Data Structures and Algorithms

Spring 2014

Cesar Agustin Garcia Vazquez

March 10, 2014

1 Problem Set 04

1. Suppose we have an array A containing n numbers, some of which may be negative. We wish to find indices i and j so that

$$\sum_{k=i}^{j} A[k]$$

is maximized. Find an algorithm that runs in time $O(n)$.

2. A challenge that arises in databases is how to summarize data in easy-to-display formats, such as a histogram. A problem in this context is the minimal imbalance problem. Again suppose we have an array A containing n numbers, this time all positive, and another input k. Consider k indices $j_1, j_2 \ldots, j_k$ that partition the array into $k+1$ subarrays $A[1, j_1], A[j_1+1, j_2], \ldots, A[j_k+1, n]$. The weight $w(i)$ of the ith subarray is the sum of its entries. The imbalance of the partition is

$$\max_i \left| w(i) - \frac{\sum_{l=1}^{n} A[k]}{k+1} \right|$$

That is, the imbalance is the maximum deviation any partition has from the average size.

   a) Give an algorithm for determining the partition with the minimal imbalance given A, n, and k. (This corresponds to finding a histogram with k breaking points, giving $k+1$ bars, as close to equal as possible, in some sense.)

   Answer: Also, an upper bound for the maximum imbalance could happen when $w(i) = -\sum_{l=1}^{n} A[k]$, so an upper bound could be defined as in equation (1).

$$\left| w(i) - \frac{\sum_{l=1}^{n} A[k]}{k+1} \right| = \left| -\sum_{l=1}^{n} A[k] - \frac{\sum_{l=1}^{n} A[k]}{k+1} \right|$$

$$
\begin{aligned}
&= \left|\frac{-(k+1)\sum_{l=1}^{n} A[k] - \sum_{l=1}^{n} A[k]}{k+1}\right| \\
&= \left|\frac{-(k+2)\sum_{l=1}^{n} A[k]}{k+1}\right| \\
&= \frac{(k+2)\sum_{l=1}^{n} A[k]}{k+1} \\
&= \frac{k+2}{k+1}\sum_{l=1}^{n} A[k] \qquad (1)
\end{aligned}
$$

Algorithm 1: Algorithm to determine the minimum imbalance

```
1  funcion minimumImbalance(A, n, k)
2  w = Array of size k + 1 // Weights for each partition
3  t = 0 // Global total
4  j = 0 // Index of each partition
5  for i = 0; i < A.size; i++ do
6  |   t += A[i]
7  |   w[j] += A[i]
8  |   if j == k[j] then
9  |   |   j++
10 m = t + 1 // Upper bound
11 a = t / (k.size + 1);
```

b) Explain how your algorithm would change if the imbalance was redefined to be

$$\sum_{i} \left| w(i) - \frac{\sum_{l=1}^{n} A[k]}{k+1} \right|$$

3. Suppose we want to print a paragraph neatly on a page. The paragraph consists of words of length $l_1, l_2, \ldots, l_n$. The maximum line length is M. (Assume $l_i \leq M$ always.) We define a measure of neatness as follows. The extra space on a line (using one space between words) containing words l_i through l_j is $M - j + 1 - \sum_{k=1}^{j} l_k$. The penalty is the sum over all lines **except the last** of the **cube** of the extra space at the end of the line. This has been proven to an effective heuristic for neatness in practice. Find a dynamic programing algorithm to determine the neatest way to print a paragraph. Of course you should provider a recursive definition of the value of the optimal solution that motivates your algorithm.
   For this problem, besides explaining/proving your algorithms as for other problems on the set, yous hold also code up your algorithm to print an optimal division of words into lines. The output should be theft split into lines appropriately, and the numerical value of the penalty. You can use any coding language you wish. You should assume that a *word* in this context is any contiguous sequence of characters

not including blank spaces.
After coding your algorithm, determine the minimal penalty for the following review of Season 1 Buffy DVD, apparently written by Ryan Crackell for the Apollo Guide, for the cases where $M = 40$ and $M = 72$. We will try to put the text of the review on the class page as well.
Buffy the Vampire Slayer fans are sure to get their fix with the DVD release of the show's first season. The three-disc collection includes all 12 episodes as well as many extras. There is a collection of interviews by the show's creator Joss Whedon in which he explains his inspiration for the show as well as comments on the various cast members. Much of the same material is covered in more depth with Wheadon's commentary track for the show's first two episodes that make up the Buffy The Vampire Slayer pilot. The most interesting points of Whedon's commentary come from his explanation of the learning curve he encountered shifting from blockbuster films like Toy Story to a much lower-budget television series. The first disc also includes a short interview with David Boreanaz who plays the role of Angel. Other features include the script for the pilot episodes, a trailer, a large photo gallery of publicity shots and in-depth biographies of Whedon and several of the show's stars, including Sarah Michelle Geller, Alyson Hannigan and Nicholas Brendon.