SC1007
Data Structures and Algorithms

Analysis of Algorithms

Dr Liu Siyuan (syliu@ntu.edu.sg)

N4-02C-117a

Office Hour: Mon & Wed 4-5pm

Overview

Conduct complexity analysis of algorithms

- Time and space complexities
- Best-case, worst-case and average efficiencies
- Order of Growth
- Asymptotic notations
  - O notation
  - Ω notation (Omega)
  - Θ notation (Theta)
- Efficiency classes

Time and space complexities

- Analyze efficiency of an algorithm in two aspects
  - Time
  - Space

- Time complexity: the amount of time used by an algorithm
- Space complexity: the amount of memory units used by an algorithm

Time Complexity or Time Efficiency

1. Count the number of primitive operations in the algorithm

Time Complexity or Time Efficiency

1. Count the number of primitive operations in the algorithm

- Declaration: int x;
- Assignment: x =1;
- Arithmetic operations: +, -, *, /, % etc.
- Logic operations: ==, !=, >, <, &&, ||

These primitive operations take constant time to perform

Basically they (the time for each operation) are not related to the problem size

Time Complexity or Time Efficiency

1. Count the number of primitive operations in the algorithm

   i. Repetition Structure: for-loop, while-loop
   ii. Selection Structure: if/else statement, switch-case statement
   iii. Recursive functions

2. Express it in term of problem size (input size)

Algorithm 4 Fibonacci Sequence: A Simple Recursive Function

```
function Fibonacci_Recursive(n)
begin
if n<1 then
    return 0
if n==1 OR n==2 then
    return 1
return  Fibonacci_Recursive(n − 1)+Fibonacci_Recursive(n − 2)
end
```

Time Complexity or Time Efficiency

i. Repetition Structure: for-loop, while-loop

$j \leftarrow 1$	c_0
factorial $\leftarrow 1$	c_1
while $j \leq n$ **do**	
factorial $\leftarrow$ factorial $* j$	c_2
$j \leftarrow j + 1$	c_3

n iterations ➡ $n(c_2+c_3)$

$f(n) = c_0+c_1+n(c_2+c_3)$

The function increases linearly with n (problem size)

Time Complexity or Time Efficiency

i. Repetition Structure: for-loop, while-loop

```
for j ← 1, m do
    for k ← 1, n do
        sum ← sum + M[ j ][ k ]
```

c_1

n iterations

$n(c_1)$

m iterations

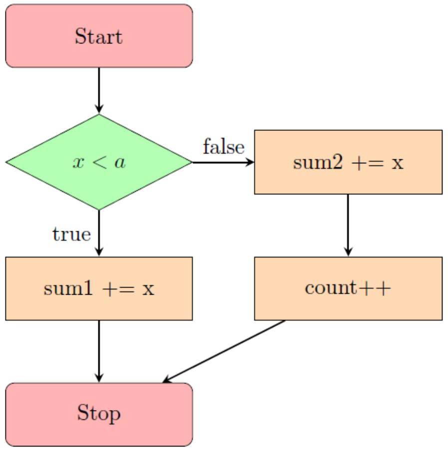
$m(n(c_1))$

The function increases quadratically with n if m==n

*Some constant time operations are ignored here.

Time Complexity or Time Efficiency

ii. Selection Structure: if/else statement, switch-case statement




```
if(x<a)
      sum1 += x;
else {
     sum2 += x;
     count ++;
     }
```

When x < a, only one primitive operation is executed
When x ≥ a, two primitive operations are executed

How do we analyze the time complexity?

1. Best-case analysis
2. Worst-case analysis
3. Average-case analysis

Time Complexity or Time Efficiency

ii. Selection Structure: if/else statement

```
if(x<a)
      sum1 += x;
else {
      sum2 += x;
      count ++;
      }
```

When $x < a$, only one primitive operation is executed
When $x \geq a$, two primitive operations are executed

How do we analyze the time complexity?

1. **Best-case analysis** c_1
2. Worst-case analysis
3. Average-case analysis

Time Complexity or Time Efficiency

ii. Selection Structure: if/else statement

```
if(x<a)
     sum1 += x;
else {
    sum2 += x;
    count ++;
    }
```

When x < a, only one primitive operation is executed
When x ≥ a, two primitive operations are executed

How do we analyze the time complexity?

1. Best-case analysis
2. **Worst-case analysis** c_2
3. Average-case analysis

Time Complexity or Time Efficiency

ii. Selection Structure: if/else statement

```
if(x<a)
      sum1 += x;
else {
      sum2 += x;
      count ++;
      }
```

When $x < a$, only one primitive operation is executed
When $x \geq a$, two primitive operations are executed

How do we analyze the time complexity?

1. Best-case analysis c_1
2. Worst-case analysis c_2
3. Average-case analysis

$$p(x < a)\, c_1 + p(x \geq a) c_2$$
$$= p(x < a)\, c_1 + (1 - p(x < a)) c_2$$

Time Complexity or Time Efficiency

ii. Selection Structure: switch-case statement

`switch(choice){`	
`case 1: compute the summation; break;`	$\dashrightarrow 5n$
`case 2: search BST; break;`	$\dashrightarrow 6\log_2 n$
`case 3: print BST; break;`	$\dashrightarrow 3n$
`case 4: search for the minimum; break;`	$\dashrightarrow 4\log_2 n$
`}`	

Time Complexity

1. Best-case analysis $\dashrightarrow C + 4\log_2 n$
2. Worst-case analysis $\dashrightarrow C + 5n$
3. Average-case analysis $\dashrightarrow C + \sum_{i=1}^{4} p(i)T_i$

Time Complexity or Time Efficiency

iii. Recursive functions

- Count the number of primitive operations in the algorithm
  - Primitive operations in each recursive call
  - Number of recursive calls

```
int factorial (int n)
{
    if(n==1) return 1;
    else return n*factorial(n-1);
}
```

c_2

c_1

- *n-1* recursive calls with the cost of c_1.
- The cost of the last call (n==1) is c_2.
- Thus, $c_1(n-1)+c_2$
- It is a linear function

Time Complexity or Time Efficiency

iii. Recursive functions

- Count the number of array[0]==a in the algorithm
  - array[0]==a in each recursive call
  - Number of recursive calls: n-1

```
int count (int array[],int n, int a)
{
    if(n==1)
          if(array[0]==a)
             return 1;
          else return 0;
    if(array[0]==a)
         return 1+ count(&array[1], n-1, a);
    else
         return count (&array[1], n-1, a);
}
```

$W_1 = 1$

$$W_n = 1 + W_{n-1}$$
$$= 1 + 1 + W_{n-2}$$

Time Complexity or Time Efficiency

iii. Recursive functions

- Count the number of array[0]==a in the algorithm
  - array[0]==a in each recursive call
  - Number of recursive calls: n-1

```
int count (int array[],int n, int a)
{
    if(n==1)
          if(array[0]==a)
             return 1;
          else return 0;
    if(array[0]==a)
         return 1+ count(&array[1], n-1, a);
    else
         return count (&array[1], n-1, a);
}
```

$$W_1 = 1$$

$$\begin{aligned}
W_n &= 1 + W_{n-1} \\
&= 1 + 1 + W_{n-2} \\
&= 1 + 1 + 1 + W_{n-3} \\
&\dots \\
&= \underbrace{1 + 1 + \dots + 1} + W_1 \\
&= (n - 1) + W_1 = n
\end{aligned}$$

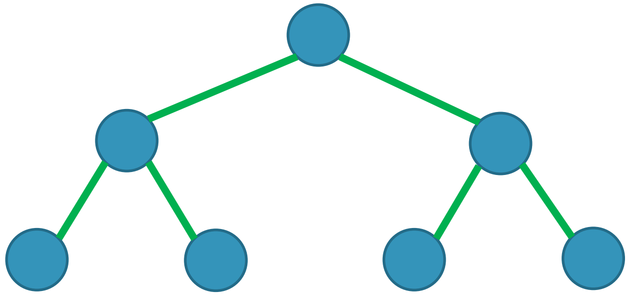
It is known as a **method of backward substitutions**

Time Complexity or Time Efficiency

iii. Recursive functions

- Count the number of multiplication operations in the algorithm

```
preorder (simple_t* tree)
{
    if(tree != NULL){
        tree->item *= 10;
        preorder (tree->left);
        preorder (tree->right);
    }
}
```



Geometric Series:

$$\begin{aligned} S_n &= a + ar + ar^2 + \ldots + ar^{n-1} \\ rS_n &= ar + ar^2 + \ldots + ar^{n-1} + ar^n \\ (1-r)S_n &= a - ar^n \\ S_n &= \frac{a(1-r^n)}{1-r} \end{aligned}$$

$W_0 = 0$

$W_1 = 1$

$W_2 = 1 + W_1 + W_1 = 3$

$W_3 = 1 + W_2 + W_2$

$= 1 + 2\,(1 + W_1 + W_1)$

$= 1 + 2\,(1 + 2)$

$= 1 + 2 + 4 = 7$

$W_{k-1} = 1 + 2\cdot W_{k-2}$

$= 1 + 2 + 4 + 8 + \ldots + 2^{k-2}$

$W_k = 1 + 2\cdot W_{k-1} = 1+2+4+8+\ldots+2^{k-1}$

$= \frac{1-2^k}{1-2} = 2^k - 1$

Prove the hypothesis can be done by mathematical induction

It is known as a **method of forward substitutions**

Series

- Geometric Series

$$G_n = \frac{a(1-r^n)}{1-r}$$

- Arithmetic Series

$$A_n = \frac{n}{2}[2a + (n-1)d] = \frac{n}{2}[a_0 + a_{n-1}]$$

- Arithmetico-geometric Series

$$\sum_{t=1}^{k} t2^{t-1} = 2^k(k-1) + 1$$

- Faulhaber's Formula for the sum of the p-th powers of the first n positive integers

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{k=1}^{n} k^3 = \frac{n^2(n+1)^2}{4}$$

*Derivation is in note section 0.7.4.1

Order of Growth

Order of growth: an approximation of the time required to run a computer program as the input size increases. The order of growth ignores the constant factor needed for fixed operations and focuses instead on the operations that increase proportional to input size

Algorithm	1	2	3	4	5	6
Operation (µsec)	$13n$	$13n\log_2 n$	$13n^2$	$130n^2$	$13n^2+10^2$	2^n

Problem size (n)

10						
100						
10^4						
10^6						

Order of Growth

Order of growth: an approximation of the time required to run a computer program as the input size increases. The order of growth ignores the constant factor needed for fixed operations and focuses instead on the operations that increase proportional to input size

Algorithm	1	2	3	4	5	6
Operation (μsec)	$13n$	$13n\log_2 n$	$13n^2$	$130n^2$	$13n^2+10^2$	2^n

Problem size (n)

10	.00013	.00043	.0013	.013	.0014	.001024
100	.0013					
10^4	.13					
10^6	13					

Order of Growth

Order of growth: an approximation of the time required to run a computer program as the input size increases. The order of growth ignores the constant factor needed for fixed operations and focuses instead on the operations that increase proportional to input size

Algorithm	1	2	3	4	5	6
Operation (μsec)	$13n$	$13n\log_2 n$	$13n^2$	$130n^2$	$13n^2+10^2$	2^n

Problem size (n)

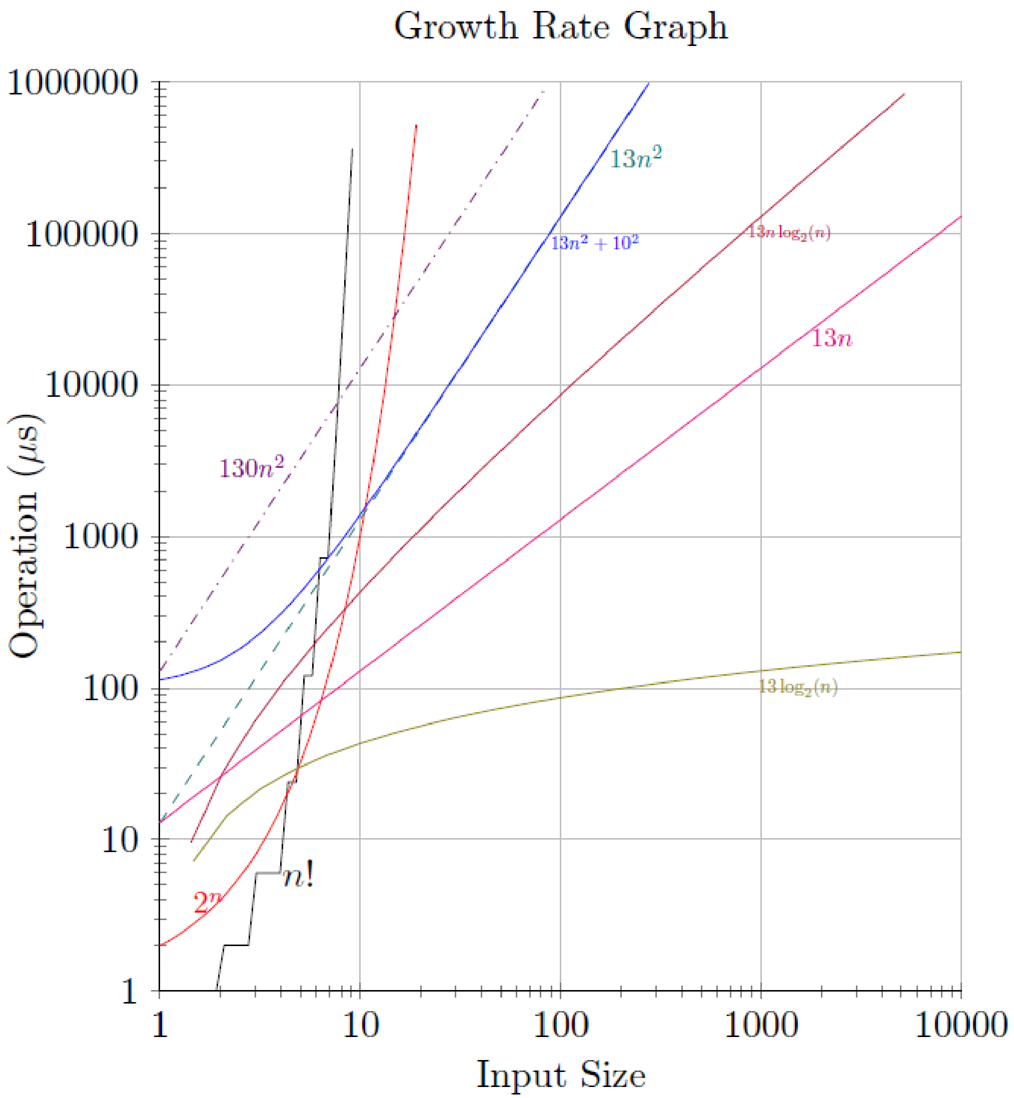
10	.00013	.00043	.0013	.013	.0014	.001024
100	.0013	.0086				
10^4	.13	.173				
10^6	13	259				

Order of Growth

Order of growth: an approximation of the time required to run a computer program as the input size increases. The order of growth ignores the constant factor needed for fixed operations and focuses instead on the operations that increase proportional to input size

Algorithm	1	2	3	4	5	6
Operation (μsec)	$13n$	$13n\log_2 n$	$13n^2$	$130n^2$	$13n^2+10^2$	2^n
Problem size (n)						
10	.00013	.00043	.0013	.013	.0014	.001024
100	.0013	.0086	.13	1.3	.1301	4×10^{16} years
10^4	.13	.173	22 mins	3.61hrs	22mins	
10^6	13	259	150 days	1505 days	150days	

Order of Growth

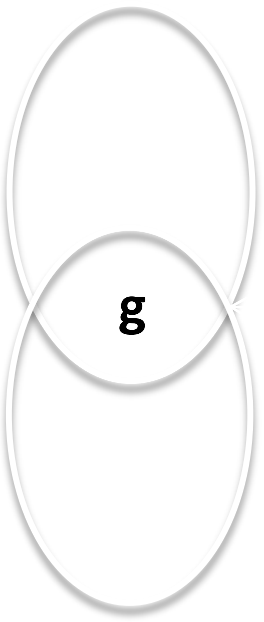


- $n!$ is the fastest growth
- 2^n is the second
- 13n is linear
- $13\log_2 n$ is the slowest

- 10^2 can be ignored when n is large
- $13n^2$ and $130n^2$ have similar growth
  - $130n^2$ slightly faster

Asymptotic Notations

- Big-Oh (**O**) , Big-Omega (**Ω**) and Big-Theta (**Θ**) are asymptotic (set) notations used for describing the order of growth of a given function.



$f \in \Omega(g)$ Set of functions that grow at higher or same rate as **g**

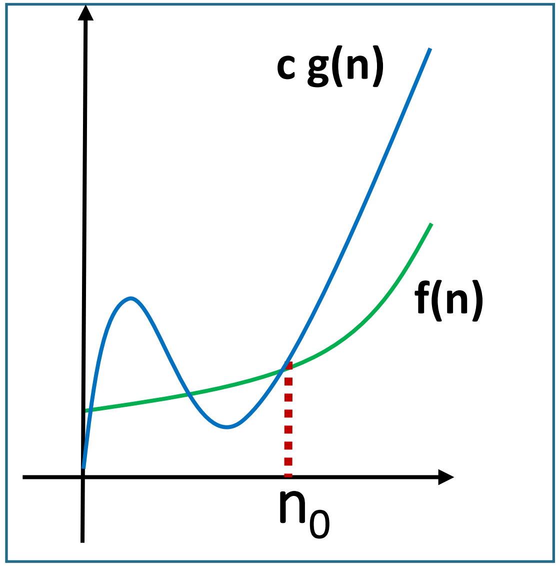
$f \in \Theta(g)$ Set of functions that grow at same rate as **g**

$f \in O(g)$ Set of functions that grow at lower or same rate as **g**

Big-Oh Notation (**O**)

Definition 3.1 *$\mathcal{O}$-notation: Let f and g be two functions such that $f(n): \mathbb{N} \to \mathbb{R}^+$ and $g(n): \mathbb{N} \to \mathbb{R}^+$, $f(n)$ is said to be in $\mathcal{O}(g(n))$, denoted $f(n) \in \mathcal{O}(g(n))$, if $f(n)$ is **bounded above** by some constant multiple of $g(n)$ for all large n, i.e., the set of functions can be defined as*

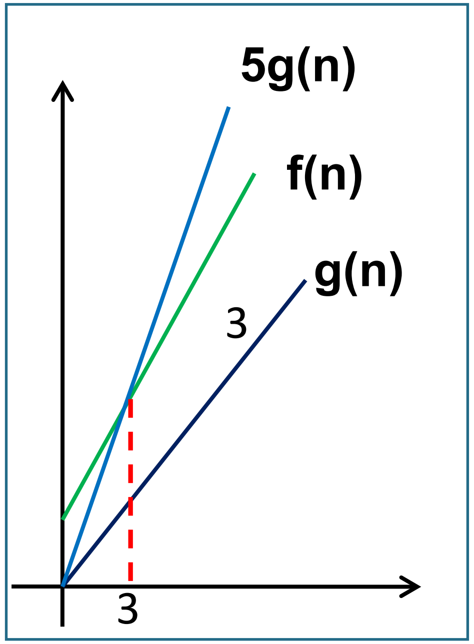
$$\mathcal{O}(g(n)) = \{f(n) : \exists \text{positive constants, } c \text{ and } n_0 \text{ such that } 0 \le f(n) \le cg(n) \quad \forall n \ge n_0\}$$



$f(n) = 4n + 3$, and $g(n) = n$

Let $c = 5, n_0 = 3$

$$f(n) = 4n + 3$$
$$4n + 3 \le 5n \qquad \forall n \ge 3$$
$$f(n) \le 5g(n) \qquad \forall n \ge 3$$
$$\Rightarrow \quad f(n) = O(g(n)) \qquad i.e.\ 4n + 3 \in O(n)$$



Big-Oh Notation (**O**)

Definition 3.1 *$\mathcal{O}$-notation: Let f and g be two functions such that $f(n) : \mathbb{N} \to \mathbb{R}^+$ and $g(n) : \mathbb{N} \to \mathbb{R}^+$, $f(n)$ is said to be in $\mathcal{O}(g(n))$, denoted $f(n) \in \mathcal{O}(g(n))$, if $f(n)$ is* ***bounded above*** *by some constant multiple of $g(n)$ for all large n, i.e., the set of functions can be defined as*

$$\mathcal{O}(g(n)) = \{f(n) : \exists \text{positive constants, } c \text{ and } n_0 \text{ such that } 0 \leq f(n) \leq cg(n) \quad \forall n \geq n_0\}$$

$f(n) = 4n + 3$ and $g(n) = n^3$

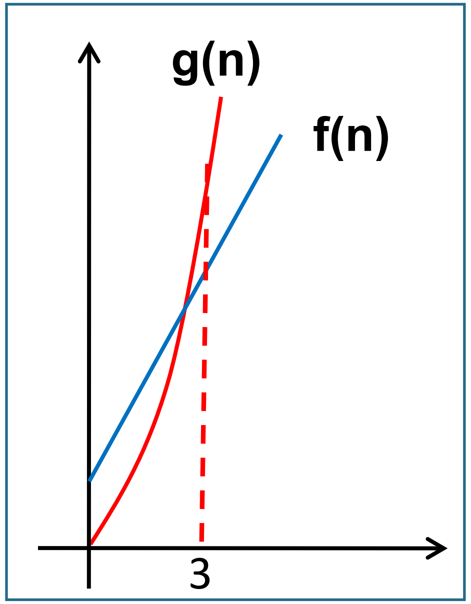
Let $c = 1, n_0 = 3$

$$f(n) = 4n + 3$$

$$4n + 3 \leq n^3 \qquad \forall n \geq 3$$

$$f(n) \leq g(n) \qquad \forall n \geq 3$$

$$\Rightarrow \quad f(n) = O(g(n)) \qquad i.e. \, 4n + 3 \in O(n^3)$$



If $f(n) = O(g(n))$, we say

$g(n)$ is asymptotic upper bound of $f(n)$

Big-Oh Notation (**O**) – Alternative definition

Definition 3.2 *$\mathcal{O}$-notation: Let f and g be two functions such that $f(n) : \mathbb{N} \to \mathbb{R}^+$ and $g(n) : \mathbb{N} \to \mathbb{R}^+$, if $\lim_{n\to\infty} \frac{f(n)}{g(n)} < \infty$, then $f(n) \in \mathcal{O}(g(n))$ or $f(n) = \mathcal{O}(g(n))$.*

$f(n) = 4n + 3$ and $g(n) = n$

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{4n+3}{n} = 4 < \infty$$

$\Rightarrow$ $f(n) = O(g(n))$ $\quad i.e. 4n + 3 \in O(n)$

$f(n) = 4n + 3$ and $g(n) = n^3$

$$\lim_{n\to\infty} \frac{f(n)}{g(n)} = \lim_{n\to\infty} \frac{4n+3}{n^3} = 0 < \infty$$

$\Rightarrow$ $f(n) = O(g(n))$ $\quad i.e. 4n + 3 \in O(n^3)$

Big-Omega Notation (Ω)

Definition 3.3 *Ω-notation: Let f and g be two functions such that $f(n) : \mathbb{N} \rightarrow \mathbb{R}^+$ and $g(n) : \mathbb{N} \rightarrow \mathbb{R}^+$, $f(n)$ is said to be in $\Omega(g(n))$, denoted $f(n) \in \Omega(g(n))$, if $f(n)$ is* **bounded below** *by some constant multiple of $g(n)$ for all large n, i.e., the set of functions can be defined as*

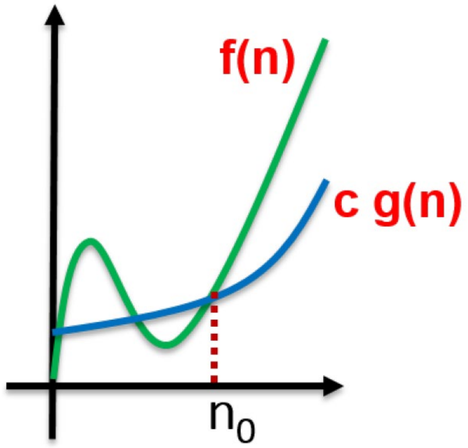
$$\Omega(g(n)) = \{f(n) : \exists \text{positive constants, } c \text{ and } n_0 \text{ such that } 0 \leq cg(n) \leq f(n) \quad \forall n \geq n_0\}$$

Definition 3.4 *Ω-notation: Let f and g be two functions such that $f(n) : \mathbb{N} \rightarrow \mathbb{R}^+$ and $g(n) : \mathbb{N} \rightarrow \mathbb{R}^+$, if $\lim_{n\to\infty} \frac{f(n)}{g(n)} > 0$, then $f(n) \in \Omega(g(n))$ or $f(n) = \Omega(g(n))$.*

$f(n) = 4n + 3$, and $g(n) = n$

Let $c = 1/5, n_0 = 0$

$$f(n) \geq (1/5)g(n) \quad \forall n \geq n_0$$
$$4n + 3 \geq (1/5)5n \quad \forall n \geq 0$$



If $f(n) = \Omega(g(n))$, we say

$g(n)$ is asymptotic lower bound of $f(n)$

Big-Theta Notation (Θ)

Definition 3.5 Θ*-notation: Let f and g be two functions such that $f(n) : \mathbb{N} \rightarrow \mathbb{R}^+$ and $g(n) : \mathbb{N} \rightarrow \mathbb{R}^+$, $f(n)$ is said to be in $\Theta(g(n))$, denoted $f(n) \in \Theta(g(n))$, if $f(n)$ is* ***bounded both above and below*** *by some constant multiples of $g(n)$ for all large n, i.e., the set of functions can be defined as*

$$\Theta(g(n)) = \{f(n) : \exists \text{positive constants, } c_1, \, c_2 \text{ and } n_0 \text{ such that } c_1 g(n) \leq f(n) \leq c_2 g(n) \quad \forall n \geq n_0\}$$

Definition 3.6 Θ*-notation: Let f and g be two functions such that $f(n) : \mathbb{N} \rightarrow \mathbb{R}^+$ and $g(n) : \mathbb{N} \rightarrow \mathbb{R}^+$, if $\lim_{n\rightarrow\infty} \frac{f(n)}{g(n)} = c$ where $0 < c < \infty$, then $f(n) \in \Theta(g(n))$ or $f(n) = \Theta(g(n))$.*

If $f(n) = \Theta(g(n))$, we say

$g(n)$ is asymptotic tight bound of $f(n)$

Summary of Limit Definition

$\lim_{n\to\infty} \frac{f(n)}{g(n)}$	$f(n) \in O(g(n))$	$f(n) \in \Omega(g(n))$	$f(n) \in \Theta(g(n))$
$\mathbf{0}$	✓		
$0 < C < \infty$	✓	✓	✓
∞		✓	

Common Complexity Classes

Order of Growth	Class	Example
1	Constant	Finding midpoint of an array
$\log_2 n$	Logarithmic	Binary Search
n	Linear	Linear Search
$n\log_2 n$	Linearithmic	Merge Sort
n^2	Quadratic	Insertion Sort
n^3	Cubic	Matrix Inversion (Gauss-Jordan Elimination)
2^n	Exponential	The Tower of Hanoi Problem
n!	Factorial	Travelling Salesman Problem

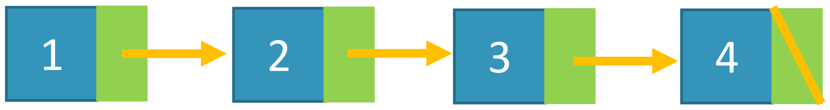
When time complexity of algorithm A grows faster than algorithm B for the same problem, we say A is inferior to B.

Time Complexity of Sequential Search

```
pt=head;
while (pt->key != a){
    pt = pt->next;
    if(pt == NULL) break;
}
```

c_1

c_2 (n-1) iterations



Assume that the search key ***a*** is always in the list

1. Best-case analysis: c_1 when **a** is the first item in the list => Θ (1)
2. Worst-case analysis: $c_2 \cdot (n-1) + c_1$ => Θ (n)
3. Average-case analysis

$$\boldsymbol{p_1 c_1 + p_2(c_1 + c_2) + p_2(c_1 + 2c_2) + \cdots + p_n(c_1 + (n-1)c_2)}$$

- Assumed that every item in the list has an equal probability as a search key

$$\boldsymbol{\frac{1}{n}[c_1 + (c_1 + c_2) + (c_1 + 2c_2) + \cdots + (c_1 + (n-1)c_2)] = \frac{1}{n}\sum_{i=1}^{n}(c_1 + c_2(i-1))}$$

$$= \frac{1}{n}\left[nc_1 + c_2\sum_{i=1}^{n}(i-1)\right]$$

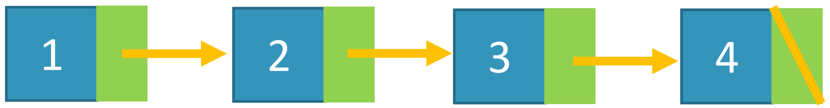
$$= c_1 + \frac{c_2}{n}\cdot\frac{n}{2}(0 + (n-1)) = c_1 + \frac{c_2(n-1)}{2} = \Theta\ (n)$$

Time Complexity of Sequential Search

```
pt=head;
while (pt->key != a){
    pt = pt->next;
    if(pt == NULL) break;
}
```

c_1

c_2 n iterations



3. Average-case analysis
   - Assumed that every item in the list has an equal probability as a search key

$$\frac{1}{n}[c_1 + (c_1 + c_2) + (c_1 + 2c_2) + \cdots + (c_1 + (n-1)c_2)] = \frac{1}{n}\sum_{i=1}^{n}(c_1 + c_2(i-1))$$

$$= \frac{1}{n}[nc_1 + c_2 \sum_{i=1}^{n}(i-1)\ ]$$

$$= c_1 + \frac{c_2}{n} \cdot \frac{n}{2}(0 + (n-1)) = c_1 + \frac{c_2(n-1)}{2}$$

If the search key, a, is not in the list, then the time complexity is

$$c_1 + nc_2 = \Theta\ (n)$$

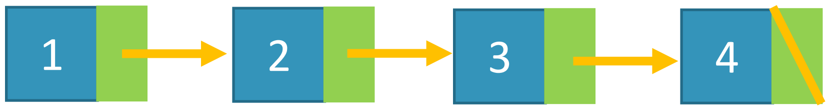
Since the probability of the search key is in the list is unknown, we only can have

$$T(n) = P(a\ in\ the\ list)(c_1 + \frac{c_2(n-1)}{2}) + (1 - P(a\ in\ the\ list))(c_1 + nc_2)$$

Hence, it is a linear function. **Θ (n)**

Time Complexity of Sequential Search

```
pt=head;
while (pt->key != a){
    pt = pt->next;
    if(pt == NULL) break;
}
```



- The data is stored in unordered
- To search a key, every element is required to read and compare
- This is a brute-force approach or a näive algorithm
- Its time complexity is **O(n)**
- How can we improve it?

Asymptotic Notation in Equations

When an asymptotic notation appears in an equation, we interpret it as standing for some anonymous function that we do not care to name.

Examples:

- $2n^2 + 3n + 1 = 2n^2 + \Theta(n)$
- $T(n) = T(n/2) + \Theta(n)$
- $2n^2 + 3n + 1 = 2n^2 + \Theta(n) = \Theta(n^2)$

Simplification Rules for Asymptotic Analysis

1. If $f(n) = O\big(cg(n)\big)$ for any constant c > 0, then $f(n) = O(g(n))$
2. If $f(n) = O\big(g(n)\big)$ and $\mathrm{g}(n) = O\big(h(n)\big)$, then $f(n) = O\big(h(n)\big)$

   e.g., $f(n) = 2n$, $\mathrm{g}(n) = n^2$, $\mathrm{h}(n) = n^3$

3. If $f_1(n) = O\big(g_1(n)\big)$ and $f_2(n) = O\big(g_2(n)\big)$,

   then $f_1(n) + f_2(n) = O(\max(g_1(n), g_2(n)))$

   e.g., $5\mathrm{n} + 3\log_2 n = O(n)$

4. If $f_1(n) = O\big(g_1(n)\big)$ and $f_2(n) = O\big(g_2(n)\big)$

   then $f_1(n)f_2(n) = O\big(g_1(n)g_2(n)\big)$

   e.g., $f_1(n) = 3n^2 = O(n^2)$, $f_2(n) = \log_2 n = O(\log_2 n)$

   Then $3n^2 \log_2 n = O(n^2 \log_2 n)$

Properties of Asymptotic Notation

- Reflexive of O, Ω and Θ

$$f(n) = O(f(n))$$
$$f(n) = \Omega(f(n))$$
$$f(n) = \Theta(f(n))$$

- Symmetric of Θ

$$f(n) = \Theta(g(n))$$
$$\Longrightarrow g(n) = \Theta(f(n))$$

- Transitive of O, Ω and Θ

$f(n) = O(g(n))$ and $g(n) = O(h(n))$
$$\Longrightarrow f(n) = O(h(n))$$

$f(n) = \Omega(g(n))$ and $g(n) = \Omega(h(n))$
$$\Longrightarrow f(n) = \Omega(h(n))$$

$f(n) = \Theta(g(n))$ and $g(n) = \Theta(h(n))$
$$\Longrightarrow f(n) = \Theta(h(n))$$

Space Complexity

- Determine number of entities in problem (also called problem size)
- Count number of basic units in algorithm

- Basic units
- Things that can be represented in a constant amount of storage space
- E.g. integer, float and character.

Space Complexity

- Space requirements for an array of n integers - $\Theta(n)$
- If a matrix is used to store edge information of a graph,

  i.e. $G[x][y] = 1$ if there exists an edge from x to y,

  space requirement for a graph with n vertices is $\Theta(n^2)$

Space/time tradeoff principle

- Reduction in time can be achieved by sacrificing space and vice-versa.

Problem for you to think about

- Given two arrays num1 and num2. Both are sorted in ascending order. The length is m and n, respectively. Please find the median of the two arrays. The time complexity need to be $O(\log(m+n))$.