第二章　排序

2.1　练习

Problem 2.1.1

(a) 证明下面的算法是对输入进行排序的。

Algorithm 9: STOOGE-SORT($A[0,...,n-1]$)

if $n = 2$ *and* $A[0] > A[1]$ **then**
　swap the values of $A[0]$ and $A[1]$;
else if $n > 2$ **then**
　$m \leftarrow \lceil 2n/3 \rceil$;
　STOOGE-SORT(A[0,...,m-1]);
　STOOGE-SORT(A[n-m,...,n-1]);
　STOOGE-SORT(A[0,...,m-1]);

(b) 如果将$m \leftarrow \lceil 2n/3 \rceil$换成$m \leftarrow \lfloor 2n/3 \rfloor$，算法还正确吗？请说明。

(c) 请给出算法中比较次数总数的递归表达式。

(d) 求解递归表达式，并证明你的解是正确的。

(e) 证明算法中$swap$操作执行的次数最多是$\begin{pmatrix} n \\ 2 \end{pmatrix}$。

Problem 2.1.2

给定正整数数组$A = [3, 1, 4, 15, 9, 2, 6, 53, 58, 97, 93, 23, 8, 46, 26, 43, 6, 97]$。分别使用插入排序，堆排序，快速排序，合并排序对该数组进行排序。请图示各个算法执行的详细过程。

Problem 2.1.3

请修改$QuickSort$算法中的$Partition$部分对数组中的n个元素重新排序使得所有负数排在非负数的前面。

Problem 2.1.4

请给出一个时间为$O(nlogk)$用来将k个已排序链表合成一个排序链表的算法。这里n表示所有输入链表中元素的总数。

Problem 2.1.5

首先在二叉树中定义节点的深度（depth）和树与节点的高度（height）的概念。

- 根节点的深度是0。非根节点的深度是其父节点的深度加1。
- 二叉树的高度是其所有叶节点的深度的最大值。二叉树中一个结点的高度就是以该结点为根的子树的高度。

请证明：在一个有n个结点的堆中，所有结点的高度之和最多为$n-1$。并请说明在何种情况堆中所有节点的高度之和正好为$n-1$。

Problem 2.1.6
证明:对于所有整数$h \geq 1, \lceil \lg(\lfloor \frac{1}{2}h \rfloor + 1) \rceil + 1 = \lceil \lg(h+1) \rceil$。

Problem 2.1.7
假设两个待合并的数组长度分别为k与m，其中k远小于m。试给出一个合并算法，它能够利用k远小于m这个特点，使得其比较次数至多为$(k+m)/2$。在最坏情况下，为了获得上述的边界，k必须多小？k是否存在一个范围，使得边界达到$\sqrt{k+m}$？你能否得出在这些情况下元素的移动次数？

Problem 2.1.8
有两个有序数组A与B，它们的长度分别为k，m，其中$k+m=n$，证明：将这两个数组进行合并时，合并的不同形式的个数为$\binom{n}{k} = \binom{n}{m}$。

Problem 2.1.9
如果先对一个矩阵中的每一列进行排序，然后再对每一行进行排序，请证明每一列的数据仍然是有序的。

Problem 2.1.10
令M是一个$m \times n$的矩阵，矩阵中的每一行元素从左到右按照升序排列，每一列的元素从上到下按照升序排列。请给出高效的算法在矩阵中查找给定的元素x（x可能不存在），并给出最坏情况下所需要的比较次数。

Problem 2.1.11
1. 给出一个算法，它在最坏情况下可以只利用五次比较来对四个元素进行排序。
2. 给出一个算法，使得在最坏情况下，对五个元素进行排序时，该算法有最优的表现。

Problem 2.1.12
请画出为四个元素x_1, x_2, x_3和x_4排序过程的决策树。

2.2 问题

Problem 2.2.1
(a) 请描述算法通过调用子过程SQRTSORT(k)对输入的数组$A[1...n]$进行排序。SQRTSORT(k) 是对子数组$A[k+1...k+\sqrt{n}]$ 进行排序的原地（in place）算法，$k(0 \leq k \leq n-\sqrt{n})$是算法的输入（为了简化问题，这里假定$\sqrt{n}$为整数）。给出的算法只允许调用SQRTSORT检查或修改数组，即你的算法不能直接比较，移动或者复制数组元素。请回答最坏情况下，你的算法所需要调用SQRTSORT的次数。

(b) 证明你的解决(a)的算法在常数因子的意义下是最优的，也就是说，如果你的算法调用SQRTSORT的次数是$f(n)$，证明没有算法能够调用$o(f(n))$次SQRTSORT完成排序。

(c) 假设SQRTSORT通过调用你解决(a)的算法递归实现。例如，在递归的第二层，算法对数组中大约$n^{1/4}$个元素进行排序。最坏情况下，算法的时间复杂度是多少？

Problem 2.2.2

给定n个螺钉，并且他们的尺寸互不相同，另外给定n个螺母，并且每个螺钉都与唯一的螺母匹配。你可以用一个螺钉旋入一个螺母来判断他们是否匹配，或者螺母比螺钉大，又或者螺母比螺钉小。但是你不可以将两个螺钉（螺母）进行比较。请给出复杂度为$\Theta(n \log n)$的算法找出所有匹配的螺钉和螺母对。

Problem 2.2.3 (The *d*-ary Heap)

一个$d-ary$堆可以用一个一维数组表示。根节点存放在$A[1]$，它的子女节点依次放在$A[2],...,A[d+1]$，这些子女节点的子女节点存放在$A[d+2],...,A[d^2+d+1]$，依次类推。请证明下面两个分别计算节点下标为i的节点的父节点和它的第j（$1 \leq j \leq d$）个子女节点的过程的正确性。

- D-ARY-PARENT(i)

  return $\lfloor \frac{i-2}{d} + 1 \rfloor$;

- D-ARY-CHILD(i, j)

  return d(i-1) + j +1;

Problem 2.2.4 (Modified BubbleSort)

我们可以修改简单的冒泡排序算法来避免在数组尾部进行的不必要的比较操作，做法是记录每次for循环最后发生交换的位置[1]。请说明这样的做法是否会影响算法的平均时间复杂度。

Problem 2.2.5

假设现在有$2n$个人参加一个游戏，每个人都有一个评分表示他擅长游戏的程度。现在需要将这$2n$个人分为两队，每队有n个人，请给出算法能够在$O(nlogn)$的时间内进行最不公平的划分（两队的评分和之差最大）。

Problem 2.2.6

现在有$2n$个实数值，请给出复杂度为$O(nlogn)$的算法将这$2n$个数分为n对，并且使得所求得的划分的数对和的最大值在所有划分的数对和最大值中是最小的。例如，$(1,3,5,9)$ 可以分为$((1,3)(5,9))$，$((1,5)(3,9))$和$((1,9)(3,5))$，这三种划分的数对和组成是$(4,14),(6,12),(10,8)$，第三种划分中10是它的数对和的最大值，并且它也是三种划分方法的数对和最大值$14,12,10$中的最小值。

Problem 2.2.7 (计算逆序对的个数)

假定数组$A[1..n]$中存放着n个各不相同的整数。我们定义下标二元组(i,j)为一个逆序对，如果$i<j$，且$A[i]>A[j]$。

1. 请设计一个$O(n \log n)$的算法计算给定数组中所有逆序对的个数。

2. 我们可以将逆序对的定义进行推广：我们定义下标二元组(i,j)为一个广义逆序对，如果对于预先给定的正整数常数C，$i<j$，且$A[i]>C \cdot A[j]$。显然当$C=1$时，该定义退化为传统逆序对。现给定$C=2$，请设计一个算法计算数组中所有的广义逆序对。

[1]请参考Sara Baase and Allen Van Gelder. Computer Algorithms-Introduction to Design and Analysis(算法设计与分析)第208页题目4.4

Problem 2.2.8

回文是指一个单词改变其中的字母的顺序就会构成另外一个单词。例如，“ate”，“tea”和“eat”属于一个回文集合。请给出算法找出一篇篇幅很大的英文文件中的回文集合。描述原理，而略去细节。

Problem 2.2.9 (空闲时间)

假设有一台并行计算机在处理N个任务，现在给定一组任务的开始和结束时间，请设计算法用来发现机器的最长空闲时间以及最长的非空闲时间。