Chapter 1 : Basic Concepts

Data Structures Lecture Note
Prof. Sungwon Jung
Big Data Processing Laboratory
Dept. of Computer Science and Engineering
Sogang University

Goals

- To provide the tools and techniques necessary to design and implement large-scale computer systems.
  - solid foundation in data abstraction, algorithm specification and performance analysis and measurement provides the necessary methodology.

1.1 SYSTEM LIFE CYCLE

- **Requirement**
  - a set of specifications that define the purpose of the project.
  - input/output
- **Analysis**
  - break the problems down into manageable pieces.
  - bottom-up / top-down
- **Design**
  - creation of abstract data types
  - specification of algorithms and consideration of algorithm design strategies.
    (* language independent *)

- **Refinement and Coding**
  - choose representations for data objects and write algorithms for each operation on them.
  - data object's representation can determine the efficiency of the algorithms related to it.
- **Verification**
  - Developing correctness proof for the program
  - Testing the program with a variety of input data
  - Error removal
  - Performance analysis
    - running time
    - amount of memory used

1.2 ALGORITHM SPECIFICATION

1.2.1 Introduction

- ***Definition:***

  An algorithm is a finite set of instructions that, if followed, accomplishes particular task.

  All algorithms must satisfy the following criteria:

  (1) Input
  (2) Output
  (3) Definiteness
  (4) Finiteness
  (5) Effectiveness

  algorithm / program (procedure)

- **How to describe an algorithm**

  natural language
  flowchart
  programming language

- **Example 1.1 [Selection Sort]**

  Sorting a set of n ≥ 1 integers

  From those integers that are currently unsorted,
  find the smallest and place it next in the sorted list.

- **[Program 1.1 Selection sort algorithm]**

```
for (i=0; i<n; i++) {
    Examine list[i] to list[n-1]
     and suppose that the smallest integer is at list[min];
    Interchange list[i] and list[min];
}
```

- **first task : finding the smallest integer;**
- **second task : exchange;**

  either a function or a macro

- **[Program 1.2 swap function]**

```
void swap(int *x, int *y)
/* both parameters are pointers to ints */
{
        int temp = *x; /* declare temp as an int and assign to it
                          the contents of what x points to */
        *x = *y;        /* stores what y points to into the location
                           where x points  */
        *y = temp;      /* place the contents of temp in the location
                           pointed to by y */
}
```

Call -- swap(&a, &b)

- **macro version of swap -**

  #define SWAP(x,y,t) ((t) = (x), (x) = (y), (y) = (t))

- **[Program 1.3 Selection sort]**

```
void sort (int list[], int n)
{
        int i, j, min, temp;
        for (i=0; i<n-1; i++)  {
                min = i;
                for (j = i+1; j < n; j++)
                if (list[j] < list[min])
                        min = j;
                SWAP(list[i], list[min], temp);
        }
}
```


- **Theorem 1.1:**

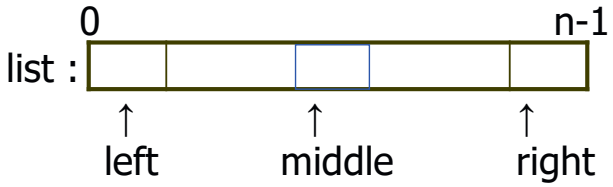
Function *sort*(*list,n*) correctly sorts a set of n≥1 integers. The result remains in list[0], . . ., list[n-1] such that list[0]≤list[1]≤ . . . ≤list[n-1].

proof : consider *loop invariant.*

- **Example 1.2 [Binary Search]**

  Find out if an integer *searchnum* is in a list.
  If so, return i such that list[i] = *searchnum,*
  Otherwise, return -1.

  For a sorted list (in ascending order)



  middle = (left + right) / 2

- **Compare list[*middle*] with *searchnum***
  - ***searchnum* < list[*middle*]**
    if *searchnum* is present, it must be in the position
    between *left* and *middle-1.*
    set *right* to *middle-1.*
  - ***searchnum* = list[*middle*]**
    return *middle.*
  - ***searchnum* > list[*middle*]**
    if *searchnum* is present, it must be in the position
    between *middle+1* and *right.*
    set *left* to *middle+1*

- **Implementing this search strategy :**

```
while (there are more integers to check) {
        middle = (left + right) / 2;
        if (searchnum < list[middle])
                right = middle - 1;
        else if (searchnum == list[middle])
                return middle;
        else  left = middle + 1;
}
```


- **Handling the comparisons:**

<	**returns**	**-1**
=		**0**
>		**1**

- ***function* -**

```
int compare (int x, int y)
{
        /* compare x and y, return -1 for less than,
        0 for equal, 1 for greater */
        if (x < y)  return -1;
        else  if  (x == y)  return 0;
        else  return 1;
}
```

- ***macro* -**

```
# define COMPARE (x,y) ((x) < (y)) ? -1: ((x) == (y)) ? 0: 1)
```


- **[Program 1.6]**

```
int binsearch(int list[], int searchnum, int left, int right)
{
  /* search list[0] <= list[1] <= . . . <= list[n-1] for    searchnum.
     Return its position if found.  Otherwise return -1 */
  int middle;
  while (left <= right)  {
        middle = (left + right)/2;
        switch (COMPARE(list[middle], searchnum)) {
                case -1 : left = middle + 1;
                          break;
                case 0 : return middle;
                case 1 : right = middle - 1;
        }
   }
   return -1;
}
```

1.2.2 Recursive Algorithms

Direct recursion
Indirect recursion

- Recursion is a general control scheme.
- Often recursive function is easier to understand than its iterative counterpart.
- Many problems can be defined recursively in natural way.

- **[Binomial Coefficients]**

$$\begin{bmatrix} n \\ m \end{bmatrix} = \frac{n!}{m!(n-m)!}$$

can be recursively computed by the formula:

$$\begin{bmatrix} n \\ m \end{bmatrix} = \begin{bmatrix} n-1 \\ m \end{bmatrix} + \begin{bmatrix} n-1 \\ m-1 \end{bmatrix}$$

- **Examples :**
  - **[factorial]**

$$n! = \begin{cases} n * (n-1)! & \text{if } n>1 \\ 1 & \text{if } n=1 \end{cases}$$

  - **[Binary search]**

$$\text{Bsrch[key, left, right]} = \begin{cases} \text{Bsrch[key, left, middle-1]} & \text{if key < list[middle]} \\ \text{list[middle]} & \text{if key = list[middle]} \\ \text{Bsrch[key, middle+1, right]} & \text{if key > list[middle]} \end{cases}$$

  - **[Fibonacci numbers]**

$$f_n = \begin{cases} 0 & \text{if } n=0 \\ 1 & \text{if } n=1 \\ f_{n-1} + f_{n-2} & \text{if } n>1 \end{cases}$$

  - **[Permutations]**

We can construct the set of permutations by printing:

(1) *a* followed by all permutations of (*b, c, d*)
(2) *b* followed by all permutations of (*a, c, d*)
(3) *c* followed by all permutations of (*a, b, d*)
(4) *d* followed by all permutations of (*a, b, c*)

Iterative function

```
int fibo(int n)
{
   int g, h, f, i;
   if (n>1) {
      g = 0;
      h = 1;
      for (i = 2; i<= n; i++) {
          f = g+h;
          g = h;
          h = f;
      }
   }
   else   f = n;
   return  f;
}
```

Recursive function

```
int rfibo (int n)
{
   if (n > 1)
      return rfibo(n-1) + rfibo(n-2);
   else
      return n;
}
```


- **[Program 1.7]**

```
int binsearch(int list[], int searchnum, int left, int right)
{
 /* search list[0] <= list[1] <= . . . <= list[n-1] for searchnum.
     Return its position if found.  Otherwise return -1 */

 int middle;
 if (left <= right)  {
    middle = (left + right)/2;
    switch (COMPARE(list[middle], searchnum)) {
        case -1 : return  binsearch(list, searchnum, middle + 1, right);
        case 0 : return middle;
        case 1 : return  binsearch(list, searchnum, left, middle - 1);
    }
 }
 return -1;
}
```

- **[Program 1.8]**

```
void perm(char *list, int i, int n)
{
    /* generate all the permutations of list[i] to list[n] */
    int j, temp;
    if (i == n)  {
        for (j=0; j<=n; j++) printf("%c", list[j]);
        printf("     ");
    }
    else  {
    /* list[i] to list[n] has more than one permutation,
        generate these recursively  */
          for (j=i; j<=n; j++)  {
              SWAP(list[i], list[j], temp);
              perm(list, i+1, n);
              SWAP(list[i], list[j], temp);
          }
    }
}
```


1.3 DATA ABSTRACTION

- **basic data types of C :**
  char, int, float, double, . . .
  short, long, unsigned

- **mechanisms for grouping data together :**
  Arrays and Structs

```
int  list[5];

struct student {
        char last_name[10];
        int  student_id;
        char grade;
};
```

- **pointer data type :**

for every basic data type
there is a corresponding pointer data type, such as
pointer-to-an-int,
pointer-to-a-real,
pointer-to-a-char,
and pointer-to-a-float.

```
int   i,  *pi;
```

predefined data types / user-defined data types

"What is a data type?"

- **Definition :**
  A *data type* is a collection of *objects* and a set of *operations* that act on those objects.

  - **specification of objects**
    e.g., type *int*,
    {0, +1, -1, +2, -2, . . ., INT_MAX, INT_MIN}
    specification of operations

  - **representation of objects**
    implementation of operations

- **Definition :**

  An *abstract data type* (ADT) is a data type that is organized in such a way that the specification of the objects and the specification of the operations on the objects is separated from the representation of the objects and the implementation of the operations.

- ***an abstract data type is implementation independent.***

  Specification of operations consists of the names of operations, the type of its arguments, and the type of its result. Also a description what the function does without appealing to internal representation details.

  ***package* in Ada**

  ***class* in C++**

- **Categories to classify the operations of a data types:**
  - Creator/constructor
  - Transformers
  - Observers/reporters

- **Example 1.5 [Abstract data type *Natural_Number*]**

ADT *Natural_Number* is
object: an ordered subrange of the integers starting at zero and ending at the maximum integer (*INT_MAX*) on the computer
functions:
for all x, y IN *Nat_Number*, *TRUE, FALSE* IN *Boolean*
and where +, -, <, and == are usual integer operations

Nat_Number Zero() ::= 0
Boolean Is_Zero(*x*) ::= if (*x*) return *FALSE*
else return *TRUE*
Nat_Number Add(*x, y*) ::= if ((*x+y*)<= *INT_MAX*) return *x+y*
else return *INT_MAX*
Boolean Equal(*x, y*) ::= if (*x==y*) return *TRUE*
else return *FALSE*
Nat_Number Successor(*x*) ::= if (*x* == *INT_MAX*) return *x*
else return *x+1*
Nat_Number Subtract(*x,y*) ::= if (*x<y*) return 0
else return *x-y*

end *Natural_Number*

1.4 PERFORMANCE ANALYSIS

- ***Criteria of judging a program:***
  1. Does the program meet the original specification of the task?
  2. Does it work correctly?
  3. Is the program well documented?
  4. Does the program effectively use functions to create logical units?
  5. Is the program's code readable?

  [Performance Evaluation]
  6. Does the program efficiently use primary and secondary storage?
  7. Is the program's running time acceptable for the task?

- ***Performance Analysis*** **:**
  estimates of time and space that are machine independent.

- ***Performance Measurement*** **:**
  obtaining machine-dependent running times.
  used to identify inefficient code segments.

- **Definition :**
  The *space complexity* of a program is the amount of memory that it needs to run to completion.
  The *time complexity* of a program is the amount of computer time that it needs to run to completion.

1.4.1 Space Complexity

- ***Fixed space requirements*** **:**
  independent from the number and size of the program's inputs and outputs, e.g., the instruction space, space for simple variables, fixed-size structured variables, and constants.

- ***Variable space requirements*** **:**
  space needed by structured variables whose size depends on the particular instance, I, of the problem being solved.

$$S(P) = c + S_P(I)$$

- **Example 1.6 : [simple arithmetic function]**

$S_{abc}(I) = 0.$

[Program 1.9]

```
float abc (float a, float b, float c)
{
   return a+b+b*c + (a+b-c)/(a+b) + 4.00;
}
```


- **Example 1.7 : [*adding a list of numbers iteratively*]**

[Program 1.10]

```
float sum(float list[], int n)
{
      float  tempsum = 0;
      int i;
      for (i=0; i<n; i++)
                tempsum += list[i];
      return tempsum;
}
```

$S_{sum}(n) = n$ if parameters are passed by value.
$S_{sum}(n) = 0$ if parameters are passed by reference

- **Example 1.8 : [*adding a list of numbers recursively*]**

  [Program 1.11]

```
float rsum(float list[], int n)
{
        if (n) return rsum(list, n-1) + list[n-1];
        return list[0];
}
```

$$S_{rsum}(n) = 12*n$$

Type	Name	Number of bytes
parameter: array pointer parameter: integer return address: (used internally)	list[] n	4 4 4
TOTAL per recursive call		12

Figure 1.1 : Space needed for one recursive call of program 1.11

1.4.2 Time Complexity

(1) *Compile Time*

(2) *Execution (Running) Time*

We are really concerned only with the program's execution time.

- ***Determining the execution time*** **:**
  - the times needed to perform each operation.
  - the number of each operation performed for the given instance (dependent on the compiler).

$$T_P(n) = c_a ADD(n) + c_s SUB(n) + c_l LDA(n) + c_{st} STA(n)$$

- Obtaining such a detailed estimate of running time is rarely worth the effort.
- Counting the number of operations the program performs gives us a machine-independent estimate.

- **Definition :**

  A *program step* is a syntactically or semantically meaningful program segment whose execution time is independent of the instance characteristics.

  Determining the number of steps that a program or a function needs to solve a particular problem instance by creating a global variable, *count*, and inserting statements that increment count

- **[Example 1.9] [*Iterative summing of a list of numbers*]**

[Program 1.12]

```
float sum(float list[], int n)
{
    float  tempsum = 0;  count++;    /*for assignment*/
    int i;
    for (i=0; i<n; i++)  {
       count++;   /*for the for loop */
       tempsum += list[i];  count++;  /*for assignment*/
    }
    count++;  /* last execution of for */
    count++;  /* for return */  return tempsum;
}
```

[Program 1.13]

```
float sum(float list[], int n)
{
      float  tempsum = 0;
      int i;
      for (i=0; i<n; i++)
         count += 2;   /*for the for loop */
      count += 3;
      return 0;
}
```

If the initial value of count is 0, its final value will be $2n+3$.

- **[Example 1.10] [*Recursive summing of a list of numbers*]**

[Program 1.14]

```
float rsum(float list[], int n)
{
    count++;    /* for if conditional */
    if (n) {
        count++;            /* for return and rsum invocation */
        return rsum(list, n-1) + list[n-1];
    }
    count++;
    return 0;
}
```

the step count is $2n+2$.

- **[Example 1.11] : [*Matrix addition*]**

[Program 1.15]

```
void add(int a[][MAX_SIZE], int b[][MAX_SIZE],
         int c[][MAX_SIZE], int rows, int cols)
{
    int i, j;
    for (i=0; i<rows; i++)
      for (j=0; j<cols; j++)
     c[i][j] = a[i][j] + b[i][j];
}
```


[Program 1.16]

```
void add(int a[][MAX_SIZE], int b[][MAX_SIZE],
         int c[][MAX_SIZE], int rows, int cols)
{
    int i, j;
    for (i=0; i<rows; i++)   {
         count++;    /* for i for loop */
         for (j=0; j<cols; j++)  {
              count++;     /* for j for loop */
              c[i][j] = a[i][j] + b[i][j];
              count++;     /* for assignment statement */
         }
         count++;  /* last time of j for loop */
    }
    count++;  /* last time of i for loop */
}
```

[Program 1.17]

```
void add(int a[][MAX_SIZE], int b[][MAX_SIZE],
                    int c[][MAX_SIZE], int rows, int cols)
{
  int i, j;
  for (i=0; i<rows; i++)  {
    for (j=0; j<cols; j++)
        count += 2;
    count += 2;
  }
  count++;
}
```

The step count will be 2 *rows* *cols* + 2*rows* +1

- **Tabular method: *steps/execution***

[Figure 1.2]

Statement	s/e	Frequency	Total steps
float sum(float list[], int n)	0	0	0
{	0	0	0
float tempsum=0;	1	1	1
int i;	0	0	0
for (i=0; i<n; i++)	1	n+1	n+1
tempsum += list[i];	1	n	n
return tempsum;	1	1	1
}	0	0	0
Total			2n+3

- **[Example 1.13]**

[Figure 1.3]

Statement	s/e	Frequency	Total steps
float rsum(float list[], int n)	0	0	0
{	0	0	0
if (n)	1	n+1	n+1
return rsum(list, n-1)+list[n-1];	1	n	n
return 0;	1	1	1
}	0	0	0
Total			2n+2

- **[Example 1.14]**

[Figure 1.4]

Statement	s/e	Frequency	Total Steps
void add(int a[][MAX_SIZE])	0	0	0
{	0	0	0
int i, j;	0	0	0
for (i=0; i<rows; i++)	1	rows + 1	rows + 1
for (j=0; j<cols; j++)	1	rows·(cols+1)	rows·cols + rows
c[i][j]=a[i][j]+b[i][j];	1	rows·cols	rows·cols
}	0	0	0
Total			2rows·cols + 2rows + 1

Summary

- Time complexity of a program is given by the number of steps taken by the program to compute the function it was written for.
- The number of steps is itself a function of the instance characteristics.

  e.g., the number of inputs, the number of outputs,
  the magnitudes of the inputs and outputs, etc.
- Before the step count of a program can be determined, we need to know exactly which characteristics of the problem are to be used.

- For many programs, the time complexity is not dependent solely on the characteristics specified.
- The step count varies for different inputs of the same size.

  Best case
  Worst case
  Average

Examples :

Binary Search
Insertion Sort

1.4.3 Asymptotic Notation (O, Ω, Θ)

- **Our motivation to determine step counts:**

  to compare the time complexities of two programs for the same function, and

  to predict the growth in run time as the instance characteristics change.

- Determining the exact step count (either worst case or average) of a program can prove to be an exceedingly difficult task.
- Expending immense effort to determine the step count exactly isn't a worthwhile endeavor as the notion of a step is itself inexact.

  (e.g., x = y and x = y+z+(x/y)+(x*y*z-x/t) count as one step)
- Because of the inexactness of what a step stands for, the exact step count isn't very useful for comparative purposes.

- For most situations, step counts can be represented as a function of instance characteristics, such as $c_1 n^2 \leq T_P(n) \leq c_2 n^2$ or $T_Q(n, m) = c_1 n + c_2 m$.

  What if the difference of two step counts are large?
  e.g., 3n+3 versus 100n+10.

  What if two step counts are of different orders?
  e.g., $c_1 n^2 + c_2 n$ versus $c_3 n$.

- ***break even point*** :

  C_1=1, C_2=2, C_3=100

  $C_1 n^2 + C_2 n \leq C_3 n$, $n \leq 98$

  $C_1 n^2 + C_2 n > C_3 n$, $n > 98$ Break even point : n = 98

  The exact break even point cannot be determined analytically.
  The programs have to be run on a computer in order to determine the break even point.

Some terminology :

- **Definition : [Big "*oh*"]**

  f(n) = O(g(n))

  iff there exist positive constants c and n_0 such that f(n) $\leq$ c g(n)
  for all n, $n \geq n_0$.

$3n + 2 = O(n)$

$100n + 6 = O(n)$

$1000n^2 + 100n - 6 = O(n^2)$

$3n + 3 = O(n)$

$10n^2 + 4n + 2 = O(n^2)$

$6*2^n + n^2 = O(2^n)$

$3n + 3 = O(n^2)$

$3n + 2 \neq O(1)$

$10n^2 + 4n + 2 = O(n^4)$

$10n^2 + 4n + 2 \neq O(n)$

$O(1)$	a constant	$O(n^2)$	quadratic
$O(\log n)$	logarithm	$O(n^3)$	cubic
$O(n)$	linear	$O(2^n)$	exponential

In order for the statement $f(n) = O(g(n))$ to be informative, $g(n)$ should be as small a function of n as one can come up with for which $f(n) = O(g(n))$.

- **Theorem 1.2 :**
  If $f(n) = a_m n^m + \ldots + a_1 n + a_0$ **then** $f(n) = O(n^m)$.

- **Proof :**

$$f(n) \leq \sum_{i=0}^{m} |a_i| n^i$$

$$\leq n^m \sum_{i=0}^{m} |a_i| n^{i-m}$$

$$\leq n^m \sum_{i=0}^{m} |a_i| , \quad \text{for } n \geq 1.$$

SO, $f(n) = O(n^m)$

- **Definition : [Omega]**

$$f(n) = \Omega(g(n))$$

iff there exist positive constants c and n_0 such that $f(n) \geq cg(n)$ for all n, $n \geq n_0$.

- **Example 1.16 :**

$3n + 2 = \Omega(n)$ $\quad$ $3n + 3 = \Omega(n)$

$100n + 6 = \Omega(n)$ $\quad$ $10n^2 + 4n + 2 = \Omega(n^2)$

$6*2^n + n^2 = \Omega(2^n)$

$10n^2 + 4n + 2 = \Omega(n)$ $\quad$ $10n^2 + 4n + 2 = \Omega(1)$

$6*2^n + n^2 = \Omega(n^{100})$ $\quad$ $6*2^n + n^2 = \Omega(n^2)$

$6*2^n + n^2 = \Omega(n)$ $\quad$ $6*2^n + n^2 = \Omega(1)$

In order for the statement $f(n) = \Omega(g(n))$ to be informative, $g(n)$ should be as large a function of n as possible for which $f(n) = \Omega(g(n))$ is true.

- **Theorem 1.3 :**

  If $f(n) = a_m n^m + \ldots + a_1 n + a_0$ **and** $a_m > 0$ **,**
  then $f(n) = \Omega(n^m)$ **.**

- **Definition : [Theta]**

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

iff there exist positive constants c_1, c_2 and n_0 such that $c_1 g(n) \leq f(n) \leq c_2 g(n)$ for all n, $n \geq n_0$.

- **Example 1.17 :**

$3n + 2 = \Theta(n)$

$3n + 3 = \Theta(n)$

$10n^2 + 4n + 2 = \Theta(n^2)$

$6*2^n + n^2 = \Theta(2^n)$

$10*\log n + 4 = \Theta(\log n)$

$3n + 2 \neq \Theta(1)$

$3n + 2 \neq \Theta(n^2)$

$10n^2 + 4n + 2 \neq \Theta(n)$

$10n^2 + 4n + 2 \neq \Theta(1)$

$6*2^n + n^2 \neq \Theta(n^{100})$

$6*2^n + n^2 \neq \Theta(n^2)$

$6*2^n + n^2 \neq \Theta(1)$

- **Theorem 1.4 :**
  If $f(n) = a_m n^m + \ldots + a_1 n + a_0$ **and** $a_m > 0$ **,**
  then $f(n) = \Theta(n^m)$ **.**

- **Example 1.18*: [Complexity of matrix addition]***

Statement	Asymptotic complexity
void add(int a[][MAX_SIZE] ...)	0
{	0
int i, j;	0
for (i=0; i<rows; i++)	Θ(rows)
for (j=0; j<cols; j++)	Θ(rows · cols)
c[i][j] = a[i][j] + b[i][j];	Θ(rows · cols)
}	0
Total	Θ(rows · cols)

- **Example 1.19 : *[Binary Search]***

[Program 1.6]

The instance characteristic -- number of elements in the list.

Each iteration of *while* loop takes Θ(1) time.

The *while* loop is iterated at most $\lceil \log_2(n+1) \rceil$ times.

Worst case - the loop is iterated **Θ(log n)** times

Best case - **Θ(1)**.

- **Example 1.21 : *[Magic square]***

  The magic square is an $n \times n$ matrix of integers from 1 to n^2 such that the sum of each row and column and two major diagonals is the same.

  When n=5 : the common sum is 65.

15	8	1	24	17
16	14	7	5	23
22	20	13	6	4
3	21	19	12	10
9	2	25	18	11

- **Coxeter's rule :**

  Put a one in the middle of the top row. Go up and left assigning numbers in increasing order to empty boxes. If your move cause you to jump off the square (that is, you go beyond the square's boundaries), figure out where you would be if you landed on a box on the opposite side of the square. Continue with this box. If a box is occupied, go down instead of up and continue.

[Program 1.22]

```
#include <stdio.h>
#define MAX_SIZE  15  /* maximum size of square */

void main(void)
/* construct a magic square, iteratively */
{
  static int square[MAX_SIZE] [MAX_SIZE];
  int i, j, row, column;       /* indices */
  int count;                   /* counter */
  int size;                    /* Square size */
```



```
printf ("Enter the size of the square: ");
scanf("%d', &size);
 /* check for input errors */
 if (size<1 || size>MAX_SIZE+1) {
            fprintf(stderr, "Error! Size is out of range\n");
            exit(1);
    }
 if (!(size % 2))   {
            fprintf(stderr, "Error! Size is even");
            exit(1);
    }
 for (i=0; i<size; i++)
    for (j=0; j<size; j++)
            square[i][j] = 0;
 square[0][(size-1)/2] = 1;  /* middle of first row */
```

```
/* i and j are current position  */
   i = 0;
   j = (size-1) / 2;
   for (count = 2; count <= size * size; count++)  {
           row = (i-1 < 0) ? (size-1) : (i-1);    /* up */
           column = (j-1 < 0) ? (size-1) : (j-1);  /* left */
           if (square[row][column])          /* down  */
             i = (++i) % size;
           else  {                    /* square is unoccupied */
             i = row;
             j = column;
           }
           square[i][j] = count;
     }
```



```
/* output the magic square */
   printf("Magic Square of the size %d : \n\n", size);
   for (i = 0; i < size; i++)   {
     for (j = 0; j < size; j++)
             printf ("%5d", square[i][j];
     printf("\n");
    }
  printf("\n \ n");
}
```

instance characteristic -- *n* denoting the size of the magic square.

the nested for loops -- $\Theta(n^2)$
next for loop -- $\Theta(n^2)$
Others --- $\Theta(1)$
Total asymptotic complexity is $\Theta(n^2)$.

1.4.4 Practical Complexities

- The time complexity of a program is generally some function of the instance characteristics.
- This complexity function:
  - is very useful in determining how the time requirements vary as the instance characteristics changes, and
  - may also be used to compare two programs P and Q that perform the same task.

Assume that program P has complexity $\Theta(n)$ and
program Q has complexity $\Theta(n^2)$.

We can assert that
P is faster than program Q for *sufficiently large* n.

How the various functions grow with n?

Instance characteristic *n*							
Time	Name	1	2	4	8	16	32
1	Constant	1	1	1	1	1	1
log n	Logarithmic	0	1	2	3	4	5
n	Linear	1	2	4	8	16	32
nlog n	Log linear	0	2	8	24	64	160
n^2	Quadratic	1	4	16	64	256	1024
n^3	Cubic	1	8	64	512	4096	32768
2^n	Exponential	2	4	16	256	65536	4294967296
n!	Factorial	1	2	24	40326	20922789888000	26313×10^{33}

Figure 1.7 Function values

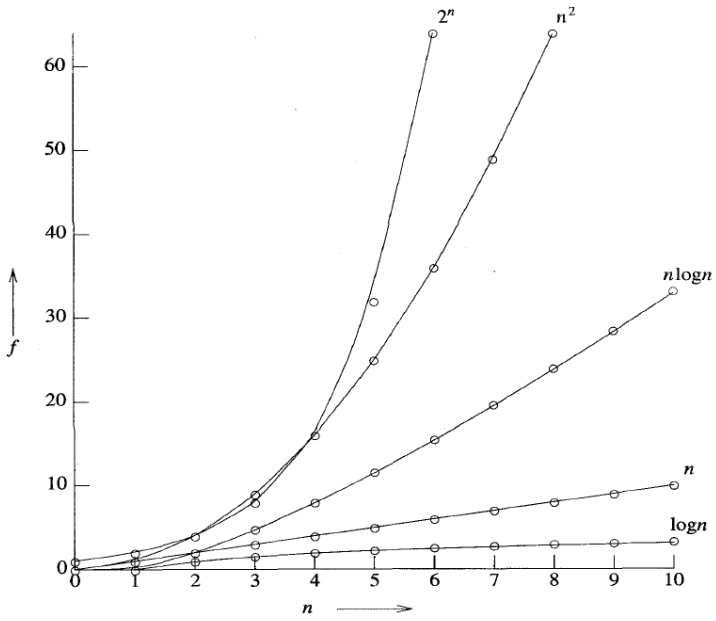


Figure 1.8 Plot of function values

	$f(n)$						
n	n	$n\log_2 n$	n^2	n^3	n^4	n^{10}	2^n
10	.01 μs	.03 μs	.1 μs	1 μs	10 μs	10 s	1 μs
20	.02 μs	.09 μs	.4 μs	8 μs	160 μs	2.84 h	1 ms
30	.03 μs	.15 μs	.9 μs	27 μs	810 μs	6.83 d	1 s
40	.04 μs	.21 μs	1.6 μs	64 μs	2.56 ms	121 d	18 m
50	.05 μs	.28 μs	2.5 μs	125 μs	6.25 ms	3.1 y	13 d
100	.10 μs	.66 μs	10 μs	1 ms	100 ms	3171 y	$4*10^{13}$ y
10^3	1 μs	9.96 μs	1 ms	1 s	16.67 m	$3.17*10^{13}$ y	$32*10^{283}$ y
10^4	10 μs	130 μs	100 ms	16.67 m	115.7 d	$3.17*10^{23}$ y	
10^5	100 μs	1.66 ms	10 s	11.57 d	3171 y	$3.17*10^{33}$ y	
10^6	1 ms	19.92 ms	16.67 m	31.71 y	$3.17*10^7$ y	$3.17*10^{43}$ y	

μs = microsecond = 10^{-6} seconds; ms = milliseconds = 10^{-3} seconds
s = seconds; m = minutes; h = hours; d = days; y = years

Figure 1.9 Times on a 1 billion instruction per second computer

1.5 PERFORMANCE MEASUREMENT

- **How to measure real execution time.**
  - Use of C's standard library.
    Functions are accessed through the statement:
    #include <time.h>.
  - Inaccurate results can be produced for small data

(e.g. if the value of CLK_TCK is 18 on our computer, the number of clock ticks for n < 500 is less than 10)

	Method 1	Method 2
Start timing	Start=clock();	Start=time(NULL);
Stop timing	Stop=clock();	Stop=time(NULL);
Type returned	Clock_t	Time_t
Result in seconds	Duration= ((double)(stop-start))/ CLOCKS_PER_SEC;	Duration= (double) difftime(stop, start);

Figure 1.10: Event timing in C

```c
#include <stdio.h>
#include <time.h>
#include "selectionSort.h"
#define MAX_SIZE 1001
void main(void)
{
   int i, n, step = 10;
   int a[MAX_SIZE];
   double duration;
   clock_t start;

   /* times for n = 0, 10, ..., 100, 200, ..., 1000 */
   printf("    n     time\n");
   for (n = 0; n <= 1000; n += step)
   {/* get time for size n */

      /* initialize with worst-case data */
      for (i = 0; i < n; i++)
         a[i] = n - i;

      start = clock( );
      sort(a, n);
      duration = ((double) (clock() - start))
                              / CLOCKS_PER_SEC;
      printf("%6d   %f\n", n, duration);
      if (n == 100) step = 100;
   }
}
```

Program 1.24: First timing program for selection sort


```c
#include <stdio.h>
#include <time.h>
#include "selectionSort.h"
#define MAX_SIZE 1001
void main(void)
{
   int i, n, step = 10;
   int a[MAX_SIZE];
   double duration;

   /* times for n = 0, 10, ..., 100, 200, ..., 1000 */
   printf("    n    repetitions     time\n");
   for (n = 0; n <= 1000; n += step)
   {
      /* get time for size n */
      long repetitions = 0;
      clock_t start = clock( );
      do
      {
         repetitions++;

         /* initialize with worst-case data */
         for (i = 0; i < n; i++)
            a[i] = n - i;

         sort(a, n);
      } while (clock( ) - start < 1000);
           /* repeat until enough time has elapsed */

      duration = ((double) (clock() - start))
                              / CLOCKS_PER_SEC;
      duration /= repetitions;
      printf("%6d  %9d   %f\n", n, repetitions, duration);
      if (n == 100) step = 100;
   }
}
```

Program 1.25: More accurate timing program for selection sort

n	repetitions	time
0	8690714	0.000000
10	2370915	0.000000
20	604948	0.000002
30	329505	0.000003
40	205605	0.000005
50	145353	0.000007
60	110206	0.000009
70	85037	0.000012
80	65751	0.000015
90	54012	0.000019
100	44058	0.000023
200	12582	0.000079
300	5780	0.000173
400	3344	0.000299
500	2096	0.000477
600	1516	0.000660
700	1106	0.000904
800	852	0.001174
900	681	0.001468
1000	550	0.001818

Figure 1.11: Worst case performance of selection sort (in seconds)

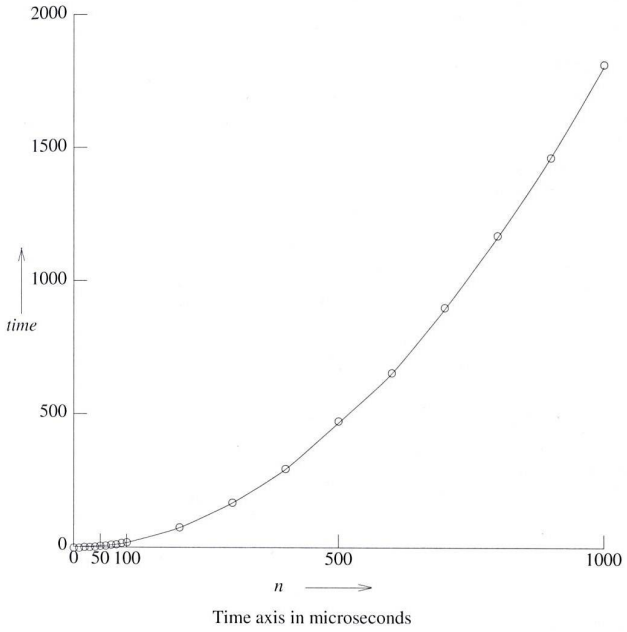


Figure 1.12: Graph of worst case performance of selection sort

Generating Test Data

- Generating a data set that results in the worst case performance of a program isn't always easy.
- We may generate a suitably large number of random test data.
- Obtaining average case data is usually much harder.
- It is desirable to analyze the algorithm being tested to determine classes of data that should be generated for the experiment - algorithm specific task.