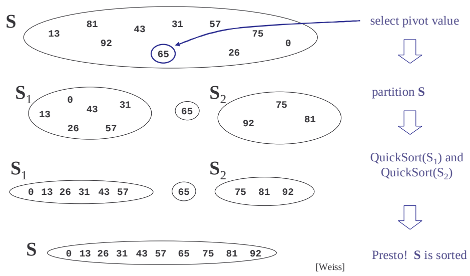
Sorting: *The Big Picture*

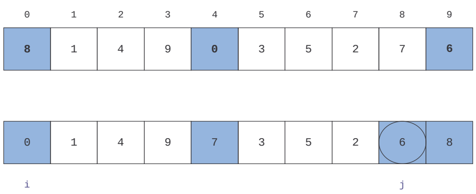
Given *n* comparable elements in an array, sort them in an increasing (or decreasing) order.

Simple algorithms: $O(n^2)$	Fancier algorithms: $O(n \log n)$	Comparison lower bound: $\Omega(n \log n)$	Specialized algorithms: $O(n)$	Handling huge data sets
Insertion sort Selection sort Bubble sort Shell sort …	Heap sort Merge sort Quick sort …		Bucket sort Radix sort	External sorting

The steps of QuickSort

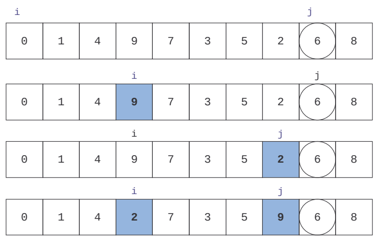


QuickSort Example



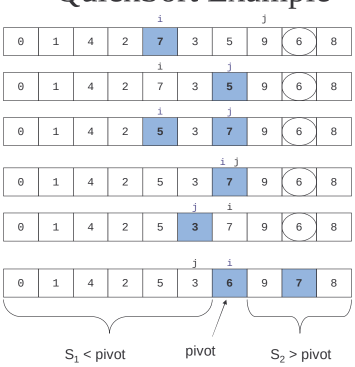
- Choose the pivot as the median of three.
- Place the pivot and the largest at the right and the smallest at the left

QuickSort Example



- Move i to the right to be larger than pivot.
- Move j to the left to be smaller than pivot.
- Swap

QuickSort Example



Recursive Quicksort

```
Quicksort(A[]: integer array, left,right : integer): {
pivotindex : integer;
if left + CUTOFF ≤ right then
  pivot := median3(A,left,right);
  pivotindex := Partition(A,left,right-1,pivot);
  Quicksort(A, left, pivotindex – 1);
  Quicksort(A, pivotindex + 1, right);
else
  Insertionsort(A,left,right);
}
```

Don't use quicksort for small arrays.
CUTOFF = 10 is reasonable.

QuickSort:
Best case complexity

7

QuickSort:
Worst case complexity

8

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.

9

Features of Sorting Algorithms

- In-place
  - Sorted items occupy the same space as the original items. (No copying required, only O(1) extra space if any.)
- Stable
  - Items in input with the same value end up in the same order as when they began.

10

Sort Properties

Are the following:	**stable?**		**in-place?**	
Insertion Sort?	No	Yes	Can Be No	Yes
Selection Sort?	No	Yes	Can Be No	Yes
MergeSort?	No	Yes	Can Be No	Yes
QuickSort?	No	Yes	Can Be No	Yes

11

How fast can we sort?

- Heapsort, Mergesort, and Quicksort all run in O(N log N) best case running time
- Can we do any better?
- No, if the basic action is a comparison.

12

Sorting Model

- Recall our basic assumption: we can <u>only compare two elements at a time</u>
  - we can only reduce the possible solution space by half each time we make a comparison
- Suppose you are given N elements
  - Assume no duplicates
- How many possible orderings can you get?
  - This is the number of potential inputs the algorithm must separate

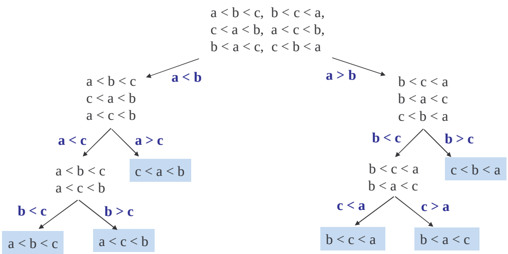
13

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 = 3·2·1 = 3!
  - 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)□(2)(1)= <u>N! possible orderings</u>

14

Decision Tree



The leaves contain all the possible orderings of a, b, c

15

Lower bound on Height

- A binary tree of height h has **at most** *how many* leaves?

L

- A binary tree with L leaves has height **at least**:

h

- The decision tree has how many leaves:
- So the decision tree has height:

h

16

log(*N*!) is Ω(*N*log*N*)

$$\log(N!) = \log\big(N\cdot(N-1)\cdot(N-2)\cdots(2)\cdot(1)\big)$$
$$= \log N + \log(N-1) + \log(N-2) + \cdots + \log 2 + \log 1$$

select just the first N/2 terms

$$\geq \log N + \log(N-1) + \log(N-2) + \cdots + \log\frac{N}{2}$$

each of the selected terms is ≥ logN/2

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

17

Ω(N log N)

- Run time of any comparison-based sorting algorithm is **Ω(N log N)**
- Can we do better if we don't use comparisons?

18

BucketSort (aka BinSort)

If all values to be sorted are *known* to be between 1 and *K*, create an array `count` of size *K*, **increment** counts while traversing the input, and finally output the result.

Example *K*=5. Input = (5,1,3,4,3,2,1,1,5,4,5)

`count` array	
1	
2	
3	
4	
5	

Running time to sort n items?

19

BucketSort Complexity: O(*n*+*K*)

- Case 1: *K* is a constant
  – BinSort is linear time
- Case 2: *K* is variable
  – Not simply linear time
- Case 3: *K* is constant but large (e.g. 2^{32})
  – ???

20

Fixing impracticality: RadixSort

- Radix = "The base of a number system"
  – We'll use 10 for convenience, but could be anything

- Idea: BucketSort on each **digit**,
  least significant to most significant
  (lsd to msd)

21

Radix Sort Example (1st pass)

Input data: 478, 537, 9, 721, 3, 38, 123, 67

Bucket sort by 1's digit

0	1	2	3	4	5	6	7	8	9
	721		3 123				537 67	478 38	9

After 1st pass: 721, 3, 123, 537, 67, 478, 38, 9

This example uses B=10 and base 10 digits for simplicity of demonstration. Larger bucket counts should be used in an actual implementation.

22

Radix Sort Example (2nd pass)

After 1st pass: 721, 3, 123, 537, 67, 478, 38, 9

Bucket sort by 10's digit

0	1	2	3	4	5	6	7	8	9
03 09		721 123	537 38			67	478		

After 2nd pass: 3, 9, 721, 123, 537, 38, 67, 478

23

Radix Sort Example (3rd pass)

After 2nd pass: 3, 9, 721, 123, 537, 38, 67, 478

Bucket sort by 100's digit

0	1	2	3	4	5	6	7	8	9
003 009 038 067	123			478	537		721		

After 3rd pass: 3, 9, 38, 67, 123, 478, 537, 721

Invariant: after k passes the low order k digits are sorted.

24

Radixsort: Complexity

- How many passes?
- How much work per pass?
- Total time?
- Conclusion?
- In practice
  - RadixSort only good for large number of elements with relatively small values
  - Hard on the cache compared to MergeSort/QuickSort

25

Internal versus External Sorting

- Need sorting algorithms that minimize disk/tape access time
- **External sorting –** Basic Idea:
  - Load chunk of data into RAM, sort, store this “run” on disk/tape
  - Use the Merge routine from Mergesort to merge runs
  - Repeat until you have only one run (one sorted chunk)
  - Text gives some examples

26