PROBLEM # 1

Checking for element uniqueness in an array:

Consider an array of elements and determine if every element in the array is unique.

21	2	15	99	260	20	15	80	No
180	60	30	1	2	4	12	24	Yes

Left

Have we already seen this problem? If yes, how did we solve it?

SOLUTION

ALGORITHM *UniqueElements*$(A[0..n-1])$

```
//Determines whether all the elements in a given array are distinct
//Input: An array A[0..n − 1]
//Output: Returns "true" if all the elements in A are distinct
//                and "false" otherwise
for i ← 0 to n − 2 do
    for j ← i + 1 to n − 1 do
        if A[i] = A[j] return false
return true
```

What is the design strategy and time efficiency of this algorithm?

SOLUTION

- ✓ This algorithm is designed using Brute Force design strategy.
- ✓ Its worst – case efficiency is $\Theta(n^2)$.
- ✓ ***Can we improve the efficiency of this algorithm from $\Theta(n^2)$ to $\Theta(n)$?***
- ✓ If so, how?

PRESORTING

ALGORITHM *PresortElementUniqueness*$(A[0..n-1])$

//Solves the element uniqueness problem by sorting the array first
//Input: An array $A[0..n-1]$ of orderable elements
//Output: Returns "true" if A has no equal elements, "false" otherwise
sort the array A
for $i \leftarrow 0$ **to** $n-2$ **do**
    if $A[i] = A[i+1]$ **return false**
return true

PRESORTING

The worst – case efficiency of this algorithm is:

$$T(n) = T_{sort}(n) + T_{scan}(n)$$

$$\in \Theta(n \log n) + \Theta(n)$$

$$= \Theta (n \log n)$$

Thus, this algorithm is better than the brute – force algorithm if a ***n log n*** sorting algorithm is used.

PRESORTING

- ✓ This algorithm solves the problem of Element Uniqueness by first transforming the given input unsorted array to a sorted array and then checks only consecutive elements: if the array has equal elements, a pair of them must be next to each other and vice versa.

- ✓ This algorithm design strategy of transforming the given input instance to something more amenable to the solution and then solving it is called Transform and Conquer.

TRANSFORM AND CONQUER

THE IDEA

- ✓ The ***transform-and-conquer*** algorithms work as two-stage procedures.
- ✓ First, in the **transformation stage**, the problem's instance is modified to be, for one reason or another, more amenable to solution.
- ✓ Then, in the second or **conquering stage**, it is solved.
- ✓ The *transform-and-conquer* idea has three major variations.

VARIATIONS

Instance Simplification

Transformation to a simpler or more convenient instance of the same problem

Representation Change

Transformation to a different representation of the same instance

Problem Reduction

Transformation to an instance of a different problem for which an algorithm is already available

PRESORTING

- ✓ The ***presorting*** idea is an example of the Instance Simplification variety.
- ✓ Many questions about lists are easier to answer if the lists are sorted.
- ✓ The time efficiency of the algorithms that involve sorting will depend on the time efficiency of the sorting algorithm.

PROBLEM # 2

Computing a mode:

A *mode* is a value that occurs most often in a given list of numbers.

1	2	2	4	1	5	4	4	Mode = 4

What is the Brute Force way of solving this problem?

SOLUTION

- ✓ Scan the list and compute the frequencies of all its distinct values, then find the value with the largest frequency.
- ✓ In order to implement this idea, we can store the values already encountered, along with their frequencies, in a separate list.
- ✓ On each iteration, the i^{th} element of the original list is compared with the values already encountered by traversing this auxiliary list.
- ✓ If a matching value is found, its frequency is incremented; otherwise, the current element is added to the list of distinct values seen so far with a frequency of 1.

EFFICIENCY

- ✓ Worst – case input: A list with no matching elements.
- ✓ For such a list, its i^{th} element is compared with i – 1 elements of the auxiliary list of distinct values seen so far before being added to the list with a frequency of 1.
- ✓ The number of comparisons made by this algorithm is:

$$C(n) = \sum_{i=1}^{n} (i-1) = 0 + 1 + \cdots + (n-1) = \frac{(n-1)n}{2} \in \Theta(n^2).$$

PRESORTING SOLUTION

ALGORITHM *PresortMode*$(A[0..n-1])$

```
//Computes the mode of an array by sorting it first
//Input: An array A[0..n − 1] of orderable elements
//Output: The array's mode
sort the array A
i ← 0                          //current run begins at position i
modefrequency ← 0              //highest frequency seen so far
while i ≤ n − 1 do
    runlength ← 1; runvalue ← A[i]
    while i+runlength ≤ n − 1 and A[i+runlength] = runvalue
        runlength ← runlength+1
    if runlength > modefrequency
        modefrequency ← runlength; modevalue ← runvalue
    i ← i+runlength
return modevalue
```

PRESORTING

The worst – case efficiency of this algorithms is:

$$T(n) = T_{sort}(n) + T_{scan}(n)$$

$$\in \Theta(n \log n) + \Theta(n)$$

$$= \Theta (n \log n)$$

Thus, this algorithm is better than the brute – force algorithm if a n log n sorting algorithm is used.

PROBLEM # 3

Searching:

Searching for a given value *v* in a given array of *n* sortable items.

14	32	12	4	100	56	84	93	v = 40	Not Found

What is the Brute Force way of solving this problem?

BRUTE FORCE SOLUTION

- ✓ The Brute Force solution to this problem is Sequential Search.
- ✓ The worst – case efficiency of this algorithms is $\Theta(n)$.

PRESORTING SOLUTION

- ✓ The presorting solution to this problem is Binary Search whose running time will be:

$$T(n) = T_{sort}(n) + T_{search}(n) = \Theta(n \log n) + \Theta(\log n) = \Theta(n \log n)$$

- ✓ This efficiency is inferior to sequential search.

BALANCED SEARCH TREES

- ✓ A Binary Search Tree is a binary tree whose nodes contain elements of a set of orderable items, one element per node, so that all elements in the left subtree are smaller than the element in the subtree's root, and all the elements in the right subtree are greater than it.

- ✓ Example of Representation Change.

BALANCED SEARCH TREES

- ✓ The time efficiency of searching, insertion, and deletion, which are all in Θ(log *n*), but only in the average case.
- ✓ In the worst case, these operations are in Θ(n) because the tree can degenerate into a severely unbalanced one with its height equal to *n* - 1.
- ✓ Computer scientists have come up with two approaches which preserve the logarithmic efficiency of dictionary operations but avoids its worst case degeneracy.

AVL TREES

- ✓ An ***AVL tree*** is a binary search tree in which the ***balance factor*** of every node, which is defined as the difference between the heights of the node's left and right subtrees, is either 0 or + 1 or -1. (The height of the empty tree is defined as -1.)

- ✓ If an insertion of a new node makes an AVL tree unbalanced, we transform the tree by a rotation.

- ✓ A ***rotation*** in an AVL tree is a local transformation of its subtree rooted at a node whose balance has become either +2 or -2; if there are several such nodes, we rotate the tree rooted at the unbalanced node that is the closest to the newly inserted leaf.

AVL TREES

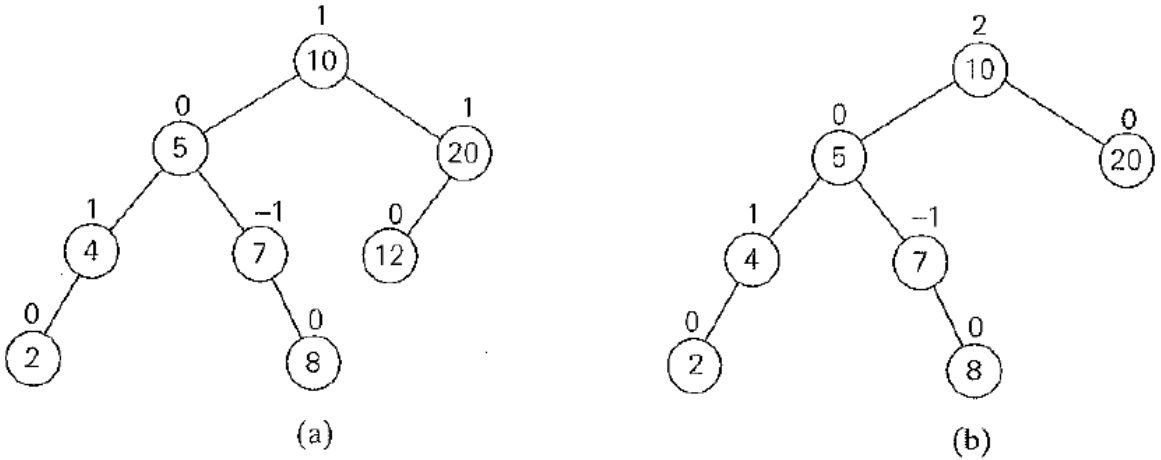


FIGURE 6.2 (a) AVL tree. (b) Binary search tree that is not an AVL tree. The number above each node indicates that node's balance factor.

AVL TREES - ROTATIONS

✓ There are only four types of rotations

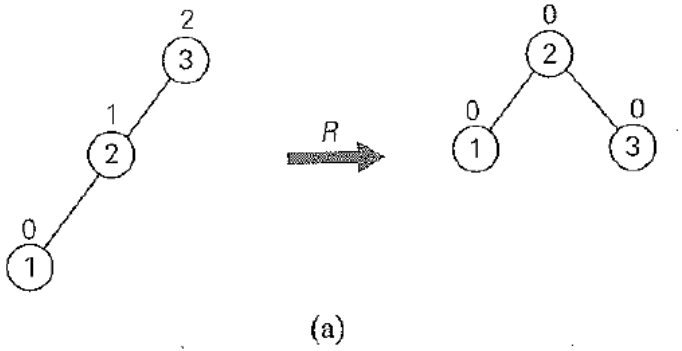
- *Single Right Rotation or R Rotation*
- *Single Left Rotation or L Rotation*
- *Double Left Right Rotation or LR Rotation*
- *Double Right Left Rotation or RL Rotation*

AVL TREES - ROTATIONS

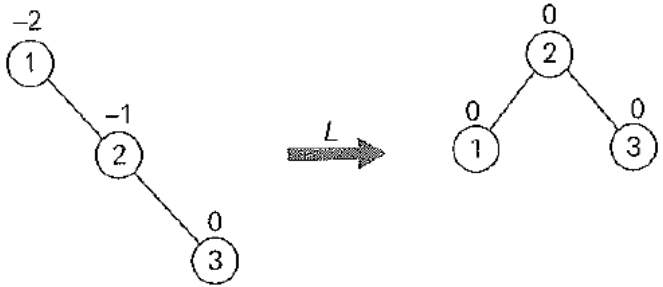
✓ There are only four types of rotations

- *Single Right Rotation or R Rotation*
- *Single Left Rotation or L Rotation*
- *Double Left Right Rotation or LR Rotation*
- *Double Right Left Rotation or RL Rotation*

AVL TREES - ROTATIONS



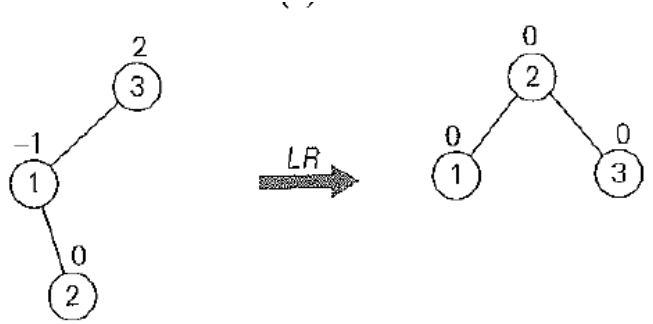
(a)



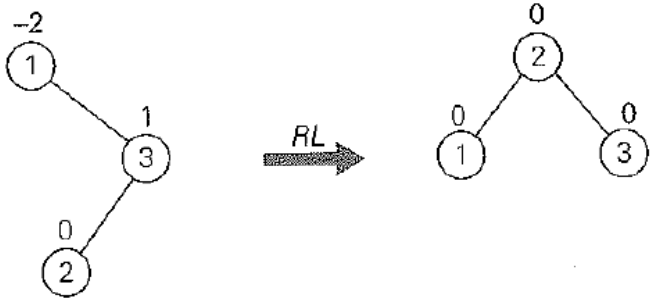
(b)

a) R Rotation
b) L Rotation

AVL TREES - ROTATIONS



(c)



(d)

c) LR Rotation
d) RL Rotation

AVL TREES - EFFICIENCY

- ✓ The operations of insertion, deletion and searching is logarithmic i.e., it belongs to Θ(log n).

2 – 3 TREES

- ✓ A *2-3 tree* is a tree that can have nodes of two kinds: 2-nodes and 3-nodes.
- ✓ A *2-node* contains a single key *K* and has two children: the left child serves as the root of a subtree whose keys are less than K and the right child serves as the root of a subtree whose keys are greater than *K*.
- ✓ A ***3-node*** contains two ordered keys K1 and *K2* (K1 < *K2)* and has three children. The leftmost child serves as the root of a subtree with keys less than K1, the middle child serves as the root of a subtree with keys between K1 and *K2,* and the rightmost child serves as the root of a subtree with keys greater than *K2*.

2 – 3 TREES

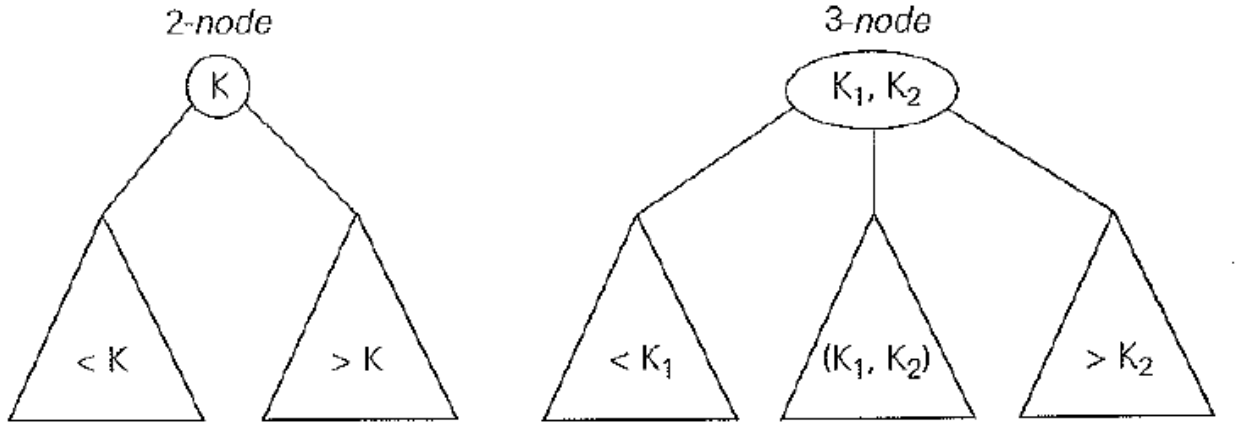


FIGURE 6.7 Two kinds of nodes of a 2-3 tree

2 – 3 TREES

- ✓ The last requirement of the 2-3 tree is that all its leaves must be on the same level, i.e., a 2-3 tree is always perfectly height-balanced: the length of a path from the root of the tree to a leaf must be the same for every leaf.

- ✓ The operations of insertion, deletion and searching can be applied to 2 – 3 Trees.

2 - 3 TREES - EFFICIENCY

- ✓ For any 2-3 tree of height *h* with *n* nodes, we get the inequality

$$n \geq 1 + 2 + \cdots + 2^h = 2^{h+1} - 1,$$

- ✓ And hence:

$$h \leq \log_2(n+1) - 1.$$

2 - 3 TREES - EFFICIENCY

- ✓ On the other hand, a 2 – 3 tree of height h with the largest number of keys is a full tree of 3 – nodes, each with two keys and three children. Therefore:

$$n \leq 2 \cdot 1 + 2 \cdot 3 + \cdots + 2 \cdot 3^h = 2(1 + 3 + \cdots + 3^h) = 3^{h+1} - 1,$$

- ✓ And hence:

$$h \geq \log_3(n+1) - 1.$$

- ✓ The lower and upper bounds on height h:

$$\log_3(n+1) - 1 \leq h \leq \log_2(n+1) - 1,$$

2 - 3 TREES - EFFICIENCY

- ✓ These upper and lower bounds indicate that the searching, insertion and deletion are all in $\Theta(\log n)$.

Red Black Trees

A red-black tree is a self balancing binary search tree in which:

- ✓ **Each node has a color (red or black) associated with it (in addition to its key and left and right children)**

And, the following 3 properties hold:

- ✓ **(root property) The root of the red-black tree is black**
- ✓ **(red property) The children of a red node are black.**
- ✓ **(black property) For each node with at least one null child, the number of black nodes on the path from the root to the null child is the same.**

Red Black Trees - Insert

- ✓ The goal of the insert operation is to insert key K into tree T, maintaining T's red-black tree properties.
- ✓ A special case is required for an empty tree.
- ✓ If T is empty, replace it with a single black node containing K. This ensures that the root property is satisfied.
- ✓ If T is a non-empty tree, then we do the following:

  use the BST insert algorithm to add K to the tree
  color the node containing K red
  restore red-black tree properties (if necessary)

Red Black Trees - Insert

Case 1: K's parent P is black

- ✓ If K's parent P is black, then the addition of K did not result in the red property being violated, so there's nothing more to do.

Red Black Trees - Insert

Case 2: K's parent P is red

- ✓ If K's parent P is red, then P now has a red child, which violates the red property. Note that P's parent, G, (K's grandparent) must be black (why?).
- ✓ In order to handle this double-red situation, we will need to consider the color of G's other child, that is, P's sibling, S. (Note that S might be null, i.e., G only has one child and that child is P.)
- ✓ We have two cases:

Red Black Trees - Insert

Case 2a: P's sibling S is black or null

If P's sibling S is black or null, then we will do a trinode restructuring of K (the newly added node), P (K's parent), and G (K's grandparent).

There are four possibilities for the relative ordering of K, P, and G.

Red Black Trees - Insert

Case 2b: P's sibling S is red

If P's sibling S is red, then we will do a recoloring of P, S, and G: the color of P and S is changed to black and the color of G is changed to red (unless G is the root, in which case we leave G black to preserve the root property).

B Trees

- ✓ A B-tree is a self-balancing tree data structure that maintains sorted data and allows searches, sequential access, insertions, and deletions in logarithmic time.
- ✓ The B-tree is a generalization of a binary search tree in that a node can have more than two children.
- ✓ Unlike self-balancing binary search trees, the B-tree is well suited for storage systems that read and write relatively large blocks of data, such as discs. It is commonly used in databases and file systems.

B Trees - Definition

According to Knuth's definition, a B-tree of order m is a tree which satisfies the following properties:

- ✓ Every node has at most m children.
- ✓ Every non-leaf node (except root) has at least [m/2] child nodes.
- ✓ The root has at least two children if it is not a leaf node.
- ✓ A non-leaf node with k children contains k − 1 keys.
- ✓ All leaves appear in the same level.

B Trees – Best and Worst Case Heights

$$h_{\min} = \lceil \log_m (n+1) \rceil - 1$$

$$h_{\max} = \left\lfloor \log_d \frac{n+1}{2} \right\rfloor .$$

B Trees - Insertion

All insertions start at a leaf node. To insert a new element, search the tree to find the leaf node where the new element should be added. Insert the new element into that node with the following steps:

- ✓ If the node contains fewer than the maximum allowed number of elements, then there is room for the new element. Insert the new element in the node, keeping the node's elements ordered.

B Trees - Insertion

- ✓ Otherwise the node is full, evenly split it into two nodes so:
- ❑ A single median is chosen from among the leaf's elements and the new element. Values less than the median are put in the new left node and values greater than the median are put in the new right node, with the median acting as a separation value.
- ❑ The separation value is inserted in the node's parent, which may cause it to be split, and so on. If the node has no parent (i.e., the node was the root), create a new root above this node (increasing the height of the tree).

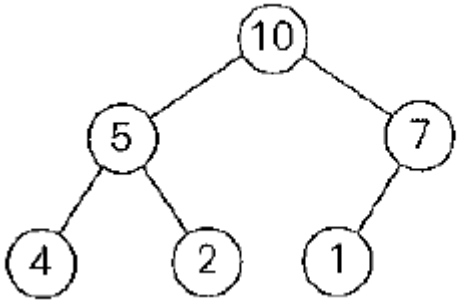
HEAPS AND HEAPSORT

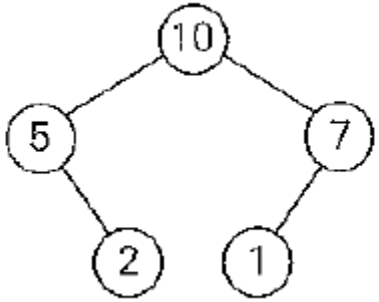
Definition:
A *heap* can be defined as a binary tree with keys assigned to its nodes (one key per node) provided the following two conditions are met:

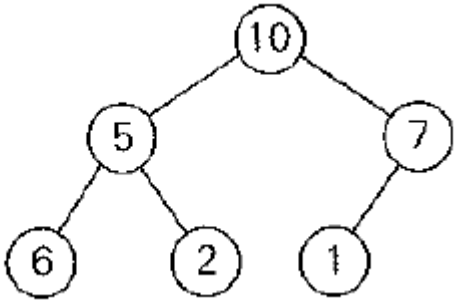
1. ***The tree's shape requirement*** - The binary tree is *essentially complete* (or simply *complete),* that is, all its levels are full except possibly the last level, where only some rightmost leaves may be missing.

2. ***The parental dominance requirement*** - The key at each node is greater than or equal to the keys at its children. (This condition is considered automatically satisfied for all leaves.)

HEAPS AND HEAPSORT







Only the leftmost tree is a heap. Why?

PROPERTIES OF HEAPS

1. There exists exactly one essentially complete binary tree with *n* nodes. Its height is equal to floor($\log_2 n$).

2. The root of a heap always contains its largest element.

3. A node of a heap considered with all its descendants is also a heap.

4. A heap can be implemented as an array by recording its elements in the topdown, left-to-right fashion. It is convenient to store the heap's elements in positions 1 though *n* of such an array, leaving H[O] either unused or putting there a sentinel whose value is greater than every element in the heap. In such a representation,

PROPERTIES OF HEAPS

a) the parental node keys will be in the first floor(n/2) positions of the array, while the leaf keys will occupy the last ceil(n/2) positions;

b) the children of a key in the array's parental position *i* (1 <= *I* <= *floor(*n/2)) will be in positions *2i* and *2i* + 1, and, correspondingly, the parent of a key in position *i* (2 <= *i* <= *n)* will be in position floor(i /2).

HEAP CONSTRUCTION – BOTTOM UP

ALGORITHM *HeapBottomUp*$(H[1..n])$

```
//Constructs a heap from the elements of a given array
// by the bottom-up algorithm
//Input: An array H[1..n] of orderable items
//Output: A heap H[1..n]
for i ← ⌊n/2⌋ downto 1 do
    k ← i; v ← H[k]
    heap ← false
    while not heap and 2 * k ≤ n do
        j ← 2 * k
        if j < n    //there are two children
            if H[j] < H[j + 1] j ← j + 1
        if v ≥ H[j]
            heap ← true
        else H[k] ← H[j]; k ← j
    H[k] ← v
```

HEAP CONSTRUCTION – BOTTOM UP (EFFICIENCY)

$$C_{worst}(n) = \sum_{i=0}^{h-1} \sum_{\text{level } i \text{ keys}} 2(h-i) = \sum_{i=0}^{h-1} 2(h-i)2^i = 2(n - \log_2(n+1)),$$

HEAP CONSTRUCTION – TOP DOWN

1. First, attach a new node with key *K* in it after the last leaf of the existing heap.
2. Then sift *K* up to its appropriate place in the new heap as follows.
3. Compare *K* with its parent's key: if the latter is greater than or equal to *K*, stop (the structure is a heap);
4. otherwise, swap these two keys and compare *K* with its new parent.
5. This swapping continues until *K* is not greater than its last parent or it reaches the root.
6. In this algorithm, too, we can sift up an empty node until it reaches its proper position, where it will get *K* 's value.

HEAP CONSTRUCTION – TOP DOWN EFFICIENCY

Efficiency of insertion is O(log n)

HEAP DELETION

Maximum Key Deletion from a heap

Step 1 Exchange the root's key with the last key K of the heap.

Step 2 Decrease the heap's size by 1.

Step 3 "Heapify" the smaller tree by sifting K down the tree exactly in the same way we did it in the bottom-up heap construction algorithm. That is, verify the parental dominance for K: if it holds, we are done; if not, swap K with the larger of its children and repeat this operation until the parental dominance condition holds for K in its new position.

The efficiency of deletion is O(log n).

HEAP SORT

Stage 1 (heap construction): Construct a heap for a given array.

Stage 2 (maximum deletions): Apply the root-deletion operation $n - 1$ times to the remaining heap.

HEAP SORT – EFFICIENCY

The number of key comparisons:

$$C(n) \le 2\lfloor \log_2(n-1) \rfloor + 2\lfloor \log_2(n-2) \rfloor + \cdots + 2\lfloor \log_2 1 \rfloor \le 2 \sum_{i=1}^{n-1} \log_2 i$$

$$\le 2 \sum_{i=1}^{n-1} \log_2(n-1) = 2(n-1)\log_2(n-1) \le 2n \log_2 n.$$

C(n) belongs to O(n log n) for the second stage of Heap Sort.

HEAP SORT – EFFICIENCY

For both stages:

O(n) + O(n log n) belongs to O(n log n).

HORNER'S RULE AND BINARY EXPONENTIATION

PROBLEM

Computing the value of a polynomial:

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$$

at a given point x and its special case of computing x^n.

HORNER'S METHOD

- ✓ Is an algorithm for calculating polynomials.
- ✓ Named after British mathematician, ***William George Horner***, who published it in the 19^{th} century.
- ✓ An example of Representation Change variety of Transform – and – Conquer.

HORNER'S METHOD – Representation Change

$$p(x) = (\ldots(a_n x + a_{n-1})x + \ldots)x + a_0.$$

For example, for the polynomial $p(x) = 2x^4 - x^3 + 3x^2 + x - 5$, we get

$$\begin{aligned} p(x) &= 2x^4 - x^3 + 3x^2 + x - 5 \\ &= x(2x^3 - x^2 + 3x + 1) - 5 \\ &= x(x(2x^2 - x + 3) + 1) - 5 \\ &= x(x(x(2x - 1) + 3) + 1) - 5. \end{aligned}$$

HORNER'S METHOD – Algorithm

ALGORITHM *Horner*$(P[0..n], x)$

//Evaluates a polynomial at a given point by Horner's rule
//Input: An array $P[0..n]$ of coefficients of a polynomial of degree n
// (stored from the lowest to the highest) and a number x
//Output: The value of the polynomial at x
$p \leftarrow P[n]$
for $i \leftarrow n-1$ **downto** 0 **do**
  $p \leftarrow x * p + P[i]$
return p

HORNER'S METHOD – Pen and Pencil Evaluation

EXAMPLE 1 Evaluate $p(x) = 2x^4 - x^3 + 3x^2 + x - 5$ at $x = 3$.

coefficients	2	−1	3	1	−5
$x = 3$	2	$3 \cdot 2 + (-1) = 5$	$3 \cdot 5 + 3 = 18$	$3 \cdot 18 + 1 = 55$	$3 \cdot 55 + (-5) = 160$

HORNER'S METHOD – Number of Multiplications and Divisions

$$M(n) = A(n) = \sum_{i=0}^{n-1} 1 = n.$$

Comparison: Using Brute Force design strategy, just to compute a single term a^n would have required n multiplications.

PROBLEM

Computing a^n.

The Horner's Rule for computing a^n degenerates to the Brute Force multiplication of a by itself.

This can be improvised.

We see two algorithms for computing a^n which use the binary representation of the exponent.

1. The Left – to – right Binary Exponentiation
2. The Right – to – left Binary Exponentiation

SOLUTION

Computing a^n.

Let $n = b_1 \ldots b_i \ldots B_0$ be a bit string representing a positive integer n in the Binary Number System.

Value of n can be computed as the polynomial:

$$p(x) = b_l x^l + \cdots + b_i x^i + \cdots + b_0$$

SOLUTION

Let us now compute this polynomial by Horner's Rule:

Horner's rule for the binary polynomial $p(2)$	**Implications for** $a^n = a^{p(2)}$
$p \leftarrow 1$ //the leading digit is always 1 for $n \geq 1$ **for** $i \leftarrow I-1$ **downto** 0 **do** $\quad p \leftarrow 2p + b_i$	$a^p \leftarrow a^1$ **for** $i \leftarrow I-1$ **downto** 0 **do** $\quad a^p \leftarrow a^{2p+b_i}$

But:

$$a^{2p+b_i} = a^{2p} \cdot a^{b_i} = (a^p)^2 \cdot a^{b_i} = \begin{cases} (a^p)^2 & \text{if } b_i = 0 \\ (a^p)^2 \cdot a & \text{if } b_i = 1 \end{cases}.$$

Left – To – Right Binary Exponentiation

ALGORITHM *LeftRightBinaryExponentiation*$(a, b(n))$

//Computes a^n by the left-to-right binary exponentiation algorithm
//Input: A number a and a list $b(n)$ of binary digits $b_l, \ldots, b_0$
// in the binary expansion of a positive integer n
//Output: The value of a^n
$product \leftarrow a$
for $i \leftarrow l - 1$ **downto** 0 **do**
  $product \leftarrow product * product$
  if $b_i = 1$ $product \leftarrow product * a$
return *product*

Left – To – Right Binary Exponentiation Efficiency

The number of multiplications M(n) is given by:

$$(b - 1) \leq M(n) \leq 2(b - 1),$$

$b - 1 = \text{floor}(\log_2 n)$

The efficiency of this algorithm is logarithmic.

YET ANOTHER SOLUTION

Computing a^n.

$$a^n = a^{b_l 2^l + \cdots + b_i 2^i + \cdots + b_0} = a^{b_l 2^l} \cdot \ldots \cdot a^{b_i 2^i} \cdot \ldots \cdot a^{b_0}.$$

Thus, a^n can be computed as the product of the terms

$$a^{b_i 2^i} = \begin{cases} a^{2^i} & \text{if } b_i = 1 \\ 1 & \text{if } b_i = 0 \end{cases},$$

Right – To – Left Binary Exponentiation

ALGORITHM *RightLeftBinaryExponentiation*$(a, b(n))$

//Computes a^n by the right-to-left binary exponentiation algorithm
//Input: A number a and a list $b(n)$ of binary digits $b_l, \ldots, b_0$
// in the binary expansion of a nonnegative integer n
//Output: The value of a^n
$term \leftarrow a$ //initializes a^{2^l}
if $b_0 = 1$ $product \leftarrow a$
else $product \leftarrow 1$
for $i \leftarrow 1$ **to** l **do**
    $term \leftarrow term * term$
    if $b_i = 1$ $product \leftarrow product * term$
return $product$

Left – To – Right Binary Exponentiation Efficiency

The efficiency of this algorithm is logarithmic.