Divide and conquer - part 2

September 20, 2017

Overview

Selection in worst-case $O(n)$

- $\mathrm{Sel}(A, n, k)$
  1. Divide A into (disjoint) sets of size 5: $F_1, F_2, \ldots, F_{\frac{n}{5}}$.
  $$(\underbrace{a_1, \ldots, a_5}_{F_1}, \ldots, \underbrace{a_{n-4}, \ldots, a_n}_{F_{\frac{n}{5}}})$$
  2. Find the median of each of these sets. Let $b_i = \mathrm{median}(F_i)$ and denote: $B = (b_1, b_2, \ldots, b_{\frac{n}{5}})$
  3. Find the median of B (recursively):
  $$x = \mathrm{Sel}(B, \frac{n}{5}, \frac{n}{10})$$
  4. Find the sets:
  $$C = \{a_i \in A \mid a_i < x\} \quad , \quad D = \{a_i \in A \mid a_i > x\}$$
  5. $$\mathrm{Sel}(A, n, k) = \begin{cases} Sel(C, |C|, k), & |C| \geq k; \\ x, & |C| = k - 1; \\ Sel(D, |D|, k - |C| - 1), & |C| \leq k - 2. \end{cases}$$

Selection in worst-case $O(n)$

- $\mathrm{Sel}(A, n, k)$
  1. Divide A into (disjoint) sets of size 5: $F_1, F_2, \ldots, F_{\frac{n}{5}}$.

$$(\underbrace{a_1, \ldots, a_5}_{F_1}, \ldots, \underbrace{a_{n-4}, \ldots, a_n}_{F_{\frac{n}{5}}})$$

  2. Find the median of each of these sets.
     Let $b_i = \mathrm{median}(F_i)$ and denote: $B = (b_1, b_2, \ldots, b_{\frac{n}{5}})$
  3. Find the median of B (recursively):

$$x = \mathrm{Sel}(B, \frac{n}{5}, \frac{n}{10})$$

  4. Find the sets:

$$C = \{a_i \in A \mid a_i < x\} \quad , \quad D = \{a_i \in A \mid a_i > x\}$$

  5. $$\mathrm{Sel}(A, n, k) = \begin{cases} Sel(C, |C|, k), & |C| \geq k; \\ x, & |C| = k-1; \\ Sel(D, |D|, k - |C| - 1), & |C| \leq k-2. \end{cases}$$

Selection in worst-case $O(n)$

- $\text{Sel}(A, n, k)$
  1. Divide A into (disjoint) sets of size 5: $F_1, F_2, ..., F_{\frac{n}{5}}$.

$$(\underbrace{a_1, ..., a_5}_{F_1}, ..., \underbrace{a_{n-4}, ..., a_n}_{F_{\frac{n}{5}}})$$

  2. Find the median of each of these sets.
     Let $b_i = \text{median}(F_i)$ and denote: $B = (b_1, b_2, ..., b_{\frac{n}{5}})$
  3. Find the median of B (recursively):

$$x = \text{Sel}(B, \frac{n}{5}, \frac{n}{10})$$

  4. Find the sets:

$$C = \{a_i \in A \mid a_i < x\} \quad , \quad D = \{a_i \in A \mid a_i > x\}$$

  5. $$\text{Sel}(A, n, k) = \begin{cases} Sel(C, |C|, k), & |C| \geq k; \\ x, & |C| = k - 1; \\ Sel(D, |D|, k - |C| - 1), & |C| \leq k - 2. \end{cases}$$

Selection in worst-case $O(n)$

- $\mathrm{Sel}(A, n, k)$
  1. Divide A into (disjoint) sets of size 5: $F_1, F_2, \ldots, F_{\frac{n}{5}}$.

     $$(\underbrace{a_1, \ldots, a_5}_{F_1}, \ldots, \underbrace{a_{n-4}, \ldots, a_n}_{F_{\frac{n}{5}}})$$

  2. Find the median of each of these sets.
     Let $b_i = \mathrm{median}(F_i)$ and denote: $B = (b_1, b_2, \ldots, b_{\frac{n}{5}})$
  3. Find the median of B (recursively):

     $$x = \mathrm{Sel}(B, \frac{n}{5}, \frac{n}{10})$$

  4. Find the sets:

     $$C = \{a_i \in A \mid a_i < x\} \quad , \quad D = \{a_i \in A \mid a_i > x\}$$

  5. $\mathrm{Sel}(A, n, k) = \begin{cases} Sel(C, |C|, k), & |C| \geq k; \\ x, & |C| = k-1; \\ Sel(D, |D|, k - |C| - 1), & |C| \leq k-2. \end{cases}$

Selection in worst-case $O(n)$

- $\mathrm{Sel}(A, n, k)$
  1. Divide A into (disjoint) sets of size 5: $F_1, F_2, ..., F_{\frac{n}{5}}$.

$$(\underbrace{a_1, ..., a_5}_{F_1}, ..., \underbrace{a_{n-4}, ..., a_n}_{F_{\frac{n}{5}}})$$

  2. Find the median of each of these sets.
     Let $b_i = \mathrm{median}(F_i)$ and denote: $B = (b_1, b_2, ..., b_{\frac{n}{5}})$
  3. Find the median of B (recursively):

$$x = \mathrm{Sel}(B, \frac{n}{5}, \frac{n}{10})$$

  4. Find the sets:

$$C = \{a_i \in A \mid a_i < x\} \quad , \quad D = \{a_i \in A \mid a_i > x\}$$

  5. $\mathrm{Sel}(A, n, k) = \begin{cases} Sel(C, |C|, k), & |C| \geq k; \\ x, & |C| = k-1; \\ Sel(D, |D|, k - |C| - 1), & |C| \leq k-2. \end{cases}$

Selection in worst-case $O(n)$

- $\mathrm{Sel}(A, n, k)$
  1. Divide A into (disjoint) sets of size 5: $F_1, F_2, \ldots, F_{\frac{n}{5}}$.

$$(\underbrace{a_1, \ldots, a_5}_{F_1}, \ldots, \underbrace{a_{n-4}, \ldots, a_n}_{F_{\frac{n}{5}}})$$

  2. Find the median of each of these sets.
     Let $b_i = \mathrm{median}(F_i)$ and denote: $B = (b_1, b_2, \ldots, b_{\frac{n}{5}})$
  3. Find the median of B (recursively):

$$x = \mathrm{Sel}(B, \frac{n}{5}, \frac{n}{10})$$

  4. Find the sets:

$$C = \{a_i \in A \mid a_i < x\} \quad , \quad D = \{a_i \in A \mid a_i > x\}$$

  5. $$\mathrm{Sel}(A, n, k) = \begin{cases} Sel(C, |C|, k), & |C| \geq k; \\ x, & |C| = k-1; \\ Sel(D, |D|, k - |C| - 1), & |C| \leq k-2. \end{cases}$$

Complexity of Sel(A, n, k)

- Denote the complexity of Sel(A, n, k) by $C_{Sel(n,k)}$
  Denote $S(n) = \max_k C_{Sel(n,k)}$
- We will need the following fact to show that $S(n) = O(n)$.
  Claim: For the sets C and D (from stage 4 of the algorithm) we have:

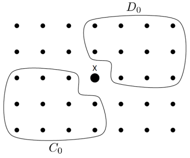
$$|C|, |D| \leq \frac{7n}{10}$$

Complexity of Sel(A, n, k)

- Denote the complexity of Sel(A, n, k) by $C_{Sel(n,k)}$
  Denote $S(n) = \max_k C_{Sel(n,k)}$
- We will need the following fact to show that $S(n) = O(n)$.
  Claim: For the sets C and D (from stage 4 of the algorithm) we have:

$$|C|, |D| \leq \frac{7n}{10}$$

Complexity of Sel(A, n, k)



Proof:(of the claim) Let:

$$C_0 = \bigcup_{i \text{ s.t } b_i < x} \{y \in F_i \mid y \leq b_i\}$$

$$D_0 = \bigcup_{i \text{ s.t } b_i > x} \{y \in F_i \mid y \geq b_i\}$$

Clearly $C_0 \subseteq C$ and $D_0 \subseteq D$.
Since $|C_0|, |D_0| \geq \frac{3n}{10}$, it follows that:

$$|C| \leq n - \frac{3n}{10} = \frac{7n}{10} \quad , \quad |D| \leq n - \frac{3n}{10} = \frac{7n}{10}$$

Complexity of Sel(A, n, k)

- Let us now "count the number of comparisons in each step of the algorithm.
  1. Step 1: No comparisons
  2. Step 2: We can find the median of 5 elements using 10 comparisons, so we need $\frac{10n}{5} = 2n$ comparisons.
  3. Step 3: $C_{Sel(\frac{n}{5}, \frac{n}{10})} < S(\frac{n}{5})$
  4. Step 4: n comparisons. Each element against x.
  5. Step 5: We use the algorithm Sel on sets smaller than $\frac{7n}{10}$ so the complexity is less than $S(\frac{7n}{10})$.
- It follows that:

$$C_{Sel(n,k)} \leq 2n + S\left(\frac{n}{5}\right) + n + S\left(\frac{7n}{10}\right)$$

- Which means that: $S(n) \leq 3n + S\left(\frac{n}{5}\right) + S\left(\frac{7n}{10}\right)$

Complexity of Sel(A, n, k)

- Let us now "count the number of comparisons in each step of the algorithm.
  1. Step 1: No comparisons
  2. Step 2: We can find the median of 5 elements using 10 comparisons, so we need $\frac{10n}{5} = 2n$ comparisons.
  3. Step 3: $C_{Sel(\frac{n}{5}, \frac{n}{10})} < S(\frac{n}{5})$
  4. Step 4: n comparisons. Each element against x.
  5. Step 5: We use the algorithm Sel on sets smaller than $\frac{7n}{10}$ so the complexity is less than $S(\frac{7n}{10})$.
- It follows that:

$$C_{Sel(n,k)} \leq 2n + S\left(\frac{n}{5}\right) + n + S\left(\frac{7n}{10}\right)$$

- Which means that: $S(n) \leq 3n + S\left(\frac{n}{5}\right) + S\left(\frac{7n}{10}\right)$

Complexity of Sel(A, n, k)

- Let us now "count the number of comparisons in each step of the algorithm.
  1. Step 1: No comparisons
  2. Step 2: We can find the median of 5 elements using 10 comparisons, so we need $\frac{10n}{5} = 2n$ comparisons.
  3. Step 3: $C_{Sel(\frac{n}{5}, \frac{n}{10})} < S(\frac{n}{5})$
  4. Step 4: n comparisons. Each element against x.
  5. Step 5: We use the algorithm Sel on sets smaller than $\frac{7n}{10}$ so the complexity is less than $S(\frac{7n}{10})$.
- It follows that:

$$C_{Sel(n,k)} \leq 2n + S\left(\frac{n}{5}\right) + n + S\left(\frac{7n}{10}\right)$$

- Which means that: $S(n) \leq 3n + S\left(\frac{n}{5}\right) + S\left(\frac{7n}{10}\right)$

Complexity of Sel(A, n, k)

- Let us now "count the number of comparisons in each step of the algorithm.
  1. Step 1: No comparisons
  2. Step 2: We can find the median of 5 elements using 10 comparisons, so we need $\frac{10n}{5} = 2n$ comparisons.
  3. Step 3: $C_{Sel(\frac{n}{5}, \frac{n}{10})} < S(\frac{n}{5})$
  4. Step 4: n comparisons. Each element against x.
  5. Step 5: We use the algorithm Sel on sets smaller than $\frac{7n}{10}$ so the complexity is less than $S(\frac{7n}{10})$.
- It follows that:

$$C_{Sel(n,k)} \leq 2n + S\left(\frac{n}{5}\right) + n + S\left(\frac{7n}{10}\right)$$

- Which means that: $S(n) \leq 3n + S\left(\frac{n}{5}\right) + S\left(\frac{7n}{10}\right)$

Complexity of Sel(A, n, k)

- Let us now "count the number of comparisons in each step of the algorithm.
  1. Step 1: No comparisons
  2. Step 2: We can find the median of 5 elements using 10 comparisons, so we need $\frac{10n}{5} = 2n$ comparisons.
  3. Step 3: $C_{Sel(\frac{n}{5}, \frac{n}{10})} < S(\frac{n}{5})$
  4. Step 4: n comparisons. Each element against x.
  5. Step 5: We use the algorithm Sel on sets smaller than $\frac{7n}{10}$ so the complexity is less than $S(\frac{7n}{10})$.
- It follows that:

$$C_{Sel(n,k)} \leq 2n + S\left(\frac{n}{5}\right) + n + S\left(\frac{7n}{10}\right)$$

- Which means that: $S(n) \leq 3n + S\left(\frac{n}{5}\right) + S\left(\frac{7n}{10}\right)$

Complexity of Sel(A, n, k)

- Let us now "count the number of comparisons in each step of the algorithm.
  1. Step 1: No comparisons
  2. Step 2: We can find the median of 5 elements using 10 comparisons, so we need $\frac{10n}{5} = 2n$ comparisons.
  3. Step 3: $C_{Sel(\frac{n}{5}, \frac{n}{10})} < S(\frac{n}{5})$
  4. Step 4: n comparisons. Each element against x.
  5. Step 5: We use the algorithm Sel on sets smaller than $\frac{7n}{10}$ so the complexity is less than $S(\frac{7n}{10})$.
- It follows that:

$$C_{Sel(n,k)} \leq 2n + S\left(\frac{n}{5}\right) + n + S\left(\frac{7n}{10}\right)$$

- Which means that: $S(n) \leq 3n + S\left(\frac{n}{5}\right) + S\left(\frac{7n}{10}\right)$

Complexity of Sel(A, n, k)

- Let us now "count the number of comparisons in each step of the algorithm.
    1. Step 1: No comparisons
    2. Step 2: We can find the median of 5 elements using 10 comparisons, so we need $\frac{10n}{5} = 2n$ comparisons.
    3. Step 3: $C_{Sel(\frac{n}{5}, \frac{n}{10})} < S(\frac{n}{5})$
    4. Step 4: n comparisons. Each element against x.
    5. Step 5: We use the algorithm Sel on sets smaller than $\frac{7n}{10}$ so the complexity is less than $S(\frac{7n}{10})$.
- It follows that:

$$C_{Sel(n,k)} \leq 2n + S\left(\frac{n}{5}\right) + n + S\left(\frac{7n}{10}\right)$$

- Which means that: $S(n) \leq 3n + S\left(\frac{n}{5}\right) + S\left(\frac{7n}{10}\right)$

Complexity of Sel(A, n, k)

- Let us now "count the number of comparisons in each step of the algorithm.
  1. Step 1: No comparisons
  2. Step 2: We can find the median of 5 elements using 10 comparisons, so we need $\frac{10n}{5} = 2n$ comparisons.
  3. Step 3: $C_{Sel(\frac{n}{5}, \frac{n}{10})} < S(\frac{n}{5})$
  4. Step 4: n comparisons. Each element against x.
  5. Step 5: We use the algorithm Sel on sets smaller than $\frac{7n}{10}$ so the complexity is less than $S(\frac{7n}{10})$.
- It follows that:

$$C_{Sel(n,k)} \leq 2n + S\left(\frac{n}{5}\right) + n + S\left(\frac{7n}{10}\right)$$

- Which means that: $S(n) \leq 3n + S\left(\frac{n}{5}\right) + S\left(\frac{7n}{10}\right)$

Complexity of Sel(A, n, k)

- We will show by induction over n that:

$$S(n) \leq \frac{3}{1 - (\frac{1}{5} + \frac{7}{10})} \cdot n = 30n$$

- Proof: We have:

$$S(n) \leq 3n + S\left(\overbrace{\frac{n}{5}}^{<n}\right) + S\left(\overbrace{\frac{7n}{10}}^{<n}\right)$$

so by the induction assumption we get:

$$S(n) \leq 3n + 30 \cdot \frac{n}{5} + 30 \cdot \frac{7n}{10} = (3 + 6 + 21)n = 30n \quad \blacksquare$$

Complexity of Sel(A, n, k)

- We will show by induction over n that:

$$S(n) \leq \frac{3}{1 - (\frac{1}{5} + \frac{7}{10})} \cdot n = 30n$$

- Proof: We have:

$$S(n) \leq 3n + S\left(\overbrace{\frac{n}{5}}^{<n}\right) + S\left(\overbrace{\frac{7n}{10}}^{<n}\right)$$

so by the induction assumption we get:

$$S(n) \leq 3n + 30 \cdot \frac{n}{5} + 30 \cdot \frac{7n}{10} = (3 + 6 + 21)n = 30n \quad \blacksquare$$