

Médiane en temps linéaire

Réfé. Cormen

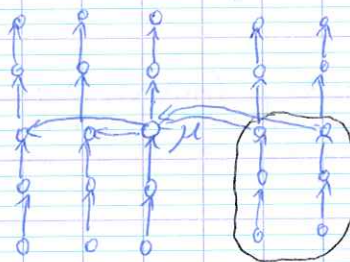
Th: On peut trouver la médiane dans une liste non triée en temps linéaire

Démo • On généralise le problème de la médiane en le problème de $\text{SELECTION}(i)$ et l'idée est d'utiliser le paradigme diviser pour régner pour construire l'algo suivant:

$\text{SELECTION}(i, L)$

```
g ← CHOIX_PIVOT(L)
S- := {x ∈ L, x < g}
S+ := {x ∈ L, g < x}
SI |S-| > i ALORS RENVoyer SELECTION(i, S-)
SINON SI |S+| < n - i ALORS RENVoyer g
SINON RENVoyer SELECTION(i - |S-|, S+)
```

- On veut trouver un choix de pivot qui sépare à peu près en deux S_+ et S_- pour éviter le pire cas du choix randomisé, mais on veut que ce pivot soit facile à calculer (sinon autant calculer la médiane directement!), i.e. en $O(n)$
- On regroupe les éléments cinq par cinq : $S_1, \dots, S_{n/5}$ et on conserve dans chacun le milieu : $\mu_1, \dots, \mu_{n/5}$ et finalement on appelle l'algorithme récursivement pour trouver la médiane μ de $\mu_1, \dots, \mu_{n/5}$. Combien d'éléments au moins dans S_+ et S_- ?



$$|S_+| \geq 3 \left(\left\lceil \frac{1}{2} \left\lceil \frac{n}{5} \right\rceil \right\rceil - 2 \right) \geq \frac{3n}{10} - 6$$

3 éls plus
grand par paquet

moins des
paquets

derniers paquets
si $n \neq 0 \pmod{5}$

De même, on a $|S_-| \leq \frac{7n}{10} - 6$

Ainsi $|S_-|, |S_+| \leq \frac{7n}{10} + 6$

- Si on note $T(n)$ la complexité de `selection(i)`, on établit la relation de récurrence suivante : $T(n) \leq T(\lceil \frac{n}{5} \rceil) + T(\frac{7n}{10} + 6) + O(n)$

qu'on ne peut pas résoudre avec le MT car les appels récursifs ne sont pas de même taille.

- Montrons que $T(n) \leq cn$ pour c bien choisi :

$$T(n) \leq c \lceil \frac{n}{5} \rceil + c \left(\frac{7n}{10} + 6 \right) + an$$

$$\leq \frac{cn}{5} + c + \frac{7cn}{10} + 6c + an$$

$$\leq \frac{9cn}{10} + 7c + an = cn + \left(7c + an - \frac{cn}{10} \right)$$

on veut donc $7c + an - \frac{cn}{10} \leq 0$

soit $10an \leq c(n - 70)$

$$c \geq \frac{10an}{n-70} \quad \text{pour } n > 70$$

or $\frac{n}{n-70} \leq 2$ pour $n \geq 140$, donc en prenant $c \geq 20a$,

on obtient l'inégalité voulue et donc $T(n) = \Theta(n)$.

□

Rq: on aurait aussi pu faire des paquets de 7 mais pas des paquets de 3 car la taille de S_+ et S_- seraient trop grandes