RESEARCH ANNOUNCEMENTS

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 23, Number 1, July 1990

TOWARDS COMPUTERIZED PROOFS OF IDENTITIES

HERBERT S. WILF AND DORON ZEILBERGER

1. INTRODUCTION

Many of the seemingly trivial facts that we take for granted are really theorems, like the fact that $11 \times 12 = 132$, or that $\frac{1}{2} + \frac{1}{3} = \frac{5}{6}$. The reason that these are no longer thought of as theorems is that nowadays quite routine algorithms perform these tasks. The same is true, thanks to modern computer algebra programs, for the "theorem" that $(a+b)^{20} = a^{20} + \cdots$, or, thanks to the recent completion by Risch and others [14, 15] of the problem of finding a complete algorithm for integration in finite terms, the same is true for the algorithmic evaluation of indefinite integrals of "elementary functions."

The purpose of this note is to announce a number of results and algorithms which, collectively, do the same for large classes of identities that occur in combinatorics and in the theory of special functions.

Historically, even the binomial theorem itself was considered to require a custom-made proof, but we will show that it, along with a very large class of identities, can be proved by computers. It should be remarked that one does not need to "trust the computer" blindly. Although the proofs are *discovered* by the computer, it

Received by the editors August 24, 1989 and, in revised form, October 19, 1989.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 05A19.
The first author was supported in part by the U.S. Office of Naval Research.
The second author was supported in part by the U.S. National Science Foundation.

produces a proof certificate that can easily be *checked* by hand, if so desired.

There are many, many binomial coefficient identities, and a great deal of effort has gone into developing proofs of these. However, more recently it has become widely appreciated that "almost all" known binomial coefficient identities are special cases of relatively few hypergeometric identities such as those of Pfaff-Saalschütz, Dixon, and Dougall (see, e.g., Andrews [1] and Graham et al. [12]).

However, more and more hypergeometric identities are still being conjectured and proved [9], so the need remains for mechanizing the proofs of even these. A decisive step in this direction was taken by Gosper [10] in 1978, who gave a finite algorithm that completely solves the problem of *indefinite* hypergeometric summation in closed form.

In this announcement we will describe some new steps that have been taken towards algorithmically deriving and proving *definite* hypergeometric identities and the more general class of *holonomic* function identities.

2. Holonomic functions

The theory of holonomic systems and functions was initiated by Joseph N. Bernstein [3] and is currently a very active area with many applications. Here we will describe some far-reaching implications of Bernstein's revolutionary ideas to the theory of special functions and combinatorial identities.

A holonomic discrete (resp. continuous) function $\{a_n\}$ (resp. $f(x)$) of one variable is simply any solution of a homogeneous, linear difference (resp. differential) equation with polynomial coefficients. By analogy, a function of several continuous and/or discrete variables is holonomic if it is a solution of a "maximally overdetermined" system of homogeneous linear partial differential-recurrence equations with polynomial coefficients (see [3, 6, 19]). Although a given holonomic function may be described in many ways, it can be finitely decided whether or not two such presentations describe the same function.

Theorem 1. *The class of holonomic functions is closed under finite addition and multiplication* [4] *and infinite summation or integration over any variable* [19].

If we are given presentations (encodings) of holonomic f and g, there are algorithms for finding the encodings of fg, $f+g$, $\int f$, and $\sum f$. Hence if we are given some holonomic function identity, it can be written in the form "holonomic function =0," and so it can be proved by verifying that the left side is indeed an encoding of 0.

Corollary. *Any identity that involves only sums, products, integrals, and summation signs acting on holonomic functions is verifiable in finitely many steps (though possibly a great many of them!).*

It is easy to see that all of the classical orthogonal polynomials [13] are holonomic in all their variables and parameters. It follows that most of the identities in, say, the book by Erdélyi and associates [8], are provable in a finite number of steps, by a single well-defined argument.

For holonomic functions $F(n, k)$ of two discrete variables, more is true.

Theorem 2. *For a two-variable holonomic function $F(n, k)$ there exist an integer L, polynomials $s_0(n), \ldots, s_L(n)$, and a holonomic function $G(n, k)$, such that*

$$(1)\quad s_0(n)F(n, k) + s_1(n)F(n+1, k) + \cdots + s_L(n)F(n+L, k) = G(n, k+1) - G(n, k),$$

i.e., such that the left side is indefinitely summable w.r.t. k ([19], §6.3).

To summarize, if we wish to discover a simple form, if one exists, for a given summation $a(n) = \sum_k F(n, k)$, where F is holonomic, we could proceed by first finding the difference equation whose existence is guaranteed by Theorem 2. The summation of (1) over k would then yield a recurrence for the unknown sum $a(n)$. If that recurrence is of first order, or otherwise admits some simple solution, then our problem is solved. Otherwise we must be content with only the recurrence formula that is satisfied by the $a(n)$'s, which for computational or asymptotic purposes may be just as good [18]. The method extends to "q" identities. It has yielded a three-line proof of the Rogers-Ramanujan identities [7].

3. The Fast Algorithm

There is a much faster algorithm for the important special case in which the summand $F(n, k)$ is closed form, i.e.,

$$F(n, k) = \frac{\prod_{i=1}^{A}(a_i n + a_i' k + a_i'')!}{\prod_{i=1}^{B}(b_i n + b_i' k + b_i'')!} z^k,$$

where the a's and b's are constant, specific integers, rather than variable parameters.

The proof of Theorem 2 can be extended to show that then $G(n, k)$ is also of closed form. We now know, thanks to the general theory of holonomic functions, that for every closed form $F(n, k)$, there is another closed form $G(n, k)$ such that (1) is true. Furthermore, the slow algorithm of [19] will exhibit the $s_0, \dots, s_L$, $G(n, k)$ explicitly, but it will take a very long time.

Suppose that by some miracle we already knew the $s_i(n)$. Is there a quick way of finding $G(n, k)$? Yes, there is. It is easy to see that the left side of (1) is always of closed form, and finding such a $G(n, k)$ can be done by Gosper's algorithm w.r.t. k. What if we don't know what the s_i's are? Then Gosper's algorithm can be modified [20] to manufacture both the s_i's and the $G(n, k)$. Thus we have an algorithm that by itself is elementary, and whose description does not require the heavy artillery of Bernstein's theory. But the proof that the algorithm always works still needs that theory.

For most closed form $F(n, k)$, the sum w.r.t. k of $F(n, k)$ is not a solution of a *first order* equation. Whenever it is, one has an identity.

4. Rational function certification

It turns out that a strikingly large percentage of hypergeometric and binomial coefficient identities can be proved by a method that is in many ways a very special case of the above. This leads to the certification of the truth of an identity by simply giving a rational function of n and k.

The method of rational function certification [17] was stimulated by the holonomic system theory above, but is *formally* independent of it. The theory is entirely self-contained and is surprisingly simple.

The main idea is this. Suppose we have two functions $F(n, k)$, $G(n, k)$, defined for integer k and integer $n \geq 0$, and suppose the following equation is satisfied:

$$F(n+1,k)-F(n,k)=G(n,k+1)-G(n,k) \tag{2}$$

for nonnegative integer n and integer k. We will call such a pair (F, G) a WZ pair. Then under certain additional boundary conditions ((F1), (G1), (G2) below) we obtain a simple evaluation of the sum

$$\sum_k F(n,k) \qquad (n=0,1,2,\dots). \tag{3}$$

We also obtain a simple evaluation of the associated sum

$$\sum_{n\geq 0} G(n,k). \tag{4}$$

Thus we may obtain two identities, one for each member of the pair. The proofs of the identities will consist in simply verifying that the condition (2) is satisfied, along with the following boundary conditions:

(F1) For each integer k, the limit

$$f_k = \lim_{n\to\infty} F(n,k) \tag{5}$$

exists and is finite.

(G1) For each integer $n \geq 0$, $\lim_{k\to\pm\infty} G(n,k)=0$.

(G2) We have $\lim_{L\to\infty}\sum_{n\geq 0} G(n,-L)=0$.

Theorem 3. *Let (F, G) be a WZ pair. If* (G1) *holds then we have the identity*

$$\sum_k F(n,k) = \mathit{const.} \qquad (n=0,1,2,\dots), \tag{6}$$

where "const." is found by putting $n=0$. Further, if (F1), (G2) *hold, then we have the identity*

$$\sum_{n\geq 0} G(n,k) = \sum_{j\leq k-1} (f_j - F(0,j)), \tag{7}$$

where f is defined by (5).

Example. To prove the identity of Pfaff-Saalschütz we need only check that the pair

$$F(n,k)=\binom{n}{k}\frac{(a)_k(b)_k(c-a-b)_{n-k}(c)_n}{(c)_k(c-a)_n(c-b)_n},$$

$$G(n,k)=-\frac{(b+k-1)(a+k-1)}{(c-b+n)(c-a+n)}F(n,k-1),$$

satisfy the conditions. Then we have a proof of Pfaff-Saalschütz's identity (which is (6) in this case)

$$\sum_k \binom{n}{k} \frac{(a)_k (b)_k (c-a-b)_{n-k}}{(c)_k} = \frac{(c-a)_n (c-b)_n}{(c)_n}.$$

Remark. The WZ pair is always of the form

$$(F(n, k), R(n, k)F(n, k-1)),$$

where R is a rational function. Hence we can state an identity in the conventional form $\sum_k U(n, k) = rhs(n)$, and provide also the rational function $R(n, k)$. The reader can then find $F(n, k) = U(n, k)/rhs(n)$, $G(n, k) = R(n, k)F(n, k-1)$, and check the conditions (2), (F1), (G2) to complete the certification (proof) process. The following strikingly short proof of a hypergeometric identity of Dixon illustrates this process.

Example. The identity of Dixon is

$$\sum_k (-1)^k \binom{n+b}{n+k} \binom{n+c}{c+k} \binom{b+c}{b+k} = \frac{(n+b+c)!}{n!b!c!}.$$

Proof. Take

$$R(n, k) = \frac{(c+1-k)(b+1-k)}{2(n+k)(n+b+c+1)}. \quad \square$$

The method works on the identities of Dougall, Clausen, on the "strange" identities of Gessel-Stanton [9], on "q" identities, etc., etc., just as easily as on those above. One-line proofs of many hypergeometric identities, using the rational function certification method, are in [17]. Over 50 computer-generated one-line proofs of binomial coefficient identities are tabulated in [16]. In many cases the companion identity (7) is "new," in the sense that it is not a special case of a known hypergeometric identity.

The way one would constructively use the method to certify an identity is this. Given $\sum_k H(n, k) = rhs(n)$ as the identity to be proved. Divide by the right side to get the standard form $\sum_k F(n, k) = 1$. Then use Gosper's algorithm to find the WZ mate G, if it exists. Having F, G, it will always be true that $F(n, k) = R(n, k)G(n, k-1)$. Thus the single rational function $R(n, k)$ certifies the identity since from it the full proof can be constructed.

REFERENCES

1. G. E. Andrews, *Applications of basic hypergeometric functions,* SIAM Rev. **16** (1974), 441–484.
2. W. N. Bailey, *Generalized hypergeometric series,* Cambridge Tracts in Mathematics, vol. 32, Cambridge University Press, London, 1935. (Reprinted by Hafner, New York, 1964.)
3. I. N. Bernstein, *Modules over a ring of differential operators, study of the fundamental solutions of equations with constant coefficients,* Functional Anal. Appl. **5** (1974), 89–101.
4. ——, *The analytic continuation of generalized functions with respect to a parameter,* Functional Anal. Appl. **6** (1972), 273–285.
5. G. D. Birkhoff and W. J. Trjitzinsky, *Analytic theory of singular difference equations,* Acta Math. **60** (1932), 1–8.
6. J. E. Björk, *Rings of differential operators,* North-Holland, Amsterdam, 1979.
7. S. B. Ekhad and S. Tre, *A purely verification proof of the first Rogers-Ramanujan identity,* J. Combin. Theory Ser. A. (to appear).
8. A. Erdélyi et al, *The higher transcendental functions,* (3 vols.), McGraw-Hill, New York, 1953.
9. I. Gessel and D. Stanton, *Strange evaluations of hypergeometric series,* SIAM J. Math. Anal. **13** (1982), 295–308.
10. R. W. Gosper, Jr., *Decision procedure for indefinite hypergeometric summation,* Proc. Nat. Acad. Sci. USA **75** (1978), 40–42.
11. H. W. Gould, *Combinatorial identities,* Morgantown, WV, 1972.
12. R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete mathematics,* Addison-Wesley, Reading, MA, 1989.
13. E. D. Rainville, *Special functions,* Chelsea, Bronx, New York, 1971. (Originally published by Macmillan, 1960.)
14. R. H. Risch, *The solution of the problem of integrating in finite terms,* Bull. Amer. Math. Soc. **76** (1970), 605–608.
15. B. M. Trager, *On the integration of algebraic functions,* Ph.D. thesis, MIT, Cambridge, MA, 1985.
16. H. S. Wilf, *54 computer-generated proofs of binomial coefficient identities* (to appear).
17. H. S. Wilf and D. Zeilberger, *Rational functions certify combinatorial identities,* J. Amer. Math. Soc. **3** (1990), 147–158.
18. J. Wimp and D. Zeilberger, *Resurrecting the asymptotics of linear recurrences,* J. Math. Anal. Appl. **111** (1985), 162–177.
19. D. Zeilberger, *A holonomic systems approach to special function identities,* J. Comp. Appl. Math. (to appear).
20. ——, *The method of creative telescoping,* J. Symbolic Computation (to appear).

UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PENNSYLVANIA 19104-6395

DREXEL UNIVERSITY, PHILADELPHIA, PENNSYLVANIA 19104