The Theory of Computation

Bernard M.E. Moret

Addison-Wesley-Longman, 1998

Solution Manual
Chapter 2, Problem 2.30

The prefix sum at any given point is simply the difference between the number of left (opening) parentheses and the number of right (closing) parentheses. That number must be nonnegative everywhere because every closing parenthesis must match an opening one that preceded it. That number must be 0 at the last parenthesis, because by then we must have seen exactly the same number of each. Thus it is clear that a string of properly nested parentheses will yield the proper prefix sum—the conditions are necessary.

We need to prove that the conditions are sufficient—that any string of parentheses with a prefix sum that has these two properties must be properly nested. It is simplest to proceed by induction.

Clearly, there is only one string of length 2 that is nested, namely (); its prefix sum is as indicated. Assume then that all strings of length $2n$ that obey the properties of the prefix sum are properly nested and consider a string of length $2(n+1)$ that also obeys these properties. (There is of course no point in considering strings of odd length: no such string can ever have a last value of 0—nor could any such string be balanced.) The first entry is nonnegative, so the first character must be an opening parenthesis; the last entry is 0 and the one before non-negative, so the last character must be a closing parenthesis; however, we cannot just remove the first and last character and proceed inductively, because the first and last parenthesis need not form a matched pair—we could have the string ()(), for instance. So we scan the prefix sum for any appearance of a value of 0 (we know there is at least one, at the last entry). The first such value (scanning from the first entry) is the one we want because it is the closing parenthesis that matches the first symbol: that it is a closing parenthesis follows from its value and the fact that the previous entry cannot be negative, while the fact that it matches the first symbol is simply due to the fact that it is the first place where the number of opening parentheses exactly matches the number of closing ones. Thus we can in fact treat any substring ending in 0 independently of the others. We remove this entry and the first entry from the string and decrease all intervening values by 1 (to reflect that fact that we removed an opening parenthesis). Since all intervening values are larger than 0, the resulting string of prefix sum values still obeys the first condition. What about the second condition—after all, we just removed the last entry with its 0 value! But note that the entry preceding the 0 must have been a 1 (no other choice is possible), so that it is now, after the decrease by 1, a 0, as needed to obey the second condition. Thus the inductive hypothesis applies, the string between the opening and closing pair must be balanced and the opening and closing pair obviously maintains that property, so that the whole string (or each substring, if there were several 0 values) is balanced.