

To resolve the question by a careful enumeration of solutions via trial and error, continued Gauss, would take only an hour or two. Apparently such inelegant work held little attraction for Gauss, for he does not seem to have carried it out, despite outlining in detail how to go about it.

— Paul Campbell, “Gauss and the Eight Queens Problem: A Study in Miniature of the Propagation of Historical Error” (1977)

I dropped my dinner, and ran back to the laboratory. There, in my excitement, I tasted the contents of every beaker and evaporating dish on the table. Luckily for me, none contained any corrosive or poisonous liquid.

— Constantine Fahlberg on his discovery of saccharin, *Scientific American* (1886)

7 Backtracking

In this lecture, I want to describe another recursive algorithm strategy called **backtracking**. A backtracking algorithm tries to build a solution to a computational problem incrementally. Whenever the algorithm needs to decide between multiple alternatives to the next component of the solution, it simply tries all possible options recursively.

7.1 n Queens

The prototypical backtracking problem is the classical **n Queens Problem**, first proposed by German chess enthusiast Max Bezzel in 1848 (under his pseudonym “Schachfreund”) for the standard 8×8 board and by François-Joseph Eustache Lionnet in 1869 for the more general $n \times n$ board. The problem is to place n queens on an $n \times n$ chessboard, so that no two queens can attack each other. For readers not familiar with the rules of chess, this means that no two queens are in the same row, column, or diagonal.

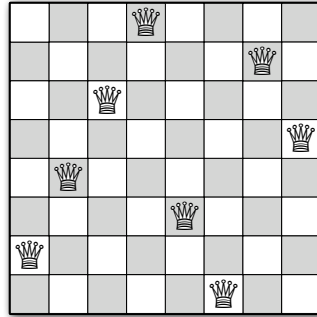
Obviously, in any solution to the n -Queens problem, there is exactly one queen in each row. So we will represent our possible solutions using an array $Q[1..n]$, where $Q[i]$ indicates which square in row i contains a queen, or 0 if no queen has yet been placed in row i . To find a solution, we put queens on the board row by row, starting at the top. A *partial* solution is an array $Q[1..n]$ whose first $r - 1$ entries are positive and whose last $n - r + 1$ entries are all zeros, for some integer r .

The following recursive algorithm, essentially due to Gauss (who called it “methodical groping”), recursively enumerates all complete n -queens solutions that are consistent with a given partial solution. The input parameter r is the first empty row. Thus, to compute all n -queens solutions with no restrictions, we would call $\text{RECURSIVENQUEENS}(Q[1..n], 1)$.

```

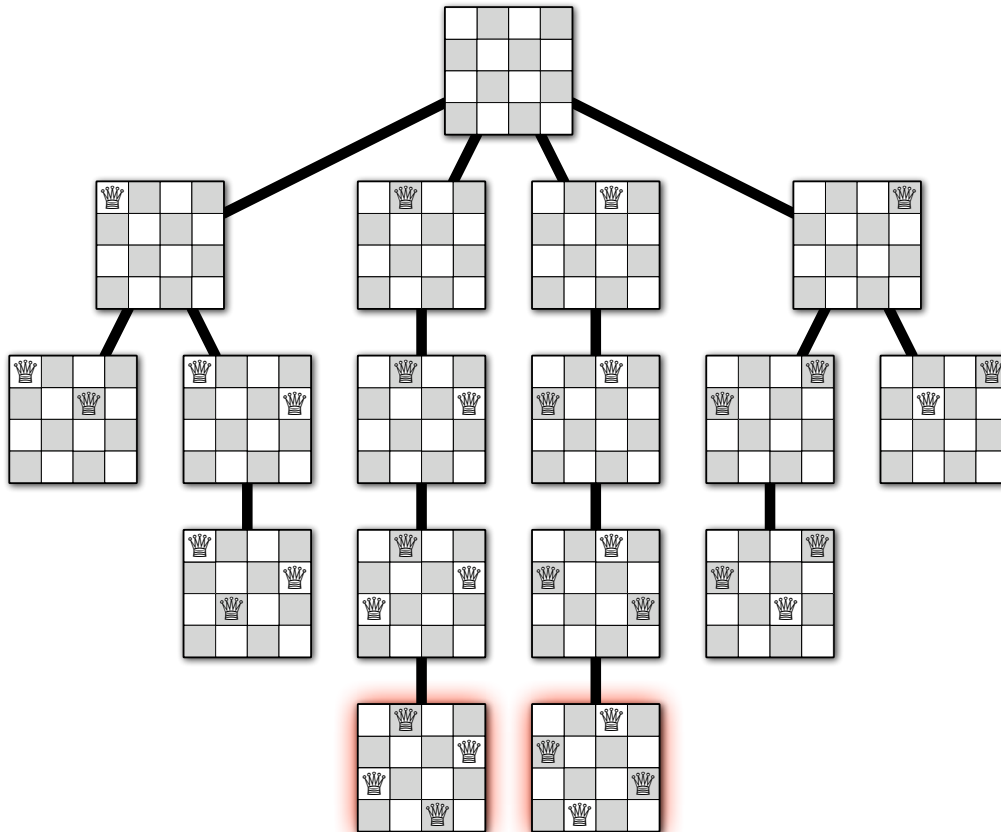
RECURSIVENQUEENS(Q[1..n], r):
  if r = n + 1
    print Q
  else
    for j ← 1 to n
      legal ← TRUE
      for i ← 1 to r - 1
        if (Q[i] = j) or (Q[i] = j + r - i) or (Q[i] = j - r + i)
          legal ← FALSE
      if legal
        Q[r] ← j
        RECURSIVENQUEENS(Q[1..n], r + 1)

```



One solution to the 8 queens problem, represented by the array [4,7,3,8,2,5,1,6]

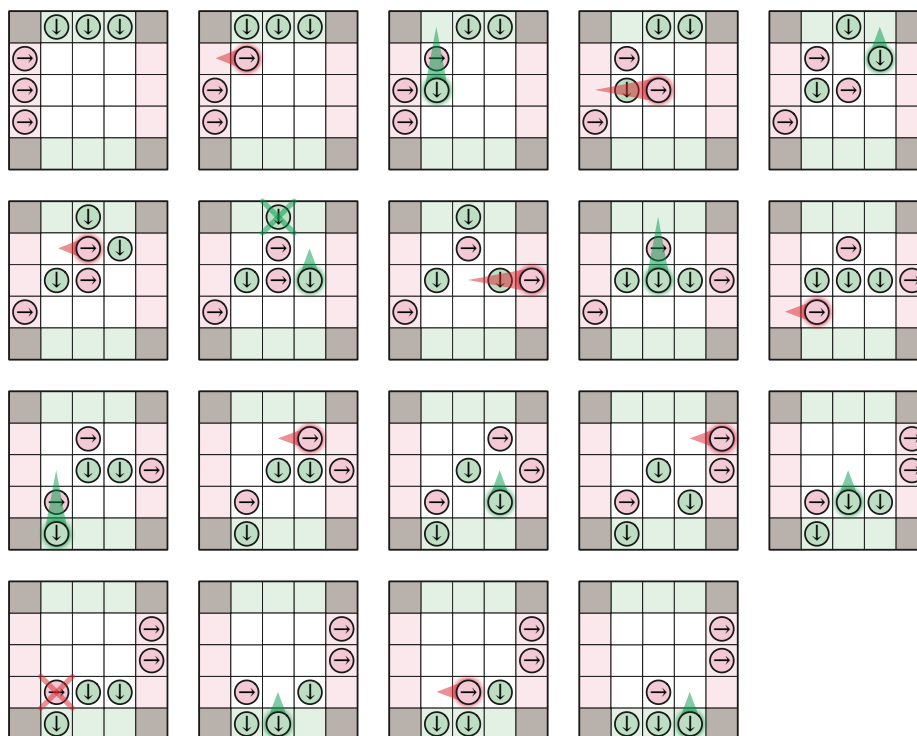
Like most recursive algorithms, the execution of a backtracking algorithm can be illustrated using a *recursion tree*. The root of the recursion tree corresponds to the original invocation of the algorithm; edges in the tree correspond to recursive calls. A path from the root down to any node shows the history of a partial solution to the n -Queens problem, as queens are added to successive rows. The leaves correspond to partial solutions that cannot be extended, either because there is already a queen on every row, or because every position in the next empty row is in the same row, column, or diagonal as an existing queen. The backtracking algorithm simply performs a depth-first traversal of this tree.



The complete recursion tree for our algorithm for the 4 queens problem.

7.2 Game Trees

Consider the following simple two-player game played on an $n \times n$ square grid with a border of squares; let's call the players Horatio Fahlberg-Remsen and Vera Rebaudi.¹ Each player has n tokens that they move across the board from one side to the other. Horatio's tokens start in the left border, one in each row, and move to the right; symmetrically, Vera's tokens start in the top border, one in each column, and move down. The players alternate turns. In each of his turns, Horatio either *moves* one of his tokens one step to the right into an empty square, or *jumps* one of his tokens over exactly one of Vera's tokens into an empty square two steps to the right. However, if no legal moves or jumps are available, Horatio simply passes. Similarly, Vera either moves or jumps one of her tokens downward in each of her turns, unless no moves or jumps are possible. The first player to move all their tokens off the edge of the board wins.



Vera wins the 3×3 game.

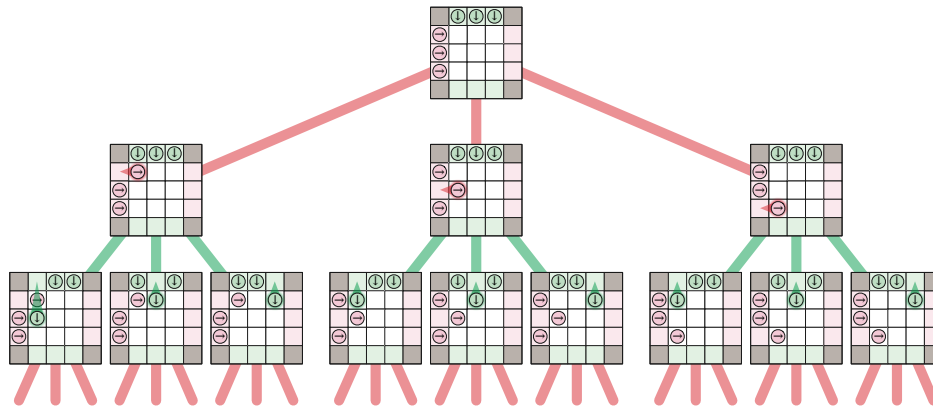
We can use a simple backtracking algorithm to determine the best move for each player at each turn. The *state* of the game consists of the locations of all the pieces and the player whose turn it is. We recursively define a game state to be *good* or *bad* as follows:

- A game state is *bad* if all the opposing player's tokens have reached their goals.
- A game state is *good* if the current player can move to a state that is bad for the opposing player.
- A configuration is *bad* if every move leads to a state that is good for the opposing player.

¹I don't know what this game is called, or even if I'm remembering the rules correctly. I learned it (or something like it) from Lenny Pitt, who recommended playing it with sweetener packets at restaurants.

Constantin Fahlberg and Ira Remsen synthesized saccharin for the first time in 1878, while Fahlberg was a postdoc in Remsen's lab investigating coal tar derivatives. In 1900, Ovidio Rebaudi published the first chemical analysis of *ka'a he'ê*, a medicinal plant cultivated by the Guaraní for more than 1500 years, now more commonly known as *Stevia rebaudiana*.

This recursive definition immediately suggests a recursive backtracking algorithm to determine whether a given state of the game is good or bad. Moreover, for any good state, the backtracking algorithm finds a move leading to a bad state for the opposing player. Thus, by induction, any player that finds the game in a good state on their turn can win the game, even if their opponent plays perfectly; on the other hand, starting from a bad state, a player can win only if their opponent makes a mistake.



The first two levels of the game tree.

All computer game players are ultimately based on this simple backtracking strategy. However, since most games have an enormous number of states, it is not possible to traverse the entire game tree in practice. Instead, game programs employ other heuristics² to *prune* the game tree, by ignoring states that are obviously good or bad (or at least obviously better or worse than other states), and/or by cutting off the tree at a certain depth (or *ply*) and using a more efficient heuristic to evaluate the leaves.

7.3 Subset Sum

Let's consider a more complicated problem, called SUBSETSUM: Given a set X of positive integers and *target* integer T , is there a subset of elements in X that add up to T ? Notice that there can be more than one such subset. For example, if $X = \{8, 6, 7, 5, 3, 10, 9\}$ and $T = 15$, the answer is TRUE, thanks to the subsets $\{8, 7\}$ or $\{7, 5, 3\}$ or $\{6, 9\}$ or $\{5, 10\}$. On the other hand, if $X = \{11, 6, 5, 1, 7, 13, 12\}$ and $T = 15$, the answer is FALSE.

There are two trivial cases. If the target value T is zero, then we can immediately return TRUE, because empty set is a subset of *every* set X , and the elements of the empty set add up to zero.³ On the other hand, if $T < 0$, or if $T \neq 0$ but the set X is empty, then we can immediately return FALSE.

For the general case, consider an arbitrary element $x \in X$. (We've already handled the case where X is empty.) There is a subset of X that sums to T if and only if one of the following statements is true:

- There is a subset of X that *includes* x and whose sum is T .
- There is a subset of X that *excludes* x and whose sum is T .

In the first case, there must be a subset of $X \setminus \{x\}$ that sums to $T - x$; in the second case, there must be a subset of $X \setminus \{x\}$ that sums to T . So we can solve SUBSETSUM(X, T) by reducing it to two simpler instances: SUBSETSUM($X \setminus \{x\}, T - x$) and SUBSETSUM($X \setminus \{x\}, T$). Here's how the resulting recursive algorithm might look if X is stored in an array.

²A heuristic is an algorithm that doesn't work.

³There's no base case like the vacuous base case!

```

SUBSETSUM( $X[1..n], T$ ):
  if  $T = 0$ 
    return TRUE
  else if  $T < 0$  or  $n = 0$ 
    return FALSE
  else
    return (SUBSETSUM( $X[1..n-1], T$ )  $\vee$  SUBSETSUM( $X[1..n-1], T - X[n]$ ))

```

Proving this algorithm correct is a straightforward exercise in induction. If $T = 0$, then the elements of the empty subset sum to T , so TRUE is the correct output. Otherwise, if T is negative or the set X is empty, then no subset of X sums to T , so FALSE is the correct output. Otherwise, if there is a subset that sums to T , then either it contains $X[n]$ or it doesn't, and the Recursion Fairy correctly checks for each of those possibilities. Done.

The running time $T(n)$ clearly satisfies the recurrence $T(n) \leq 2T(n-1) + O(1)$, which we can solve using either recursion trees or annihilators (or just guessing) to obtain the upper bound $T(n) = O(2^n)$. In the worst case, the recursion tree for this algorithm is a complete binary tree with depth n .

Here is a similar recursive algorithm that actually *constructs* a subset of X that sums to T , if one exists. This algorithm also runs in $O(2^n)$ time.

```

CONSTRUCTSUBSET( $X[1..n], T$ ):
  if  $T = 0$ 
    return  $\emptyset$ 
  if  $T < 0$  or  $n = 0$ 
    return NONE
   $Y \leftarrow$  CONSTRUCTSUBSET( $X[1..n-1], T$ )
  if  $Y \neq$  NONE
    return  $Y$ 
   $Y \leftarrow$  CONSTRUCTSUBSET( $X[1..n-1], T - X[n]$ )
  if  $Y \neq$  NONE
    return  $Y \cup \{X[n]\}$ 
  return NONE

```

7.4 The General Pattern

Find a small choice whose correct answer would reduce the problem size. For each possible answer, temporarily adopt that choice and recurse. (Don't try to be clever about which choices to try; just try them all.) The recursive subproblem is often more general than the original target problem; in each recursive subproblem, we must consider *only* solutions that are consistent with the choices we have already made.

7.5 NFA acceptance

Recall that a nondeterministic finite-state automaton, or NFA, can be described as a directed graph, whose edges are called *states* and whose edges have *labels* drawn from a finite set Σ called the *alphabet*. Every NFA has a designated *start* state and a subset of *accepting* states. Any walk in this graph has a label, which is a string formed by concatenating the labels of the edges in the walk. A string w is *accepted* by an NFA if and only if there is a walk from the start state to one of the accepting states whose label is w .

More formally (or at least, more symbolically), an NFA consists of a finite set Q of states, a start state $s \in Q$, a set of accepting states $A \subseteq Q$, and a transition function $\delta : Q \times \Sigma \rightarrow 2^Q$. We recursively extend

the transition function to strings by defining

$$\delta^*(q, w) = \begin{cases} \{q\} & \text{if } w = \varepsilon, \\ \bigcup_{r \in \delta(q, a)} \delta^*(r, x) & \text{if } w = ax. \end{cases}$$

The NFA accepts string w if and only if the set $\delta^*(s, w)$ contains at least one accepting state.

We can express this acceptance criterion more directly as follows. We define a boolean function $Accepts?(q, w)$, which is TRUE if the NFA would accept string w if we started in state q , and FALSE otherwise. This function has the following recursive definition:

$$Accepts?(q, w) := \begin{cases} \text{TRUE} & \text{if } w = \varepsilon \text{ and } q \in A \\ \text{FALSE} & \text{if } w = \varepsilon \text{ and } q \notin A \\ \bigvee_{r \in \delta(q, a)} Accepts?(r, x) & \text{if } w = ax \end{cases}$$

The NFA accepts w if and only if $Accepts?(s, w) = \text{TRUE}$.

In the magical world of non-determinism, we can imagine that the NFA always magically makes the right decision when faces with multiple transitions, or perhaps spawns off an independent parallel thread for each possible choice. Alas, real computers are neither clairvoyant nor (despite the increasing use of multiple cores) infinitely parallel. To simulate the NFA's behavior directly, we must recursively explore the consequences of each choice explicitly.

The recursive definition of $Accepts?$ translates directly into the following recursive backtracking algorithm. Here, the transition function δ and the accepting states A are represented as global boolean arrays, where $\delta[q, a, r] = \text{TRUE}$ if and only if $r \in \delta(q, a)$, and $A[q] = \text{TRUE}$ if and only if $q \in A$.

```

ACCEPTS?(q, w[1..n]):
  if n = 0
    return A[q]
  for all states r
    if δ[q, w[1], r] and ACCEPTS?(r, w[2..n])
      return TRUE
  return FALSE

```

To determine whether the NFA accepts a string w , we call $ACCEPTS?(\delta, A, s, w)$.

The running time of this algorithm satisfies the recursive inequality $T(n) \leq O(|Q|) \cdot T(n-1)$, which immediately implies that $T(n) = O(|Q|^n)$.

7.6 Longest Increasing Subsequence

Now suppose we are given a sequence of integers, and we want to find the longest subsequence whose elements are in increasing order. More concretely, the input is an array $A[1..n]$ of integers, and we want to find the longest sequence of indices $1 \leq i_1 < i_2 < \dots < i_k \leq n$ such that $A[i_j] < A[i_{j+1}]$ for all j .

To derive a recursive algorithm for this problem, we start with a recursive definition of the kinds of objects we're playing with: sequences and subsequences.

A sequence of integers is either empty
or an integer followed by a sequence of integers.

This definition suggests the following strategy for devising a recursive algorithm. If the input sequence is empty, there's nothing to do. Otherwise, we only need to figure out what to do with the first element of the input sequence; the Recursion Fairy will take care of everything else. We can formalize this strategy somewhat by giving a recursive definition of subsequence (using array notation to represent sequences):

The only *subsequence* of the empty sequence is the empty sequence.
 A *subsequence* of $A[1..n]$ is either a subsequence of $A[2..n]$
 or $A[1]$ followed by a subsequence of $A[2..n]$.

We're not just looking for just *any* subsequence, but a *longest* subsequence with the property that elements are in *increasing* order. So let's try to add those two conditions to our definition. (I'll omit the familiar vacuous base case.)

The *LIS* of $A[1..n]$ is
 either the LIS of $A[2..n]$
 or $A[1]$ followed by the LIS of $A[2..n]$ with elements larger than $A[1]$,
 whichever is longer.

This definition is correct, but it's not quite recursive—we're defining the object 'longest increasing subsequence' in terms of the slightly *different* object 'longest increasing subsequence with elements larger than x ', which we haven't properly defined yet. Fortunately, this second object has a very similar recursive definition. (Again, I'm omitting the vacuous base case.)

If $A[1] \leq x$, the LIS of $A[1..n]$ with elements larger than x is
 the LIS of $A[2..n]$ with elements larger than x .
 Otherwise, the LIS of $A[1..n]$ with elements larger than x is
 either the LIS of $A[2..n]$ with elements larger than x
 or $A[1]$ followed by the LIS of $A[2..n]$ with elements larger than $A[1]$,
 whichever is longer.

The longest increasing subsequence without restrictions can now be redefined as the longest increasing subsequence with elements larger than $-\infty$. Rewriting this recursive definition into pseudocode gives us the following recursive algorithm.

$LIS(A[1..n])$:
 return LISBIGGER($-\infty, A[1..n]$)

$LISBIGGER(prev, A[1..n])$:
 if $n = 0$
 return 0
 else
 $max \leftarrow LISBIGGER(prev, A[2..n])$
 if $A[1] > prev$
 $L \leftarrow 1 + LISBIGGER(A[1], A[2..n])$
 if $L > max$
 $max \leftarrow L$
 return max

The running time of this algorithm satisfies the recurrence $T(n) \leq 2T(n-1) + O(1)$, which as usual implies that $T(n) = O(2^n)$. We really shouldn't be surprised by this running time; in the worst case, the algorithm examines each of the 2^n subsequences of the input array.

The following alternative strategy avoids defining a new object with the “larger than x ” constraint. We still only have to decide whether to include or exclude the first element $A[1]$. We consider the case where $A[1]$ is excluded exactly the same way, but to consider the case where $A[1]$ is included, we remove any elements of $A[2..n]$ that are larger than $A[1]$ *before* we recurse. This new strategy gives us the following algorithm:

```

FILTER( $A[1..n]$ ,  $x$ ):
   $j \leftarrow 1$ 
  for  $i \leftarrow 1$  to  $n$ 
    if  $A[i] > x$ 
       $B[j] \leftarrow A[i]$ ;  $j \leftarrow j + 1$ 
  return  $B[1..j]$ 

```

```

LIS( $A[1..n]$ ):
  if  $n = 0$ 
    return 0
  else
     $max \leftarrow$  LIS( $prev, A[2..n]$ )
     $L \leftarrow 1 +$  LIS( $A[1],$  FILTER( $A[2..n], A[1]$ ))
    if  $L > max$ 
       $max \leftarrow L$ 
  return  $max$ 

```

The FILTER subroutine clearly runs in $O(n)$ time, so the running time of LIS satisfies the recurrence $T(n) \leq 2T(n-1) + O(n)$, which solves to $T(n) \leq O(2^n)$ by the annihilator method. This upper bound pessimistically assumes that FILTER never actually removes any elements; indeed, if the input sequence is sorted in increasing order, this assumption is correct.

7.7 Optimal Binary Search Trees

Retire this example? It's not a *bad* example, exactly—certainly it's infinitely better than the execrable matrix-chain multiplication problem from Aho, Hopcroft, and Ullman—but it's not the best *first* example of tree-like backtracking. Minimum-ink triangulation of convex polygons is both more intuitive (geometry FTW!) and structurally equivalent. CFG parsing and regular expression matching (really just a special case of parsing) have similar recursive structure, but are a bit more complicated.

Our next example combines recursive backtracking with the divide-and-conquer strategy. Recall that the running time for a successful search in a binary search tree is proportional to the number of ancestors of the target node.⁴ As a result, the worst-case search time is proportional to the depth of the tree. Thus, to minimize the worst-case search time, the height of the tree should be as small as possible; by this metric, the ideal tree is perfectly balanced.

In many applications of binary search trees, however, it is more important to minimize the total cost of several searches rather than the worst-case cost of a single search. If x is a more ‘popular’ search target than y , we can save time by building a tree where the depth of x is smaller than the depth of y , even if that means increasing the overall depth of the tree. A perfectly balanced tree is *not* the best choice if some items are significantly more popular than others. In fact, a totally unbalanced tree of depth $\Omega(n)$ might actually be the best choice!

This situation suggests the following problem. Suppose we are given a sorted array of **keys** $A[1..n]$ and an array of corresponding **access frequencies** $f[1..n]$. Our task is to build the binary search tree that minimizes the *total* search time, assuming that there will be exactly $f[i]$ searches for each key $A[i]$.

Before we think about how to solve this problem, we should first come up with a good recursive definition of the function we are trying to optimize! Suppose we are also given a binary search tree T with n nodes. Let v_i denote the node that stores $A[i]$, and let r be the index of the root node. Ignoring

⁴An *ancestor* of a node v is either the node itself or an ancestor of the parent of v . A *proper* ancestor of v is either the parent of v or a proper ancestor of the parent of v .

constant factors, the cost of searching for $A[i]$ is the number of nodes on the path from the root v_r to v_i . Thus, the total cost of performing all the binary searches is given by the following expression:

$$\text{Cost}(T, f[1..n]) = \sum_{i=1}^n f[i] \cdot \#\text{nodes between } v_r \text{ and } v_i$$

Every search path includes the root node v_r . If $i < r$, then all other nodes on the search path to v_i are in the left subtree; similarly, if $i > r$, all other nodes on the search path to v_i are in the right subtree. Thus, we can partition the cost function into three parts as follows:

$$\begin{aligned} \text{Cost}(T, f[1..n]) &= \sum_{i=1}^{r-1} f[i] \cdot \#\text{nodes between } \text{left}(v_r) \text{ and } v_i \\ &\quad + \sum_{i=1}^n f[i] \\ &\quad + \sum_{i=r+1}^n f[i] \cdot \#\text{nodes between } \text{right}(v_r) \text{ and } v_i \end{aligned}$$

Now the first and third summations look exactly like our original expression (*) for $\text{Cost}(T, f[1..n])$. Simple substitution gives us our recursive definition for Cost :

$$\text{Cost}(T, f[1..n]) = \text{Cost}(\text{left}(T), f[1..r-1]) + \sum_{i=1}^n f[i] + \text{Cost}(\text{right}(T), f[r+1..n])$$

The base case for this recurrence is, as usual, $n = 0$; the cost of performing no searches in the empty tree is zero.

Now our task is to compute the tree T_{opt} that minimizes this cost function. Suppose we somehow magically knew that the root of T_{opt} is v_r . Then the recursive definition of $\text{Cost}(T, f)$ immediately implies that the left subtree $\text{left}(T_{\text{opt}})$ must be the optimal search tree for the keys $A[1..r-1]$ and access frequencies $f[1..r-1]$. Similarly, the right subtree $\text{right}(T_{\text{opt}})$ must be the optimal search tree for the keys $A[r+1..n]$ and access frequencies $f[r+1..n]$. **Once we choose the correct key to store at the root, the Recursion Fairy automatically constructs the rest of the optimal tree.** More formally, let $\text{OptCost}(f[1..n])$ denote the total cost of the optimal search tree for the given frequency counts. We immediately have the following recursive definition.

$$\text{OptCost}(f[1..n]) = \min_{1 \leq r \leq n} \left\{ \text{OptCost}(f[1..r-1]) + \sum_{i=1}^n f[i] + \text{OptCost}(f[r+1..n]) \right\}$$

Again, the base case is $\text{OptCost}(f[1..0]) = 0$; the best way to organize no keys, which we will plan to search zero times, is by storing them in the empty tree!

This recursive definition can be translated mechanically into a recursive algorithm, whose running time $T(n)$ satisfies the recurrence

$$T(n) = \Theta(n) + \sum_{k=1}^n (T(k-1) + T(n-k)).$$

The $\Theta(n)$ term comes from computing the total number of searches $\sum_{i=1}^n f[i]$.

Yeah, that's one ugly recurrence, but it's actually easier to solve than it looks. To transform it into a more familiar form, we regroup and collect identical terms, subtract the recurrence for $T(n-1)$ to get rid of the summation, and then regroup again.

$$\begin{aligned} T(n) &= \Theta(n) + 2 \sum_{k=0}^{n-1} T(k) \\ T(n-1) &= \Theta(n-1) + 2 \sum_{k=0}^{n-2} T(k) \\ T(n) - T(n-1) &= \Theta(1) + 2T(n-1) \\ T(n) &= 3T(n-1) + \Theta(1) \end{aligned}$$

The solution $T(n) = \Theta(3^n)$ now follows from the annihilator method.

Let me emphasize that this recursive algorithm does *not* examine all possible binary search trees. The number of binary search trees with n nodes satisfies the recurrence

$$N(n) = \sum_{r=1}^{n-1} (N(r-1) \cdot N(n-r)),$$

which has the closed-form solution $N(n) = \Theta(4^n / \sqrt{n})$. Our algorithm saves considerable time by searching *independently* for the optimal left and right subtrees. A full enumeration of binary search trees would consider all possible *pairings* of left and right subtrees; hence the product in the recurrence for $N(n)$.

7.8 CFG Parsing

Our final example is the *parsing* problem for context-free languages. Given a string w and a context-free grammar G , does w belong to the language generated by G ? Recall that a context-free grammar over the alphabet Σ consists of a finite set Γ of *non-terminals* (disjoint from Σ) and a finite set of *production rules* of the form $A \rightarrow w$, where A is a nonterminal and w is a string over $\Sigma \cup \Gamma$.

Real-world applications of parsing normally require more information than just a single bit. For example, compilers require parsers that output a parse tree of the input code; some natural language applications require the *number* of distinct parse trees for a given string; others assign probabilities to the production rules and then ask for the *most likely* parse tree for a given string. However, these more general problems can be solved using relatively straightforward generalizations of the following decision algorithm.

★★★★ Backtracking recurrence behind CYK

Exercises

- Let $A[1..m]$ and $B[1..n]$ be two arbitrary arrays. A *common subsequence* of A and B is both a subsequence of A and a subsequence of B . Give a simple recursive definition for the function $lcs(A, B)$, which gives the length of the *longest* common subsequence of A and B .
 - Let $A[1..m]$ and $B[1..n]$ be two arbitrary arrays. A *common supersequence* of A and B is another sequence that contains both A and B as subsequences. Give a simple recursive definition for the function $scs(A, B)$, which gives the length of the *shortest* common supersequence of A and B .

- (c) Call a sequence $X[1..n]$ *oscillating* if $X[i] < X[i + 1]$ for all even i , and $X[i] > X[i + 1]$ for all odd i . Give a simple recursive definition for the function $los(A)$, which gives the length of the longest oscillating subsequence of an arbitrary array A of integers.
- (d) Give a simple recursive definition for the function $sos(A)$, which gives the length of the shortest oscillating supersequence of an arbitrary array A of integers.
- (e) Call a sequence $X[1..n]$ *accelerating* if $2 \cdot X[i] < X[i - 1] + X[i + 1]$ for all i . Give a simple recursive definition for the function $lxs(A)$, which gives the length of the longest accelerating subsequence of an arbitrary array A of integers.

For more backtracking exercises, see the next two lecture notes!