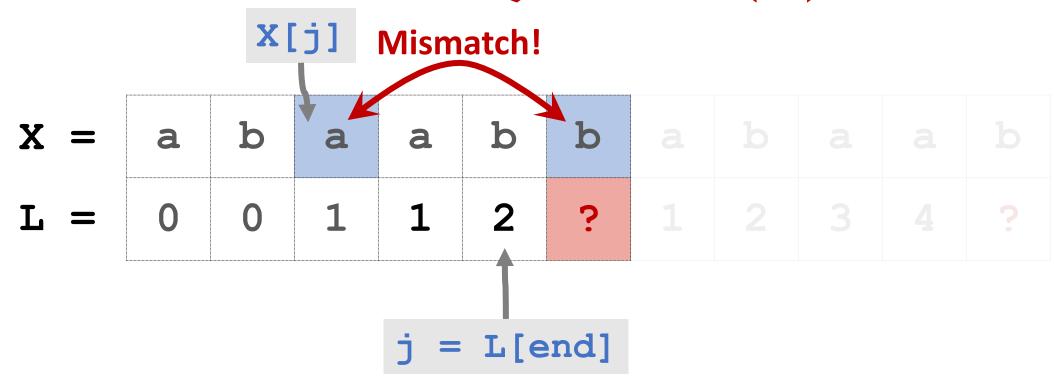
- Case 1: the new char is equal to X[j].
- Then the new value is j+1.



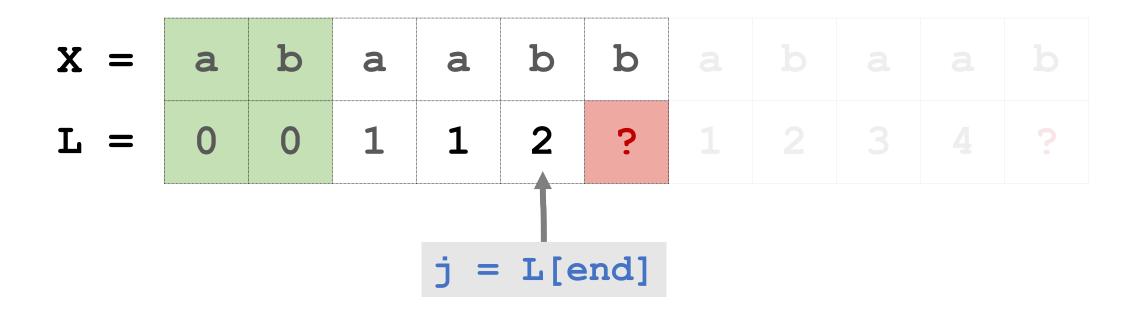
- Case 1: the new char is equal to X[j].
- Then the new value is j+1.

| x = | | | | | | | | | |
|-----|---|---|---|---|---|---|--|--|--|
| L = | 0 | 0 | 1 | 1 | 2 | • | | | |



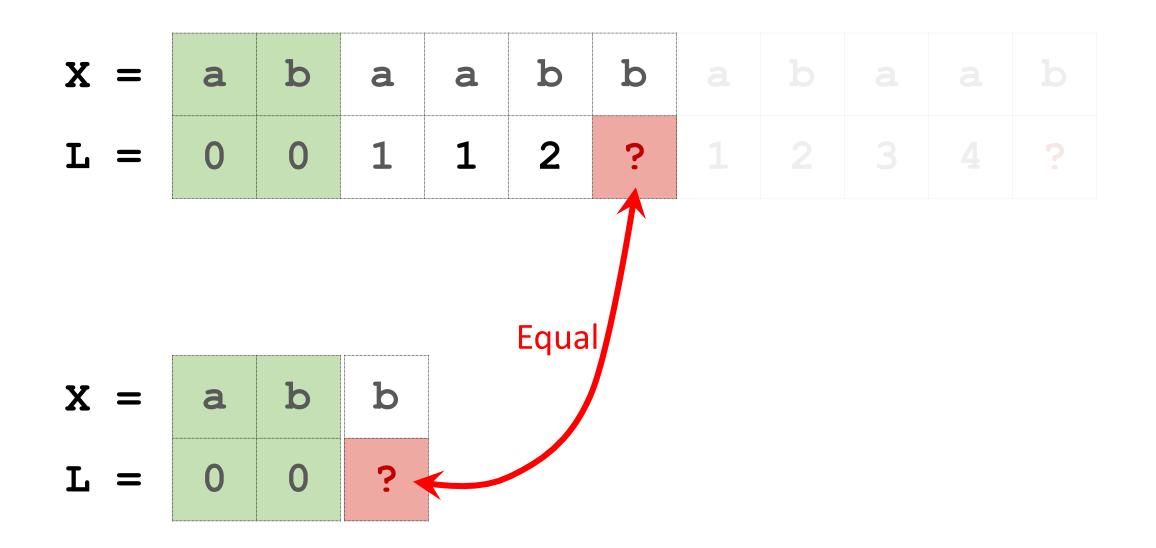


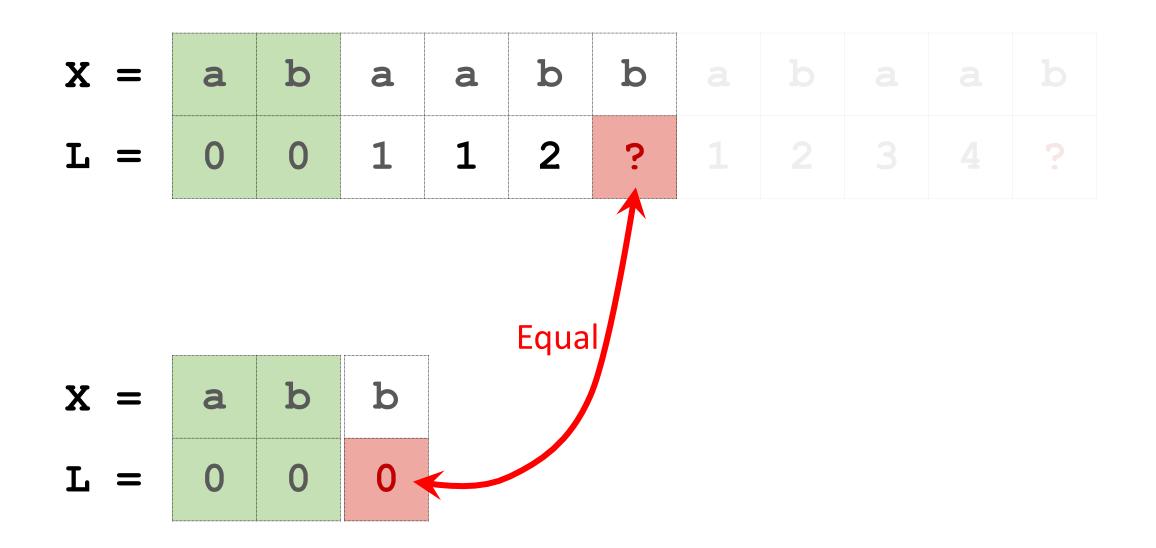
- This is Case 3.
- Reduce the problem to a smaller problem.

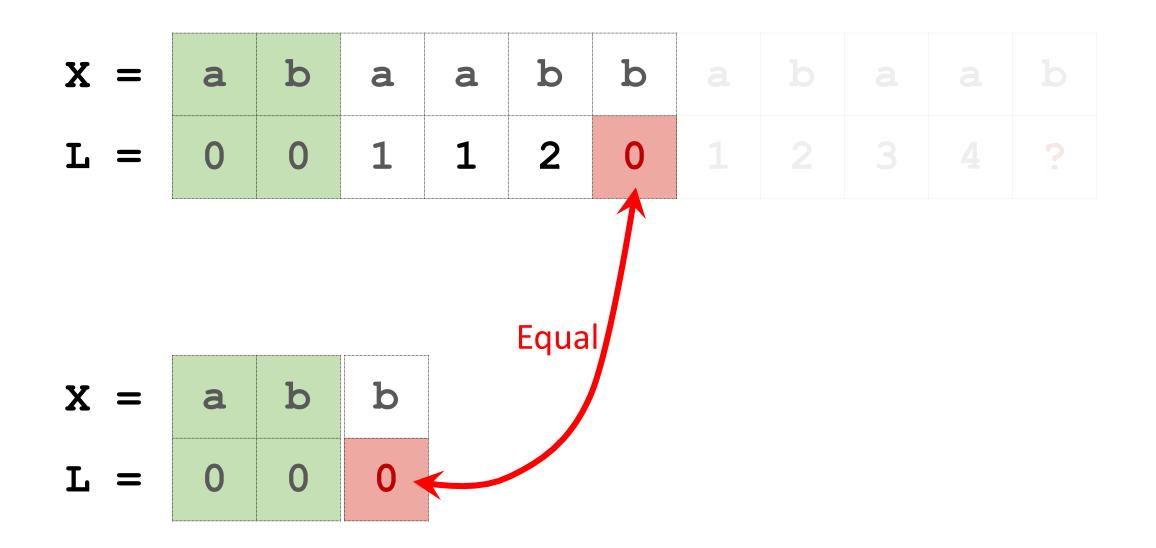


- This is Case 3.
- Reduce the problem to a smaller problem.

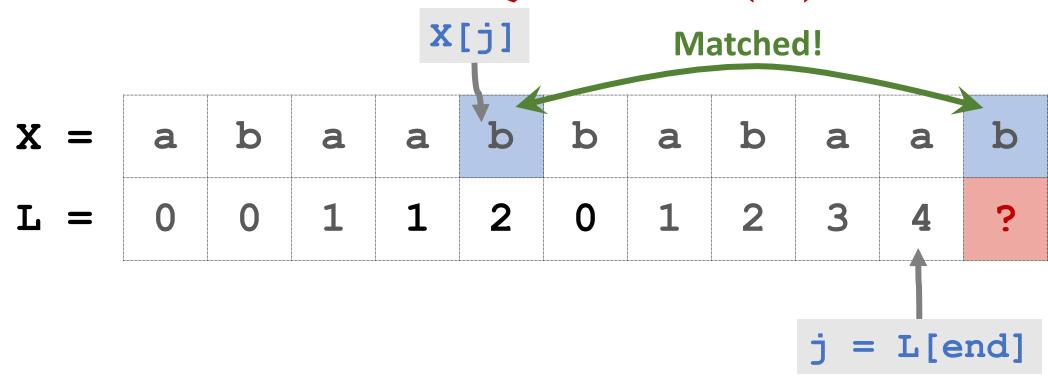
| x = | | | | | | | | | |
|-----|---|---|---|---|---|---|--|--|--|
| L = | 0 | 0 | 1 | 1 | 2 | ? | | | |







| x = | a | b | a | a | b | b | a | b | a | a | b |
|-----|---|---|---|---|---|---|---|---|---|---|---|
| L = | 0 | 0 | 1 | 1 | 2 | 0 | 1 | 2 | 3 | 4 | ? |



- Case 1: the new char is equal to X[j].
- Then the new value is j+1.

- Case 1: the new char is equal to X[j].
- Then the new value is j+1.