Analyzing the Average Complexity of QuickSelect

Karl Stroetmann

May 10, 2022

1 The Algorithm QuickSelect

The goal of QuickSelect is to compute the `k` smallest elements of a given list `L`, i.e. the function `quickSelect` takes two arguments:

(a) `L` is a list of numbers,

(b) `k` is a natural number such that $2 \leq$ `k` $\leq$ `len(L)`.

The function call `quickSelect(L, k)` returns a list of length `k` that contains the `k` smallest elements of `L`, i.e. it satisfies the following specification:

- `len(quickSelect(L, k)) = k`,
- `set(quickSelect(L, k)) = set(sorted(L)[:k])`.

  Note that the list returned by `quickSelect(L, k)` is not required to be sorted. It can contain the `k` smallest elements of `L` in any order.

The function can be implemented recursively via the following equations:

1. $\texttt{len}(L) < k \rightarrow \texttt{quickSelect}(L, k) = \Omega$,

   because if the length of `L` is less than `k`, then there is no way to select the `k` smallest elements from `L`.

2. $\texttt{len}(L) = k \rightarrow \texttt{quickSelect}(L, k) = L$,

   because if the list `L` has exactly `k` elements, then `L` itself is a list containing the `k` smallest elements of `L`.

3. Otherwise we assume that `L = [x] + R` and partition `L` as in QuickSort, i.e. we define

   $$S := [y \in L \mid y \leq x] \quad \text{and} \quad B := [y \in L \mid y > x].$$

   Then there are three cases:

   (a) $k \leq \texttt{len}(S) \rightarrow \texttt{quickSelect}(L, k) = \texttt{quickSelect}(S, k)$,

   (b) $k = \texttt{len}(S) + 1 \rightarrow \texttt{quickSelect}(L, k) = S + [x]$,

   (c) $k > \texttt{len}(S) + 1 \rightarrow \texttt{quickSelect}(L, k) = S + [x] + \texttt{quickSelect}(B, k - \texttt{len}(S) - 1)$.

2 Analysis of the Average Complexity of `quickSelect`

Let us define $d_{n,k}$ as the average number of comparisons of list elements that are performed when `quickSelect(L, k)` is evaluated with a list `L` of length n. We proceed to construct a recurrence equation for $d_{n,k}$.

(a) $d_{k,k} = 0$,

(b) $d_{n+1,k} = n + \frac{1}{n+1} \cdot \left(\sum_{i=k}^{n} d_{i,k} + \sum_{i=0}^{k-2} d_{n-i,k-i-1} \right)$.

We will not be able to solve this recurrence equation exactly, but we will be able to establish an upper bound for $d_{n,k}$ that is independent of k. To this end, we need the following Lemma.

Lemma 1 Assume that $n \in \mathbb{N}$ and the function $f : \mathbb{R} \to \mathbb{R}$ is defined as

$$f(x) = 4 \cdot (x-1) \cdot (n+1-x)$$

Then we have

$$\forall x \in \mathbb{R} : f(x) \leq n^2.$$

Proof: The function f is quadratic in x and the coefficient of x^2 is negative. By completing the square we find the following chain of equations and inequations:

$$\begin{array}{lcl}
f(x) & = & 4 \cdot (x-1) \cdot (n+1-x) \\
 & = & 4 \cdot \big(n \cdot x + x - x^2 - n - 1 + x\big) \\
 & = & -4 \cdot \big(x^2 - (n+2) \cdot x + n + 1\big) \\
 & = & -4 \cdot \Big(x^2 - (n+2) \cdot x + \big(\frac{n+2}{2}\big)^2 + n + 1 - \big(\frac{n+2}{2}\big)^2\Big) \\
 & = & -4 \cdot \Big(x - \big(\frac{n+2}{2}\big)\Big)^2 - 4 \cdot \Big(n + 1 - \big(\frac{n+2}{2}\big)^2\Big) \\
 & \leq & -4 \cdot \Big(n + 1 - \big(\frac{n+2}{2}\big)^2\Big) \\
 & = & -4 \cdot n - 4 + (n+2)^2 \\
 & = & -4 \cdot n - 4 + n^2 + 4 \cdot n + 4 \\
 & = & n^2
\end{array}$$

Hence we have shown that $f(x) \leq n^2$. □

Theorem 2 For all $k \in \mathbb{N}$ such $k \geq 2$ and all $n \in \mathbb{N}$ such that $k \leq n$ we have

$$d_{n,k} \leq 4 \cdot n.$$

Hence the complexity of QuickSelect is linear in the length of the list and the constant of linearity is independent of k.

Proof: We prove the claim by induction on n.

Base case: $n = k$. In this case we have

$$d_{n,k} = d_{k,k} = 0 \leq 4 \cdot n. \ \checkmark$$

Induction step: $0, 1, \cdots, n \mapsto n+1$. Then we have the following chain of equations and inequations:

$$\begin{array}{lcl}
d_{n+1,k} & = & n + \frac{1}{n+1} \cdot \left(\sum_{i=k}^{n} d_{i,k} + \sum_{i=0}^{k-2} d_{n-i,k-i-1} \right) \\
 & \stackrel{\text{ih}}{\leq} & n + \frac{1}{n+1} \cdot \left(\sum_{i=k}^{n} 4 \cdot i + \sum_{i=0}^{k-2} 4 \cdot (n-i) \right)
\end{array}$$

In order to prove our claim we proceed to show that

$$\sum_{i=k}^{n} 4 \cdot i + \sum_{i=0}^{k-2} 4 \cdot (n-i) \leq 3 \cdot (n+1)^2$$

This claim is shown as follows:

$$\begin{aligned}
\sum_{i=k}^{n} 4 \cdot i + \sum_{i=0}^{k-2} 4 \cdot (n-i) &= 4 \cdot \left(\sum_{i=0}^{n} i - \sum_{i=0}^{k-1} i + (k-1) \cdot n - \sum_{i=0}^{k-2} i \right) \\
&= 4 \cdot \left(\sum_{i=0}^{n} i - 2 \cdot \sum_{i=0}^{k-2} i - (k-1) + (k-1) \cdot n \right) \\
&= 2 \cdot n \cdot (n+1) - 4 \cdot (k-2) \cdot (k-1) + 4 \cdot (k-1) \cdot (n-1)
\end{aligned}$$

In order to show that this expression is less or equal than $3 \cdot (n+1)^2$ we use Lemma 1 where we have shown that for $f(x) = 4 \cdot (x-1) \cdot (n+1-x)$ the inequation $f(x) \leq n^2$ holds for all $x \in \mathbb{R}$. We have

$$\begin{aligned}
& & f(k) &\leq n^2 \\
&\Rightarrow & 4 \cdot (k-1) \cdot (n+1-k) &\leq n^2 \\
&\Rightarrow & -4 \cdot (k-1) \cdot (k-n-1) &\leq n^2 \\
&\Rightarrow & -4 \cdot (k-1) \cdot \big((k-2) - (n-1)\big) &\leq n^2 \\
&\Rightarrow & -4 \cdot (k-2) \cdot (k-1) + 4 \cdot (k-1) \cdot (n-1) &\leq n^2 \\
&\Rightarrow & -4 \cdot (k-2) \cdot (k-1) + 4 \cdot (k-1) \cdot (n-1) &\leq (n+1)^2 \\
&\Rightarrow & 2 \cdot n \cdot (n+1) - 4 \cdot (k-2) \cdot (k-1) + 4 \cdot (k-1) \cdot (n-1) &\leq 3 \cdot (n+1)^2
\end{aligned}$$

Combining this with our previous result we now have

$$\begin{aligned}
d_{n+1,k} &= n + \frac{1}{n+1} \cdot \left(\sum_{i=k}^{n} d_{i,k} + \sum_{i=0}^{k-2} d_{n-i,k-i-1} \right) \\
&\overset{\text{ih}}{\leq} n + \frac{1}{n+1} \cdot \left(\sum_{i=k}^{n} 4 \cdot i + \sum_{i=0}^{k-2} 4 \cdot (n-i) \right) \\
&\leq n + \frac{1}{n+1} \cdot 3 \cdot (n+1)^2 \\
&\leq n + 3 \cdot (n+1) \\
&< (n+1) + 3 \cdot (n+1) \\
&= 4 \cdot (n+1) \quad \checkmark
\end{aligned}$$

Hence we have shown that $d_{n+1,k} \leq 4 \cdot (n+1)$ and the claim is proven by induction. $\square$