CSCI 6212 – Design and Analysis of Algorithms (Fall 2023)  
Project Assignment 3  
Abde Manaaf Ghadiali – G29583342, Gehna Ahuja – G35741419, Venkatesh Shanmugam – G00000000, Manikanta Allanki – G00000000

Problem Statement - Maximum Value but Limited Neighbours

You are given an array a[1…n] of positive numbers and an integer k. You must produce an array b[1…n], such that:

1. For each j, b[j] is 0 or 1,
2. Array b has adjacent 1s at most k times, and
3. sum\_{j=1 to n} a[j]\*b[j] is maximized.

For example, given an array [100, 300, 400, 50] and integer k = 1, the array b can be: [0 1 1 0], which maximizes the sum to be 700. Or, given an array [10, 100, 300, 400, 50, 4500, 200, 30, 90] and k = 2, the array b can be [1, 0, 1, 1, 0, 1, 1, 0, 1] which maximizes the sum to 5500. To be precise about the definition of adjacency: sequence [0, 1, 0, 1, 0, 1, 1, 1] has two adjacent 1s. Sequence [0, 1, 0, 0, 1, 1, 1, 1] has 3 adjacent 1s. Sequence [1, 0, 1, 1, 0, 1, 1, 1] also has 3 adjacent 1s.

Theoretical Analysis

**Notation** – Let MVLN[i][j] be the maximum sum of a[j]\*b[j] for 1 <= j <= i, subject to the constraints that there are at most k adjacent 1s in b[1…i].

**Proof of Optimality** – We can prove that MVLN[i][j] is the optimal solution by induction on i. The base case is when i = 1, in which case MVLN[1][j] = a[1]\*b[1] if j = 1 and 0 otherwise. This is clearly optimal for the subproblem with only one element.

For the inductive step, assume that MVLN[i-1][j] is optimal for any 0 <= j <= i. Then we can show that MVLN[i][j] is optimal by considering two cases:

1. If k = 0, then MVLN[i][j] = MVLN[i - 1][j] + MVLN[i - 1][j], which is the maximum sum we can get by appending a 0 to any optimal solution for the subproblem with i - 1 elements; if k = 1, then MVLN[i][j] = MAX(MVLN[i - 1][j], MVLN[i - 2][j]) + a[i], which is the maximum sum we can get by appending a 1 to any optimal solution for the subproblem with i - 1 elements that ends with a 0 or two 0s.

Therefore, by induction, MVLN[i][j] is optimal for the original problem.

**Recurrence Relation** – MAX(MVLN[i - 1][j], MVLN[i - 1][j - 1] + a[i], MVLN[i - 2][j - 2] + a[i])

**Algorithm** –

**Time Complexity** – O(n \* k) or O(nk)

Experimental Analysis

* **GitHub Project Repository Link –** [**https://github.com/default741/CSCI\_6212\_Course\_Notes/blob/main/project-files/project-03/msnn.py**](https://github.com/default741/CSCI_6212_Course_Notes/blob/main/project-files/project-03/msnn.py)
* **Program Listing**

I have executed the code for values of n ranging from 100 to 4900 with increments of 200 and k being a random number between 1 to √n, which can be seen in the following section of Output Numerical Data (due to the restriction in # of pages, have only included 15 rows).

* **Data Normalization Notes**

To Normalize the theoretical time, we take the average of the experimental time and divide that by the average of the theoretical time, that give us the scaling constant which we multiply with the theoretical time. This gets the theoretical and experimental values in the same range. But the values of n are still too large to compare with the execution times. So, we log the values of n as well as we log the values of theoretical and experimental execution Times.

* **Output Numerical Data**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Value of 'n' | Value of 'k' | Theoretical Time Taken (O(n \* k)) | Experimental Time Taken (ns) | Scaling Constant | Theoretical Time x Scaling Constant | Log based 10 for (Theoretical Time x Scaling Constant) | Log based 10 for Experimental Time (ns) |
| 100 | 7 | 7.00E+02 | 4.64E+05 | 631.48 | 4.42E+05 | 5.6455 | 5.6668 |
| 300 | 11 | 3.30E+03 | 2.10E+06 | 631.48 | 2.08E+06 | 6.3189 | 6.3224 |
| 500 | 3 | 1.50E+03 | 1.15E+06 | 631.48 | 9.47E+05 | 5.9764 | 6.0602 |
| 700 | 14 | 9.80E+03 | 6.19E+06 | 631.48 | 6.19E+06 | 6.7916 | 6.7914 |
| 900 | 22 | 1.98E+04 | 1.24E+07 | 631.48 | 1.25E+07 | 7.0970 | 7.0922 |
| 1100 | 8 | 8.80E+03 | 5.82E+06 | 631.48 | 5.56E+06 | 6.7448 | 6.7647 |
| 1300 | 3 | 3.90E+03 | 2.94E+06 | 631.48 | 2.46E+06 | 6.3914 | 6.4691 |
| 1500 | 32 | 4.80E+04 | 2.97E+07 | 631.48 | 3.03E+07 | 7.4816 | 7.4725 |
| 1700 | 11 | 1.87E+04 | 1.20E+07 | 631.48 | 1.18E+07 | 7.0722 | 7.0802 |
| 1900 | 26 | 4.94E+04 | 3.07E+07 | 631.48 | 3.12E+07 | 7.4941 | 7.4873 |
| 2100 | 34 | 7.14E+04 | 4.46E+07 | 631.48 | 4.51E+07 | 7.6541 | 7.6498 |
| 2300 | 24 | 5.52E+04 | 3.47E+07 | 631.48 | 3.49E+07 | 7.5423 | 7.5408 |
| 2500 | 5 | 1.25E+04 | 8.70E+06 | 631.48 | 7.89E+06 | 6.8973 | 6.9397 |
| 2700 | 38 | 1.03E+05 | 6.37E+07 | 631.48 | 6.48E+07 | 7.8115 | 7.8043 |
| 2900 | 31 | 8.99E+04 | 5.64E+07 | 631.48 | 5.68E+07 | 7.7541 | 7.7516 |

* **Graph**

* **Graph Observations**

We can visually observe from the graph that the Theoretical Execution Time and Experimental Execution Time follow a similar path for same values of n. Since we assume that the variable assignment and conditional checking and expression evaluation takes constant time for theoretical values, it does not hold true for experimental values. Hence, we see the Bumps in the graph for the experimental values.

Conclusion

Hence, we can say that the dynamic programming algorithm has the same experimental time as the theoretical time of O(nk) to find the maximum value but with limited neighbours.