

Lecture 12: Runtime

Introduction

When designing data structures and algorithms, it is crucial to write efficient functions, particularly in terms of time complexity. The *runtime* of a function refers to the number of computational steps required to complete a task. Efficient functions minimize these steps.

How Computers Execute Code

To understand runtime, it's important to recognize how computers execute code. Every statement in a program is translated into *machine language commands*, which are fundamental instructions a computer understands. The number of machine instructions required to execute a statement determines its execution time, referred to as the *processing cost* of the statement. Therefore, a function's runtime is the sum of the processing cost of each statement multiplied by the number of times that statement executes during the function's execution.

Summation Properties

Summations (or *series*) help express the total runtime of an algorithm. A summation is written as:

$$\sum_{i=a}^b f(i) = f(a) + f(a+1) + \cdots + f(b)$$

where a and b define the summation's range and $f(i)$ is the function being summed. Useful summation properties include:

- **Splitting a Summation**

$$\sum_{i=a}^b f(i) = \sum_{i=a}^c f(i) + \sum_{i=c+1}^b f(i)$$

where $a \leq c < b$.

- **Sum/Difference & Scalar Products**

$$\sum_{i=a}^b c(f(i) \pm g(i)) = c \left(\sum_{i=a}^b f(i) \pm \sum_{i=a}^b g(i) \right)$$

where c is a number.

- **Sum of Ones**

$$\sum_{i=a}^b 1 = (b - a + 1)$$

- **Consecutive Integer Sum**

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}$$

where n is a positive integer.

These properties assist in analyzing the runtime of loops and recursive functions.

Runtime Function & Factor

A *runtime function* expresses an algorithm's execution time as a function of the input size (or *runtime factor*)—the specific aspect of the input that influences statement execution. It is computed as:

$$T(n) = \sum_{i=1}^m c_i \cdot t_i$$

where:

- m is the number of unique statements in the function,
- c_i is the processing cost of the i th statement,
- t_i is the number of times the i th statement executes,
- n represents the size of the input data.

Since exact processing costs vary across systems, abstract cost values are used.

Example:

<pre> 01 ulg C(ulg x) 02 { 03 ulg n = 0; 04 05 while(x > 0) 06 { 07 x = x / 10; 08 n += 1; 09 } 10 return n; 11 }</pre>	<pre> 01 bool P(ulg n) 02 { 03 for(ulg i = 2; i*i <= n; i += 1) 04 { 05 if(n % i == 0) 06 { 07 return false; 08 } 09 } 10 return (n > 1); 11 }</pre>	<pre> 01 void R(Array<int>& dt, ulg n, ulg m) 02 { 03 while(n < m) 04 { 05 int t = dt[n]; 06 dt[n] = dt[m]; 07 dt[m] = t; 08 n += 1; 09 m -= 1; 10 } 11 }</pre>
--	---	---

where ‘ulg’ is an abstract data type for unsigned long and *Array* is a container class for an array. The runtime analysis of each algorithm is:

- C() - The loop executes for each digit of *x*; hence, the runtime factor is the number of digits in the parameter.
- P() - The loop checks if *n* is divisible by integers at most its square root; hence, the runtime factor is the parameter’s value.
- R() - The loop swaps values between *n* and *m*; hence, the runtime factor is the difference of the integer parameters.

Worst-Case, Average-Case, & Best-Case Analysis

When analyzing runtime, different scenarios must be considered since the input varies the statement executions. The scenarios are:

- **Best-case:** the function executes the minimum number of steps.
- **Average-case:** the function executes an average number of steps.
- **Worst-case:** the function executes the maximum number of steps.

The worst-case scenario is often used because it provides an upper bound on runtime.

Example:

For each algorithm, its worst-case scenario is:

- C() - All cases are the same.
- P() - The worst-case scenario occurs when the parameter is prime.
- R() - The worst-case scenario occurs when integer parameters are the end indices of the array parameter.

Analyzing Function Runtime with a Runtime Table

Constructing a *runtime table*—a table that records the cost and number of executions (time) for each distinct statement in an algorithm—simplifies determining an algorithm’s runtime function. The cost of a statement is represented as an abstract value, typically 0 or 1, depending on the aspects of the analyzed algorithm. The time is expressed as either a fixed integer or a function of the algorithm’s runtime factor. When the time is an expression involving the runtime factor, the floor and ceiling functions are used to ensure proper calculations.

Example:

The runtime tables for each algorithm is:

C():			P():			R():		
line	cost	time	line	cost	time	line	cost	time
03	c_1	1	03	c_1	1	03	c_1	$\lceil \frac{n}{2} \rceil + 1$
05	c_2	$n + 1$	03	c_2	$\lfloor \sqrt{n} \rfloor$	05	c_2	$\lceil \frac{n}{2} \rceil$
07	c_3	n	05	c_3	$\lfloor \sqrt{n} \rfloor - 1$	06	c_3	$\lceil \frac{n}{2} \rceil$
08	c_4	n	07	c_4	0	07	c_4	$\lceil \frac{n}{2} \rceil$
10	c_5	1	03	c_5	$\lfloor \sqrt{n} \rfloor - 1$	08	c_5	$\lceil \frac{n}{2} \rceil$
			10	c_6	1	09	c_6	$\lceil \frac{n}{2} \rceil$

- The loop condition in Algorithm C() (line 05) is evaluated once for each digit of x , plus one additional time when the condition evaluates to false, resulting in a total of $n + 1$ evaluations. The loop body executes only when the condition is true, meaning its statements run n times—one less than the total condition evaluations. Therefore, assuming all statement costs are 1, the runtime function is:

$$T(n) = 3n + 3$$

- The loop condition in Algorithm P() (line 03) is evaluated once for each integer upto \sqrt{n} starting from 2 plus one additional time when the condition evaluates to false, resulting in a total of $\lfloor \sqrt{n} \rfloor$ evaluations. Since for the worst-case scenario, n is prime, the return statement on line 07 will never execute; whereas, all other statements in the body run $\lfloor \sqrt{n} \rfloor - 1$ times as expected. Therefore, assuming all statement costs are 1, the runtime function is:

$$T(n) = 3\lfloor \sqrt{n} \rfloor$$

- The loop condition in Algorithm R() (line 03) is evaluated for $\lceil \frac{n}{2} \rceil$ since both endpoints are adjusted by 1 for each cycle plus one additional time when the condition evaluates to false, resulting in a total of $\lceil \frac{n}{2} \rceil + 1$ evaluations. The body statements run $\lceil \frac{n}{2} \rceil$ times as expected. Therefore, assuming all statement costs are 1, the runtime function is:

$$T(n) = 6\left\lceil \frac{n}{2} \right\rceil + 1$$

Order of Growth & Big-O Notation

Asymptotic notations are used to compare algorithms by providing a mathematical boundary for their runtime functions. These notations define an upper bound, a lower bound, or both, helping to analyze the efficiency of algorithms as input size grows.

- **Big-Oh:** upper bound on runtime growth denoted $O(f(n))$ which is $\{g(n) : 0 < g(n) \leq cf(n) \text{ whenever } n \geq n_0 \text{ for some } c > 0, n_0 > 0\}$
- **Big-Omega:** lower bound on runtime growth denoted $\Omega(f(n))$ which is $\{g(n) : 0 < cf(n) \leq g(n) \text{ whenever } n \geq n_0 \text{ for some } c > 0, n_0 > 0\}$
- **Big-Theta:** tight bound (both upper and lower bound) denoted $\Theta(f(n))$ which is $\{g(n) : c_1f(n) \leq g(n) \leq c_2f(n) \text{ whenever } n \geq n_0 \text{ for some } c_2 \geq c_1 > 0, n_0 > 0\}$

Example:

- $O(1)$: constant time (independent of input size).
- $O(\log n)$: logarithmic time (e.g., binary search).
- $O(n)$: linear time (e.g., iterating over an array).
- $O(n^2)$: quadratic time (e.g., nested loops).
- $O(2^n)$: exponential time (e.g., brute-force recursive algorithms).

In addition, although a function may have multiple big-oh bounds, the tightest bound is preferred. For instance, $3n + 5 = O(n^2)$ is valid, but $3n + 5 = O(n)$ is tighter and more accurate.