Quicksort

Quicksort uses a divide and conquer strategy, but does not require the $O(N)$ extra space that MergeSort does.

Here’s the idea for sorting array $S$:
1. Pick an element $v$ in $S$. This is the *pivot* value.
2. Partition $S\setminus\{v\}$ into two disjoint subsets, $S_1$ and $S_2$, such that:
   - elements in $S_1$ are all $\leq v$
   - elements in $S_2$ are all $\geq v$
3. Return concatenation of QuickSort($S_1$), $v$, QuickSort($S_2$)

Recursion ends when QuickSort($S$) receives an array of length 0 or 1.

The steps of Quicksort

QuickSort Example
Pivot Picking and Partitioning

The tricky pieces are:

- **Picking the pivot**
  - Goal: pick a pivot value that will cause $|S_1|$ and $|S_2|$ to be roughly equal in size.
- **Partitioning**
  - Preferably in-place
  - Dealing with duplicates.

Picking the Pivot

- random number generation can be costly
- middle
  - not too bad
- median
  - too expensive
- average
  - worst
  - median of three $A[\text{left} ], A[\text{right}], A[(\text{left} + \text{right})/2]$

Median of Three Pivot

```
0 1 2 3 4 5 6 7 8 9
8 1 4 9 6 3 5 2 7 0
```

```
medianOf3Pivot(...) 
```

```
0 1 4 9 7 3 5 2 6 8
```

Choose the pivot as the median of three.

Place the pivot and the largest at the right and the smallest at the left.

Quicksort Partitioning

- Need to partition the array into left and right sub-arrays such that:
  - elements in left sub-array are $\leq$ pivot
  - elements in right sub-array are $\geq$ pivot
- Can be done in-place with another “two pointer method”
  - Sounds like mergesort, but here we are *partitioning*, not sorting...
  - ... and we can do it in-place.
Partitioning In-place

Setup: \( i = \text{start and } j = \text{end of un-partitioned elements:}\)

\[
\begin{array}{cccccccc}
0 & 1 & 4 & 9 & 7 & 3 & 5 & 2 & 6 & 8 \\
\end{array}
\]

Advance \( i \) until element \( \geq \) pivot:

\[
\begin{array}{cccccccc}
0 & 1 & 4 & 9 & 7 & 3 & 5 & 2 & 6 & 8 \\
\end{array}
\]

Advance \( j \) until element \( \leq \) pivot:

\[
\begin{array}{cccccccc}
0 & 1 & 4 & 9 & 7 & 3 & 5 & 2 & 6 & 8 \\
\end{array}
\]

If \( j > i \), then swap:

\[
\begin{array}{cccccccc}
0 & 1 & 4 & 2 & 7 & 3 & 5 & 9 & 6 & 8 \\
\end{array}
\]

Partition Pseudocode

\[
\text{Partition}(A[], \text{left, right}) \{
\begin{align*}
& v = A[\text{right}]: \quad \text{// Assumes pivot value currently at right} \\
& i = \text{left}: \quad \text{// Initialize left side, right side pointers} \\
& j = \text{right}-1; \\
& \text{// Do } i++ \text{, } j-- \text{ until they cross, swapping values as needed} \\
& \quad \text{while } (i < j) \{ \\
& \quad \quad \text{while } (A[i] < v) \; i++; \\
& \quad \quad \text{while } (A[j] > v) \; j--; \\
& \quad \quad \text{if } (i < j) \{ \\
& \quad \quad \quad \text{Swap}(A[i], A[j]); \\
& \quad \quad \quad i++; \; j--; \\
& \quad \quad \} \\
& \quad \text{else} \\
& \quad \quad \text{break;} \\
& \}
\}
\]

QuickSort Pseudocode

Putting the pieces together:

\[
\text{QuickSort}(A[], \text{left, right}) \{
\begin{align*}
& \text{if } (\text{left} < \text{right}) \{ \\
& \quad \text{medianOf3Pivot}(A, \text{left, right}); \\
& \quad \text{pivotIndex} = \text{Partition}(A, \text{left+1, right-1}); \\
& \quad \text{QuickSort}(A, \text{left}, \text{pivotIndex} - 1); \\
& \quad \text{QuickSort}(A, \text{pivotIndex} + 1, \text{right}); \\
& \}
\}
\]

Complexity for input size \( n \)? \( O(n) \)
QuickSort:
Best case complexity

QuickSort(A[], left, right) {
    if (left < right) {
        medianOf3Pivot(A, left, right);
        pivotIndex = Partition(A, left+1, right-1);
        QuickSort(A, left, pivotIndex - 1);
        QuickSort(A, pivotIndex + 1, right);
    }
    T(1) = 1
    T(n) = 2T(\lceil n/2 \rceil) + n
         = O(n \log n)
    \Rightarrow \text{best case } \Omega(n \log n)
}

QuickSort:
Worst case complexity

QuickSort(A[], left, right) {
    if (left < right) {
        medianOf3Pivot(A, left, right);
        pivotIndex = Partition(A, left+1, right-1);
        QuickSort(A, left, pivotIndex - 1);
        QuickSort(A, pivotIndex + 1, right);
    }
    T(1) = 1
    T(n) = n + T(n-1)
         = n + \sum_{i=0}^{n-2} \frac{n-i}{2}
         = \frac{n}{2} + 2 + \cdots + n
         = \Theta(n^2)

QuickSort:
Average case complexity

Turns out to be $O(n \log n)$.

See Section 7.7.5 for an idea of the proof. 
*Don’t need to know proof details for this course.*

Many Duplicates?

An important case to consider is when an array has many duplicates.

```
0 1 2 3 4 5 6 7 8 9
8 6 6 6 6 6 6 6 6 0
```

medianOf3Pivot(…)

```
0 6 6 6 6 6 6 6 6 8
```
Partitioning with Duplicates

Setup: \( i = \text{start and } j = \text{end of un-partitioned elements:} \)

\[
\begin{array}{ccccccccccc}
0 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 8 \\
\end{array}
\]

Advance \( i \) until element \( \geq \) pivot:

\[
\begin{array}{ccccccccccc}
0 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 8 & 8 \\
\end{array}
\]

Advance \( j \) until element \( \leq \) pivot:

\[
\begin{array}{ccccccccccc}
0 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 8 & 8 \\
\end{array}
\]

If \( j > i \), then swap:

\[
\begin{array}{ccccccccccc}
0 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 8 & 8 & 8 \\
\end{array}
\]

Start \( i = \text{start and } j = \text{end of un-partitioned elements:} \)

\[
\begin{array}{ccccccccccc}
0 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 8 & 8 \\
\end{array}
\]

Advance \( i \) until element \( > \) pivot (and in bounds):

\[
\begin{array}{ccccccccccc}
0 & 6 & 6 & 6 & 6 & 6 & 6 & 8 & 6 & 6 & 8 \\
\end{array}
\]

Advance \( j \) until element \( < \) pivot (and in bounds):

\[
\begin{array}{ccccccccccc}
0 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 8 & 8 \\
\end{array}
\]

Finish:

\[
\begin{array}{ccccccccccc}
0 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 6 & 8 & 8 \\
\end{array}
\]

Partitioning with Duplicates: Take Two

Partitioning with Duplicates: Upshot

It’s better to stop advancing pointers when elements are equal to pivot, and then just do swaps.

Complexity of quicksort on an array of identical values? \( O(n \log n) \)

Can we do better?
Important Tweak

Insertion sort is actually better than quicksort on small arrays. Thus, a better version of quicksort:

```java
QuickSort(A[], left, right) {
    if (right - left ≥ CUTOFF) {
        medianOf3Pivot(A, left, right);
        pivotIndex = Partition(A, left+1, right-1);
        QuickSort(A, left, pivotIndex - 1);
        QuickSort(A, pivotIndex + 1, right);
    } else {
        InsertionSort(A, left, right);
    }
}
```

CUTOFF = 10 is reasonable.

Properties of Quicksort

- $O(N^2)$ worst case performance, but $O(N \log N)$ average case performance.
- Pure quicksort not good for small arrays.
- No iterative version (without using a stack).
- “In-place,” but uses auxiliary storage because of recursive calls.
- Stable? No
- Used by Java for sorting arrays of primitive types.

How fast can we sort?

Heapsort, Mergesort, and Binary Tree Sort all have $O(N \log N)$ worst case running time.

These algorithms, along with Quicksort, also have $O(N \log N)$ average case running time.

Can we do any better?

Permutations

- Suppose you are given $N$ elements
  - Assume no duplicates
- How many possible orderings can you get?
  - Example: a, b, c ($N = 3$)
    - $N \cdot (N-1) \cdot (N-2) \cdot \ldots \cdot 1 = N!$
Permutations

- How many possible orderings can you get?
  - Example: a, b, c \((N = 3)\)
  - \((a \ b \ c), (a \ c \ b), (b \ a \ c), (b \ c \ a), (c \ a \ b), (c \ b \ a)\)
  - 6 orderings = 3\cdot2\cdot1 = 3! \ (i.e., \text{“}3 \text{ factorial}”) 
  - All the possible permutations of a set of 3 elements
- For \(N\) elements
  - \(N\) choices for the first position, \((N-1)\) choices for the second position, ..., \(2\) choices, \(1\) choice
  - \(N(N-1)(N-2)\cdots(2)(1) = N! \text{ possible orderings}\)

Sorting Model

Recall our basic sorting assumption:

We can only compare two elements at a time.

These comparisons prune the space of possible orderings.

We can represent these concepts in a...

Decision Tree

A Decision Tree is a Binary Tree such that:
- Each node = a set of orderings
  - i.e., the remaining solution space
- Each edge = 1 comparison
- Each leaf = 1 unique ordering
- How many leaves for \(N\) distinct elements? \(N!\)

- Only 1 leaf has the ordering that is the desired correctly sorted arrangement

The leaves contain all the possible orderings of a, b, c.
Decision Tree Example

Decision Trees and Sorting

- Every sorting algorithm corresponds to a decision tree
  - Finds correct leaf by choosing edges to follow
    - ie, by making comparisons
- We will focus on worst case run time.

Observations:
- Worst case run time is $\geq$ maximum number of comparisons.
- Maximum number of comparisons is the length of the longest path in the decision tree, i.e. the height of the tree.

How many leaves on a tree?

Suppose you have a binary tree of height $h$. How many leaves in a perfect tree?

$2^h$

We can prune a perfect tree to make any binary tree of same height. Can # of leaves increase?

Lower bound on Height

- A binary tree of height $h$ has at most $2^h$ leaves
  - Can prove formally by induction
- A decision tree has $N!$ leaves. What is its minimum height of that tree?

$$N! = \# \text{leaves} \leq 2^h$$

$$N! \leq 2^h$$

$$\Rightarrow h \geq \log_2(N!)$$
An Alternative Explanation

At each decision point, one branch has $\leq \frac{1}{2}$ of the options remaining, the other has $\geq \frac{1}{2}$ remaining.

Worst case: we always end up with $\geq \frac{1}{2}$ remaining.

Best algorithm, in the worst case: we always end up with exactly $\frac{1}{2}$ remaining.

Thus, in the worst case, the best we can hope for is halving the space $d$ times (with $d$ comparisons), until we have an answer, i.e., until the space is reduced to size = 1.

The space starts at $N!$ in size, and halving $d$ times means multiplying by $1/2^d$, giving us a lower bound on the worst case:

$$\frac{N!}{2^d} = 1 \Rightarrow N! = 2^d \Rightarrow d = \log_2(N!)$$

Lower Bound on $\log(N!)$

$$\log(N!) = \log(N \cdot (N-1) \cdot (N-2) \cdot \ldots \cdot 2 \cdot 1)$$

$$= \log N + \log(N-1) + \log(N-2) + \ldots$$

$$= \sum_{i=2}^{N} \log i$$

$$= \log \frac{N!}{2^{\frac{N}{2}+1}}$$

$$\geq \frac{N}{2} \log \frac{N}{2} = \frac{N}{2}(\log N - \log 2)$$

$$\geq \frac{N}{2} \log N - \frac{N}{2}$$

$$\Omega(N \log N)$$

Worst case run time of any comparison-based sorting algorithm is $\Omega(N \log N)$.

Can also show that average case run time is also $\Omega(N \log N)$.

Can we do better if we don’t use comparisons? (Huh?)