A Coloring Problem

Ira M. Gessel[1]
Department of Mathematics
Brandeis University
Waltham, MA 02254
Revised May 4, 1989

Introduction. A well-known algorithm for coloring the vertices of a graph is the "greedy algorithm": given a totally ordered set of colors, each vertex of the graph (taken in some order) is colored with the least color not already used to color an adjacent vertex. When applied to a path graph with at least two vertices, the algorithm uses either 2 or 3 colors, depending on the order in which the vertices are colored.

I. Bouwer and Z. Star [1] solved the problem of counting the number of vertex orderings for a path of n vertices (out of $n!$ possible vertex orderings) for which the greedy algorithm uses only two colors. They expressed the number of 2-color vertex orderings in terms of the number $O(n)$ of 2-color vertex orderings in which the first vertex to be colored occurs in an odd position. They then found recurrences for $O(2m)$ and $O(2m+1)$ that led to differential equations for the exponential generating functions

$$G(t) = \sum_{m=0}^{\infty} O(2m+1) \frac{t^{2m+1}}{(2m+1)!}$$

and

$$H(t) = \sum_{m=0}^{\infty} O(2m) \frac{t^{2m}}{(2m)!}$$

(with $O(0) = O(1) = 1$) which they solved to obtain

$$\frac{1}{G(t)} = \frac{1}{t}\left(t - t^2 + \frac{2t}{e^{2t}-1}\right) \tag{1}$$

and

$$H(t) = e^{t^2/2} \exp\left(\int_0^t \tau^2 G(\tau)\, d\tau\right). \tag{2}$$

When a simple generating function can be found by solving a differential equation, it can often be found more directly by a combinatorial argument. Although (1) and (2) don't look particularly simple, (1) can be rewritten as

$$\begin{aligned} G(t) &= \frac{\sinh t}{\cosh t - t \sinh t} \\ &= \sum_{m=0}^{\infty} \frac{t^{2m+1}}{(2m+1)!} \Bigg/ \left(1 - \sum_{k=1}^{\infty} (2k-1) \frac{t^{2k}}{(2k)!}\right) \end{aligned} \tag{3}$$

[1] partially supported by NSF grant DMS-8703600
1980 *Mathematics Subject Classification.* 05A15

and (2) can be evaluated to give

$$H(t) = \frac{1}{\cosh t - t \sinh t} = 1 \Big/ \left(1 - \sum_{k=1}^{\infty} (2k-1) \frac{t^{2k}}{(2k)!}\right). \tag{4}$$

We shall give direct combinatorial proofs of (3) and (4).

1. Permutations. We may assume that the vertices of the n-path are 1, 2, ..., n, in that order along the path. To any vertex ordering we associate the permutation π for which vertex i is the $\pi(i)$th vertex to be colored; in other words, the vertices are colored in the order $\pi^{-1}(1)$, $\pi^{-1}(2)$, ..., $\pi^{-1}(n)$. If only two colors are used, they must alternate along the path, and this will be the case if and only if all vertices colored with color 1 have the same parity. In the coloring associated to the permutation π, a "mistake" will occur only if some vertex i is colored 1 when it should be colored 2. This happens when vertex i is colored before its neighbors but i has the opposite parity from $\pi^{-1}(1)$, the first vertex to be colored. Thus π yields a 2-coloring if and only if π has the property that all its *valleys* have the same parity, where a valley of π is defined to be an i such that $\pi(i-1) > \pi(i) < \pi(i+1)$ (with $\pi(0) = \pi(n+1) = \infty$). Bouwer and Star showed that the number $P(n)$ of permutations of $1, 2, \ldots, n$ in which all valleys have the same parity is related to the number $O(n)$ of permutations in which every valley is odd by $P(2m) = 2O(2m)$ and $P(2m+1) = 2m(2m+1)O(2m-1)$ for $m \geq 1$. Thus we need only count permutations in which every valley is odd. We call these *coloring permutations*.

2. Derangement numbers and the hook factorization. If π is a permutation of $1, 2, \ldots, n$, we call i a *rise* (or *ascent*) of π if $\pi(i) < \pi(i+1)$ and a *fall* (or *descent*) of π if $\pi(i) > \pi(i+1)$. There is an extensive theory of of counting permutations with respect to various aspects of their rises and falls, of which a comprehensive account has been given by Goulden and Jackson [3, Chapter 4]. Nearly any counting formula involving the rises and falls (and in particular, the valleys) of permutations can be derived from the general theory. However, a very simple, and apparently new, decomposition for permutations which we call the *hook factorization* enables us to count coloring permutations directly.

We introduce this decomposition by first applying it to a similar, but simpler, problem. Let $d(n)$ be the number of *derangements* of $1, 2, \ldots, n$, i.e., permutations π such that for all i, $\pi(i) \neq i$. It is well known that

$$\sum_{n=0}^{\infty} d(n) \frac{t^n}{n!} = \frac{e^{-t}}{1-t},$$

which may be written as

$$\sum_{n=0}^{\infty} d(n) \frac{t^n}{n!} = 1 \Big/ \left(1 - \sum_{k=2}^{\infty} (k-1) \frac{t^k}{k!}\right) \tag{5}$$

since $(1-t)e^t = \sum_{k=0}^{\infty} t^k/k! - \sum_{k=0}^{\infty} kt^k/k!$. Equation (5) seems difficult to interpret directly in terms of derangements. However, there is another class of permutations counted by $d(n)$, found by Désarménien [2], for which (5) has a simple combinatorial interpretation. Let us identify a permutation π with its sequence of values $\pi(1)\,\pi(2)\cdots\pi(n)$. We shall call π a *D-permutation* if its longest final[1] increasing subsequence $\pi(n-i+1)\ \pi(n-i+2)\cdots\pi(n)$ has even length. Thus 531246 is a D-permutation ($i=4$), but 531264 is not ($i=1$).

To prove that the right side of (5) counts D-permutations, let us define a *hook* to be a sequence $h_1\,h_2\ldots h_k$, with $k \ge 2$, satisfying $h_1 > h_2 > \cdots > h_{k-1} < h_k$. (If $k=2$ this reduces to $h_1 < h_2$.) If a sequence of distinct numbers is not decreasing, it has a unique left factor (under concatenation) which is a hook. It follows that every permutation has a unique factorization, which we call the *hook factorization*, of the form $\alpha_1\alpha_2\cdots\alpha_m\beta$, where the α_i are hooks and β, which may be empty, is a decreasing sequence, called the *tail*. For example, the hook factorization of 976813452 is $\alpha_1 = 9768$, $\alpha_2 = 13$, $\alpha_3 = 45$, and $\beta = 2$. It is not difficult to see that a permutations is a D-permutation if and only if its tail is empty.

Next we count permutations whose hooks have given lengths. First note that k distinct numbers may be arranged into a hook in $k-1$ ways: the last element of the hook may be any of the numbers except the smallest, and once the last element is chosen, the others must be arranged in decreasing order. So the number of permutations of $1, 2, \ldots, n = k_1 + \cdots + k_r + m$ whose hooks have lengths $k_1,\ k_2,\ \ldots,\ k_r$ and whose tail has length m is

$$\binom{k_1+\cdots+k_r+m}{k_1,\cdots,k_r,m}(k_1-1)\cdots(k_r-1). \tag{6}$$

Thus if $D(n)$ is the number of D-permutations of $1, 2, \ldots, n$, then $\sum_{n=0}^{\infty} D(n)t^n/n!$ may be obtained by setting $m=0$ in (6), multiplying by $t^{k_1+\cdots+k_r}/(k_1+\cdots+k_r)!$ and summing on $k_1, \ldots, k_r \ge 2$ and $r \ge 0$. We obtain

$$\sum_{r=0}^{\infty}\left(\sum_{k=2}^{\infty}(k-1)\frac{t^k}{k!}\right)^r$$

which is the right side of (5). (Désarménien showed that $D(n) = d(n)$ by constructing a bijection between derangements and D-permutations.)

3. Coloring permutations There is a close connection between the valleys of a permutation and the lengths in its hook factorization. It is easy to verify the following facts: A valley in a hook can occur only in the next-to-last position of the hook. This position will always be a valley in a hook of length greater than 2 and may or may not be a valley in a hook of length 2. A valley in the tail of a permutation can occur only at the end of the tail. If the tail has length greater than 1, the end will always be a valley, and if the tail has length 1, the end may or may not be a valley.

[1] Désarménien considered the longest initial (rather than final) increasing subsequence, but the cardinalities are the same by symmetry.

Although the lengths in the hook factorization do not completely determine the set of valleys, they do determine whether or not every valley is odd: the observations of the previous paragraph imply that a permutation π of $1, 2, \ldots, n$ is a coloring permutation if and only if the following two conditions are satisfied:

(i) Every hook in the hook factorization of π has even length.

(ii) If n is even, then π has an empty tail.

Thus by (5), the number of coloring permutations whose hooks have lengths $2k_1, \ldots, 2k_r$ and whose tail has length $2m+1$ is

$$\binom{2k_1 + \cdots + 2k_r + 2m + 1}{2k_1, \cdots, 2k_r, 2m+1}(2k_1 - 1)\cdots(2k_r - 1). \tag{7}$$

As before, multiplying (7) by $t^{2k_1+\cdots+2k_r+2m+1}/(2k_1 + \cdots + 2k_r + 2m + 1)!$ and summing on $k_1, \ldots, k_r \geq 1, m \geq 0, r \geq 0$ yields (3), and (4) can be derived similarly.

References

1. I. Bouwer and Z. Star, A question of protocol, Amer. Math. Monthly 95 (1988), 118–121.
2. J. Désarménien, Une autre interprétation du nombre des dérangements, Actes 8e Séminaire Lotharingien, ed. D. Foata, Publ. I.R.M.A. Strasbourg, 1984, 11–16.
3. I. P. Goulden and D. M. Jackson, Combinatorial Enumeration, Wiley, 1983.

Addendum for combinatorialists. Since this paper is intended as a *Monthly* note, I had to keep it short and to the point. But I would like to say a little more about derangement numbers and the hook factorization in this section (which will not be included in the published paper).

There are many interesting examples of exponential generating functions f, g, and h related by

$$f(x) = \frac{1}{1-g(x)} = e^{h(x)}.$$

The archetypal example is, of course, $g(x) = x$, $h(x) = \log(1-x)^{-1}$, $f(x) = (1-x)^{-1}$. If we start with any exponential generating function g which counts something then the corresponding $h = \log(1-g)^{-1}$ and $f = (1-g)^{-1}$ have simple combinatorial interpretations: h counts cycles of g-things, or equivalently, permutations (i.e., linear arrangements) of g-things in which the smallest (or perhaps largest) label occurs in the first (or perhaps last) g-thing; and f counts sets of h-things, or equivalently, permutations of g-things. It is often instructive to start with f and try to find combinatorial interpretations for the corresponding g and h (especially if they are known to have nonnegative coefficients!). For example, if f is the generating function for (labeled) graphs then h is clearly the generating function for connected graphs. However, an interpretation for the corresponding g is not so obvious. To interpret g a different model for f seems to be necessary: f is also the generating function for tournaments and h is the generating function for initially connected tournaments (those in which there is a directed path from the vertex with the smallest label to every other vertex), since any tournament is uniquely determined by its initially connected components. Moreover a tournament can be identified with a linear arrangement of its strongly connected components, so g is the generating function for strongly connected tournaments.

Since the exponential generating function $f(x) = D(x)$ for derangements can be expressed as $(1-g)^{-1}$ where g has positive coefficients, it was natural to try to find a combinatorial interpretation, and this led to the hook factorization. However, we might also look at the corresponding h. From the usual interpretation of $D(x)$, h is the generating function for cycles of length greater than 1. From the D-permutation interpretation, we see that h also counts cycles of hooks. There is a very easy bijection between these two kinds of objects: any cycle of length greater than 1 has a unique circular factorization as a cycle of hooks. By combining this observation with Foata's "fundamental transformation" between permutations and sets of cycles, we recover Désarménien's bijection between derangements and D-permutations: given a derangement as a set of cycles of length greater than 1, factor each cycle into a cycle of hooks. Now use Foata's correspondence to transform the set of cycles of hooks into a linear arrangement of hooks, which is a D-permutation.

One final remark: everything in this paper (suitably modified) also works for permutations of a multiset.