ZEILBERGER'S ALGORITHM - HOW THE ALGORITHM WORKS

MAT625 SEMINAR: AUTOMATIC PROOFS OF BINOMIAL SUM IDENTITIES

BETTINA WOHLFENDER

1. Introduction

The topic of my presentation is the derivation of Zeilberger's algorithm, also known as "the method of creative telescoping". This algorithm will give us a recurrence formula for F(n,k) in telescoped form. F(n,k) is a proper hypergeometric term.
In the last presentation, it was shown, how this result can be used to compute a closed form for $f(n)$ where $f(n) = \sum_{k=0}^{\infty} F(n,k)$. In the next presentation, hold by Alain Schmid, we will see, that this recurrence always exists. The existence result is very important to derive the algorithm but the methods will be different from the methods used in the existence proof.
The exact form of the recurrence formula is the following:

$$\sum_{j=0}^{J} a_j(n) F(n+j,k) = G(n,k+1) - G(n,k). \tag{1}$$

The unknowns are:

(i) on the left side the coefficients $a_0, ..., a_J$ and the order of recurrence J
(ii) and on the right side the function G.

2. First part of the algorithm: reshaping and rearranging terms

We will assume that the order of recurrence is J and fix this integer. Then we try to find a recurrence of order J and if there is none, we try to find one of the order $J+1$ and so on. Thanks to the existence result, we'll finally come to an end.
We define the left side of (1) as t_k.
Now we consider the quotient

$$\frac{t_{k+1}}{t_k} = \frac{\sum_{j=0}^{J} a_j(n) F(n+j,k+1) / F(n,k+1)}{\sum_{j=0}^{J} a_j(n) F(n+j,k) / F(n,k)} \cdot \frac{F(n,k+1)}{F(n,k)} \tag{2}$$

We divide and multiply in the numerator by $F(n,k+1)$ and in the denominator by $F(n,k)$. The second factor on the right, i. e. $\frac{F(n,k+1)}{F(n,k)}$, is a rational function (if $F(n,k)$ is a hypergeomteric term in k, then $\frac{F(n,k+1)}{F(n,k)}$ is a rational function) depending on n and k.

EMAIL: BETTINA.WOHLFENDER@UZH.CH

We define

$$\frac{F(n,k+1)}{F(n,k)} = \frac{r_1(n,k)}{r_2(n,k)}. \tag{3}$$

r_1 and r_2 are polynomials (as a rational function can be written as the quotient of two polynomials). Furthermore, we define

$$\frac{F(n,k)}{F(n-1,k)} = \frac{s_1(n,k)}{s_2(n,k)}. \tag{4}$$

As $F(n,k)$ is also hypergeomteric in n, we have another rational function and s_1, s_2 are polynomials, too.
Next we do some reshaping: It holds

$$\frac{F(n+j,k)}{F(n,k)} = \prod_{i=0}^{j-1} \frac{F(n+j-i,k)}{F(n+j-i-1,k)} = \prod_{i=0}^{j-1} \frac{s_1(n+j-i,k)}{s_2(n+j-i,k)} \tag{5}$$

The first equality holds since in the middle term everything but the first factor in the numerator and the last factor in the denominator will be reduced. The second equality holds by definition of s_1 and s_2.
Now we can do some further reshaping and get:

$$\frac{t_{k+1}}{t_k} = \frac{\sum_{j=0}^{J} a_j \left(\prod_{i=0}^{j-1} \frac{s_1(n+j-i,k+1)}{s_2(n+j-i,k+1)}\right) \cdot r_1(n,k)}{\sum_{j=0}^{J} a_j \left(\prod_{i=0}^{j-1} \frac{s_1(n+j-i,k)}{s_2(n+j-i,k)}\right) \cdot r_2(n,k)} \tag{6}$$

We get this by substituting (3) and (5) into (2). (6) is equal to:

$$\frac{\sum_{j=0}^{J} a_j \left(\prod_{i=0}^{j-1} s_1(n+j-i,k+1) \prod_{r=j+1}^{J} s_2(n+r,k+1)\right)}{\sum_{j=0}^{J} a_j \left(\prod_{i=0}^{j-1} s_1(n+j-i,k) \prod_{r=j+1}^{J} s_2(n+r,k)\right)} \cdot \frac{r_1(n,k) \prod_{r=1}^{J} s_2(n+r,k)}{r_2(n,k) \prod_{r=1}^{J} s_2(n+r,k+1)}. \tag{7}$$

How we get from (6) to (7)? There are certain factors in equations (6) and (7) which are identical. These are $\prod_{i=0}^{j-1} s_1(n+j-i,k+1)$, $\prod_{i=0}^{j-1} s_1(n+j-i,k)$, $r_1(n,k)$ and $r_2(n,k)$.
Now I comment on the differences of (6) and (7). In (7), instead of dividing by $\prod_{i=0}^{j-1} s_2(n+j-i,k+1)$ as in (6), we divide by $\prod_{r=1}^{J} s_2(n+r,k+1)$. So we divide too much and therefore multiply with $\prod_{r=j+1}^{J} s_2(n+r,k+1)$. Analogously, instead of multiplying $\prod_{i=0}^{j-1} s_2(n+j-i,k)$ we multiply by $\prod_{r=1}^{J} s_2(n+r,k)$. So we multiply too much and therefore we divide by $\prod_{r=j+1}^{J} s_2(n+r,k)$.
(7) is now an expression of the form

$$\frac{t_{k+1}}{t_k} = \frac{p_0(k+1)}{p_0(k)} \frac{r(k)}{s(k)} \tag{8}$$

where $p_0\left(k\right) = \sum_{j=0}^{J} a_j \left(\prod_{i=0}^{j-1} s_1\left(n+j-i,k\right) \prod_{r=j+1}^{J} s_2\left(n+r,k\right)\right)$,

$$r\left(k\right) = r_1\left(n,k\right) \prod_{r=1}^{J} s_2\left(n+r,k\right),$$

$$s\left(k\right) = r_2\left(n,k\right) \prod_{r=1}^{J} s_2\left(n+r,k+1\right).$$

Important remark: the (unknown) coefficients $\{a_j\}_{j=0}^{J}$ just appear in $p_0\left(k\right)$. They don't appear in $r\left(k\right)$ or $s\left(k\right)$.

3. Second part of the algorithm: use of the previous chapter about Gosper's algorithm

Next we use theorem 5.3.1 of the book which states:

Theorem. *Let K be a field of characteristic zero and $r \in K[n]$ a nonzero rational function. Then there exist polynomials $a, b, c \in K[n]$ such that b, c are monic and $r\left(n\right) = \frac{a(n)}{b(n)} \frac{c(n+1)}{c(n)}$, where*

(i) $gcd\left(a\left(n\right), b\left(n+k\right)\right) = 1$ for every nonnegative integer k,
(ii) $gcd\left(a\left(n\right), c\left(n\right)\right) = 1$,
(iii) $gcd\left(b\left(n\right), c\left(n+1\right)\right) = 1$.

Thanks to this theorem we can write $\frac{r(k)}{s(k)}$ as

$$\frac{r\left(k\right)}{s\left(k\right)} = \frac{p_1\left(k+1\right)}{p_1\left(k\right)} \cdot \frac{p_2\left(k\right)}{p_3\left(k\right)} \tag{9}$$

such that the three assumptions of the theorem are fulfilled.
Now we define $p\left(k\right)$:

$$p(k) = p_0\left(k\right) p_1\left(k\right). \tag{10}$$

By the equations (8), (9) and (10) we get

$$\frac{t_{k+1}}{t_k} = \frac{p\left(k+1\right)}{p\left(k\right)} \cdot \frac{p_2\left(k\right)}{p_3\left(k\right)}. \tag{11}$$

We can again use results of chapter 5 about Gosper's algorithm as equation (11) is a setup for which we can apply Gosper's algorithm. I won't discuss any further details concerning this since it isn't part of my topic. Explanations can be found in chapter 5.2 of the book. From there we can conclude that t_k is an indefinitely summable hypergeometric term if and only if the recurrence formula

$$p_2\left(k\right) b\left(k+1\right) - p_3\left(k-1\right) b\left(k\right) = p\left(k\right) \tag{12}$$

has a polynomial $b\left(k\right)$ as solution.
Recall: A hypergeometric term t_k is indefinitely summable if $\sum_{k=0}^{a} t_k$ has a closed form, where a is an arbitrary upper bound.

Important remark: the coefficients of $p_2(k)$ and $p_3(k)$ in equation (12) are independent of the (unknown) coefficients $\{a_j\}_{j=0}^{J}$ and the right hand side, namely $p(k)$, depends on $\{a_j\}_{j=0}^{J}$ (only) linearly. (This is a direct consequence of the definitions of p_0, p_1, p_2 and p_3.)
For finding a polynomial solution of (12), we can again use a part of chapter 5, namely page 85, to find an upper bound on the degree of the polynomial $b(k)$. I won't discuss any further details of this process here.

4. Third part of the algorithm: solve system of linear equations

After having found an upper bound for $b(k)$, we call it A, we proceed as follows. We write $b(k)$ as a general polynomial of degree A, where the coefficients $\{c_l\}_{l=0}^{A}$ aren't known yet. So

$$b(k) = \sum_{l=0}^{A} c_l k^l. \tag{13}$$

We can substitute the equality (13) into the equality (12) and we get a system of linear equations (by doing coefficient comparison) with the unknown coefficients $\{a_j\}_{j=0}^{J}$ and $\{c_l\}_{l=0}^{A}$. The system is linear since the coefficients of $p_2(k)$ and $p_3(k)$ are independent of the (unknown) coefficients $\{a_j\}_{j=0}^{J}$ and the right hand side, namely $p(k)$ depends on $\{a_j\}_{j=0}^{J}$ (only) linearly, as I've already mentioned.
Now we try to solve this system for the unknown coefficients. If there is no solution for it, then there doesn't exist a recurrence in the telescoped form (1) of the assumed order J. So we need to check if there is one of order J+1 and so on. Thanks to the existence theorem, we know that for a certain order there has to exist a solution for the system of linear equations.
If we find a polynomial solution of (12), then we get the coefficients $\{a_j\}_{j=0}^{J}$ of the recurrence formula (1). Additionally, we can again use a section of chapter 5 to get G(n,k). Equation (5.2.5) in chapter 5.2 is the following: $y(n) = \frac{b(n-1)x(n)}{c(n)}$. Therefore we can conclude

$$G(n,k) = \frac{p_3(k-1)}{p(k)} b(k)\, t_k.$$

That was the complete algorithm to find the recurrence formula (1).
Remark: The system of linear equations has usually not a unique solution but the solution is determined up to a constant. This is enough and the closed form $f(n)$ is independent of this constant, since if we have a recurrence formula like (1) and then sum over all k, the right side telescopes to 0 and we can multiply the equation with every constant (not equal to 0) that we want.

5. Example

To finish my presentation, I'd like to determine the coefficients $\{a_j\}_{j=0}^{J}$ of the recurrence formula for a specific example by hand with the help of the algorithm. We'd like to find a closed form of
$f(n) = \sum_{k=0}^{\infty} F(n,k)$ with $F(n,k) = \binom{2n}{2k}$.
We first assume that the order of recurrence J is equal to 1. First we do some reshaping in order to get the form we wanted for $\frac{t_{k+1}}{t_k}$. It holds

$$\frac{t_{k+1}}{t_k} = \frac{a_0\binom{2n}{2k+2} + a_1\binom{2n+2}{2k+2}}{a_0\binom{2n}{2k} + a_1\binom{2n+2}{2k}} = \frac{a_0\frac{2n!}{(2k+2)!(2n-2k-2)!} + a_1\frac{(2n+2)!}{(2k+2)!(2n-2k)!}}{a_0\frac{2n!}{(2k)!(2n-2k)!} + a_1\frac{(2n+2)!}{2k!(2n-2k+2)!}}.$$

After reducing the fractions we get

$$\frac{a_0\frac{1}{(2k+1)(2k+2)} + a_1\frac{(2n+1)(2n+2)}{(2k+1)(2k+2)(2n-2k-1)(2n-2k)}}{a_0\frac{1}{(2n-2k-1)(2n-2k)} + a_1\frac{(2n+1)(2n+2)}{(2n-2k-1)(2n-2k)(2n-2k+1)(2n-2k+2)}}.$$

Now we expand the fraction with $(2n-2k-1)(2n-2k)$ and get

$$\frac{a_0\frac{(2n-2k-1)(2n-2k)}{(2k+1)(2k+2)} + a_1\frac{(2n+1)(2n+2)}{(2k+1)(2k+2)}}{a_0 + a_1\frac{(2n+1)(2n+2)}{(2n-2k+1)(2n-2k+2)}}.$$

As a next step we can write $\frac{t_{k+1}}{t_k}$ in the form of equation (8):

$$\frac{a_0(2n-2k-1)(2n-2k) + a_1(2n+1)(2n+2)}{a_0(2n-2k+1)(2n-2k+2) + a_1(2n+1)(2n+2)} \cdot \frac{(2n-2k+1)(2n-2k+2)}{(2k+1)(2k+2)}$$

So it holds

$$\begin{aligned} p_0(k) &= a_0(2n-2k+1)(2n-2k+2) + a_1(2n+1)(2n+2), \\ r(k) &= (2n-2k+1)(2n-2k+2), \\ s(k) &= (2k+1)(2k+2). \end{aligned}$$

Now we have to find the canonical form (9) for $\frac{r(k)}{s(k)}$: it holds

$$r(k) = (2n-2k+1)(2n-2k+2) = 4n^2 + 4k^2 + 2 - 8kn + 6n - 6k$$

$$s(k) = (2k+1)(2k+2) = 4k^2 + 6k + 2$$

The two terms $r(k)$ and $s(k)$ aren't coprime. We divide both terms by 2 and get the form of equation (9):

$$\frac{r(k)}{s(k)} = \frac{1}{1} \cdot \frac{2n^2 + 2k^2 + 1 - 4kn + 3n - 3k}{2k^2 + 3k + 1}$$

So we have

$$\begin{aligned} p_1(k) &= 1, \\ p_1(k+1) &= 1, \\ p_2(k) &= 2n^2 + 2k^2 + 1 - 4kn + 3n - 3k, \end{aligned}$$

$$p_3\left(k\right) = 2k^2 + 3k + 1.$$

We write $\frac{t_{k+1}}{t_k}$ as in equation (11) with

$$p\left(k\right) = p_0\left(k\right) p_1\left(k\right) = a_0\left(2n - 2k + 1\right)\left(2n - 2k + 2\right) + a_1\left(2n + 1\right)\left(2n + 2\right)$$

$$\frac{t_{k+1}}{t_k} = \frac{a_0\left(2n - 2k - 1\right)\left(2n - 2k\right) + a_1\left(2n + 1\right)\left(2n + 2\right)}{a_0\left(2n - 2k + 1\right)\left(2n - 2k + 2\right) + a_1\left(2n + 1\right)\left(2n + 2\right)} \cdot \frac{2n^2 + 2k^2 + 1 - 4kn + 3n - 3k}{2k^2 + 3k + 1}$$

Now we have to solve the following equation (this corresponds to equation (12))

$$\left(2n^2 + 2k^2 + 1 - 4kn + 3n - 3k\right) b\left(k + 1\right) - \left(2k^2 - k\right) b\left(k\right) =$$
$$a_0\left(2n - 2k + 1\right)\left(2n - 2k + 2\right) + a_1\left(2n + 1\right)\left(2n + 2\right)$$

By page 85 in chapter 5 of the book we find out that $b\left(k\right)$ has degree 1, we write $b\left(k\right)$ as $b\left(k\right) = c + dk$.
By doing coefficient comparison, we can create a system of linear equations:

(i) $2n^2c + c + 3nc + 2n^2d + d + 3nd = 4a_0n^2 + 2a_0 + 6a_0n + 4a_1n^2 + 6a_1n + 2a_1$
(ii) $-4nkc - 3kc + 2n^2dk + dk + 3ndk - 4ndk - 3kd + kc = -8a_0kn - 6a_0k$
(iii) $2k^2c - 4nk^2d - 3k^2d + 2k^2d - 2k^2c + dk^2 = 4a_0k^2$

When we solve this with maple we get:

$$c = -3/2nd - d, a_0 = -nd, a_1 = nd/4, d = d$$

As I've already mentioned, the solution is only determined up to constants. If we choose d=4, we get $a_0 = -4n$ and $a_1 = n$. This leads to the recurrence

$$nF\left(n + 1, k\right) - 4nF\left(n, k\right) = G\left(n, k + 1\right) - G\left(n, k\right).$$

And this is exactly what we get when we use one of the computer programs.
We can now find a closed form for $f(n)$ thanks to the presentation of Noam Szyfer. We are in scenario A, as $J = 1$ and get: $f\left(n\right) = f\left(1\right)\prod_{j=1}^{n-1} -\frac{a_0}{a_1} = 2\prod_{j=1}^{n-1} -\frac{-4n}{n} = 2 \cdot 4^{n-1} = 2 \cdot 2^{2n-2} = 2^{2n-1}$ for $n \geq 1$.
Remark: This formula holds true for $n \geq 1$ since for $n = 0$, the equation would be fulfilled for an arbitrary $f(n)$ and moreover, we would divide by 0.
This was my presentation about Zeilberger's algorithm with the focus on how the algorithm works. Of course, it is quite hard to do more complicated examples by hand, so usually we use the computer programs which were presented in the last presentation by Noam Szyfer.