CSE 221: Algorithms

Quicksort

Mumit Khan
Fatema Tuz Zohora

Computer Science and Engineering
BRAC University

References

1. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms, Second Edition*. The MIT Press, September 2001.
2. Erik Demaine and Charles Leiserson, *6.046J Introduction to Algorithms*. MIT OpenCourseWare, Fall 2005. Available from: `ocw.mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/6-046JFall-2005/CourseHome/index.htm`

Last modified: January 29, 2013

Contents

Contents

Quicksort

- Proposed by C. A. R. Hoare in 1962.

Quicksort

- Proposed by C. A. R. Hoare in 1962.
- Divide and Conquer algorithm – like merge sort.

Quicksort

- Proposed by C. A. R. Hoare in 1962.
- Divide and Conquer algorithm – like merge sort.
- In-place algorithm – like insertion and heap sorts.

Quicksort

- Proposed by C. A. R. Hoare in 1962.
- Divide and Conquer algorithm – like merge sort.
- In-place algorithm – like insertion and heap sorts.
- Runs very well with tuning.

Quicksort

- Proposed by C. A. R. Hoare in 1962.
- Divide and Conquer algorithm – like merge sort.
- In-place algorithm – like insertion and heap sorts.
- Runs very well with tuning.
- Worst-case is $O(n^2)$, but for all practical purposes runs in $O(n \lg n)$.

Quicksort

- Proposed by C. A. R. Hoare in 1962.
- Divide and Conquer algorithm – like merge sort.
- In-place algorithm – like insertion and heap sorts.
- Runs very well with tuning.
- Worst-case is $O(n^2)$, but for all practical purposes runs in $O(n \lg n)$.

Why do we want to study Quicksort?

One of the most widely used, and extensively studied, sorting algorithms.

Divide and conquer

Quicksort an *n*-element array:

1. *Divide* Partition the array into subarrays around a *pivot* x such that the elements in lower subarray $\leq x \leq$ elements in the upper subarray.

2. *Conquer* Recursively sort the two subarrays.
3. *Combine* Trivial – just concatenate the lower subarray, *pivot*, and the upper subarray.

Divide and conquer

Quicksort an *n*-element array:

1. *Divide* Partition the array into subarrays around a *pivot* x such that the elements in lower subarray $\leq x \leq$ elements in the upper subarray.

2. *Conquer* Recursively sort the two subarrays.
3. *Combine* Trivial – just concatenate the lower subarray, *pivot*, and the upper subarray.

Divide and conquer

Quicksort an n-element array:

1. *Divide* Partition the array into subarrays around a *pivot* x such that the elements in lower subarray $\leq x \leq$ elements in the upper subarray.

2. *Conquer* Recursively sort the two subarrays.
3. *Combine* Trivial – just concatenate the lower subarray, *pivot*, and the upper subarray.

Divide and conquer

Quicksort an *n*-element array:

1. *Divide* Partition the array into subarrays around a *pivot* x such that the elements in lower subarray $\leq x \leq$ elements in the upper subarray.

2. *Conquer* Recursively sort the two subarrays.
3. *Combine* Trivial – just concatenate the lower subarray, *pivot*, and the upper subarray.

Divide and conquer

Quicksort an *n*-element array:

1. *Divide* Partition the array into subarrays around a *pivot* x such that the elements in lower subarray $\leq x \leq$ elements in the upper subarray.

2. *Conquer* Recursively sort the two subarrays.
3. *Combine* Trivial – just concatenate the lower subarray, *pivot*, and the upper subarray.

Key

Linear-time partitioning algorithm.

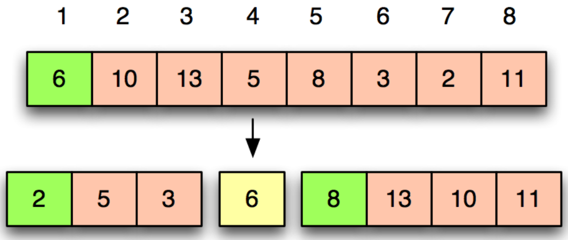
Quicksort in action

1	2	3	4	5	6	7	8
6	10	13	5	8	3	2	11

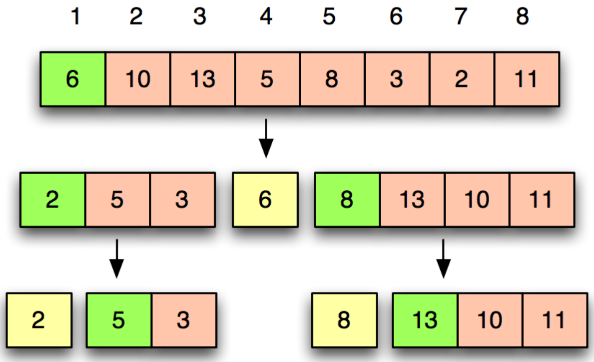
Quicksort in action

1	2	3	4	5	6	7	8
6	10	13	5	8	3	2	11

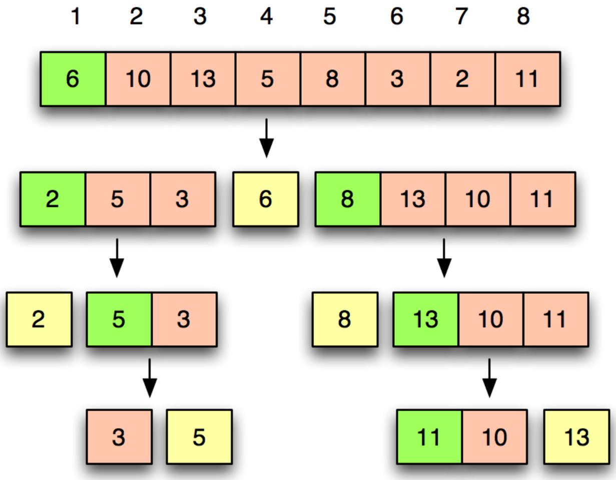
Quicksort in action



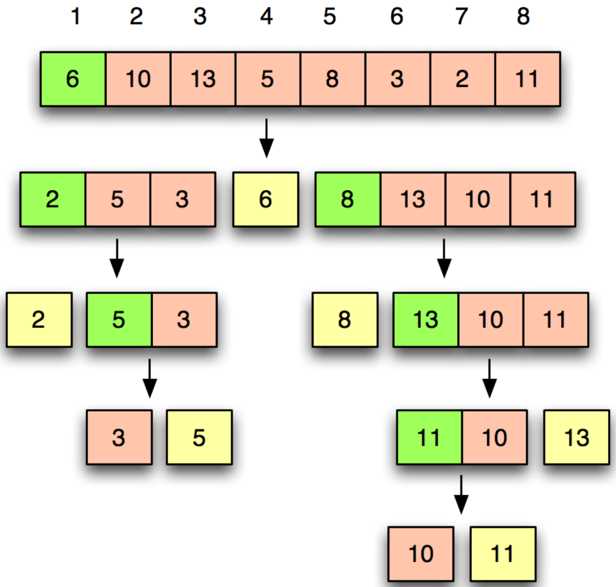
Quicksort in action



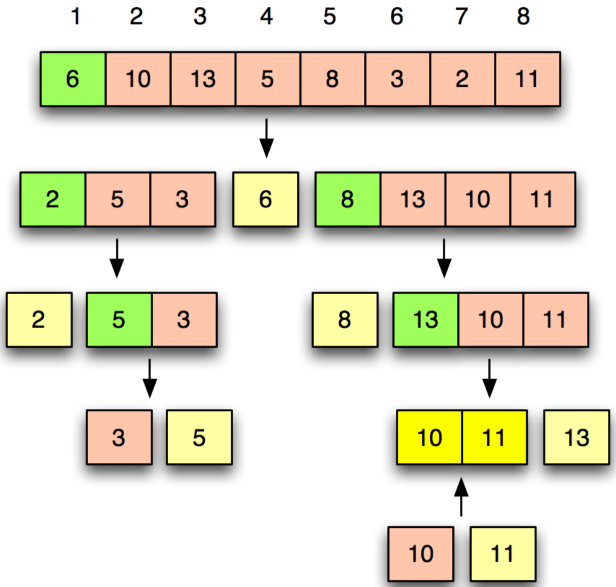
Quicksort in action



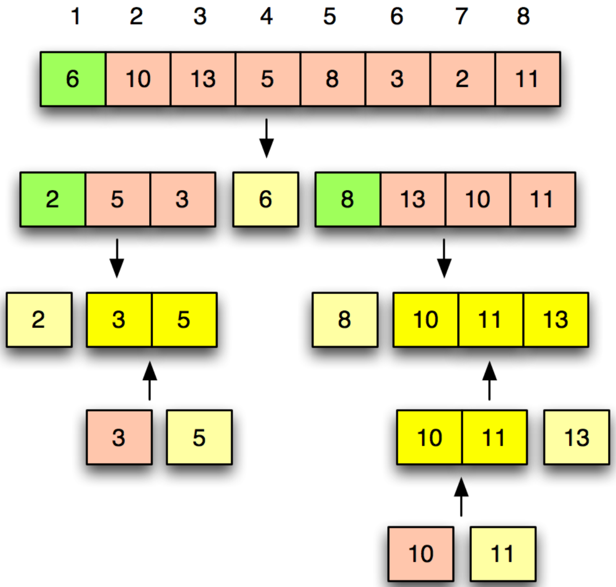
Quicksort in action



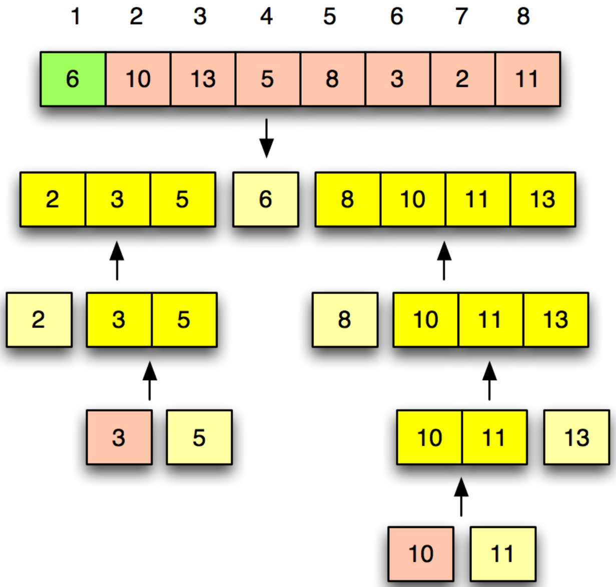
Quicksort in action



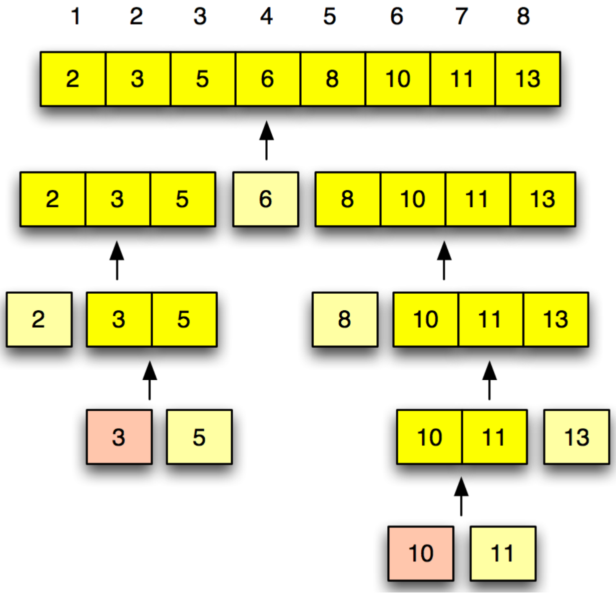
Quicksort in action



Quicksort in action



Quicksort in action



Contents

Partitioning algorithm

Algorithm

PARTITION(A, p, q) $\triangleright$ $A[p \,..\, q]$

$x \leftarrow A[p]$ $\triangleright$ pivot $= A[p]$

$i \leftarrow p$

for $j \leftarrow p + 1$ **to** q

  do if $A[j] \leq x$

    then $i \leftarrow i + 1$

      exchange $A[i] \leftrightarrow A[j]$

exchange $A[p] \leftrightarrow A[i]$

return i

Partitioning algorithm

Algorithm

$\textsc{Partition}(A, p, q) \triangleright A[p \,..\, q]$

$x \leftarrow A[p] \qquad \triangleright$ pivot $= A[p]$

$i \leftarrow p$

for $j \leftarrow p + 1$ **to** q

    do if $A[j] \leq x$

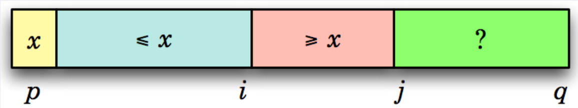
        then $i \leftarrow i + 1$

            exchange $A[i] \leftrightarrow A[j]$

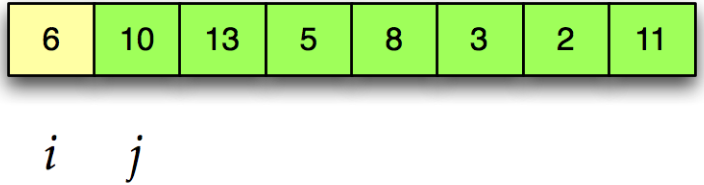
exchange $A[p] \leftrightarrow A[i]$

return i

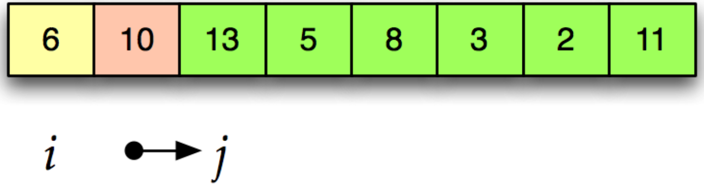
Invariant



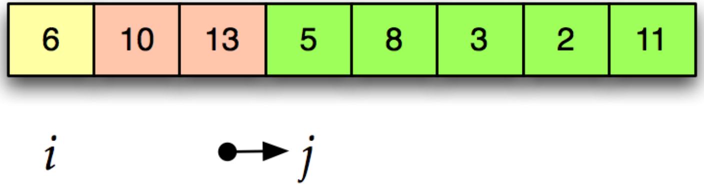
Partitioning in action



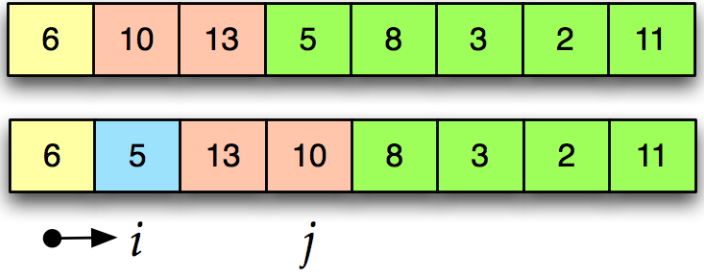
Partitioning in action



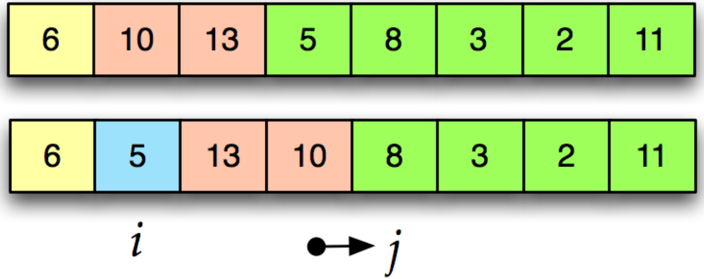
Partitioning in action



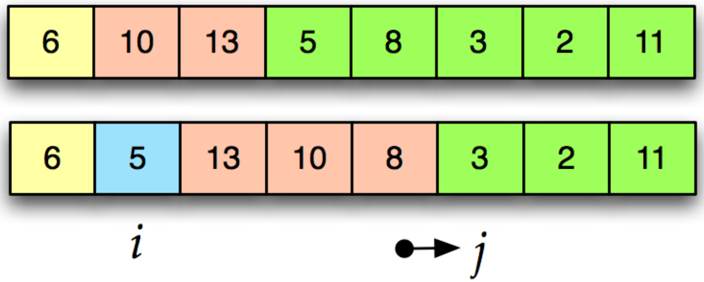
Partitioning in action



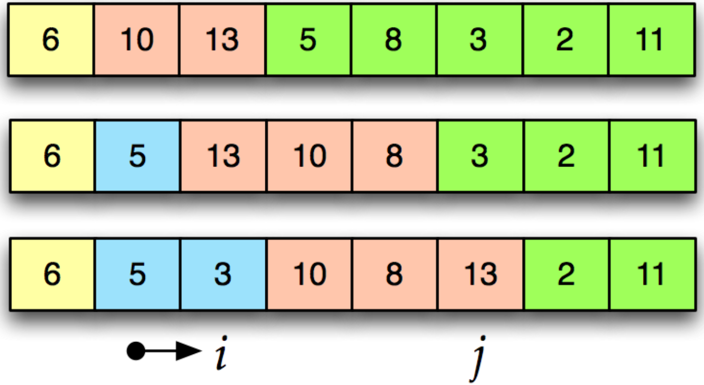
Partitioning in action



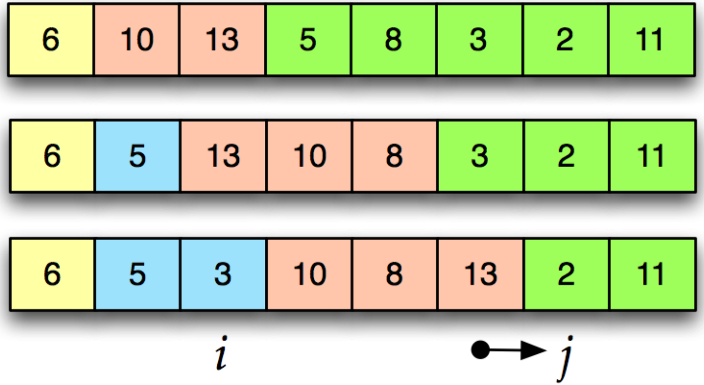
Partitioning in action



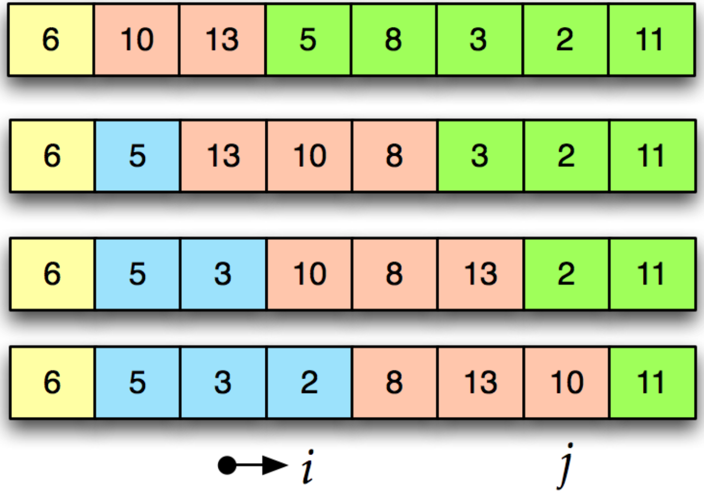
Partitioning in action



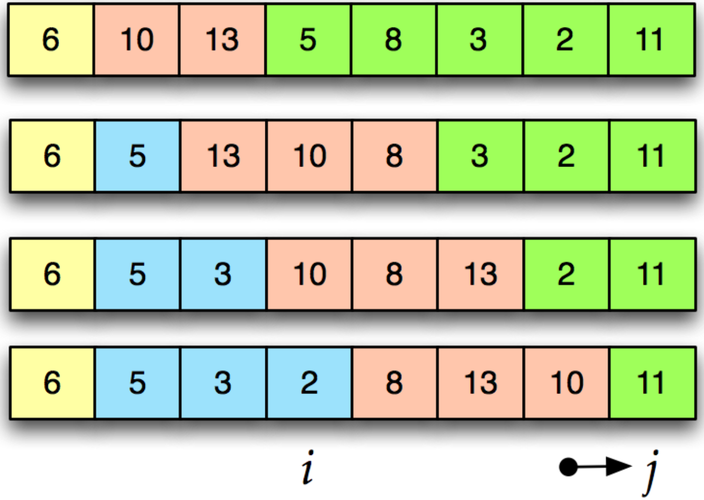
Partitioning in action



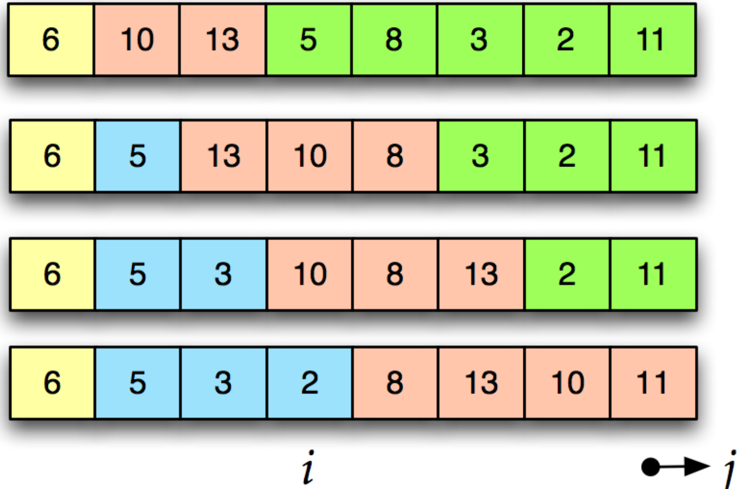
Partitioning in action



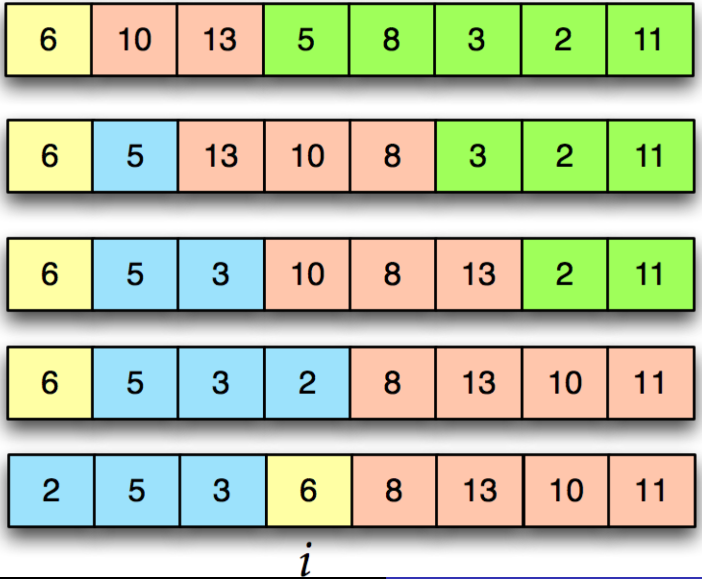
Partitioning in action



Partitioning in action



Partitioning in action



Contents

Quicksort algorithm

Algorithm

$\textsc{Quicksort}(A, p, r) \triangleright A[p \,..\, r]$

1 **if** $p < r$

2 **then** $q \leftarrow \textsc{Partition}(A, p, r)$

3 $\textsc{Quicksort}(A, p, q - 1)$

4 $\textsc{Quicksort}(A, q + 1, r)$

Quicksort algorithm

Algorithm

$\textsc{Quicksort}(A, p, r) \triangleright A[p \mathinner{..} r]$

if $p < r$
  then $q \leftarrow \textsc{Partition}(A, p, r)$
    $\textsc{Quicksort}(A, p, q-1)$
    $\textsc{Quicksort}(A, q+1, r)$

Initial call

$\textsc{Quicksort}(A, 1, n)$

Contents

Analyzing Quicksort - worst-case performance

- Worst-case happens when pivot is always the minimum or maximum element.

Analyzing Quicksort - worst-case performance

- Worst-case happens when pivot is always the minimum or maximum element.
- Result is that one of the partitions is always empty.

Analyzing Quicksort - worst-case performance

- Worst-case happens when pivot is always the minimum or maximum element.
- Result is that one of the partitions is always empty.
- When?

Analyzing Quicksort - worst-case performance

- Worst-case happens when pivot is always the minimum or maximum element.
- Result is that one of the partitions is always empty.
- When? Input sorted (either non-decreasing or non-increasing)

Analyzing Quicksort - worst-case performance

- Worst-case happens when pivot is always the minimum or maximum element.
- Result is that one of the partitions is always empty.
- When? Input sorted (either non-decreasing or non-increasing)

Worst-case analysis

(Note: the worst-case running time for partitioning is $\Theta(n)$.)

$$\begin{aligned} T(n) &= T(0) + T(n-1) + \Theta(n) \\ &= \Theta(1) + T(n-1) + \Theta(n) \\ &= T(n-1) + \Theta(n) \\ &= \boxed{\Theta(n^2)} \end{aligned}$$

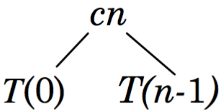
Worst-case recursion tree

$$T(n) = T(0) + T(n-1) + cn$$

$$T(n)$$

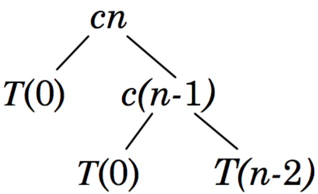
Worst-case recursion tree

$$T(n) = T(0) + T(n-1) + cn$$



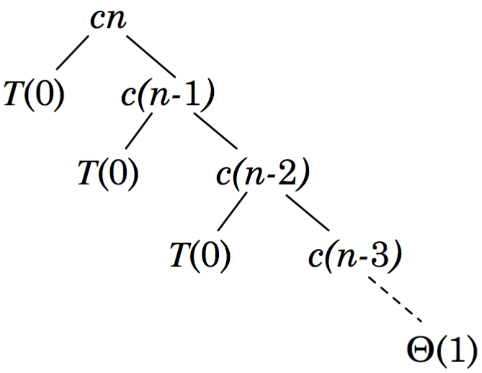
Worst-case recursion tree

$$T(n) = T(0) + T(n-1) + cn$$



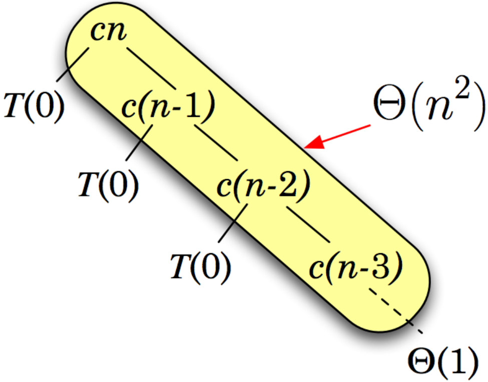
Worst-case recursion tree

$$T(n) = T(0) + T(n-1) + cn$$



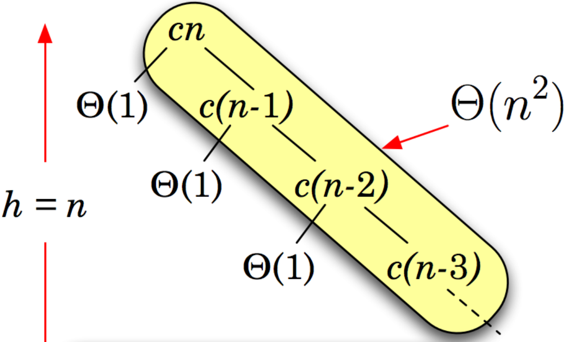
Worst-case recursion tree

$$T(n) = T(0) + T(n-1) + cn$$



Worst-case recursion tree

$$T(n) = T(0) + T(n-1) + cn$$



$$\begin{aligned} T(n) &= \Theta(n) + \Theta(n^2) \\ &= \Theta(n^2) \end{aligned}$$

Best- and almost-worst case performances

- Best-case happens when pivot is the median element, creating equal size partitions.

Best- and almost-worst case performances

- Best-case happens when pivot is the median element, creating equal size partitions.

$$T(n) = 2T(n/2) + \Theta(n) = \Theta(n \lg n)$$

Best- and almost-worst case performances

- Best-case happens when pivot is the median element, creating equal size partitions.

$$T(n) = 2T(n/2) + \Theta(n) = \Theta(n \lg n)$$

- What if the split is always $\frac{1}{10} : \frac{9}{10}$?

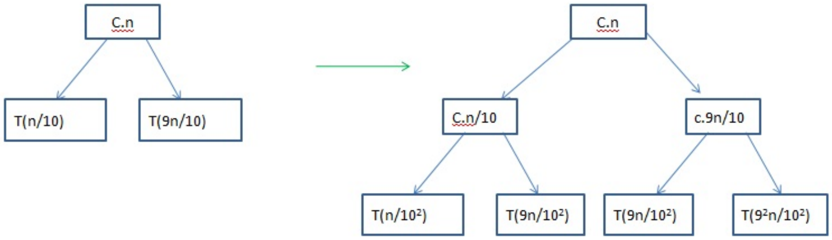
Best- and almost-worst case performances

- Best-case happens when pivot is the median element, creating equal size partitions.

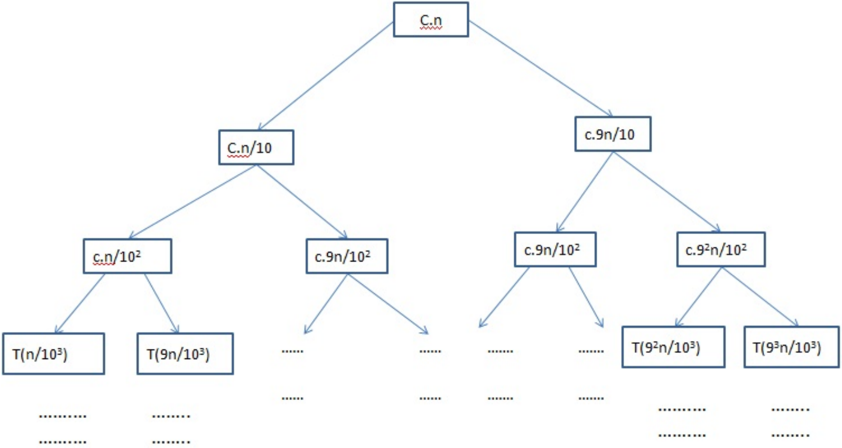
$$T(n) = 2T(n/2) + \Theta(n) = \Theta(n \lg n)$$

- What if the split is always $\frac{1}{10} : \frac{9}{10}$?
- $T(n) = T(n/10) + T(9n/10) + Cn$
- Lets draw the recursion tree for this recurrence formula.

Recursion Tree for split proportion 1/10: 9/10



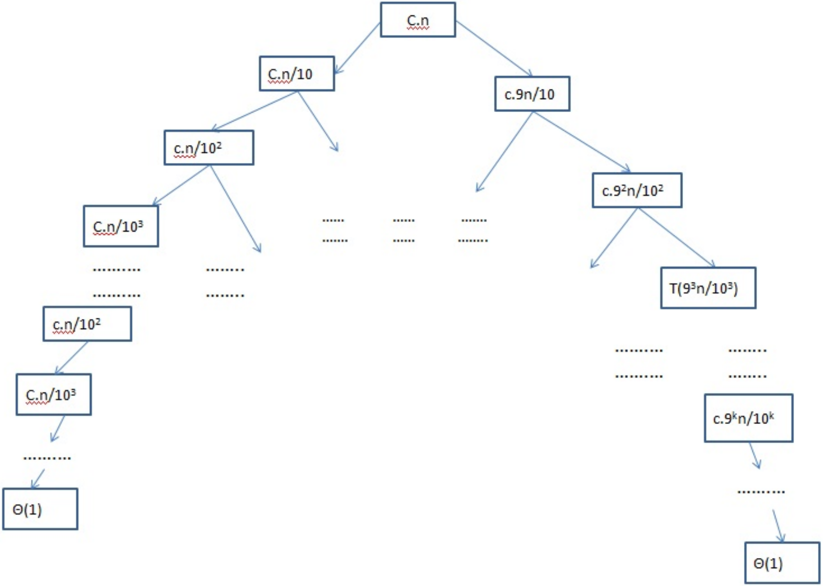
Recursion Tree for split proportion 1/10: 9/10



Recursion Tree for split proportion 1/10: 9/10

- Lets consider only left most branch and right most branch

Recursion Tree for split proportion 1/10: 9/10



Recursion Tree for split proportion 1/10: 9/10

- Lets consider only left most branch and right most branch

Recursion Tree for split proportion 1/10: 9/10

- Lets consider only left most branch and right most branch
- Note here, cost at each level is cutting down by 10 if we consider the left most branch (Cn is being divided by (10) at each level)

Recursion Tree for split proportion 1/10: 9/10

- Lets consider only left most branch and right most branch
- Note here, cost at each level is cutting down by 10 if we consider the left most branch (Cn is being divided by (10) at each level)
- Again, cost at each level is cutting down by 10/9 if we consider the right most branch (Cn is being divided by (9/10) at each level)

Recursion Tree for split proportion 1/10: 9/10

- Lets consider only left most branch and right most branch
- Note here, cost at each level is cutting down by 10 if we consider the left most branch (Cn is being divided by (10) at each level)
- Again, cost at each level is cutting down by 10/9 if we consider the right most branch (Cn is being divided by (9/10) at each level)
- If we cut down a value X by a factor of a then how long it takes me to reduce to one?

Recursion Tree for split proportion 1/10: 9/10

- Lets consider only left most branch and right most branch
- Note here, cost at each level is cutting down by 10 if we consider the left most branch (Cn is being divided by (10) at each level)
- Again, cost at each level is cutting down by 10/9 if we consider the right most branch (Cn is being divided by (9/10) at each level)
- If we cut down a value X by a factor of a then how long it takes me to reduce to one?
- answer: $\log_a X$

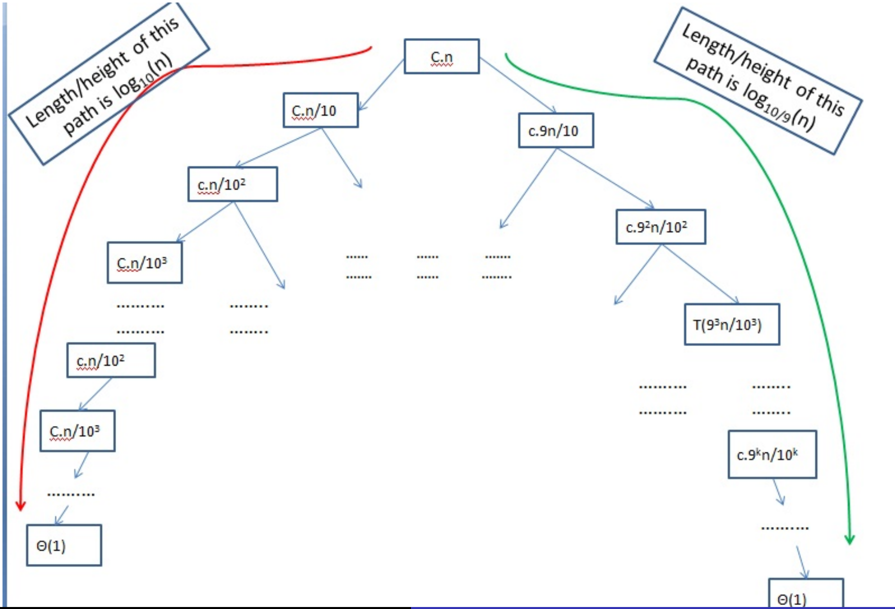
Recursion Tree for split proportion 1/10: 9/10

- Lets consider only left most branch and right most branch
- Note here, cost at each level is cutting down by 10 if we consider the left most branch (Cn is being divided by (10) at each level)
- Again, cost at each level is cutting down by 10/9 if we consider the right most branch (Cn is being divided by (9/10) at each level)
- If we cut down a value X by a factor of a then how long it takes me to reduce to one?
- answer: $\log_a X$
- Length/ height of left most branch is $\log_{10} n$

Recursion Tree for split proportion 1/10: 9/10

- Lets consider only left most branch and right most branch
- Note here, cost at each level is cutting down by 10 if we consider the left most branch (Cn is being divided by (10) at each level)
- Again, cost at each level is cutting down by 10/9 if we consider the right most branch (Cn is being divided by (9/10) at each level)
- If we cut down a value X by a factor of a then how long it takes me to reduce to one?
- answer: $\log_a X$
- Length/ height of left most branch is $\log_{10} n$
- Length/ height of right most branch is $\log_{10/9} n$

Recursion Tree for split proportion 1/10: 9/10



- $\log_{10} n \leq \log_{10/9} n$

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$
- Lets consider, all branches have the maximum height $\log_{10/9} n$

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$
- Lets consider, all branches have the maximum height $\log_{10/9} n$
- Then total cost will be $Cn \log_{10/9} n + Cn$ (total cost at each level is Cn, recall the calculation of total cost for quicksort having split proportion $1/2 : 1/2$)

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$
- Lets consider, all branches have the maximum height $\log_{10/9} n$
- Then total cost will be $Cn \log_{10/9} n + Cn$ (total cost at each level is Cn, recall the calculation of total cost for quicksort having split proportion $1/2 : 1/2$)
- But actually total cost would be less then this cost as all branches are not of maximum height

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$
- Lets consider, all branches have the maximum height $\log_{10/9} n$
- Then total cost will be $Cn \log_{10/9} n + Cn$ (total cost at each level is Cn, recall the calculation of total cost for quicksort having split proportion $1/2 : 1/2$)
- But actually total cost would be less then this cost as all branches are not of maximum height
- So, $T(n) \leq Cn \log_{10/9} n + Cn$

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$
- Lets consider, all branches have the maximum height $\log_{10/9} n$
- Then total cost will be $Cn \log_{10/9} n + Cn$ (total cost at each level is Cn, recall the calculation of total cost for quicksort having split proportion $1/2 : 1/2$)
- But actually total cost would be less then this cost as all branches are not of maximum height
- So, $T(n) \leq Cn \log_{10/9} n + Cn$
- Simplify it as $T(n) \leq Cn \log_2 n + Cn$

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$
- Lets consider, all branches have the maximum height $\log_{10/9} n$
- Then total cost will be $Cn \log_{10/9} n + Cn$ (total cost at each level is Cn, recall the calculation of total cost for quicksort having split proportion $1/2 : 1/2$)
- But actually total cost would be less then this cost as all branches are not of maximum height
- So, $T(n) \leq Cn \log_{10/9} n + Cn$
- Simplify it as $T(n) \leq Cn \log_2 n + Cn$
- $T(n) \leq Cn \lg n + Cn$

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$
- Lets consider, all branches have the maximum height $\log_{10/9} n$
- Then total cost will be $Cn \log_{10/9} n + Cn$ (total cost at each level is Cn, recall the calculation of total cost for quicksort having split proportion $1/2 : 1/2$)
- But actually total cost would be less then this cost as all branches are not of maximum height
- So, $T(n) \leq Cn \log_{10/9} n + Cn$
- Simplify it as $T(n) \leq Cn \log_2 n + Cn$
- $T(n) \leq Cn \lg n + Cn$
- $T(n) = O(n \lg n)$

- $\log_{10} n \leq \log_{10/9} n$
- Left most branch ends before all other branches, it is the path having smallest height
- Right most branch ends after all other branches, so it is the path having maximum height
- So length/height of other branches in between these two branches will be in between $\log_{10} n$ and $\log_{10} n$
- Lets consider, all branches have the maximum height $\log_{10/9} n$
- Then total cost will be $Cn \log_{10/9} n + Cn$ (total cost at each level is Cn, recall the calculation of total cost for quicksort having split proportion $1/2 : 1/2$)
- But actually total cost would be less then this cost as all branches are not of maximum height
- So, $T(n) \leq Cn \log_{10/9} n + Cn$
- Simplify it as $T(n) \leq Cn \log_2 n + Cn$
- $T(n) \leq Cn \lg n + Cn$
- $T(n) = O(n \lg n)$

Key observation

Very close to worst-case produces $O(n \lg n)$, not $O(n^2)$.

Key observation

Very close to worst-case produces $O(n \lg n)$, not $O(n^2)$.

How to ensure that we don't *usually* hit the worst-case?

Contents

Randomized Quicksort

- Pick a random pivot and partition around it.

Randomized Quicksort

- Pick a random pivot and partition around it.
- Pivot independent of input order, so no specific input produces worst-case behavior.

Randomized Quicksort

- Pick a random pivot and partition around it.
- Pivot independent of input order, so no specific input produces worst-case behavior.
- The worst-case is determined only by the output of a random number generator.

Randomized Quicksort

- Pick a random pivot and partition around it.
- Pivot independent of input order, so no specific input produces worst-case behavior.
- The worst-case is determined only by the output of a random number generator.

$\textsc{Randomized-Partition}(A, p, r) \triangleright A[p \mathinner{..} r]$

1 $i \leftarrow \textsc{Random}(p, r)$ $\triangleright i = [p \mathinner{..} r]$
2 exchange $A[p] \leftrightarrow A[i]$
3 **return** $\textsc{Partition}(A, p, r)$

Randomized Quicksort

- Pick a random pivot and partition around it.
- Pivot independent of input order, so no specific input produces worst-case behavior.
- The worst-case is determined only by the output of a random number generator.

$\textsc{Randomized-Partition}(A, p, r) \triangleright A[p \,..\, r]$

```
i ← RANDOM(p, r)                    ▷ i = [p .. r]
exchange A[p] ↔ A[i]
return PARTITION(A, p, r)
```

$\textsc{Randomized-Quicksort}(A, p, r)$

```
if p < r
   then q ← RANDOMIZED-PARTITION(A, p, r)
        RANDOMIZED-QUICKSORT(A, p, q − 1)
        RANDOMIZED-QUICKSORT(A, q + 1, r)
```

Conclusion

- One of the most widely used sorting algorithm.
- While it runs in $O(n^2)$ time in the worst-case, it runs in $O(n \lg n)$ time on the average.
- Runs almost twice as fast as merge-sort.
- Can be *tuned* substantially.
- Many programming language runtime libraries provide some variant of Quicksort (java.util.Arrays.sort() in Java, qsort() in C, etc).

Conclusion

- One of the most widely used sorting algorithm.
- While it runs in $O(n^2)$ time in the worst-case, it runs in $O(n \lg n)$ time on the average.
- Runs almost twice as fast as merge-sort.
- Can be *tuned* substantially.
- Many programming language runtime libraries provide some variant of Quicksort (java.util.Arrays.sort() in Java, qsort() in C, etc).

Questions to ask (and remember)

Conclusion

- One of the most widely used sorting algorithm.
- While it runs in $O(n^2)$ time in the worst-case, it runs in $O(n \lg n)$ time on the average.
- Runs almost twice as fast as merge-sort.
- Can be *tuned* substantially.
- Many programming language runtime libraries provide some variant of Quicksort (java.util.Arrays.sort() in Java, qsort() in C, etc).

Questions to ask (and remember)

- What are the worst, best and average case performances?

Conclusion

- One of the most widely used sorting algorithm.
- While it runs in $O(n^2)$ time in the worst-case, it runs in $O(n \lg n)$ time on the average.
- Runs almost twice as fast as merge-sort.
- Can be *tuned* substantially.
- Many programming language runtime libraries provide some variant of Quicksort (java.util.Arrays.sort() in Java, qsort() in C, etc).

Questions to ask (and remember)

- What are the worst, best and average case performances?
- Is it in-place?

Conclusion

- One of the most widely used sorting algorithm.
- While it runs in $O(n^2)$ time in the worst-case, it runs in $O(n \lg n)$ time on the average.
- Runs almost twice as fast as merge-sort.
- Can be *tuned* substantially.
- Many programming language runtime libraries provide some variant of Quicksort (java.util.Arrays.sort() in Java, qsort() in C, etc).

Questions to ask (and remember)

- What are the worst, best and average case performances?
- Is it in-place?
- Is it stable?