A PROOF OF JULIAN WEST'S CONJECTURE THAT THE NUMBER OF TWO-STACK-SORTABLE PERMUTATIONS OF LENGTH n IS 2(3n)!/((n+1)!(2n+1)!)

Doron Zeilberger *

Appeared in: Discrete Math. **102** (1992), 85-93.

A proof on the computer is just a physical experiment.— (Common sentiment among mathematicians)

Abstract: The Polya-Schutzenberger-Tutte methodology of weight enumeration, combined with about 10 hours of CPU time (of Maple running on Drexel University's Sun network) established Julian West's conjecture that 2-stack-sortable permutations are enumerated by sequence #651 in the Sloane listing.

(-1). Prologue

June 3, 1991: About a month ago, (10:30 AM, May 4, 1991, Bordeaux time, to be precise), at the *Séries formelles et combinatoire algébrique*conference, Julian West gave an enthralling talk which contained an intriguing conjecture: a certain naturally defined combinatorial family is enumerated by a certain nice formula. First I was sure that I could do it the same night. Then I was certain that it would be proved during the 8-hour plane ride back home. Well, it took longer than expected, and required about 50 mathematician-hours, 10 (Maple) programmer-hours, (the mathematician and programmer being myself), and 10 CPU-hours to *construct* the proof. Once constructed, the verification of the proof takes a few minutes of Maple CPU time (on the above computer.)

The proof would not have been possible without the generous and kind permission of Drexel's Mathematics and Computer Science Head, James C.T. Pool, to use the Drexel computing facilities.

People who detest the Appel-Haken proof of the 4 Color Theorem would probably not like the present proof either. I like both proofs very much. The human part of the present proof is very elegant, using the Polya-Schutzenberger-Tutte ([P],[S],[T]) powerful methodology of *weight-enumeration*. The machine part is very tedious, but *who cares?*. Certainly not the machine, who is always happy to be useful. Another reason why I liked working on this project is that I got to *experience* what it's like to be an experimental scientist. Both the construction of the proof, and its final verification, used the methodology of experimental science. The resulting proof is as rigorous and valid as any old fashioned proof, but the *flavor*and *spirit* of the proof are experimental, and making it rigorous amounts to just mumbling a few words. I agree with the motto if you delete the word "just", which turns it from a curse to a blessing. After all, a human proof is just a sociological-psychological act of polemics, and physics is a hard science, while sociology and psychology are soft.

0. Introduction

In his remarkable thesis [W1][W2], Julian West introduced a fascinating new kind of combinatorial objects: k-stack-sortable permutations. They may defined as follows ([W2], lemma 5). Define a mapping Π acting on permutations π of a finite set S of integers, with $n := max(S)$, by the recursive recipe:

---

* Department of Mathematics, Temple University, Philadelphia, PA19122. Supported in part by NSF grant DMS8901690.

$$\Pi(\pi^L n \pi^R) := \Pi(\pi^L)\Pi(\pi^R)n \, , \; \Pi(empty) := empty.$$

A permutation π is $k - stack - sortable$ if $\Pi^k(\pi)$ equals the identity permutation. As observed by West, the number of 1-stack-sortable permutations on n objects is well known to be Catalan's number $(2n)!/((n)!(n+1)!)$. West conjectured that the number of 2-stack sortable permutations of length n is $2(3n)!/((n+1)!(2n+1)!)$.

1. How The Proof Was Found

Step 0: Use West's[W2] characterization of 2-stack-sortable permutations as permutations avoiding such and such kind of subsequences to get a hold on them. Approach abandoned and two weeks wasted.

Step 1: This is the purely human part, described in section 2. Let W_n be the number of 2-stack-sortable permutations of length n, and let $P(x)$ be its ordinary generating function:

$$P(x) := \sum_{n=0}^{\infty} W_n x^n \; .$$

Ideally, it would have been nice to find a recurrence for the W_n, or equivalently, some functional equation for $P(x)$ directly. I was unable to do so. Instead, using the definition of Π, a bijection, and *weight enumeration*, a functional equation for a more general formal power series, $\Phi(x,t)$, was obtained, that for $t = 1$ reduced to the former $P(x)$: $\Phi(x,1) = P(x)$. Unfortunately, it was *not*a plain *algebraic* equation, and furthermore plugging in $t = 1$ resulted in the famous tautology $0 = 0$. The functional equation was of the form $G(\Phi(x,t), \Phi(x,1), x, t) \equiv 0$, for some 4-variate polynomial G given in section 2. The functional equation did however give an effective way to compute the West numbers W_n much beyond $n = 11$, that West[W2] computed by directly enumerating permutations. This corroborated West's conjecture and safely moved it outside the jurisdiction of the law of small numbers.

Step 2 : Put your faith in *notre bon maitre*, and conjecture that $\Phi(x,t)$ satisfies an algebraic equation, i.e. there exists a polynomial F in (Φ, x, t) such that $F(\Phi, x, t) \equiv 0$. Systematically I tried raising the degrees in x and Φ, until Maple produced an "awful" polynomial F of degree 6 in Φ, degree 8 in x and degree 9 in t. It was found by computing $\Phi(x,t)$ up to a sufficiently large power of x, using the functional equation of step 1, plugging into the generic F, and setting the coefficients of the powers of x to zero, until one gets enough linear equations for the coefficients of F. However, if you do it naively, you will run out of memory pretty fast. So you plug in many specific values of t and then combine them together by "Lagrange" (or rather "Pade") interpolation. $F(\Phi, x, t)$ is given in the Maple program of the appendix.

Step 3: Define $\Psi(x,t)$ as the (unique formal power series) solution of the algebraic equation $F(\Psi, x, t) \equiv 0$. Our goal is to prove that $\Psi \equiv \Phi$. A naive approach is to "solve" $F(\Psi, x, t)$ "explicitly", say by radicals, and verify that it satisfies the functional equation $G = 0$ of step 1. However, Maple was unable to do it.

The functional equation of step 1, $G(\Phi(x,t), \Phi(x,1), x, t)$ is hard to work with, because of the unwieldy $\Phi(x,1)$, which is $P(x)$. By differentiating G w.r.t t, and using the chain rule, one obtains a first order algebraic differential equation $G_1(\Phi(x,t), \Phi_t(x,t), P, x, t)$. Finding the resultant of

$G(\Phi, P, x, t)$ and $G_1(\Phi, \Phi_t, P, x, t)$, w.r.t P, eliminates P and yields an algebraic (first order) differential equation for $\Phi(x,t)$: $H(\Phi, \Phi_t, x, t) = 0$. The Maple code that produces H is given in the appendix.

Step 4: Differentiate $F(\Psi(x,t), x, t) \equiv 0$, w.r.t t, using the chain rule, to get $\Psi_t(x,t) = -F_t(\Psi, x, t)/F_\Psi(\Psi, x, t)$. Substitute it into $H(\Psi, \Psi_t, x, t)$, and find out whether it's zero. In other words, find out whether the numerator of $H(\Psi_t, \Psi, x, t)$ is an exact multiple of the polynomial $F(\Psi, x, t)$. Maple said: YES. Hence both Ψ and Φ satisfy the same algebraic differential equation $H \equiv 0$, and it follows by uniqueness that $\Phi = \Psi$. Even here we had to be clever, since a direct verification resulted in the error message: "object too large". We found out the appropriate degrees in x, t, and plugged in enough special cases. We then use the fact that "if a polynomial of degree $\leq r$ is 0 in r+1 distinct values, it is identically zero."

Step 5: Now we are on the home stretch. We need information about $P(x) := \Phi(x, 1)$, which we now know is equal to $\Psi(x, 1)$. Plugging in $t = 1$ in $F(\Psi(x,t), x, t)$, gives you $F(\Psi(x, 1), x, 1)$ and surprise! It equals:

$$x^2(x-1)^3(-1 + P + 11x - 14Px + 2P^2x + x^2 + 3Px^2 + 3P^2x^2 + P^3x^2)^2 \ .$$

Since the ring of formal power series has no zero divisors, (and hence also no nilpotents), it follows that

$$-1 + P + 11x - 14Px + 2P^2x + x^2 + 3Px^2 + 3P^2x^2 + P^3x^2 = 0 \ , \tag{1.1}$$

which is not quite yet doable by Lagrange inversion, but we are getting close.

Step 6: Now it's time to "peek at the answer at the end of the book".

$$P(x) := \sum_{n=0}^{\infty} W_n x^n \ ,$$

satisfies (1.1). We want to prove that $W_n = 2(3n)!/((2n+1)!(n+1)!)$, $n \geq 1$. We do know that the generating function $C(x)$ for ternary trees satisfies $C = 1 + xC^3$, and its coefficients T_n have the nice formula, obtainable by Lagrange inversion (and otherwise), $T_n = (3n)!/(n!(2n+1)!)$. We want to prove that $C(x) = (1 + (xP(x))')/2$. Differentiating (1.1) w.r.t x, we find an expression for $P'(x)$ in terms of $P(x)$ and x, set $D(x) := (1 + (xP(x))')/2$, evaluate $D(x) - 1 - xD(x)^3$ and verify that its numerator is a multiple of the left side of (1.1), and hence is identically zero, and hence $C(x) = D(x)$. QED.

2.The Human Part: Getting The Functional Equation Of Step 1

For any permutation π of $1, 2, ..., n$, let $i(\pi)$ be the largest integer i such that the subsequence of the "big i": $n-i+1, ..., n-1, n$ are in decreasing order. Let $W^{(i)}$, be the set of all permutations (of any length) π such that $i(\pi) = i$, and let $W^{\geq i}$ be the set of permutations π such that $i(\pi) \geq i$.

Let's analyze a typical member of $W^{\geq i}$. If its length is n, then it has the form

$$\pi \ = \ \sigma_0 n \sigma_1 (n-1) ... (n-i+1) \sigma_i \ , \tag{2.1}$$

where $\sigma_0, ..., \sigma_i$ are (possibly empty) permutations of disjoint smaller sets, the union of whose underlying sets is $1, 2, ..., n-i$. Now, by iterating the definition of Π,

$$\Pi(\pi) = \Pi(\sigma_0)\Pi(\sigma_1)...\Pi(\sigma_i)(n-i+1)(n-i+2)...n \quad ,$$

so that,

$$\Pi^2(\pi) = \Pi(\Pi(\sigma_0)\Pi(\sigma_1)...\Pi(\sigma_i))(n-i+1)...n \quad .$$

It follows that there is a 1-1 correspondence between the elements of $W^{\geq i}$ and $i+1$-tuples of permutations $\sigma_0, ..., \sigma_i$, such that $\Pi(\Pi(\sigma_0)...\Pi(\sigma_i))$ equals the "identity" (i.e. the increasing permutation), and the underlying sets of the σ's are disjoint and their union is $1, 2, ..., n-i$.

Consider now a typical element of $W^{(i)}$,*except* the following permutation of length i: $i, i-1, ..., 1$. It still has the form (2.1) *but* $n-i$ should not be in σ_i. In other words, although the "big i" are in decreasing order, the "big $i+1$" are not, so the subsequence consisting of the "big $i+1$" looks as follows, for some $0 \leq j \leq i-1$:

$$n(n-1)...(n-j+1)(n-i)(n-j)...(n-i+1) \quad . \tag{2.2}$$

Padding in the rest, we get that a typical π of length n, belonging to $W^{(i)}$ has the form

$$\pi \ = \ \sigma_0 n \sigma_1 (n-1) \sigma_2 ... \sigma_{j-1} (n-j+1) \bar{\sigma} (n-i) \sigma_j (n-j) \sigma_{j+1} ... \sigma_{i-1} (n-i+1) \sigma_i \, .$$

It follows from the definition of Π that

$$\Pi(\pi) = \Pi(\sigma_0)...\Pi(\sigma_{j-1})\Pi(\bar{\sigma}(n-i)\sigma_j(n-j)\sigma_{j+1}...\sigma_{i-1}(n-i+1)\sigma_i)\ (n-j+1)...n \ =$$

$$\Pi(\sigma_0)...\Pi(\sigma_{j-1})\Pi(\bar{\sigma}(n-i)\sigma_j)\Pi(\sigma_{j+1}(n-j-1)...\sigma_{i-1}(n-i+1)\sigma_i)(n-j)(n-j+1)...n \ =$$

$$\Pi(\sigma_0)...\Pi(\sigma_{j-1})\Pi(\bar{\sigma})\Pi(\sigma_j)\ (n-i)\ \Pi(\sigma_{j+1})...\Pi(\sigma_i)(n-i+1)...n \quad .$$

Now apply Π again, to get

$$\Pi^2(\pi) \ = \ \Pi(\Pi(\sigma_0)...\Pi(\bar{\sigma})\Pi(\sigma_j))\Pi(\Pi(\sigma_{j+1})...\Pi(\sigma_i))(n-i)(n-i+1)...n \quad .$$

It follows that every element of $W^{(i)}$, except the excluded permutation $(i, i-1, ..., 1)$, corresponds to a pair of tuples of permutation, for some $0 \leq j \leq i-1$,

$$[(\sigma_0, ..., \bar{\sigma}, \sigma_j), (\sigma_{j+1}, ..., \sigma_i)] \quad ,$$

such that both $\Pi(\Pi(\sigma_0)...\Pi(\sigma_j))$ and $\Pi(\Pi(\sigma_{j+1})...\Pi(\sigma_i))$ are the identity permutation, and the underlying sets satisfy the obvious requirements. But we saw that these correspond to members of $W^{\geq j+1}$ and $W^{\geq i-j-1}$ respectively. So we have a bijection

$$W^{(i)} \to \{(i, i-1, ..., 1)\} \cup \cup_{j=0}^{i-1} W^{\geq j+1} \times W^{\geq i-j-1} \ , \ \pi \to (\pi_1, \pi_2) \ , \tag{2.3}$$

such that $length(\pi) = length(\pi_1) + length(\pi_2) + 1$.

For each permutation π, introduce the weight:

$$weight(\pi) := x^{length(\pi)} \ ,$$

and, by abuse of notation, from now on, for any set of permutations S, let $S(x)$ be the formal power series that equals the sum of all the weights of the elements of S. By taking weights on both sides of (2.3) (the *Polya-Schutzenberger-Tutte transform*), we get

$$W^{(i)}(x) = x^i + x \sum_{j=0}^{i-1} W^{\geq j+1}(x) W^{\geq i-j-1}(x) \ . \tag{2.4}$$

Now let

$$\Phi(x,t) := \sum_{i=0}^{\infty} W^{(i)}(x) t^i \ .$$

It is easily seen that if we define

$$\bar{\Phi}(x,t) := \sum_{i=0}^{\infty} W^{\geq i}(x) t^i \ ,$$

then

$$\bar{\Phi}(x,t) = \sum_{j \geq i \geq 0} W^{(j)}(x) t^i \ = \ \sum_{j=0}^{\infty} W^{(j)}(1 + t + ... + t^j) \ = \tag{2.5}$$

$$\sum_{j=0}^{\infty} W^{(j)}(1 - t^{j+1})/(1-t) \ = \ (\Phi(x,1) - t\Phi(x,t))/(1-t) \ .$$

Now (2.4) can be written as

$$W^{(i)}(x) = x^i + x \sum_{j=0}^{i} W^{\geq j}(x) W^{\geq i-j}(x) \ - \ x W^{\geq 0}(x) W^{\geq i}(x) \ .$$

Multiplying both sides by t^i and summing from $i=0$ to ∞, realizing that the middle term on the right is a convolution, and that $W^{\geq 0} = \Phi(x,1)$, we get

$$\Phi(x,t) = \frac{1}{(1-xt)} + x\bar{\Phi}(x,t)^2 - x\Phi(x,1)\bar{\Phi}(x,t) \ ,$$

which upon substituting for $\bar{\Phi}(x,t)$ its expression (2.5) in terms of $\Phi(x,t)$, we get (recall that $\Phi(x,1) = P$)

$$\Phi \ - \ \frac{1}{1-xt} \ - \ \frac{xt(P-t\Phi)(P-\Phi)}{(1-t)^2} = 0 \ ,$$

which by clearing denominators, and taking the numerator, finally yields the functional equation $G(\Phi, P, x, t) \equiv 0$, promised in step 1 of section 1.

ω. Epilogue: How The Proof Could Have Been Found

July 2, 1991: The first proof of any conjecture is seldom the shortest. It turns out that the present proof is no exception. Ira Gessel made the brilliant observation that steps 2-5 can be replaced by the following.

Step 2': Conjecture $I(P(x), x)) = 0$ ((1.1)) empirically. To prove it rigorously, we must show that the unique $\Psi(x,t)$ that satisfies $G(\Psi(x,t), \Psi(x,1), x, t) \equiv 0$, is such that $I(\Psi(x,1), x) \equiv 0$. Let's write,

$$(i) G(\Psi(x,t), Q(x), x, t) \equiv 0 \ , \ (ii) \Psi(x,1) = Q(x) \ , \ (iii) I(Q(x), x) \equiv 0 \, .$$

We have to prove that (i)+(ii) implies (iii). But note that (i)+(ii) have a unique solution, and (i)+(iii) have a unique solution, and we must show that these are the same. So it's enough to show that (i)+(iii) implies (ii). Taking the resultant of G and I w.r.t. $Q(x)$ gives the algebraic equation $F(\Psi, x, t) \equiv 0$ found empirically, and very painfully, in step 3. Proceeding as in step 5, we see that indeed $Q(x) = P(x)$. This observation is the *leitmotif* of a paper [GZ] that Ira Gessel and I hope to write.

References

[GZ] I. Gessel and D. Zeilberger, *An empirical method for solving (rigorously) algebraic-functional equations of the form* $F(P(x,t), P(x,1), x, t) \equiv 0$(temporary title), in planning.

[P] G. Polya, *On picture writing*, Amer. Math. Monthly **63** (1956), 689-697. Reprinted in: "*CLASSICAL PAPERS IN COMBINATORICS*", edited by I. Gessel and G. -C. Rota, Birkhauser, Boston, 1987, pp. 249-258.

[S] M. P. Schutzenberger, *Context free languages and pushdown automata*, Information and Control **6**(1963), 246-264.

[T] W.T. Tutte, *On the theory of chromatic polynomials*, Canadian J. Math. **68**(1954), 101-121.

[W1] Julian West, *"Permutations with restricted subsequences and stack-sortable permutations"*, doctoral thesis, M.I.T., 1990.

[W2] ——, *Sorting twice through a stack*, Proceedings of *Séries Formelles et Combinatoire Algébrique* (M. Delest, G. Jacob, and P. Leroux, eds.) 397-406. Also to appear in J. Theoretical Computer Science.