Grundlæggende Algoritmer og Datastrukturer

Selektion i worst-case lineær tid
[CLRS, kapitel 9.3]

Manuel Blum, Robert W. Floyd, Vaughan Pratt, Ronald L. Rivest, and Robert E. Tarjan, 1973

Selektion

Find det i'te mindste element i en liste

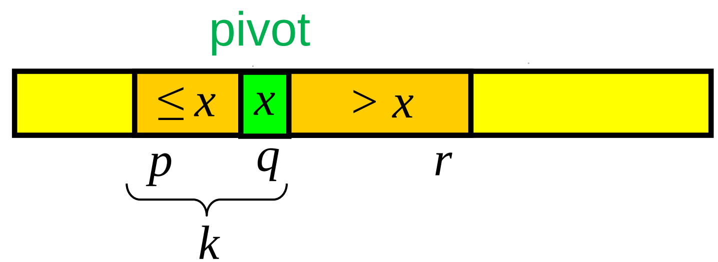
$L =$	10	5	12	3	1	7	42	9	15

$$\textsc{Select}(L, 5) = 9$$

Algoritme	Tid
Randomized-Select [CLRS, Kap. 9.2]	$\begin{cases} O(n) \text{ forventet} \\ O(n^2) \text{ worst-case} \end{cases}$
Deterministic-Select [CLRS, Kap. 9.3]	$O(n)$ worst-case

Randomized-Select:

Find det *i'te mindste element* i $A[p..r]$ $(1 \le i \le r\text{-}p\text{+}1)$




```
RANDOMIZED-SELECT(A, p, r, i)
if p == r
    return A[p]
q = RANDOMIZED-PARTITION(A, p, r)
k = q − p + 1
if i == k            // the pivot value is the answer
    return A[q]
elseif i < k
    return RANDOMIZED-SELECT(A, p, q − 1, i)
else return RANDOMIZED-SELECT(A, q + 1, r, i − k)
```

Randomized-Select 15

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
18	10	5	4	20	11	22	15	3	17	14	16	19	8	6	21	7	13	9	12	23	2	1
1	10	5	4	20	11	22	15	3	17	14	16	19	8	6	21	7	13	9	12	23	2	18
																	pivot					
	10	5	4	11	15	3	17	14	16	8	6	7	13	9	12	2	18	20	21	23	19	22
	2	5	4	11	15	3	17	14	16	8	6	7	13	9	12	10						
		5	4	3	8	6	7	9	10	15	11	17	13	14	12	16						
										15	11	13	14	12	16	17						
										11	12	13	14	15								
												13	14	15								

Randomized-Select

- **Randomiseret** algoritme (vælger pivot tilfældig)
  - pivot vælges i midterste del med en vis sandsynlighed
- Eksempel på **del-og-kombiner**
  - kun 1 mindre delproblem løses rekursivt
  - hele tiden bruges i opdelingen
    (kombination returnerer blot resultatet fra rekursionen)
- Tid: worst-case $O(n^2)$, **forventet $\mathbf{O(n)}$**
  - Analysen kan *ikke* anvende Master teoremet

Deterministic-Select

- Samme idé som Randomized-Select
  - Vælg et element som pivot
  - Opdel m.h.t. pivot
  - Lav højst ét rekursivt kald på dem der er < eller > pivot
- Ny idé
  - Rekursivt brug Select til at finde god pivot
- Analyse
  - Del-og-kombiner

$$T(n) \leq T(a \cdot n) + T(b \cdot n) + c \cdot n$$

  - Kan ikke bruge Master teoremet ☹

Deterministic-Select

SELECT(A, i)

små input

if $|A| \leq 5$
    sort A and return i'th element

beregn pivot

partition A into $G_1,\ldots,G_{\lceil n/5 \rceil}$ where $|G_i| \leq 5$
medians $= \{ g_i \mid g_i$ median of $G_i \}$
pivot = SELECT(*medians*, $\lfloor |medians|/2 \rfloor$)

max 1 rekursivt kald (som randomized select)

partition A w.r.t. *pivot* into $A_<$, $A_=$ and $A_>$
if $i < |A_<|$
    return SELECT($A_<$, i)
if $i \geq |A_<| + |A_=|$
    return SELECT($A_>$, $i - |A_<| - |A_=|$)
return *pivot*

Eksempel

A = 30, 37, 91, 78, 34,
76, 22, 72, 99, 63,
57, 57, 83, 97, 78,
44, 3, 25, 44, 86,
44, 82, 52, 26, 53,
90, 70, 17, 9, 56,
76, 89, 9, 37, 39,
80, 84, 23, 42, 97,
72, 26

betragt input som $\lceil n/5 \rceil$ grupper

G_1	G_2	G_3	G_4	G_5	···			$G_{\lceil n/5 \rceil}$
30	76	57	44	44	90	76	80	
37	22	57	3	82	70	89	84	72
91	72	83	25	52	17	9	23	26
78	99	97	44	26	9	37	42	
34	63	78	86	53	56	39	97	

sorter grupperne hver for sig

rekursivt find medianen

pivot = median(*medians*)

30	22	57	3	26	9	9	23	
34	63	57	25	44	17	37	42	26
37	72	78	44	52	56	39	80	72
78	76	83	44	53	70	76	84	
91	99	97	86	82	90	89	97	

medians

Kvaliteten af *pivot* ?

$\leq$ *pivot*

grupperne permuteret så medianerne er opdelt m.h.t. *pivot*

30	3	26	9	9	22	57	23	
34	25	44	37	17	63	57	42	26
37	44	52	39	56	72	78	80	72
78	44	53	76	70	76	83	84	
91	86	82	89	90	99	97	97	

$\sim \frac{3}{10}$ af A

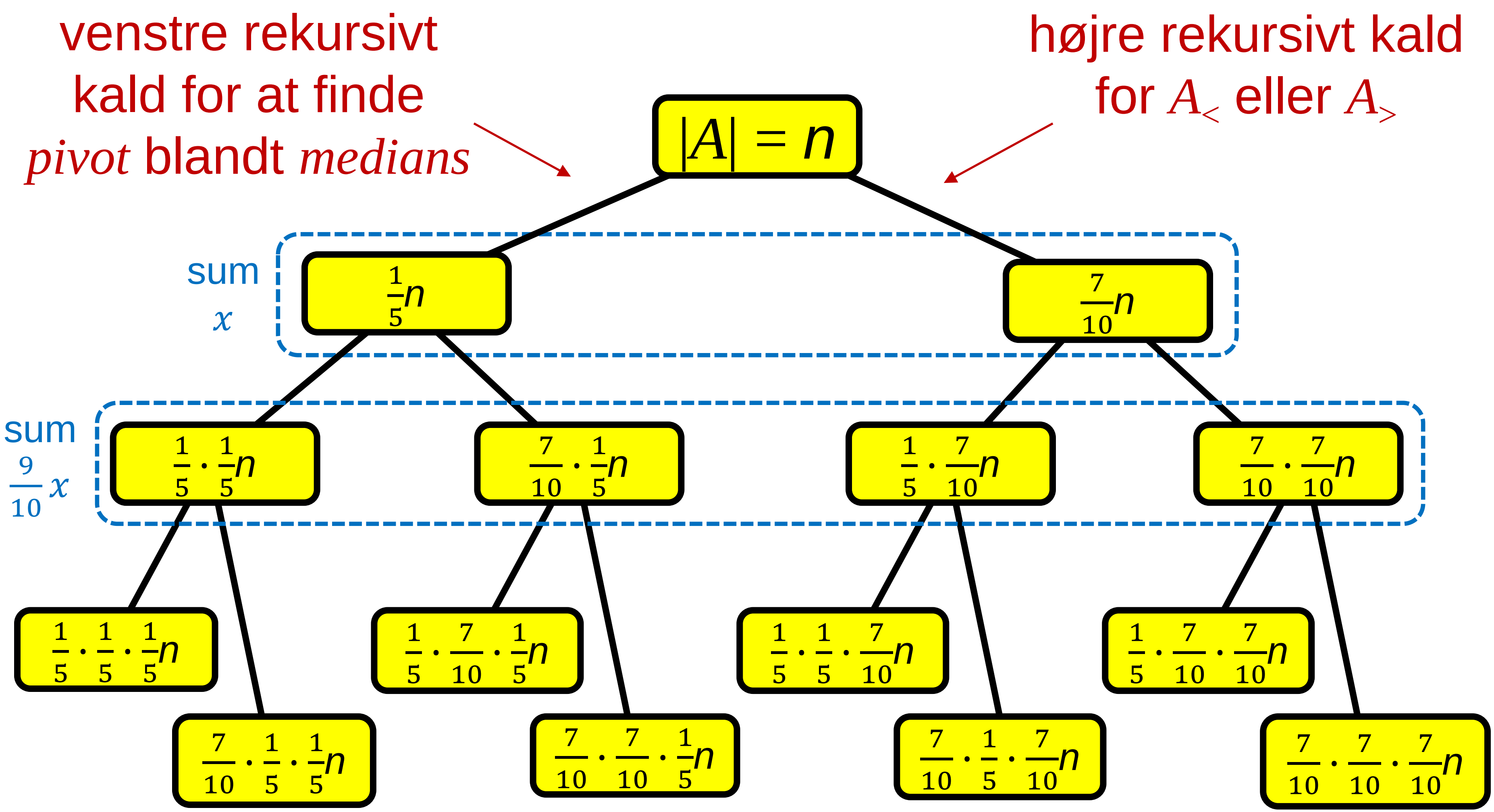
$\geq$ *pivot*

Hvor stor er $A_<$ maksimalt ?

a) ~ $3/10 \cdot |A|$
b) ~ $1/4 \cdot |A|$
c) ~ $1/2 \cdot |A|$
d) ~ $7/10 \cdot |A|$
e) ~ $3/4 \cdot |A|$
f) Ved ikke

$A_<$ er alle elementerne i A som er mindre end pivot elementet

Rekursionstræ SELECT(A, i)



Note: Beviset ignorerer at der til de rekursive kald kan være O(1) ekstra elementer når n ikke kan dividers med 5 og 10

Analyse

rekursivt fald for at bestemme *pivot*

rekursivt kald for $A_<$ eller $A_>$

tid for at finde medianen af hver af grupperne og at lave opdelingen i $A_<$, $A_=$ og $A_>$

$$T(n) \leq \begin{cases} T\left(\frac{n}{5}\right) + T\left(\frac{7n}{10}\right) + c \cdot n & \text{for } n > 5 \\ c & \text{for } n \leq 5 \end{cases}$$

tid for at sortere
≤ fem elementer

Note: Beviset ignorer at der til de rekursive kald kan være O(1) ekstra elementer når n ikke kan dividers med 5 og 10

Analyse

$$T(n) \le \begin{cases} T\left(\frac{n}{5}\right) + T\left(\frac{7n}{10}\right) + c \cdot n & \text{for } n > 5 \\ c & \text{for } n \le 5 \end{cases}$$

Bemærk i rekursionstræet er summen af størrelserne i dybde $i + 1$ højst $\frac{1}{5} + \frac{7}{10} = \frac{9}{10}$ gange størrelsen i dybde i

$$T(n) \le \sum_{i=0}^{\infty} c \cdot n \cdot \left(\frac{9}{10}\right)^i = c \cdot n \cdot \frac{1}{1 - \frac{9}{10}} = 10 \cdot c \cdot n$$

Note: Beviset ignorer at der til de rekursive kald kan være O(1) ekstra elementer når n ikke kan dividers med 5 og 10

Præcis Analyse

(tættere analyse end CLRS)

$|medians|$ $\qquad\qquad$ max{ $|A_<|$, $|A_>|$ }

$$T(n) \leq \begin{cases} T\left(\left\lceil \frac{n}{5} \right\rceil\right) + T\left(n - 3\left\lceil \frac{\lceil n/5 \rceil}{2} \right\rceil + 2\right) + c \cdot n & \text{for } n > 5 \\ c & \text{for } n \leq 5 \end{cases}$$

Løsning $\quad T(n) \leq c \cdot \max\{1, 10n - 30\}$

Bevis: For $1 \leq n \leq 15$ udregn rekursionsligningen...

n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
$T(n)/c \leq$	1	1	1	1	1	8	16	25	35	46	28	38	49	61	44
$\max\{1, 10n - 30\}$	1	1	1	10	20	30	40	50	60	70	80	90	100	110	120

(fortsættes)

Præcis Analyse (fortsat)

For $n \geq 16$ bevis ved **induktion**.

Induktionshypotese (antag at vi allerede har bevist)

$$T(k) \leq \mathrm{c} \cdot \max\{1, 10k - 30\} \text{ for } 1 \leq k \leq n-1.$$

Induktionsskridt (vis for n)

$$
\begin{aligned}
T(n) &\leq T\left(\left\lceil\frac{n}{5}\right\rceil\right) + T\left(n - 3\left\lceil\frac{\left\lceil\frac{n}{5}\right\rceil}{2}\right\rceil + 2\right) + c \cdot n \\
&\leq c \cdot \left(10\left\lceil\frac{n}{5}\right\rceil - 30 + 10\left(n - 3\left\lceil\frac{\left\lceil\frac{n}{5}\right\rceil}{2}\right\rceil + 2\right) - 30 + n\right) \\
&\leq c \cdot \left(10\left(\frac{n}{5} + 1\right) + 10\left(n - 3\frac{\left(\frac{n}{5}\right)}{2} + 2\right) - 60 + n\right) \\
&= c \cdot (10n - 30)
\end{aligned}
$$

□

Worst-case antal sammenligninger for Select for $n = 5$?

a) 1
b) 2
c) 3
d) 4
e) 5
f) 6
g) 7
h) 8
i) 9
j) Ved ikke

Endnu mere Præcis Analyse : # Sammenligninger

beregne *medians*

beregne $|A_<|$, $|A_=|$ og $|A_>|$

$$T(n) \leq \begin{cases} T\left(\left\lceil \frac{n}{5} \right\rceil\right) + T\left(n - 3\left\lceil \frac{\lceil n/5 \rceil}{2} \right\rceil + 2\right) + \frac{7}{5}n + n - 1 & \text{for } n > 5 \\ \begin{array}{c|ccccc} n & 1 & 2 & 3 & 4 & 5 \\ \hline T(n) \leq & 0 & 1 & 3 & 5 & 7 \end{array} & \text{for } n \leq 5 \end{cases}$$

Løsning $\quad T(n) \leq \max\{n, 24n - 72\}$

Bevis: $n \leq 15$ check manuelt. $n \geq 16$ som før ved induktion med $c = 12/5$.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$T(n) \leq$	0	1	3	5	7	21	38	57	79	103	66	88	112	138	104
$\max\{n, 24n - 72\}$	1	2	3	24	48	72	96	120	144	168	192	216	240	264	268

Selektion

Algoritme	Tid
Randomized-Select [CLRS, Kap. 9.2] Hoare 1961	$O(n)$ forventet $O(n^2)$ worst-case
Deterministic-Select [CLRS, Kap. 9.3] Blum et al. 1973	$O(n)$ worst-case
Median worst-case sammenligninger Dor, Zwick 1995, 1996	$\leq 2.95n$ $\geq (2+\varepsilon)n$ $\quad \varepsilon \approx 2^{-80}$