2-9 Sorting and Selection

Hengfeng Wei

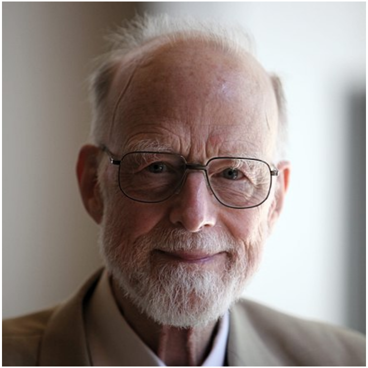
hfwei@nju.edu.cn

May 28, 2018

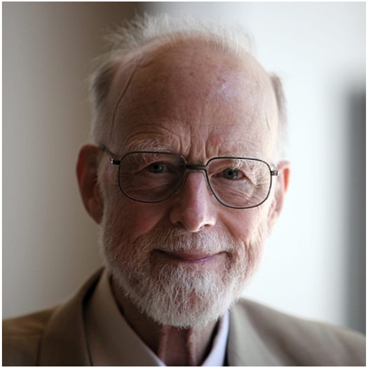
How to Argue?

Am I Alone?

QUICKSORT Invented by Tony Hoare in 1959/1960

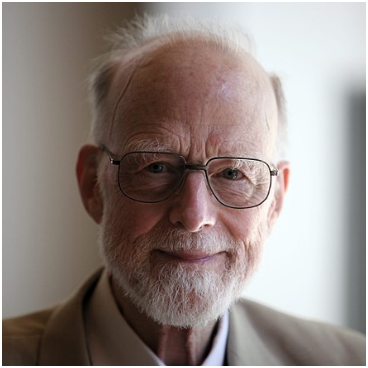


QUICKSORT Invented by Tony Hoare in 1959/1960



`null pointer`

Quicksort Invented by Tony Hoare in 1959/1960



`null pointer`

"I call it my billion-dollar mistake."

Best-Case Complexity of QUICKSORT $(7.4-2)$

Show that QUICKSORT's best-case running time is $\Omega(n \log n)$.

Best-Case Complexity of QUICKSORT $(7.4-2)$

Show that QUICKSORT's best-case running time is $\Omega(n \log n)$.

$$T(n) = 2T(\frac{n}{2}) + \Theta(n)$$

Best-Case Complexity of Quicksort $(7.4 - 2)$

Show that Quicksort's best-case running time is $\Omega(n \log n)$.

$$T(n) = 2T(\frac{n}{2}) + \Theta(n)$$

$$T(n) = \underbrace{n}_{\text{Partition}} \underbrace{\log n}_{\text{Height}} \quad \text{(Recursion Tree)}$$

Best-Case Complexity of QUICKSORT $(7.4 - 2)$

Show that QUICKSORT's best-case running time is $\Omega(n \log n)$.

$$T(n) = 2T(\frac{n}{2}) + \Theta(n)$$

$$T(n) = \underbrace{n}_{\text{PARTITION}} \underbrace{\log n}_{\text{Height}} \quad \text{(Recursion Tree)}$$

$$\boxed{T(n) = \min_{0 \le q \le n-1} \Big(T(q) + T(n-q-1) \Big) + \Theta(n)}$$

Best-Case Complexity of QUICKSORT $(7.4 - 2)$

Show that QUICKSORT's best-case running time is $\Omega(n \log n)$.

$$T(n) = 2T(\frac{n}{2}) + \Theta(n)$$

$$T(n) = \underbrace{n}_{\text{PARTITION}} \underbrace{\log n}_{\text{Height}} \quad \text{(Recursion Tree)}$$

$$\boxed{T(n) = \min_{0 \le q \le n-1} \Big(T(q) + T(n-q-1)\Big) + \Theta(n)}$$

$$T(n) = \Omega(n \log n)$$

Best-Case Complexity of $\textsc{Quicksort}$ $(7.4-2)$

Show that $\textsc{Quicksort}$'s best-case running time is $\Omega(n \log n)$.

$$T(n) = 2T(\frac{n}{2}) + \Theta(n)$$

$$T(n) = \underbrace{n}_{\textsc{Partition}} \underbrace{\log n}_{\text{Height}} \quad \text{(Recursion Tree)}$$

$$\boxed{T(n) = \min_{0 \le q \le n-1} \Big(T(q) + T(n-q-1) \Big) + \Theta(n)}$$

$$T(n) = \Omega(n \log n)$$

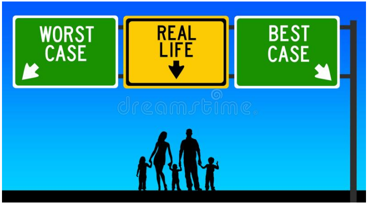
By substitution.

Median-of-3 PARTITION (Problem $7-5$)

Argue that in the $\Omega(n \log n)$ running time of QUICKSORT, the *median-of-3* method affects only the constant factor.

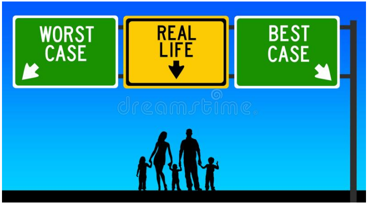
Median-of-3 PARTITION (Problem $7-5$)

Argue that in the $\Omega(n \log n)$ running time of QUICKSORT, the *median-of-3* method affects only the constant factor.



Median-of-3 PARTITION (Problem $7-5$)

Argue that in the $\Omega(n \log n)$ running time of QUICKSORT, the *median-of-3* method affects only the constant factor.

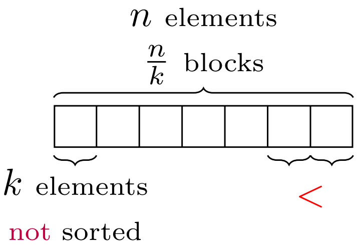


$$\boxed{T(n) = \min_{0 \le q \le n-1} \Big(T(q) + T(n-q-1) \Big) + \Theta(n)}$$

$$T(n) = \Omega(n \log n)$$

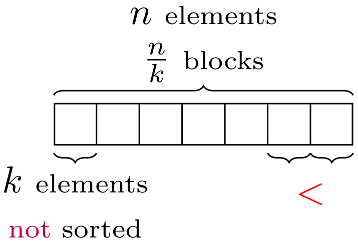
The $\frac{n}{k}$-sorted Problem (Problem $8.1 - 4$)

Sorts an already $\frac{n}{k}$-sorted array



The $\frac{n}{k}$-sorted Problem (Problem $8.1-4$)

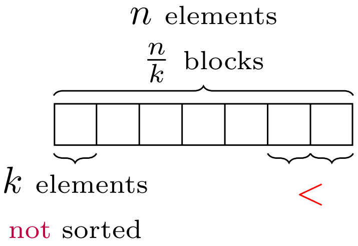
Sorts an already $\frac{n}{k}$-sorted array



$\Omega(n \log k)$

The $\frac{n}{k}$-sorted Problem (Problem $8.1 - 4$)

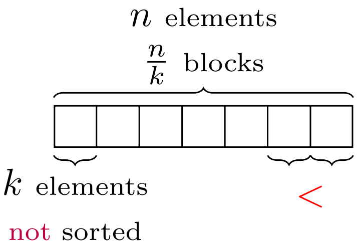
Sorts an already $\frac{n}{k}$-sorted array



$\Omega(n \log k)$ $\qquad$ $O(n \log k)$

The $\frac{n}{k}$-sorted Problem (Problem $8.1-4$)

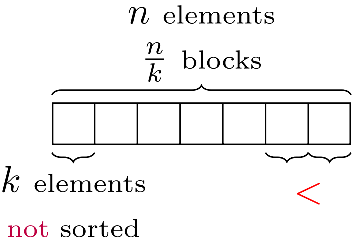
Sorts an already $\frac{n}{k}$-sorted array



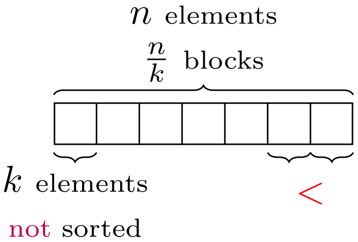
$\Omega(n \log k) \qquad O(n \log k)$

$$(k!)^{\frac{n}{k}} \leq L \leq 2^H$$

$\frac{n}{k}$-sorts an arbitrary array

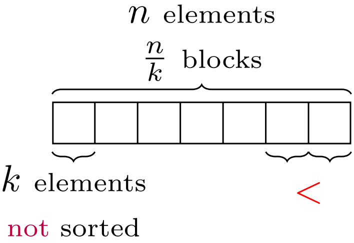


$\frac{n}{k}$-sorts an arbitrary array



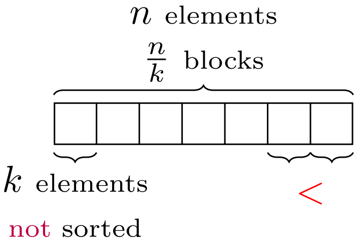
$O(?)$

$\frac{n}{k}$-sorts an arbitrary array



$O(?)$ $\Omega(?)$

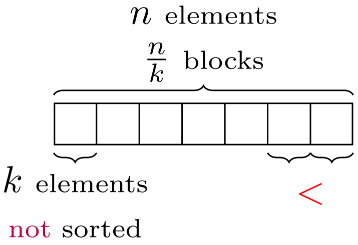
$\frac{n}{k}$-sorts an arbitrary array



$$O(?) \qquad \Omega(?)$$

$$L \geq \binom{n}{\underbrace{k, \dots, k}_{\frac{n}{k}}} = \frac{n!}{(k!)^{\frac{n}{k}}}$$

$\frac{n}{k}$-sorts an arbitrary array



$$O(?) \qquad \Omega(?)$$

$$L \geq \binom{n}{\underbrace{k, \ldots, k}_{\frac{n}{k}}} = \frac{n!}{(k!)^{\frac{n}{k}}} \implies \Omega(n \log(n/k))$$

Sorting $[0, n^3 - 1]$ (Problem $8.3 - 4$)

Sort n integers in $[0, n^3 - 1]$ in $O(n)$ time.

Sorting in Place in Linear Time (Problem $8-2$ (e))

Suppose that the n records have keys in the range $[0, k]$.
Modify COUNTING-SORT to sort them in place $(O(k))$ in $O(n+k)$ time.

Sorting in Place in Linear Time (Problem $8-2$ (e))

Suppose that the n records have keys in the range $[0, k]$.
Modify COUNTING-SORT to sort them in place ($O(k)$) in $O(n+k)$ time.

	1	2	3	4	5	6	7	8
A :	2	5	3	0	2	3	0	3

	0	1	2	3	4	5
C :	2	0	2	3	0	1
C :	2	2	4	7	7	8

	1	2	3	4	5	6	7	8
B :	0	0	2	2	3	3	3	5

Sorting in Place in Linear Time (Problem $8-2$ (e))

Suppose that the n records have keys in the range $[0, k]$.
Modify COUNTING-SORT to sort them in place ($O(k)$) in $O(n+k)$ time.

	1	2	3	4	5	6	7	8
A :	2	5	3	0	2	3	0	3

	0	1	2	3	4	5
C :	2	0	2	3	0	1
C :	2	2	4	7	7	8

	1	2	3	4	5	6	7	8
B :	0	0	2	2	3	3	3	5

	1	2	3	4	5	6	7	8
A :	2	0	3	0	2	3	3	5

	0	1	2	3	4	5
C :	1	2	4	7	7	8

Sorting in Place in Linear Time (Problem $8-2$ (e))

Suppose that the n records have keys in the range $[0, k]$.
Modify COUNTING-SORT to sort them in place ($O(k)$) in $O(n+k)$ time.

A:

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

C:

0	1	2	3	4	5
2	0	2	3	0	1

C:

0	1	2	3	4	5
2	2	4	7	7	8

B:

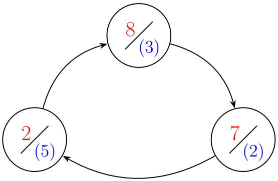
1	2	3	4	5	6	7	8
0	0	2	2	3	3	3	5

A:

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

C:

0	1	2	3	4	5
1	2	4	7	7	8



Sorting in Place in Linear Time (Problem $8-2\ (e)$)

Suppose that the n records have keys in the range $[0, k]$.
Modify COUNTING-SORT to sort them in place ($O(k)$) in $O(n+k)$ time.

	1	2	3	4	5	6	7	8
A :	2	5	3	0	2	3	0	3

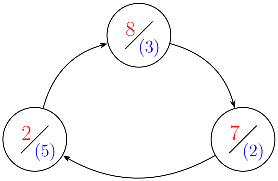
	0	1	2	3	4	5
C :	2	0	2	3	0	1
C :	2	2	4	7	7	8

	1	2	3	4	5	6	7	8
B :	0	0	2	2	3	3	3	5

	1	2	3	4	5	6	7	8
A :	2	0	3	0	2	3	3	5

	0	1	2	3	4	5
C :	1	2	4	7	7	8

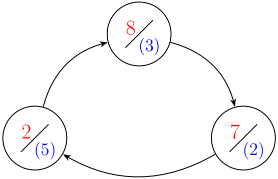
`While (`$i \geq 1$`):`



	1	2	3	4	5	6	7	8
A :	2	0	3	0	2	3	3	5

	0	1	2	3	4	5
C :	1	2	4	7	7	8

While $(i \geq 1)$:



Code here

Finding the 2nd Smallest Element (Problem $9.1-1$)

Show that the 2nd smallest of n elements can be found with $n + \lceil \log n \rceil - 2$ comparisons in the worst case.

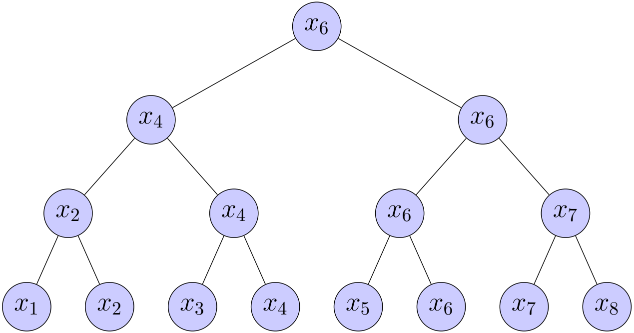
Finding the 2nd Smallest Element (Problem $9.1-1$)

Show that the 2nd smallest of n elements can be found with $n + \lceil \log n \rceil - 2$ comparisons in the worst case.

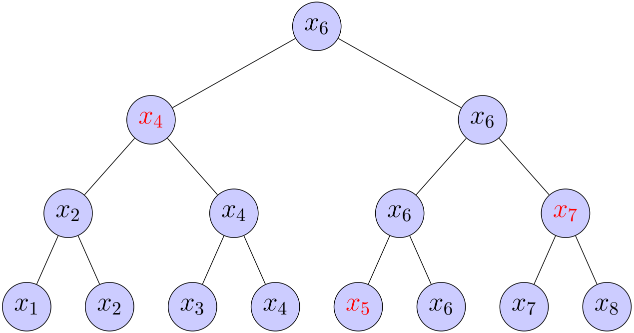
$$(n-1) + (n-1-1) = 2n-3$$

$$n + \lceil \log n \rceil - 2 = (n - 1) + (\lceil \log n \rceil - 1)$$

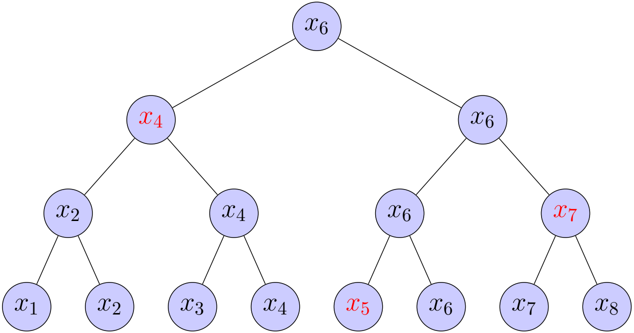
$$n + \lceil \log n \rceil - 2 = (n-1) + (\lceil \log n \rceil - 1)$$



$$n + \lceil \log n \rceil - 2 = (n-1) + (\lceil \log n \rceil - 1)$$

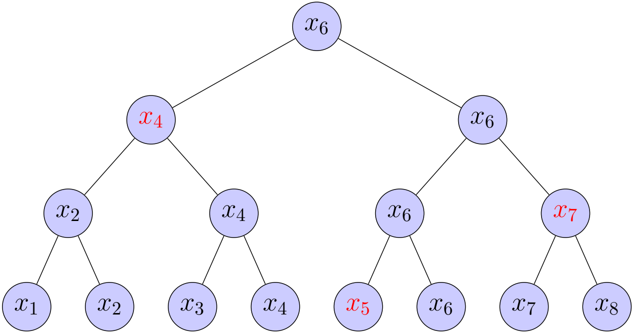


$$n + \lceil \log n \rceil - 2 = (n-1) + (\lceil \log n \rceil - 1)$$



#Potential 2nd smallest elements $\leq \lceil \log n \rceil$

$$n + \lceil \log n \rceil - 2 = (n-1) + (\lceil \log n \rceil - 1)$$



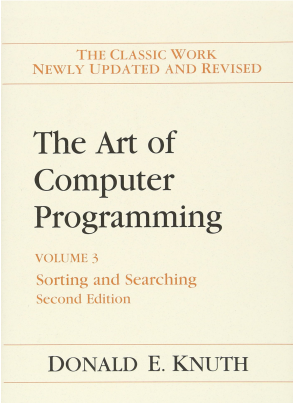
#Potential 2nd smallest elements $\leq \lceil \log n \rceil$

Q : Can we do even better?

$$\Omega = n + \lceil \log n \rceil - 2$$

$$\Omega = n + \lceil \log n \rceil - 2 = (n-1) + (\lceil \log n \rceil - 1)$$

$$\Omega = n + \lceil \log n \rceil - 2 = (n-1) + (\lceil \log n \rceil - 1)$$



TAOCP Vol 3 (Page 209, Section 5.3.3)

k Numbers Closest to the Median (Problem $9.3 - 7$)

S : n distinct numbers $\qquad k \leq n$

k Numbers Closest to the Median (Problem $9.3-7$)

S : n distinct numbers $\qquad k \leq n$

$$\frac{n}{2}-\frac{k}{2} \sim \frac{n}{2}+\frac{k}{2}$$

k Numbers Closest to the Median (Problem $9.3-7$)

$S : n$ distinct numbers $\qquad k \le n$

$$\frac{n}{2} - \frac{k}{2} \sim \frac{n}{2} + \frac{k}{2}$$

$$S = \{800, 6, 900, 50, 7\}, \quad k = 2 \implies \{6, 7\}$$

k Numbers Closest to the Median (Problem $9.3-7$)

S : n distinct numbers $\qquad k \leq n$

$$\frac{n}{2} - \frac{k}{2} \sim \frac{n}{2} + \frac{k}{2}$$

$$S = \{800, 6, 900, 50, 7\}, \quad k = 2 \implies \{6, 7\}$$

$$S - 50 = \{750, -44, 850, 0, -43\}$$

k Numbers Closest to the Median (Problem $9.3 - 7$)

S : n distinct numbers $\qquad k \le n$

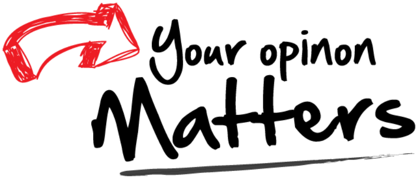
$$\frac{n}{2} - \frac{k}{2} \sim \frac{n}{2} + \frac{k}{2}$$

$$S = \{800, 6, 900, 50, 7\}, \quad k = 2 \implies \{6, 7\}$$

$$S - 50 = \{750, -44, 850, 0, -43\}$$

median + subtraction + $(k+1)$-th smallest + partition + add


Thank You!



Office 302

Mailbox: H016

hfwei@nju.edu.cn