for about $\frac{1}{6}n^3$ values of k. This means we can determine an optimum tree in $O(n^3)$ units of time, using $O(n^2)$ cells of memory.

A factor of n can actually be removed from the running time if we make use of a monotonicity property. Let $r(i,j)$ denote an element of $R(i,j)$; we need not compute the entire set $R(i,j)$, a single representative is sufficient. Once we have found $r(i,j-1)$ and $r(i+1,j)$, the result of exercise 27 proves that we may always assume that

$$r(i,j-1) \le r(i,j) \le r(i+1,j) \tag{17}$$

when the weights are nonnegative. This limits the search for the minimum, since only $r(i+1,j) - r(i,j-1) + 1$ values of k need to be examined in (16) instead of $j-i$. The total amount of work when $j-i = d$ is now bounded by the telescoping series

$$\sum_{\substack{d \le j \le n \\ i=j-d}} \big(r(i+1,j) - r(i,j-1) + 1\big) = r(n-d+1,n) - r(0,d-1) + n - d + 1 < 2n;$$

hence the total running time is reduced to $O(n^2)$.

The following algorithm describes this procedure in detail.

Algorithm K (*Find optimum binary search trees*). Given $2n+1$ nonnegative weights $(p_1,\dots,p_n; q_0,\dots,q_n)$, this algorithm constructs binary trees $t(i,j)$ that have minimum cost for the weights $(p_{i+1},\dots,p_j; q_i,\dots,q_j)$ in the sense defined above. Three arrays are computed, namely

$$c[i,j], \quad \text{for } 0 \le i \le j \le n, \quad \text{the cost of } t(i,j);$$
$$r[i,j], \quad \text{for } 0 \le i < j \le n, \quad \text{the root of } t(i,j);$$
$$w[i,j], \quad \text{for } 0 \le i \le j \le n, \quad \text{the total weight of } t(i,j).$$

The results of the algorithm are specified by the r array: If $i = j$, $t(i,j)$ is null; otherwise its left subtree is $t(i, r[i,j]-1)$ and its right subtree is $t(r[i,j], j)$.

K1. [Initialize.] For $0 \le i \le n$, set $c[i,i] \leftarrow 0$ and $w[i,i] \leftarrow q_i$ and $w[i,j] \leftarrow w[i,j-1] + p_j + q_j$ for $j = i+1,\dots,n$. Then for $1 \le j \le n$ set $c[j-1,j] \leftarrow w[j-1,j]$ and $r[j-1,j] \leftarrow j$. (This determines all the 1-node optimum trees.)

K2. [Loop on d.] Do step K3 for $d = 2,3,\dots,n$, then terminate the algorithm.

K3. [Loop on j.] (We have already determined the optimum trees of fewer than d nodes. This step determines all the d-node optimum trees.) Do step K4 for $j = d, d+1, \dots, n$.

**K4.** [Find $c[i,j]$, $r[i,j]$.] Set $i \leftarrow j - d$. Then set

$$c[i,j] \leftarrow w[i,j] + \min_{r[i,j-1] \le k \le r[i+1,j]}\big(c[i,k-1] + c[k,j]\big),$$

and set $r[i,j]$ to a value of k for which the minimum occurs. (Exercise 22 proves that $r[i,j-1] \le r[i+1,j]$.) ∎

As an example of Algorithm K, consider Fig. 15, which is based on a "keyword-in-context" (KWIC) indexing application. The titles of all articles in the

(Aquí se han eliminado páginas...)

19. [*M23*] What is the approximate average number of comparisons, (11), when the input probabilities satisfy the 80-20 law defined in Eq. 6.1–(11)?

20. [*M20*] Suppose we have inserted keys into a tree in order of decreasing frequency $p_1 \geq p_2 \geq \cdots \geq p_n$. Can this tree be substantially worse than the optimum search tree?

21. [*M20*] If p, q, r are probabilities chosen at random, subject to the condition that $p + q + r = 1$, what are the probabilities that trees I, II, III, IV, V of (13) are optimal, respectively? (Consider the relative areas of the regions in Fig. 14.)

22. [*M20*] Prove that $r[i, j-1]$ is never greater than $r[i+1, j]$ when step K4 of Algorithm K is performed.

▶ **23.** [*M23*] Find an optimum binary search tree for the case $N = 40$, with weights $p_1 = 9$, $p_2 = p_3 = \cdots = p_{40} = 1$, $q_0 = q_1 = \cdots = q_{40} = 0$. (Don't use a computer.)

24. [*M25*] Given that $p_n = q_n = 0$ and that the other weights are nonnegative, prove that an optimum tree for $(p_1, \ldots, p_n; q_0, \ldots, q_n)$ may be obtained by replacing
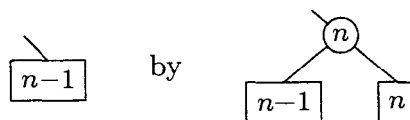
in any optimum tree for $(p_1, \ldots, p_{n-1}; q_0, \ldots, q_{n-1})$.

25. [*M20*] Let A and B be nonempty sets of real numbers, and define $A \leq B$ if the following property holds:

$$(a \in A, \ b \in B, \text{ and } b < a) \qquad \text{implies} \qquad (a \in B \text{ and } b \in A).$$

a) Prove that this relation is transitive on nonempty sets.

b) Prove or disprove: $A \leq B$ if and only if $A \leq A \cup B \leq B$.

26. [*M22*] Let $(p_1, \ldots, p_n; q_0, \ldots, q_n)$ be nonnegative weights, where $p_n + q_n = x$. Prove that as x varies from 0 to ∞, while $(p_1, \ldots, p_{n-1}; q_0, \ldots, q_{n-1})$ are held constant, the cost $c(0, n)$ of an optimum binary search tree is a concave, continuous, piecewise linear function of x with integer slopes. In other words, prove that there exist positive integers $l_0 > l_1 > \cdots > l_m$ and real constants $0 = x_0 < x_1 < \cdots < x_m < x_{m+1} = \infty$ and $y_0 < y_1 \cdots < y_m$ such that $c(0, n) = y_h + l_h x$ when $x_h \leq x \leq x_{h+1}$, for $0 \leq h \leq m$.

27. [*M33*] The object of this exercise is to prove that the sets of roots $R(i, j)$ of optimum binary search trees satisfy

$$R(i, j-1) \leq R(i, j) \leq R(i+1, j), \qquad \text{for } j - i \geq 2,$$

in terms of the relation defined in exercise 25, when the weights $(p_1, \ldots, p_n; q_0, \ldots, q_n)$ are nonnegative. The proof is by induction on $j-i$; our task is to prove that $R(0, n-1) \leq R(0, n)$, assuming that $n \geq 2$ and that the stated relation holds for $j - i < n$. [By left-right symmetry it follows that $R(0, n) \leq R(1, n)$.]

a) Prove that $R(0, n-1) \leq R(0, n)$ if $p_n = q_n = 0$. (See exercise 24.)

b) Let $p_n + q_n = x$. In the notation of exercise 26, let R_h be the set $R(0, n)$ of optimum roots when $x_h < x < x_{h+1}$, and let R'_h be the set of optimum roots when $x = x_h$. Prove that

$$R'_0 \leq R_0 \leq R'_1 \leq R_1 \leq \cdots \leq R'_m \leq R_m.$$

Hence by part (a) and exercise 25 we have $R(0, n-1) \le R(0, n)$ for all x. [*Hint:* Consider the case $x = x_h$, and assume that both the trees
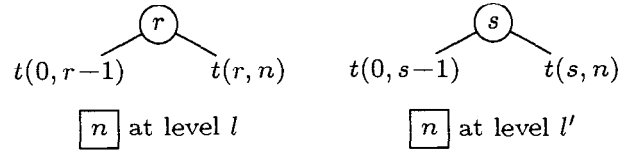
are optimum, with $s < r$ and $l \ge l'$. Use the induction hypothesis to prove that there is an optimum tree with root (r) such that $\boxed{n}$ is at level l', and an optimum tree with root (s) such that $\boxed{n}$ is at level l.]

28. [*24*] Use some macro language to define a "optimum binary search" macro, whose parameter is a nested specification of an optimum binary tree.

29. [*40*] What is the *worst* possible binary search tree for the 31 most common English words, using the frequency data of Fig. 12?

30. [*M34*] Prove that the costs of optimum binary search trees satisfy the "quadrangle inequality" $c(i, j) - c(i, j-1) \ge c(i+1, j) - c(i+1, j-1)$ when $j \ge i + 2$.

31. [*M35*] (K. C. Tan.) Prove that, among all possible sets of probabilities $(p_1, \ldots, p_n; q_0, \ldots, q_n)$ with $p_1 + \cdots + p_n + q_0 + \cdots + q_n = 1$, the most expensive minimum-cost tree occurs when $p_i = 0$ for all i, $q_j = 0$ for all even j, and $q_j = 1/\lceil n/2 \rceil$ for all odd j.

▸ **32.** [*M25*] Let $n + 1 = 2^m + k$, where $0 \le k \le 2^m$. There are exactly $\binom{2^m}{k}$ binary trees in which all external nodes appear on levels m and $m + 1$. Show that, among all these trees, we obtain one with the minimum cost for the weights $(p_1, \ldots, p_n; q_0, \ldots, q_n)$ if we apply Algorithm K to the weights $(p_1, \ldots, p_n; M+q_0, \ldots, M+q_n)$ for sufficiently large M.

33. [*M41*] In order to find the binary search tree that minimizes the running time of Program T, we should minimize the quantity $7C + C1$ instead of simply minimizing the number of comparisons C. Develop an algorithm that finds optimum binary search trees when different costs are associated with left and right branches in the tree. (Incidentally, when the right cost is twice the left cost, and the node frequencies are all equal, the Fibonacci trees turn out to be optimum; see L. E. Stanfel, *JACM* **17** (1970), 508–517. On machines that cannot make three-way comparisons at once, a program for Algorithm T will have to make two comparisons in step T2, one for equality and one for less-than; B. Sheil and V. R. Pratt have observed that these comparisons need not involve the same key, and it may well be best to have a binary tree whose internal nodes specify either an equality test *or* a less-than test but not both. This situation would be interesting to explore as an alternative to the stated problem.)

34. [*HM21*] Show that the asymptotic value of the multinomial coefficient

$$\binom{N}{p_1 N, \ p_2 N, \ \ldots, \ p_n N}$$

as $N \to \infty$ is related to the entropy $H(p_1, p_2, \ldots, p_n)$.

35. [*HM22*] Complete the proof of Theorem B by establishing the inequality (24).

▸ **36.** [*HM25*] (Claude Shannon.) Let X and Y be random variables with finite ranges $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$, and let $p_i = \Pr(X = x_i)$, $q_j = \Pr(Y = y_j)$, $r_{ij} = \Pr(X = x_i$ and $Y = y_j)$. Let $H(X) = H(p_1, \ldots, p_m)$ and $H(Y) = H(q_1, \ldots, q_n)$ be the