CS 580: Algorithm Design and Analysis

Deterministic Selection Algorithms

Goal: Find k-th smallest element of an array of n elements.

Deterministic Selection

Note: This was thought to be impossible for a long time (to do efficiently), until a 1973 paper, by Blum, Floyd, Pratt, Rivest and Tarjan proposed the "median-of-medians" algorithm.

Time Bounds for Selection*

MANUEL BLUM, ROBERT W. FLOYD, VAUGHAN PRATT,
RONALD L. RIVEST, AND ROBERT E. TARJAN

Department of Computer Science, Stanford University, Stanford, California 94305

Received November 14, 1972

The number of comparisons required to select the i-th smallest of n numbers is shown to be at most a linear function of n by analysis of a new selection algorithm—PICK. Specifically, no more than $5.430\dot{5}\, n$ comparisons are ever required. This bound is improved for extreme values of i, and a new lower bound on the requisite number of comparisons is also proved.

Deterministic Selection

1. Divide the elements into **groups of five**, where the last group may have less than five elements in case when the input array size is not a multiple of five.

2. Compute the **median** of each group (ties can be broken arbitrarily).

3. Make a recursive call to calculate the **median of the medians**. Set x to the median.

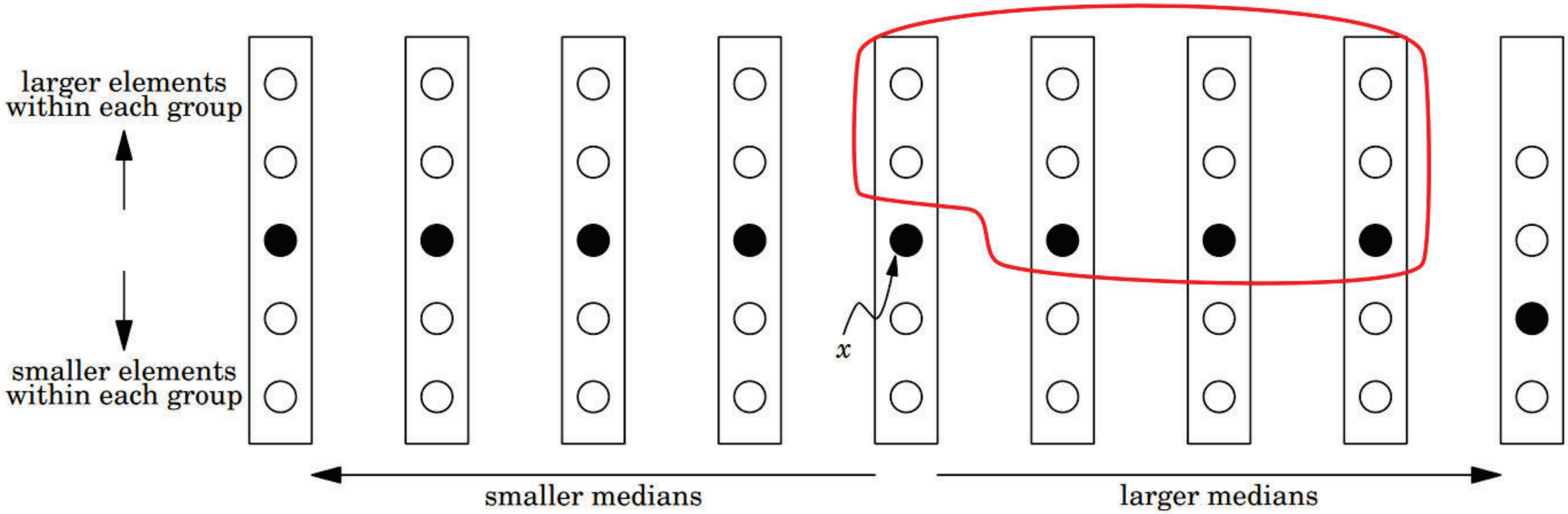
4. Use x as the pivot and partition.

5. If the pivot is not the order statistics that is searched for, recurse on the subarray that contains it.

Deterministic Selection

Use a bound B to stop recursion: If the size of the array is less than or equal to B then use brute-force search to find the desired order statics.

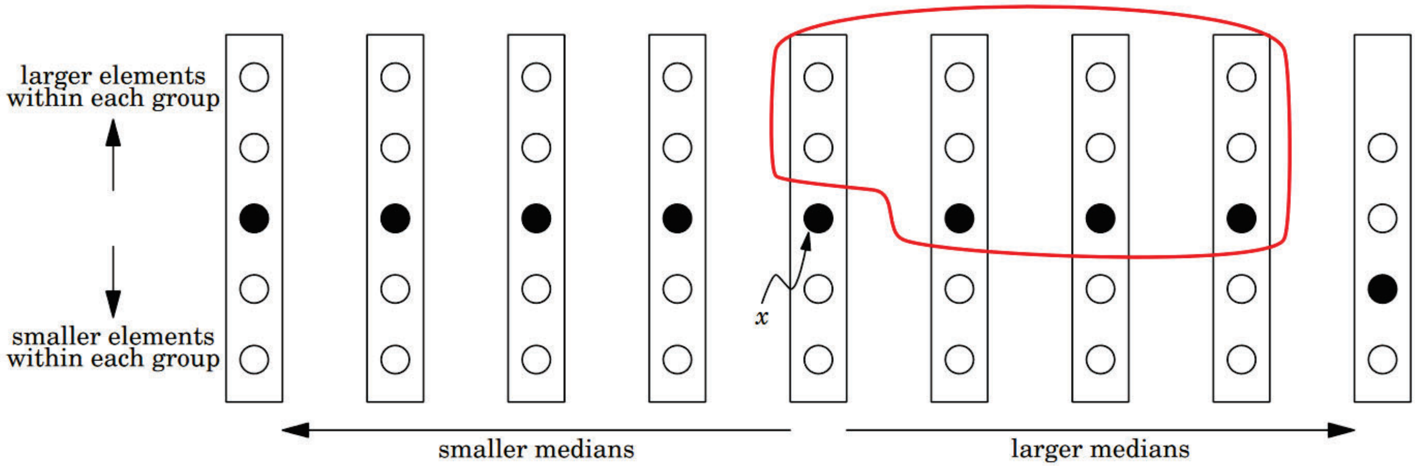
Imagine laying out the groups side by side, ordered with respect to their medians. Also imagine each group is sorted from greatest to least, top to bottom.



x : median of medians

Q: How many elements are smaller than x?

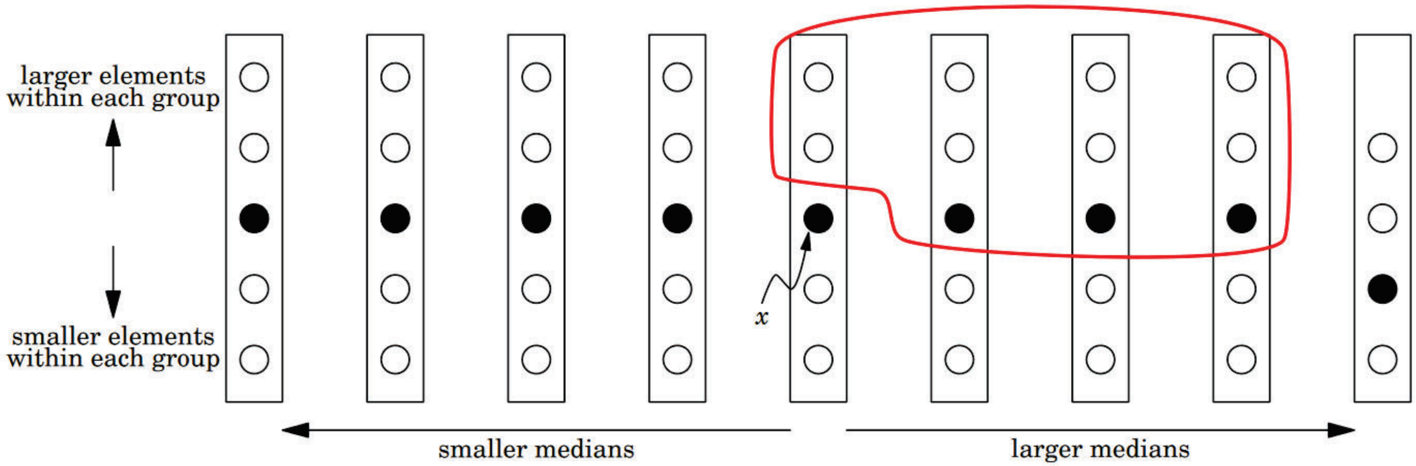
Deriving bounds on the size of the smaller subarray



Obs. 1: Each group of size 5 whose median is:

- at most x contains at least 3 elements smaller than x.
- at least x contains at least 3 elements greater than x.

Deriving bounds on the size of the smaller subarray

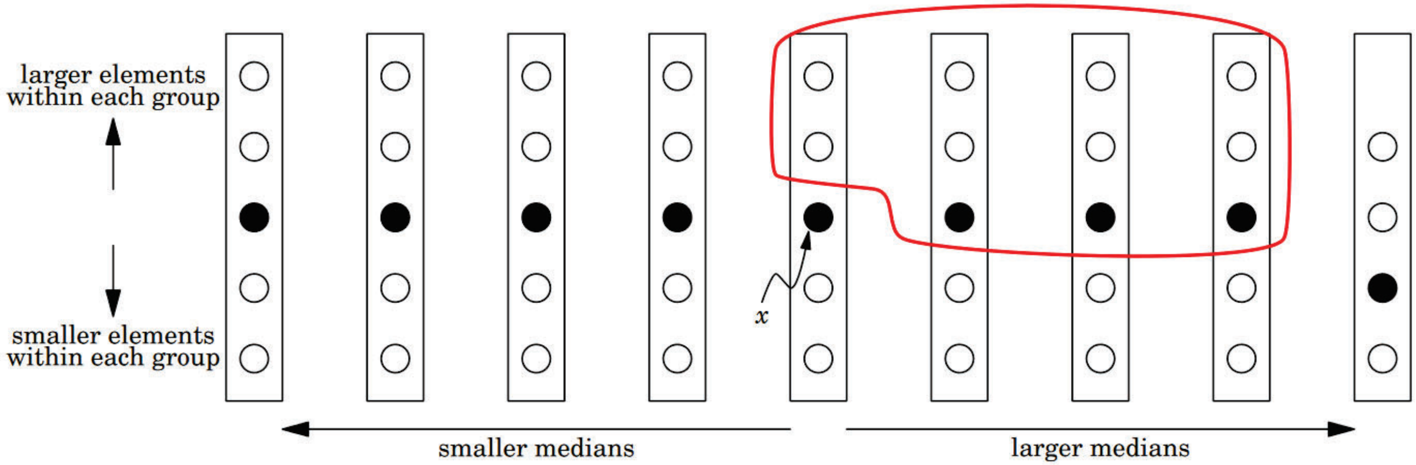


Obs. 1: Each group of size 5 whose median is:

- at most x contains at least 3 elements smaller than x.
- at least x contains at least 3 elements greater than x.

Obs. 2: There are at least $\lceil \frac{n}{5} \rceil - 1$ groups of size 5. How many have median smaller than x (roughly)? What about smaller than x?

Deriving bounds on the size of the smaller subarray



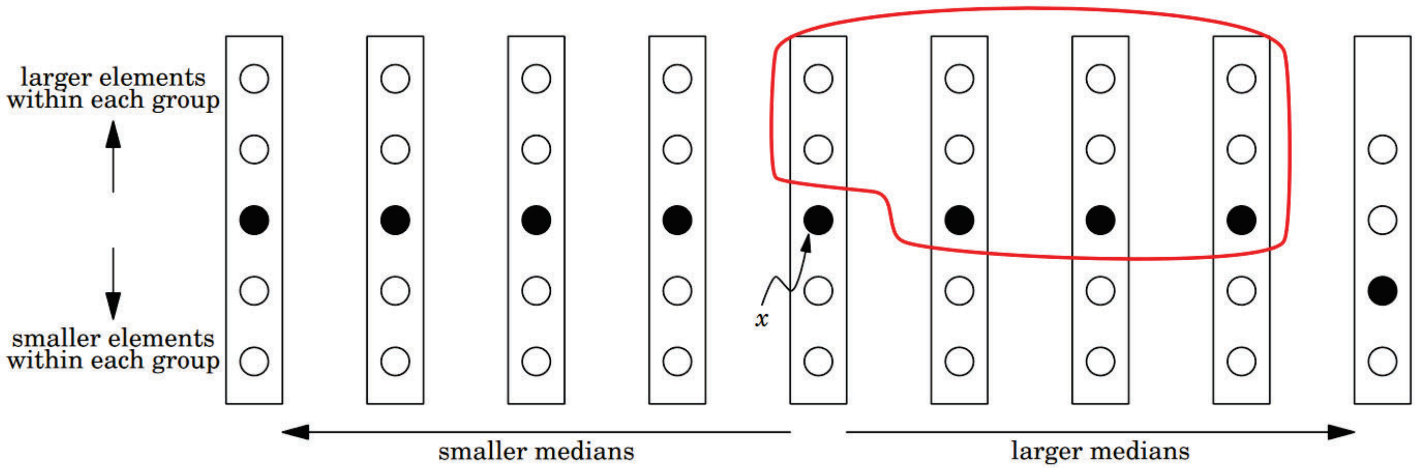
Obs. 1: Each group of size 5 whose median is:

- at most x contains at least 3 elements smaller than x.
- at least x contains at least 3 elements greater than x.

Obs. 2: There are at least $\lceil \frac{n}{5} \rceil - 1$ groups of size 5.

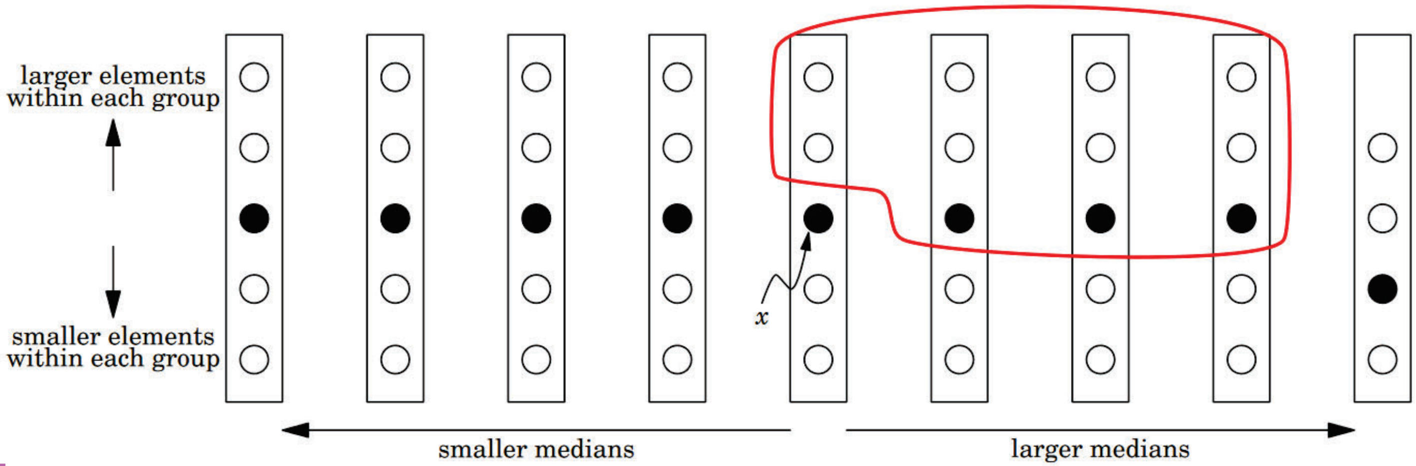
- At least $\lfloor \left(\frac{1}{2} \cdot \lceil \frac{n}{5} \rceil - 2 \right) \rfloor$ of these groups have median smaller than x;
- At least $\lfloor \left(\frac{1}{2} \cdot \lceil \frac{n}{5} \rceil - 2 \right) \rfloor$ of these groups have median larger than x.

Deriving bounds on the size of the smaller subarray



Obs. 3: The group with x as median contains two elements greater than x and two smaller than x.

Deriving bounds on the size of the smaller subarray



Summary:

Obs. 1: Each group of size 5 whose median is:

- at most x contains at least 3 elements smaller than x.
- at least x contains at least 3 elements greater than x.

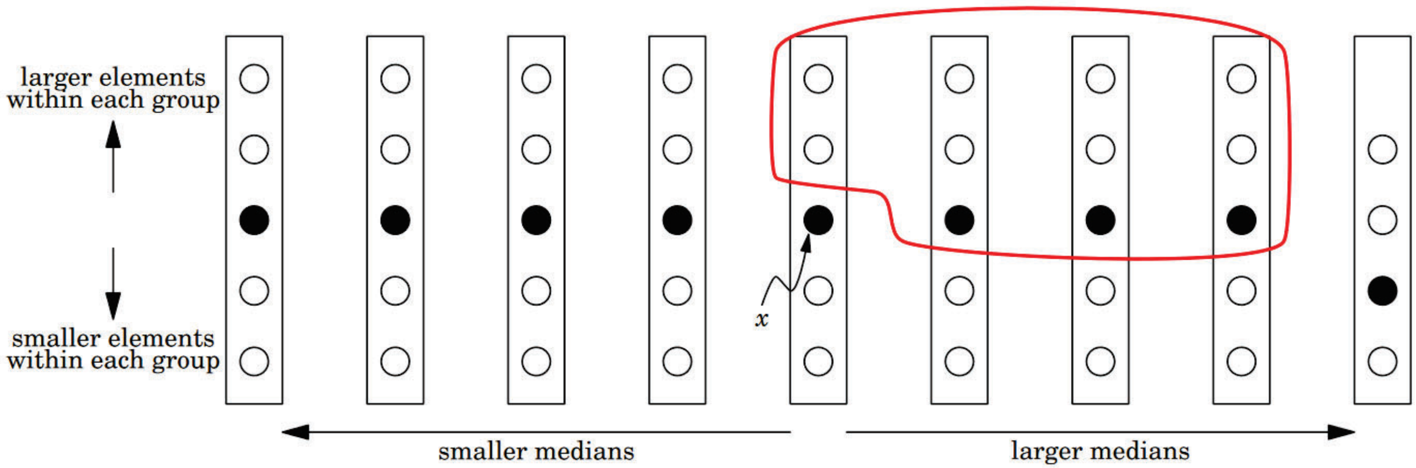
Obs. 2: There are at least $\lceil \frac{n}{5} \rceil - 1$ groups of size 5:

- At least $\lfloor \left(\frac{1}{2} \cdot \lceil \frac{n}{5} \rceil - 2 \right) \rfloor$ of these groups have median smaller than x;
- At least $\lfloor \left(\frac{1}{2} \cdot \lceil \frac{n}{5} \rceil - 2 \right) \rfloor$ of these groups have median larger than x.

Obs. 3: The group with x as median contains 2 elements greater than x and 2 smaller than x.

Combining these observations, how many elements are less than x?

Deriving bounds on the size of the smaller subarray



Combining observations (1-3), we get:

$$(\#\text{ elts of } A \text{ less than } x) \geq 3\left(\left\lfloor \frac{1}{2}\lceil n/5 \rceil - 2 \right\rfloor\right) + 2 \geq 3\left(\frac{1}{2}\lceil n/5 \rceil - 2 - \frac{1}{2}\right) + 2 > \frac{3}{10}n - 6.$$

$$(\#\text{ elts of } A \text{ greater than } x) \geq \frac{3}{10}n - 6.$$

Thus:

- the size of the smaller sub-array is at least $\frac{3n}{10} - 6$.
- the size of the larger sub-array is at most $\frac{7n}{10} + 6$.

Analysis

Assume that the input numbers are pairwise distinct. We claim that there is a constant α such that, for all $n \geq 1$, $T(n)$, is the max # of comparisons of this method, where the max is over the worst case index k (of the statistic searched for).

As long as B is set to a constant, we can adjust a value of α so that the claim holds for all $n \leq B$.

Let $n > B$. The number of medians is $\lceil \frac{n}{5} \rceil$.
So, it is at most $\leq \frac{n}{5} + 1$ and is at least $\frac{n}{5}$.
The number of medians less than x is at least $\frac{n}{10} - 2$. So, the size of the smaller subarray is at least $3(\frac{n}{10} - 2) = \frac{3n}{10} - 6$. Thus, the size of the larger subarray is at most $\frac{7n}{10} + 6$.

Let β be a constant such that the partition operation and division into groups of 5 requires at most βn. comparisons.

The max number of comparisons, $T(n)$, satisfies the inequality:

$$T(n) \leq \beta n + T\left(\frac{n}{5} + 1\right) + T\left(\frac{7n}{10} + 6\right)$$

The max number of comparisons, $T(n)$, satisfies the inequality:

$$T(n) \leq \beta n + T\left(\frac{n}{5} + 1\right) + T\left(\frac{7n}{10} + 6\right)$$

Exercise: Prove by induction an upper bound on $T(n)$ as a function of n and various constants.