Malek M. Moussa

Professor Prasanna Joeg

CMSC 441

13 May 2025

Paper 3

1. Problem Definition & Usefulness

- **Definition:** An Optimal Binary Search Tree (OBST) is a BST constructed for a set of keys (with given access probabilities) such that the expected search cost is minimized.
- **Applications:**
    - **Databases:** Faster lookups for frequent queries.
    - **Compilers:** Efficient symbol table access.
    - **Huffman-like encoding:** Optimizing weighted path lengths.

2. Size of Solution Space

- **Answer:** The number of possible BSTs is the n-th Catalan number:

$$C_n = \frac{1}{n+1}\begin{bmatrix} 2n \\ n \end{bmatrix}$$

- **Justification:** Each unique BST structure corresponds to a valid parenthesis combination (counted by Catalan numbers).

3. Brute Force Algorithm

- **Algorithm:**
    1. Generate all possible BST structures (using Catalan combinations).
    2. For each BST, compute its expected search cost.
    3. Select the BST with the minimum cost.
- **Time Complexity:** $O(n \cdot 4^n/n^{1.5})$ (exponential).
- **Justification:**
    1. $C_n \approx 4^n/n^{1.5}$ (Catalan growth).
    2. Computing cost for each BST is O(n).

4. Divide & Conquer (D&C) Approach

- **Possible?** No, because of overlapping subproblems.
- **Reason:** A pure D&C approach would recompute subtree costs repeatedly (inefficient).

5. Dynamic Programming (DP) Approach

- **Algorithm:**
    - Let dp[i][j] = min cost for keys i to j.
    - **Recurrence:**

$$dp[i][j] = min_{k=i}^{j}(dp[i][k-1] + dp[k+1][j]) + \sum_{x=i}^{j} freq[x]$$

    - **Base Case:** dp[i][i]=freq[i].
- **Time Complexity:** $O(n^3)$ (three nested loops).
- **DP vs D&C:** DP is better (avoids recomputation via memoization).

6. Greedy Method

- **Possible?** No.
- **Reason:** Selecting the highest probability root first (greedy) does not guarantee global optimality.
- **Counterexample:** A less frequent root might yield a better-balanced tree overall.

7. Backtracking & Branch & Bound (B&B)

- **Backtracking:**
    - Explore all BSTs, prune paths where cost > current best.

- **Branch & Bound:**
    - Use a priority queue (min-heap) for partial solutions.
    - **Bound:** If a subtree's lower bound cost > best known, discard it.

8. Extra Credit: Space Optimization in DP

- **Question:** Can we reduce DP space complexity?
- **Answer:**
    - **Yes**, from $O(n^2) \rightarrow O(n)$ by filling diagonally (since we only need previous diagonals).
    - **But** time remains $O(n^3)$.

Works Cited

Kleinberg, Jon, and Éva Tardos. *Algorithm Design*. Pearson, 2006.

"Program for nth Catalan Number." *GeeksforGeeks*, link.

CodeBix. "Catalan Numbers Variation Explained | Dynamic Programming Interview Questions." *YouTube*, 15 June 2020, link.