# Sorting, Searching, Selection, and Amortized Analysis

Hengfeng Wei

hfwei@nju.edu.cn

April 24, 2018



Maximal-sum Subarray (Problem 3.7)

- ightharpoonup Array  $A[1\cdots n], a_i>=<0$
- $\blacktriangleright$  To find (the sum of) an MS in A

$$A[-2,1,-3, \boxed{4,-1,2,1},-5,4]$$



 $\mathsf{MSS}[i]$ : the sum of the MS (MS[i]) in  $A[1\cdots i]$ 

 $\mathsf{MSS}[i] \text{: the sum of the MS } \big(\mathsf{MS}[i]\big) \text{ in } A[1 \cdots i]$ 

 $\mathsf{mss} = \mathsf{MSS}[n]$ 

 $\mathsf{MSS}[i] \text{: the sum of the MS } \big(\mathsf{MS}[i]\big) \text{ in } A[1 \cdots i]$ 

$$\mathsf{mss} = \mathsf{MSS}[n]$$

$$Q$$
: Is  $a_i \in \mathsf{MS}[i]$ ?

 $\mathsf{MSS}[i] :$  the sum of the MS (MS[i]) in  $A[1 \cdots i]$ 

$$\mathsf{mss} = \mathsf{MSS}[n]$$

$$Q: \mathsf{Is}\ a_i \in \mathsf{MS}[i]$$
?

$$\mathsf{MSS}[i] = \max\{\mathsf{MSS}[i-1], \red{???}\}$$

$$\mathsf{mss} = \max_{1 \leq i \leq n} \mathsf{MSS}[i]$$

Q: where does the  $\mathsf{MS}[i]$  start?

$$\mathsf{mss} = \max_{1 \leq i \leq n} \mathsf{MSS}[i]$$

Q: where does the MS[i] start?

$$\mathsf{MSS}[i] = \max\left\{\mathsf{MSS}[i-1] + a_i, 0\right\}$$

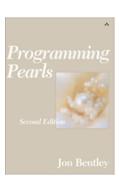
$$\mathsf{mss} = \max_{1 \leq i \leq n} \mathsf{MSS}[i]$$

Q: where does the MS[i] start?

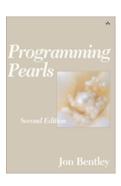
$$\mathsf{MSS}[i] = \max \left\{ \mathsf{MSS}[i-1] + a_i, 0 \right\}$$

$$\mathsf{MSS}[0] = 0$$

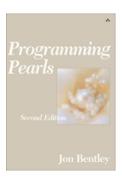
- 1: procedure  $MSS(A[1 \cdots n])$
- 2:  $MSS[0] \leftarrow 0$
- 3: for  $i \leftarrow 1$  to n do
- 4:  $\mathsf{MSS}[i] \leftarrow \max \left\{ \mathsf{MSS}[i-1] + A[i], 0 \right\}$
- 5: return  $\max_{1 \leq i \leq n} \mathsf{MSS}[i]$



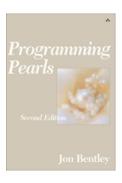
Ulf Grenander 
$$O(n^3) \implies O(n^2)$$



Ulf Grenander  $O(n^3) \implies O(n^2)$ Michael Shamos  $O(n \log n)$ , onenight



Ulf Grenander  $O(n^3) \implies O(n^2)$ Michael Shamos  $O(n \log n)$ , onenight Jon Bentley Conjecture:  $\Omega(n \log n)$ 

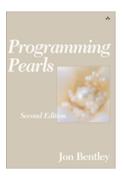


Ulf Grenander  $O(n^3) \implies O(n^2)$ 

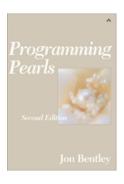
Michael Shamos  $O(n \log n)$ , onenight

Jon Bentley Conjecture:  $\Omega(n \log n)$ 

Michael Shamos Carnegie Mellon seminar



Ulf Grenander  $O(n^3) \Longrightarrow O(n^2)$ Michael Shamos  $O(n\log n)$ , onenight Jon Bentley Conjecture:  $\Omega(n\log n)$ Michael Shamos Carnegie Mellon seminar Jay Kadane O(n),



Ulf Grenander  $O(n^3) \implies O(n^2)$ Michael Shamos  $O(n \log n)$ , onenight Jon Bentley Conjecture:  $\Omega(n \log n)$ Michael Shamos Carnegie Mellon seminar Jay Kadane O(n),  $\leq 1$  minute

## Definition (K-sorting (Problem 6.8))

An array  $A[1 \cdots n]$  is **k-sorted** if it can be divided into k blocks, each of size n/k (we assume that  $n/k \in \mathbb{N}$ ), such that the elements in each block are larger than the elements in earlier blocks and smaller than elements in later blocks. The elements within each block need not be sorted.

$$n = 16, \ k = 4, \ \frac{n}{k} = 4$$

1, 2, 4, 3; 7, 6, 8, 5; 10, 11, 9, 12; 15, 13, 16, 14



1-sorted

 $1\text{-sorted} \to 2\text{-sorted}$ 

1-sorted  $\rightarrow 2$ -sorted  $\rightarrow 4$ -sorted

1-sorted o 2-sorted o 4-sorted  $o \cdots o n$ -sorted

1-sorted  $\rightarrow$  2-sorted  $\rightarrow$  4-sorted  $\rightarrow \cdots \rightarrow n$ -sorted

Quicksort (with median as pivot) stops after the  $\log k$  recursions.

1-sorted  $\rightarrow$  2-sorted  $\rightarrow$  4-sorted  $\rightarrow \cdots \rightarrow n$ -sorted

Quicksort (with median as pivot) stops after the  $\log k$  recursions.

$$\Theta(n \log k)$$

 $\Omega(n \log k)$ 

 $\Omega(n \log k)$ 

L =

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

$$L = \binom{n}{n/k, \dots, n/k}$$

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

$$L = \binom{n}{n/k, \dots, n/k} = \frac{n!}{\left(\left(\frac{n}{k}\right)!\right)^k}$$

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

$$L = \binom{n}{n/k, \dots, n/k} = \frac{n!}{\left(\left(\frac{n}{k}\right)!\right)^k}$$

$$H \ge$$

## $\Omega(n \log k)$

$$L = \binom{n}{n/k, \dots, n/k} = \frac{n!}{\left(\left(\frac{n}{k}\right)!\right)^k}$$

$$H \ge \log \left( \frac{n!}{\left( \left( \frac{n}{k} \right)! \right)^k} \right)$$

## $\Omega(n \log k)$

$$L = \binom{n}{n/k, \dots, n/k} = \frac{n!}{\left(\left(\frac{n}{k}\right)!\right)^k}$$

$$H \ge \log \left( \frac{n!}{\left( \left( \frac{n}{k} \right)! \right)^k} \right)$$

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \implies \log n! \sim n \log n$$







Quicksort



Quicksort

$$A(n) = O(n \log n)$$



### Quicksort

$$A(n) = O(n \log n)$$

### In the worst case:

- "Matching Nuts and Bolts" by Alon et al.,  $\Theta(n \log^4 n)$
- lacktriangle "Matching Nuts and Bolts Optimality" by Bradford, 1995,  $\Theta(n \log n)$



 $\Omega(n \log n)$ 



 $\Omega(n \log n)$ 

$$3^H \ge L \ge n!$$



 $\Omega(n \log n)$ 

$$\mathbf{3}^{H} \ge L \ge n! \implies H \ge \log n! \implies H = \Omega(n \log n)$$

Repeated elements (Problem 2.12)

$$R[1\dots n]$$

$$\# > \frac{n}{13}$$

#### Repeated elements (Problem 2.12)

$$R[1\dots n]$$

$$\# > \frac{n}{13}$$

check(R[i], R[j])

$$\# > \frac{n}{k}$$

$$\# > \frac{n}{k}$$

$$\# > \frac{n}{k}$$

Lower bound  $\Omega(n \log k)$ 

$$\# > \frac{n}{k}$$

Lower bound  $\Omega(n \log k)$ 

L: # of leaves?

$$\# > \frac{n}{k}$$

Lower bound  $\Omega(n \log k)$ 

L: # of leaves?

"Finding Repeated Elements" by Misra & Gries, 1982

$$a_i \in \mathbb{Z}^+$$

$$\forall i \neq j : a_i \neq a_j$$

$$a_i \in \mathbb{Z}^+$$
$$\forall i \neq j : a_i \neq a_j$$

$$A = [1, 2, 4, 5] \implies 3$$



$$a_i \in \mathbb{Z}^+$$
$$\forall i \neq j : a_i \neq a_j$$

$$A = [1, 2, 4, 5] \implies 3$$

$$T(n) = O(n)$$

$$a_i \in \mathbb{Z}^+$$
$$\forall i \neq j : a_i \neq a_j$$

$$A = [1,2,4,5] \implies 3$$

$$T(n) = O(n)$$

$$T(n) = T(\frac{n}{2}) + 1 = O(\log n)$$



$$A[1\cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

$$A[i-1] \ge A[i] \le A[i+1]$$

$$A[1 \cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

 $A[i-1] \ge A[i] \le A[i+1]$ 

Ξ

$$A[1 \cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

$$A[i-1] \ge A[i] \le A[i+1]$$

 $\exists$ 

$$Scan: T(n) = O(n)$$

$$A[1 \cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

$$A[i-1] \ge A[i] \le A[i+1]$$

 $\exists$ 

$$Scan: T(n) = O(n)$$

$$\min A: T(n) = O(n)$$



$$A[1 \cdots n]$$

$$A[0] \geq A[1] \mathrel{\wedge} A[n-2] \leq A[n-1]$$

$$A[i-1] \ge A[i] \le A[i+1]$$

$$A[1 \cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

$$A[i-1] \ge A[i] \le A[i+1]$$

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



$$A[1 \cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

$$A[i-1] \ge A[i] \le A[i+1]$$

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

$$T(n) = T(\frac{n}{2}) + 1$$



$$A[1 \cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

$$A[i-1] \ge A[i] \le A[i+1]$$

$$m = \frac{n}{2}$$

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

$$T(n) = T(\frac{n}{2}) + 1$$



$$A[1 \cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

$$A[i-1] \ge \mathbf{A[i]} \le A[i+1]$$

$$m = \frac{n}{2}$$
 
$$T(n) = O(\log n)$$
 
$$A[m-1] \ge A[m] \le A[m+1]$$
 
$$T(n) = T(\frac{n}{2}) + 1$$
 
$$A[m-1] < A[m] \lor A[m+1] < A[m]$$

16 / 34

$$A[1 \cdots n]$$

$$A[0] \ge A[1] \land A[n-2] \le A[n-1]$$

$$A[i-1] \ge \underline{A[i]} \le A[i+1]$$

$$m = \frac{n}{2}$$

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

$$T(n) = T(\frac{n}{2}) + 1$$

$$A[m-1] \ge A[m] \le A[m+1]$$

$$A[m-1] < A[m] \lor A[m+1] < A[m]$$

$$n=1 \quad \lor \quad n=2$$



 $M:m\times n$ 

Row: increasing from left to right

Col: increasing from top to down

 $x \in M$ ?

 $M:m\times n$ 

Row: increasing from left to right

Col: increasing from top to down

 $x \in M$ ?

Divide & Conquer

$$M:m\times n$$

Row: increasing from left to right

Col: increasing from top to down

$$x \in M$$
?

## Divide & Conquer

$$T(m,n) = 3T(\frac{m}{2}, \frac{n}{2}) + 1$$

$$M:m\times n$$

Row: increasing from left to right

Col: increasing from top to down

$$x \in M$$
?

#### Divide & Conquer

$$T(m,n) = 3T(\frac{m}{2}, \frac{n}{2}) + 1$$

Always checking the lower left corner.

$$M: m \times n$$

Row: increasing from left to right

Col: increasing from top to down

$$x \in M$$
?

#### Divide & Conquer

$$T(m,n) = 3T(\frac{m}{2}, \frac{n}{2}) + 1$$

Always checking the lower left corner.

$$T(m,n) = m + n - 1$$



$$W(n) \le 2n - 1$$

$$W(n) \le 2n - 1$$

$$W(n) \ge 2n - 1$$

$$W(n) \le 2n - 1$$

$$W(n) \ge 2n - 1$$

By Adversary Argument!

$$W(n) \le 2n - 1$$

$$W(n) \ge 2n - 1$$

By Adversary Argument!

$$\mbox{Diagonals: } i+j=n-1 \quad \& \quad i+j=n \\$$

$$W(n) \le 2n - 1$$

$$W(n) \ge 2n - 1$$

By Adversary Argument!

$$\mbox{Diagonals: } i+j=n-1 \quad \& \quad i+j=n \\$$

No particular ordering requirements on these two diagonals!



$$W(n) \le 2n - 1$$

$$W(n) \ge 2n - 1$$

By Adversary Argument!

$$\mbox{Diagonals: } i+j=n-1 \quad \& \quad i+j=n \\$$

No particular ordering requirements on these two diagonals!

$$i+j \le n-1 \implies x > M_{ij}$$
  
 $i+j > n-1 \implies x < M_{ij}$ 

$$A + B = c$$
 (Problem 9.9)

Sorted 
$$S[1 \cdots n], \quad c$$

$$\exists A,B:A+B=c?$$

### Left-right-pointers iteration

$$S_i + S_j > = < c$$

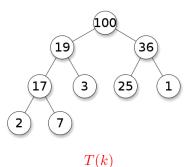


Hengfeng Wei (hfwei@nju.edu.cn)

April 24, 2018

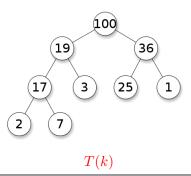
#### The k-th Largest Elements in a Heap (Problem 7.2)

 $k \ll n$ 



#### The k-th Largest Elements in a Heap (Problem 7.2)

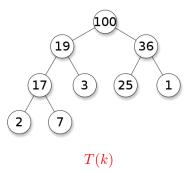




Must be in the first k layers

#### The k-th Largest Elements in a Heap (Problem 7.2)

$$k \ll n$$



Must be in the first k layers  $\implies O(2^k)$ 

20 / 34

$$O(n \log n)$$

$$O(n + k \log n)$$

$$O(n + k \log k)$$

$$O(n \log n)$$

sorting

$$O(n + k \log n)$$

$$O(n + k \log k)$$

$$O(n \log n)$$

sorting

$$O(n + k \log n)$$

max-heap

$$O(n + k \log k)$$

$$O(n \log n)$$

sorting

$$O(n + k \log n)$$

max-heap

$$O(n + k \log k)$$

k-selection + partition + sorting

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

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

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

$$O(n\log n + k)$$

$$O(n + k \log k)$$

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

$$O(n\log n + k)$$
 sorting +

$$O(n + k \log k)$$

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

$$O(n\log n + k)$$

sorting + left-right iteration

$$O(n + k \log k)$$

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

$$O(n \log n + k)$$
 sorting + left-right iteration

$$O(n + k \log k)$$
 median-selection +

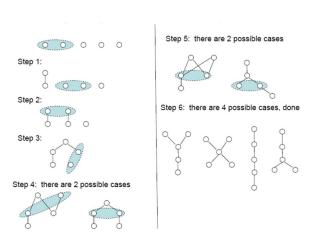
$$S = \{800, 6, 900, \frac{50}{7}\}, \quad k = 2 \implies \{6, 7\}$$

$$O(n \log n + k)$$
 sorting + left-right iteration

$$O(n + k \log k)$$
 median-selection + the smallest  $k$  elements

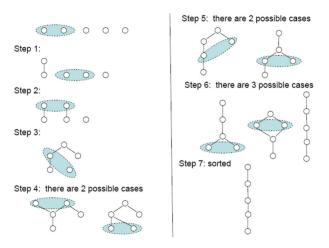


## Selecting the Median of 5 Elements using 6 Comparisons (Problem 8.2)



23 / 34

#### Sorting 5 Elements using 7 Comparisons



Amortized analysis is

an algorithm analysis technique for
analyzing a sequence of operations
irrespective of the input to show that
the average cost per operation is small, even though
a single operation within the sequence might be expensive.

#### The Summation Method

$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

#### The Summation Method

$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

$$\forall i, \ \hat{c_i} = \frac{\left(\sum\limits_{i=1}^n c_i\right)}{n}$$

```
o_i: o_1 o_2 o_3 o_4 o_5 o_6 o_7 o_8 o_9 o_{10}
c_i: 1 2 3 1 5 1 1 9 1
```

$$c_i = \begin{cases} (i-1) + 1 = i & \text{if } i-1 \text{ is an exact power of 2} \\ 1 & \text{o.w.} \end{cases}$$

$$c_i = \begin{cases} (i-1) + 1 = i & \text{if } i-1 \text{ is an exact power of } 2\\ 1 & \text{o.w.} \end{cases}$$

$$\sum_{i=1}^{n} c_i \le n + \sum_{j=0}^{\lceil \lg n \rceil - 1} 2^j = n + (2^{\lceil \lg n \rceil} - 1) \le n + 2n = 3n$$

$$c_i = \begin{cases} (i-1) + 1 = i & \text{if } i-1 \text{ is an exact power of 2} \\ 1 & \text{o.w.} \end{cases}$$

$$\sum_{i=1}^{n} c_i \le n + \sum_{j=0}^{\lceil \lg n \rceil - 1} 2^j = n + (2^{\lceil \lg n \rceil} - 1) \le n + 2n = 3n$$

$$\forall i, \ \hat{c_i} = 3$$



$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

$$a_1, a_2, \ldots, a_n$$

$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

$$a_1, a_2, \ldots, a_n$$

$$\hat{c_i} = c_i + a_i, \ a_i > = < 0$$

Amortized Cost = Actual Cost + Accounting Cost

$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

$$a_1, a_2, \ldots, a_n$$

$$\hat{c_i} = c_i + a_i, \ a_i > = < 0$$

Amortized Cost = Actual Cost + Accounting Cost

$$\forall n, \ \sum_{i=1}^{n} c_i \le \sum_{i=1}^{n} \hat{c_i}$$

$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

$$a_1, a_2, \ldots, a_n$$

$$|\hat{c_i} = c_i + a_i, \ a_i > = < 0$$

#### Amortized Cost = Actual Cost + Accounting Cost

$$\forall n, \sum_{i=1}^{n} c_i \leq \sum_{i=1}^{n} \hat{c_i} \implies \boxed{\forall n, \sum_{i=1}^{n} a_i \geq 0}$$



$$o_1, o_2, \ldots, o_n$$

$$c_1, c_2, \ldots, c_n$$

$$a_1, a_2, \ldots, a_n$$

$$|\hat{c_i} = c_i + a_i, \ a_i > = < 0$$

Amortized Cost = Actual Cost + Accounting Cost

$$\forall n, \sum_{i=1}^{n} c_i \leq \sum_{i=1}^{n} \hat{c}_i \implies \boxed{\forall n, \sum_{i=1}^{n} a_i \geq 0}$$

Key Point: Put the accounting cost on specific objects.



## The Accounting Method for Array Doubling

$$Q: \hat{c_i} = 3$$
 vs.  $\hat{c_i} = 2$ 

## The Accounting Method for Array Doubling

$$Q: \hat{c_i} = 3$$
 vs.  $\hat{c_i} = 2$ 

$$\hat{c}_i = 3 = \underbrace{1}_{\text{insert}} + \underbrace{1}_{\text{move itself}} + \underbrace{1}_{\text{help move another}}$$

# The Accounting Method for Array Doubling

$$Q: \hat{c_i} = 3$$
 vs.  $\hat{c_i} = 2$ 

$$\hat{c_i} = 3 = \underbrace{1}_{\text{insert}} + \underbrace{1}_{\text{move itself}} + \underbrace{1}_{\text{help move another}}$$

	$\hat{c_i}$	$c_i$ (actual cost)	$a_i$ (accounting cost)
Insert (normal)		1	2
Insert (expansion)	3	1+t	-t+2

# Simulating a queue Q using two stacks $S_1, S_2$ (Problem )

```
procedure \operatorname{Enq}(x)
\operatorname{Push}(S_1,x)

procedure \operatorname{Deq}()
if S_2 = \emptyset then
while S_1 \neq \emptyset do
\operatorname{Push}(S_2,\operatorname{Pop}(S_1))
\operatorname{Pop}(S_2)
```

## The Summation Method for Queue Simulation

$$\frac{\left(\sum\limits_{i=1}^{n}c_{i}\right)}{n}$$

#### The Summation Method for Queue Simulation

$$\frac{\left(\sum\limits_{i=1}^{n}c_{i}\right)}{n}$$

The operation sequence is **NOT** known.

item: Push into 
$$S_1$$
 Pop from  $S_1$  Push into  $S_2$  Pop from  $S_2$  
$$1 \qquad 1 \qquad 1 \qquad 1$$
 
$$\hat{c}_{\rm ENQ}=3$$
 
$$\hat{c}_{\rm DEQ}=1$$

$$\hat{c}_{\mathrm{ENQ}} = 3$$
  
 $\hat{c}_{\mathrm{DEQ}} = 1$ 

$$\hat{c}_{\mathrm{DEQ}} = 1$$

$$\sum_{i=1}^{n} a_i \ge 0$$

$$\hat{c}_{\text{ENQ}} = 3$$

$$\hat{c}_{\mathrm{DeQ}}=1$$

$$\sum_{i=1}^{n} a_i \ge 0 \Longleftrightarrow \sum_{i=1}^{n} a_i = \#S_1 \times 2$$

$$\hat{c}_{\text{ENQ}} = 3$$
 $\hat{c}_{\text{DEQ}} = 1$ 

$$\#S_1 = t$$

	$\hat{c_i}$	$c_i$ (actual cost)	$a_i$ (accounting cost)
Enqueue	3	1	2
Dequeue $(S_2 = \emptyset)$	1	1	0
DEQUEUE $(S_2 \neq \emptyset)$	1	1+2t	-2t

# Thank You!



Office 302

Mailbox: H016

hfwei@nju.edu.cn