STANFORD UNIVERSITY STANFORD, CALIFORNIA 94305-2140

DONALD E. KNUTH
Professor Emeritus of The Art of
Computer Programming
Department of Computer Science
Telephone [415] 723-4367

October 28, 1993

Professor Herb Wilf
Department of Mathematics DRL/E1
University of Pennsylvania
Philadelphia, Pennsylvania 19104–6395

Professor Doron Zeilberger
Department of Mathematics
Temple University
Philadelphia, PA 19122

Dear Complex Variables,

This letter is the report I promised to send about your "algorithmic proof theory" paper. You and I know that it is a wonderful paper, brimful of results, a milestone achievement. But I also must report some of the negative impressions I had while reading it; I wish it had been prepared with the extra care it deserved. The editors of *Inventiones* should have found a referee to say the things below, before they accepted it for publication. (Let's hope they do better with Andrew Wiles's draft!)

Most of my complaints are due, I guess, to a certain sloppiness. The introduction, through the middle of page 584, is excellent; by this time I'm expecting to really enjoy the paper, especially after the great joke on page 582. But then you change notational conventions without warning, putting the continuous variables $x_1, \ldots, x_n$ first while you had them last up till now. On half of page 589 and all of page 590 they go back to second position. In 6.2.2 they're first, in 6.3.1 they're last.

On page 590, the statement of the fundamental theorem suddenly uses the notations K_i and N without defining them; earlier they had been E_k and E_n. (And in the proof they are, in fact, E_k and E_n. Notice, however, that the proof refers to $P(n, E_n)$ while the statement of the theorem refers to $P(N, n)$.) When you finally do get around to defining K_i and N, on page 599 and 608, they are *down*-shift operators instead of *up*-shifts! Then they go back to up-shifts on page 618. There's no reason you couldn't have kept them up-shifts throughout, by slightly modifying the argument on pages 597–600. Having the same notation for two different things is a sure source of confusion.

Inconsistent and undefined notations are merely annoying, but incomplete mathematical reasoning is more serious. As I mentioned in earlier correspondence, I cannot see how to complete your proof of the fundamental theorem. More precisely, I guess you have proved the fundamental theorem stated on page 590, but that theorem is trivially true when P is the zero operator and $R_i = S_j = 0$. Your proof shows that we can satisfy (2.1.1) with P, R_i, S_j not all zero; but in the applications, we want to satisfy (2.1.1) with P nonzero. I cannot see how to prove this.

You claim on page 592, lines 4 and 5, that when P is zero "it is possible to use (2.1.1) to get another, nonzero operator that annihilates $f(n)$." By $f(n)$ I guess you mean $F(n, k_1, \ldots, k_r, y_1, \ldots, y_s)$, since f is undefined. But I see no way to verify your claim. For example, suppose $F(n, j, k)$ is an enormously complicated proper hypergeometric term that happens to be a function of $j + k$. Then (2.1.1) holds with $0 = (J-1)F - (K-1)F$, but how do I use this to get a nonzero recurrence $P(N, n)F$ satisfied by $(J-1)R_jF + (K-1)R_kF$?

Herb, I have your letter of October 7 in which you point out that the proofs of Theorem 3.2C and 4.2C establish a nonzero recurrence. But those proofs require additional conditions about zones of zeroes, etc.; is that what we need for a nontrivial fundamental theorem to be true? (It wipes out my example of a function of $j + k$, because arbitrarily large $|j|$ and $|k|$ will make $F(n, j, k)$ nonzero. On the other hand I do see how to handle the special case $F(n, j, k) = G(n, j + k)$; we find a recurrence $P(N, n)G(n, m) = (M - 1)R(n, m)G(n, m)$ and then get one for F.)

I fear that my two previous paragraphs are incoherent, but they do explain (I hope) why I still seek a proof of a more satisfactory fundamental theorem. I have no doubt that the hypotheses should imply (2.1.1) for some nonzero P; but I don't see how to prove it, although your remark on page 592 line 3 claims this.

Moreover, the fundamental corollary on page 592 is trivial unless you assert that P is nonzero.

And when I saw the proof of that corollary, my first reaction was to write "lean and meaningless" in the margin! Why? Because it was obvious to you, but didn't dawn on me for awhile, that you intend to sum the terms of (2.1.1) involving k_i first with respect to k_i, using a different order of summation for each i. (For example, in the double-sum case, if you sum the whole right side with respect to j first, you mess up the telescoping property with respect to k.) This trick should be explained, I think. Moreover, you need a stronger condition than "vanishing at infinity" as you have defined on page 590, in order to justify this change of summation order and change of integration order. Like absolute convergence at least.

On page 593 near the bottom you speak of shifting the starting value of n so that the leading coefficient of P does not vanish, then you "check the identity case by case" for smaller n. OK, that's a finite operation, but what do you do when the free variable is continuous? "Case by case" fails.

On page 594 you speak of $r + s$ "bonus" identities. This sounds very exciting until one looks at the examples; in practice it seems that the rational functions R_i are so complicated, the bonus identities are of no interest whatever. At least, none of the bonus identities in section 6 are of significantly greater interest than the formulas generated on a random day by monkeys who are churning out all the "theorems" of mathematics. This is in strong contrast to the WZ pairs, whose bonuses and duals are often insightful.

On page 596, the definition of proper-hypergeometric term should not require the c's and w's to be integers; I already wrote you about that. Furthermore there's an interesting issue about what it means for ξ to be a "parameter." (On page 589 line 4 you refer to "constant parameters," whatever those are.) Actually ξ can be any nonzero function of other parameters; also $P(n,k)$ and the c's and w's can depend on these other parameters. For example, I see no reason why

$$(n\cos z + k\ln w)(2n-3k+\sqrt{wz}\,)!\,\left(\frac{1+w^{\zeta(z)}}{\ln\Gamma(w)}\right)^k$$

should not be a proper hypergeometric term involving the parameters w and z. The polynomials $\alpha_{ij}(n)$ in (3.1.2) will, of course, involve these parameters too. The number q in q-hypergeometric terms might also depend on several parameters.

On page 597 there's another problem with definitions. Your example of the Pascal triangle recurrence does *not* hold "at every grid point in the plane other than the origin," by your definitions, because $n!/k!(n-k)!$ is not well-defined when $n < 0$. (With a better definition, I think Pascal's recurrence holds everywhere in the plane, including the origin.) Furthermore you apply Theorem 3.1 later as if the recurrence (3.1.2) were true also when $F(n,k) = 0$, although the theorem specifically excludes this case.

In the hypotheses to Theorem 3.2A you need to take I and J large enough to get a k-free recurrence. Moreover, this theorem is stated to hold only for a particular value of n and k, so its conclusion is trivial; you want $a_0(n)$, $a_1(n)$, etc., to be the same functions for lots of different values of n and k, but the present statement has all the a_i depending on (n,k)!

The footnote on page 599, which speaks of characterizing denominator polynomials, seems to contradict the conjecture about "holonomic-plus-hypergeometric equals proper" on the middle of page 585. (Incidentally, $2^n 3^k$ is a counterexample, but you can easily patch the conjecture.)

I wrote before about the gap in the proof of Theorem 3.2A, which I patched in the new GKP. Your generating function proof of Theorem 3.2B fixes it nicely, but only in the special case of "admissible" terms. On page 602, line 11, you say that the k-free recurrence found by Theorem 3.1A will produce a nontrivial recurrence in the inadmissible case, but for this you really need the stronger statement in Theorem 3.2A.

I didn't have time to read sections 4 and 5, and I merely skimmed the earlier material about q-identities. On page 588, I believe the displayed formula on line 18 needs to be more general, namely

$$(c)_n = \frac{(c)_\infty}{(cq^n)_\infty}\,.$$

At the bottom of that page, the q-derivative is undefined when $x = 0$ or $q = 1$; it approaches D_x as $x \to 0$ too, not only as $q \to 1$.

In Theorem 6.1.1 the sign of $8kni$ in R_1 should be $+$, not $-$. It's curious that '/4' occurs in the numerator of R_1 while 4 is in the denominator of R_2. But the real puzzle in Theorem 6.1.1 is the presence of the factor $(n+1+i)$ in the denominators of both R_1 and R_2. You claimed on page 590, bottom line, that it's possible to predict the denominators, but I do not believe that the proof of Theorem 4.1 gives the slightest hint about a denominator factor like this. None of the factorials in numerator or denominator of $F(j,m)$ has the form $n+i+\text{constant}$.

Similarly, I cannot understand how $i+j$ gets into the denominator of R_1 and R_2 in examples 6.1.2 and 6.1.3. Here there is $(i+j)!$ in the numerator of F_1, but still I don't see how the proof of 4.1 (if rewritten for forward differences instead of backward differences as actually applied here) would "predict" $(i+j)$ in the denominator. If I had been referee of this paper I would have insisted on an explanation. Looking at the way Shalosh actually works, I see that such factors were inserted by hand. How? You must explain this in the paper that you have promised to write (page 593, line 3).

On page 618, there is nothing wrong with Theorem 6.1.2, but I would have preferred to see a nicer proof in which $R_1(i,j) = R_2(j,i)$. It's easy to see that such certifiers exist when F is symmetric, hence we can significantly speed up the computation by halving the number of unknown coefficients. The solution in this case (unique, for its degree) is

$$\begin{aligned} R_1(i,j) = i^2(1-i+n)(2i+3i^2+i^3-4j-2ij+2i^2j+7j^2-ij^2 \\ -2j^3+in-i^2n-8jn+ijn+5j^2n-3jn^2)/d \end{aligned}$$

where d is the old denominator.

Example 6.3.2 is not quite complete; you need to verify the constants of proportionality. Stanton's proof would be shorter too if he had left much of it to the reader's imagination.

Typos:

Page 578, line 2, $(a+n=1) \rightarrow (a+n-1)$

Page 581, line −11, f and k should be F and $\mathbf{k}$

Page 583, line −4, prof $\rightarrow$ proof

Page 584, line 14, $D_z \rightarrow D_x$

Page 586, line 10, set $S \rightarrow$ finite set S

Page 586, line 17, sequence $\rightarrow$ term

Page 587, line 11, GR $\rightarrow$ GaRa

Page 587, line 13, AW $\rightarrow$ AsWi

Page 587, line 21, $\binom{n+k}{n} \rightarrow \binom{n+k}{k}$

Page 591, line 8, exists an $\rightarrow$ exists a nonzero

Page 591, line 15, apply F to the above operator equation $\rightarrow$

apply the above operator equation to F

Page 593, line −10, It they $\rightarrow$ If they

Page 596, (3.1.1) and following, why use "pp" and "qq" in a

mathematics paper? Must they be perfect squares??

Page 596, line 18 and page 599, line 26, hypergeomtric $\rightarrow$ hypergeometric

Page 597, line 10, subject $\rightarrow$ subsection

Page 597, (3.1.4), $\binom{n-1}{k} \rightarrow \binom{n-1}{k}^2$

Page 598, (4.1.6)(c), *ff* $\to$ ff

Page 598, (3.1.6)(e), pff $\to$ pff {

Page 599, line 8, $\epsilon^3.$ $\to$ $\epsilon.^3$

Page 603, line 9, $\alpha_{i,j}$ $\to$ α_j

Page 605, fix the definition of $\Lambda_u(a)$ when $u < 0$.

Page 608, line 19, let be $\to$ let N be

Page 610, line 6, series $\to$ polynomial

Page 618, line 15, funtions $\to$ functions

Page 618, line –5, identitfy $\to$ identity

Page 619, line –1, turn out $\to$ turns out

Page 624, line 15, troughout $\to$ throughout

Page 624, line –9 should be in italics.

Finally, a comment that I hope Doron will appreciate because English is not his first language: When Americans violate the currently conventional wisdom about proper usage of 'which' versus 'that', they always overuse 'which'. I don't recall ever seeing 'that' where the word ought to be 'which', except in your papers! And in the present paper, I found at least 18 instances. (Page 575, line –4; page 576, line 26; page 577, line –8; page 578, lines –12 and –6 (twice); . . . ; page 628, lines –7 and –5.) There's a simple general rule that would automatically correct almost all of these: After a comma or left parenthesis, use 'which'.

Well, I hope you realize that I wouldn't have taken the time to write a letter like this if I didn't have great admiration for the overwhelming majority of things I didn't complain about. Still, I felt I needed to let off a little steam, particularly since I found nothing at all amiss in [WZ1] and [WZ2]. I look for a return to that former standard of excellence when you make the next breakthrough. This letter is private between the three of us.

Best regards,

Don

Donald E. Knuth
Professor

DEK/pw