HAL open science

The ABC of Creative Telescoping — Algorithms, Bounds, Complexity

Frédéric Chyzak

To cite this version:

Frédéric Chyzak. The ABC of Creative Telescoping — Algorithms, Bounds, Complexity. Symbolic Computation [cs.SC]. Ecole Polytechnique X, 2014. tel-01069831

HAL Id: tel-01069831
https://tel.archives-ouvertes.fr/tel-01069831

Submitted on 30 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Memoir of *Habilitation à diriger des recherches (HDR)*
University Paris-Sud 11
Speciality Computer Sciences (ED 427)

The ABC of Creative Telescoping

Algorithms, Bounds, Complexity

by

FRÉDÉRIC CHYZAK

Chargé de recherche, INRIA (France)

April 14, 2014

Defended after the reviews by

- Moulay Barkatou (University of Limoges),
- Marko Petkovšek (University of Ljubljana, Slovenia),
- Nobuki Takayama (Kobe University, Japan),

in front of the committee consisting of

- Moulay Barkatou (University of Limoges),
- Jean-Charles Faugère (INRIA),
- André Galligo (University of Nice Sophia Antipolis),
- Florent Hivert (University Paris-Sud 11),
- Joris van der Hoeven (CNRS).

Contents

Chapter 1

Introduction

This memoir is concerned with the algorithmic symbolic treatment of integrals and sums that appear in various fields like the theory of special functions, combinatorics, and mathematical physics, and more generally with exact algebraic manipulations of systems of linear functional equations, whether differential or of recurrence, and of their solutions. My approach here is almost always algorithmic, whether with the goal of enlarging the class of systems and functions to which the approach applies, or with the goal of making the algorithms more efficient, as much as possible with a provably better arithmetic complexity. At the same time, we endeavour to propose a fair account on the context of the research in this domain, both ours and others', so as to provide a global and coherent view on it.

As a memoir written to obtain the French degree of *Habilitation à diriger des recherches*, it discusses my research contributions and is a picture of the scientific landscape I understand. But composing a picture is a selective process that requires one to keep disturbing details out of the view. My choice here is to focus on the part of my research concerned with the method of *Creative Telescoping*. I shall thus not cover some parts of my research topics, like applications of Gröbner bases to control theory, quasi-optimal complexity algorithms (power-series solutions of differential systems, differential equations for algebraic functions, products of linear differential operators), or my recent works on the Dynamic Dictionary of Mathematical Functions. As to my starting works on the interaction between computer algebra and formal-proofs theory about questions related to special functions and creative telescoping, this will only be alluded to in the perspectives.

Creative telescoping can be viewed as a common formalism for several, possibly surprisingly related, operations on functions, series, and formal series. I shall present creative telescoping first as a method to perform integration and summation of special functions, combinatorial sequences, orthogonal polynomials, but its introduction in combinatorics was largely motivated by other operations on generating series like extractions of constant coefficients and diagonals. It has also been applied to the evaluation of certain scalar products in the theory of the symmetric functions of combinatorics. This will be reviewed in the following text.

1.1 The Name "Creative Telescoping"

The phrase *creative telescoping* appears in an explanation by van der Poorten (1979, p. 211) of Apéry's irrationality proof of $\zeta(3)$. One of van der Poorten's steps is to establish that the sum

$$b_n = \sum_{k=0}^{n} b_{n,k}, \quad \text{where} \quad b_{n,k} = \binom{n}{k}^2 \binom{n+k}{k}^2, \tag{1.1}$$

satisfies the same second-order linear recurrence equation as another binomial sum. He mentions that, to get the recurrence explicitly, Cohen and Zagier "cleverly construct

$$B_{n,k} = 4\,(2n+1)\,\bigl(k\,(2k+1) - (2n+1)^2\bigr)\,b_{n,k}, \tag{1.2}$$

with the motive that

$$B_{n,k} - B_{n,k-1} = (n+1)^3 b_{n+1,k} - (34n^3 + 51n^2 + 27n + 5) b_{n,k} + n^3 b_{n-1,k}\text{''}, \tag{1.3}$$

so that summing over k from 0 to $n+1$ results in the wanted recurrence,

$$0 = (n+1)^3 b_{n+1} - (34n^3 + 51n^2 + 27n + 5) b_n + n^3 b_{n-1}. \tag{1.4}$$

Here, the series whose general term is the right-hand side of (1.3) reduces to the sum over k of the term $B_{n,k} - B_{n,k-1}$, which telescopes, giving the name to the method. Works by Zeilberger to design algorithms for obtaining analogues of the key relation (1.3) for general summands popularised the approach (Zeilberger, 1982, 1990a, 1991).

The method has a differential analogue, which Almkvist and Zeilberger (1990) name the *method of differentiating under the integral sign*, but this highlights only one aspect of the computation. For concreteness and to motivate my statement, I reproduce an example by Zeilberger (1982), which he borrowed from a classical textbook on integration: to evaluate the parametrised integral

$$f(b) = \int_{-\infty}^{+\infty} e^{-x^2} \cos 2bx \, dx, \tag{1.5}$$

first perform differentiation w.r.t. b under the integral sign, followed by integration by parts w.r.t. x, to get the relation

$$f'(b) = \int_{-\infty}^{+\infty} -2xe^{-x^2} \sin 2bx \, dx = \Bigl[e^{-x^2} \sin 2bx\Bigr]_{x=-\infty}^{x=+\infty} + \int_{-\infty}^{+\infty} -2be^{-x^2} \cos 2bx \, dx = -2b\, f(b).$$

Solving the induced ODE requires to know an initial condition, $f(0)$, which amounts to getting an explicit form for the integral at $b = 0$. In the end, this results in the explicit form $f(b) = \sqrt{\pi}\,\exp(-b^2)$. This calculation with integrals can be reconsidered on the level of integrands: after introducing $f(b,x) = \exp(-x^2)\cos 2bx$, the integration by parts above is a consequence of the relations

$$\frac{df}{db}(b,x) = \frac{d}{dx}\bigl(e^{-x^2} \sin 2bx\bigr) - 2b\, f(b,x) = \frac{d}{dx}\left(-\frac{1}{2x}\frac{df}{db}(b,x)\right) - 2b\, f(b,x).$$

Reorganising terms delivers the following analogue of (1.2)–(1.3), with self-explanatory notation:

$$\frac{dF}{dx}(b,x) = \frac{df}{db}(b,x) + 2b\, f(b,x) \quad \text{with} \quad F(b,x) = -\frac{1}{2x}\frac{df}{db}(b,x). \tag{1.6}$$

In view of the strong formal analogy between the recurrence case (1.3) and the differential case (1.6), and because "differentiating under the integral sign" is only one aspect of the approach, Salvy and I have started to use the same phrase "creative telescoping" in (Chyzak and Salvy, 1998) to denote these two similar situations, and specifically for the task of obtaining (1.3) or (1.6) in an algorithmic way.

As I shall describe in the historical context below, creative telescoping is related to an elimination theory applied to a (non-commutative) polynomial representation of linear differential/difference operators. The absence of algorithms for linear operators in the early 1980s—or at least their relative immaturity—may explain the slow start of the creative-telescoping theory developed by Zeilberger.

I should note here that in the present memoir I shall only consider a kind of creative telescoping where, like in (1.2) and (1.6), the auxiliary term derived from the summand or integrand is crafted *in a linear way*. By contrast, another line of research has been investigated, starting with Schneider's PhD thesis and after works by Karr (1981, 1985), by allowing non-linear expressions. More recently, a differential analogue has been developed by Raab as a continuation of Risch's algorithm (1969; 1970). On this non-linear creative telescoping, I refer the reader to the recent surveys (Raab, 2013; Schneider, 2013).

1.2 Linear Operators

It has proved very fruitful in the works that will be described in this memoir to represent the linear differential equations and linear recurrences under consideration as linear differential/difference "operators". The following notation and conventions will be used throughout the whole text.

I shall constantly consider functions of (continuous) variables x, y, etc., and sequences of (discrete) variables n, k, etc., as well as variations like sequences of functions, parametrised families of functions, etc. All such objects will collectively be called "functions", unless disambiguation is needed, and will be subject to respective derivation operators denoted by D_x, D_y, etc., and (forward) shift operators denoted by S_n, S_k, etc. When needed, a backward shift operator will be denoted as an inverse: S_n^{-1}, S_k^{-1}, etc. Composition will be denoted just by products and powers, so that, for example, $D_x^5 S_n^{-1} S_k^3$ acts on a "function" f by the rule:

$$\left(D_x^5 S_n^{-1} S_k^3 f\right)(n,k,x) = \frac{\partial^5 f}{\partial x^5}(n-1,k+3,x).$$

These operators combine with operators of multiplication by a variable to generate more general operators. Operators of multiplication will be denoted by the variable itself, so as to enforce rules like:

$$(xf)(n,x) = xf(n,x), \qquad (nf)(n,x) = nf(n,x).$$

It follows from the Leibniz rule and just the effect of substitution that

$$\begin{aligned} D_x x f = (D_x x) f = \big(D_x(xf)\big) = xD_x f + f = (xD_x + 1)f, \\ S_n n f = (S_n n) f = \big(S_n(nf)\big) = (n+1)S_n f, \end{aligned}$$

where the operator 1 denotes the identity operator. These formal rules lead us to expect the following algebraic relations between non-commutative polynomials:

$$D_x x = xD_x + 1, \qquad S_n n = (n+1)S_n.$$

A theory to make sense of such commutations has been developed in algebra, starting with Ore's work in the 1930s. Since "operators" are considered in the present memoir for their algebraic properties, and not for any topological or analytic one, I shall at times more properly speak of *skew polynomials* for the objects originally studied by Ore. In the literature, they are also known as Ore polynomials, Ore operators, pseudo-linear transformations, and pseudo-linear operators.

The theory of skew polynomials started with Ore's work on polynomials in a single derivation or shift operator: mainly with (1933), in which he developed a theory of one-sided gcd for skew polynomials, but also with (1931), in which he considered matrices of skew polynomials. Skew polynomials were later considered by Jacobson (1937) under the name *pseudo-linear transformations*, and algorithms for gcd and factorisation for general skew polynomials were discussed in (Bronstein and Petkovšek, 1994, 1996).

Ore's construction produces rings of operators, a.k.a. skew polynomial rings. We thus have, for example: the ring $\mathbb{Q}(x)\langle D_x\rangle$ of linear differential operators with coefficients in the rational-function field $\mathbb{Q}(x)$; the ring $\mathbb{Q}(n)\langle S_n\rangle$ of linear recurrence operators with coefficients in the rational-function field $\mathbb{Q}(n)$; analogues $\mathbb{Q}[x]\langle D_x\rangle$ and $\mathbb{Q}[n]\langle S_n\rangle$ when we are interested in operators with polynomial coefficients only; an analogue $\mathbb{Q}(n)\langle S_n^{-1}\rangle$ when considering backward shifts instead of forward shifts.

When several derivation/shift operators are needed in the same algebraic setting, Ore's construction could be iterated abstractly, which introduces the possibility of derivations and shifts that do not commute with one another. We viewed this as a drawback in the work coauthored with Salvy (1998), as we would have lost finiteness properties crucial for good the behaviour of algorithms. In that article, we therefore developed a theory of a kind of algebras that we named *Ore algebras*, in which derivations and shifts necessarily commute, while they need not commute with the coefficients. Coming back to the example of $D_x^5 S_n^{-1} S_k^3$, this lives for example in the algebra $\mathbb{Q}(n, k, x)\langle S_n, S_n^{-1}, S_k, S_n, D_x\rangle$, where S_n, S_n^{-1}, S_k, and D_x commute pairwise, while none of them commutes with all elements from $\mathbb{Q}(n, k, x)$.

In the present document, I chose to present neither Ore's theory nor our theory of Ore algebras and to refrain from using the corresponding more heavy notation, $\mathbb{Q}(x)[\partial; \sigma, \delta]$.

1.3 Important Classes of Functions and Sequences

A few specific classes of functions and sequences appear constantly in the present text, namely hypergeometric sequences, hyperexponential functions, D-finite functions, and ∂-finite functions, which are all solutions of linear functional operators. The distinction between those classes is based on the type of operators considered and the order of the equations that the functions solve. As creative-telescoping algorithms manipulate systems of defining equations rather than functions explicitly represented by their names, the different classes lead to different specialised algorithms, taking advantage of properties of the defining systems.

Sequences of the index n (over $\mathbb{N}$ or $\mathbb{Z}$) that solve a *first*-order linear *recurrence* equation with coefficients that are polynomial functions of n are called *hypergeometric*. Equivalently, a sequence $u = (u_n)$ is hypergeometric if there exists a rational function $R(n)$ such that the ratio u_{n+1}/u_n is equal to $R(n)$, except maybe for finitely many values of n. In this case, I shall consider that $S_n - R(n)$ annihilates u, disregarding the finite number of exceptions. This generalises to sequences $u = (u_{n_1,\dots,n_r})$ of several indices by requiring that there exist for each i a rational function $R_i(n_1, \dots, n_r)$ such that u is annihilated by each first-order operator $S_{n_i} - R_i$. Examples of hypergeometric sequences are given by the terms 2^n, $n!$, $(n_1 + n_2)!$, $1/(n_1 - n_2)!$, $1/(n_1 + n_2)$, and $\binom{n_1}{n_2}$. Rational functions of the indices are hypergeometric, as well as products of hypergeometric sequences.

The situation is analogous in the differential case, where a *hyperexponential* function is defined as a function f of continuous variables $x_1, \dots, x_r$ whose r logarithmic derivatives $D_i f/f$ are rational functions of the r variables. Hyperexponential functions solve systems of *first*-order linear *differential* equations with coefficients that are polynomial functions of the x's. Examples include rational functions, exponentials of rational functions, powers of rational functions to rational and transcendental constants. In addition, products of hyperexponential functions are hyperexponential.

The generalisation to D-finite and ∂-finite functions is best explained after introducing the notion of *annihilating ideals*: indeed, the skew polynomials (with rational-function coefficients) from some skew-polynomial algebra $A = \mathbb{Q}(x, y, \dots, m, n, \dots)\langle D_x, D_y, \dots, S_m, S_n, \dots\rangle$ that cancel a given function f constitute a left ideal, denoted $\operatorname{ann}_A f$, or more simply $\operatorname{ann} f$ when no ambiguity can arise. When acting on f, skew polynomials can be viewed modulo $\operatorname{ann} f$, since the function Pf obtained for $P \in A$ is equal to any $(P + Z)f$ for $Z \in \operatorname{ann} f$. So the Pf are actually parametrised by classes $P + \operatorname{ann} f$ from the quotient module $A/\operatorname{ann} f$, and there is in fact an isomorphism from $A/\operatorname{ann} f$ to Af defined by mapping the class $P + \operatorname{ann} f$ to Pf.

This leads to new definitions of hypergeometric and hyperexponential functions as functions f whose corresponding quotient modules $A/\operatorname{ann} f$ are vector spaces of dimension 1 over the rational-function field $\mathbb{Q}(x, y, \dots, m, n, \dots)$. The generalised notion of ∂-finite functions corresponds to finite-dimensional quotient modules of dimension possibly more than 1. In the purely differential situation, such a function is simply called *differentiably finite*, in short, *D-finite*. The special case of D-finite series was introduced and studied by Stanley (1980) in the univariate case and by Lipshitz (1989) in the multivariate case; see also Chapter 6 in the textbook (Stanley, 1999). Salvy and I introduced the more general case of ∂-*finite functions* in our work (Chyzak and Salvy, 1998), because ∂ is the symbol we used in that work to denote derivation operators like D_x and shift operators like S_n in a unified way. Each class of functions is closed under additions and products, and algorithms exist to produce a defining system for $f + g$ and fg from defining systems for f and g; see the same references.

Now, a property of any D-finite function f is that all its (infinitely many) mixed partial derivatives f, $D_x f$, $D_y f$, $\dots$, $D_x^2 f$, $D_x D_y f$, $D_y^2 f$, $\dots$, are so much linearly related that they span a finite-dimensional vector space over $\mathbb{Q}(x, y, \dots)$. The generalisation of this property to ∂-finite functions is that all shifts of mixed partial derivatives span a finite-dimensional vector space over $\mathbb{Q}(x, y, \dots, m, n, \dots)$. As a consequence, they also satisfy an ordinary differential equation w.r.t. each derivation operator, and an ordinary difference equation w.r.t. each shift operator.

1.4 q-Analogues

Many functions of the classical world of sequences and functions, like counting numbers, special functions, and orthogonal polynomials, find a generalisation commonly called q-*analogues*. These are functions of an additional parameter q that specialise to the classical functions by setting $q = 1$. A force of creative telescoping is that it can deal with these q-analogues just as it does with the original, classical counterparts. Each of the classes of functions described in Section 1.3 find a q-analogue counterpart. In this section, I briefly introduce the most fundamental examples that will appear later in Section 1.5.

Just as many enumeration problems involve factorials and binomial terms, and thus lead to recurrences that relate values at n with values at $n+1$, $n+2$, etc., using "additive shifts", so do refined enumeration problems in combinatorics, in statistical physics, and in the theory of partitions involve recurrences that relate values at x with values at qx, q^2x, etc., using "multiplicative shifts" w.r.t. some fixed base q. A change of variables to replace x with q^n will bring multiplicative shifts back to additive ones, but if the recurrences under considerations had coefficients that were polynomials in x, the change of variables produces recurrences involving q^n in the coefficients.

The simplest examples of terms that admit a q-analogue are the classical Pochhammer symbol, factorials, and binomial coefficients, which are related as follows: The Pochhammer symbol $(x)_n$ is defined for $x \in \mathbb{C}$ and $n \in \mathbb{Z}$ by

$$(x)_n = \begin{cases} x\,(x+1)\cdots(x+n-1) & \text{if } n > 0, \\ 1 & \text{if } n = 0, \\ 1/\big((x-1)\ \cdots (x+n)\big) & \text{if } n < 0, \end{cases}$$

provided the quotient is well defined. The factorial sequence is then defined by an evaluation as

$$n! = (1)_n;$$

and, binomial coefficients are defined by the well known formula

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

Note that the Pochhammer symbol satisfies the two recurrences

$$(x+1)_n = \frac{x+n}{x}\,(x)_n \quad \text{and} \quad (x)_{n+1} = (x+n)\,(x)_n,$$

from which recurrences can be derived for factorials and binomial coefficients. The q-Pochhammer symbol $(x;q)_n$ is defined for $x \in \mathbb{C}$ and $n \in \mathbb{Z}$ by

$$(x;q)_n = \begin{cases} (1-x)(1-qx)\cdots(1-q^{n-1}x) & \text{if } n > 0, \\ 1 & \text{if } n = 0, \\ \left(1-\dfrac{x}{q}\right)^{-1} \cdots \left(1-\dfrac{x}{q^n}\right)^{-1} & \text{if } n < 0; \end{cases} \tag{1.7}$$

it satisfies the two recurrences

$$(qx;q)_n = \frac{1-q^n x}{1-x}(x;q)_n \quad \text{and} \quad (x;q)_{n+1} = (1-q^n)(x;q)_n.$$

This is not a q-analogue of $(x)_n$, in the sense that setting $q = 1$ does not produce the classical Pochhammer symbol, but the q-factorial defined as $(q;q)_n$ is a q-analogue:

$$\frac{(q;q)_n}{(1-q)^n} = 1 \cdot (1+q) \cdot \cdots \cdot (1+q+\cdots+q^{n-1})$$

goes to $n!$ when q tends to 1. By way of consequence, the q-binomial coefficient defined as

$$\begin{bmatrix} n \\ k \end{bmatrix} = \frac{(q;q)_n}{(q;q)_k\,(q;q)_{n-k}}$$

becomes $\binom{n}{k}$ when q goes to 1. In the tradition, the base q is either a complex number with $|q| < 1$ or an indeterminate, whence the name of *q-series* for objects of the theory.

A nice introduction to q-analogues was written by Askey (1992), including examples of use of q-analogue functions in identities that lift classical identities.

1.5 List of Examples

The primary goal of creative telescoping is the evaluation of integrals and sums involving combinatorial numbers and special functions, especially of hypergeometric/hyperexponential or D-finite/∂-finite type, and the proof of identities involving such sums and integrals. This contains and extends to:

- Binomial sums, as the equality

$$\sum_{k=0}^{n} \binom{n}{k}^2 \binom{n+k}{k}^2 = \sum_{k=0}^{n} \binom{n}{k}\binom{n+k}{k} \sum_{j=0}^{k} \binom{k}{j}^3$$

  between binomial sums, which appears in connection to Apéry's proof of the irrationality of $\zeta(3)$ and has been proved by creative telescoping by Strehl (1994), or the evaluation obtained by Blodgett, Andrews, Paule, and Peck (1990)

$$\sum_{i=0}^{n}\sum_{j=0}^{n} \binom{i+j}{i}^2 \binom{4n-2i-2j}{2n-2i} = (2n+1)\binom{2n}{n}^2;$$

- Integrals of the theory of special functions, like the example of an integral

$$\int_0^{+\infty} x\,J_1(ax)\,I_1(ax)\,Y_0(x)\,K_0(x)\,dx = -\frac{\ln(1-a^4)}{2\pi a^2}$$

  involving the four types of Bessel functions and first considered by Glasser and Montaldi (1994), or the double integral

$$\int_0^{\infty}\int_0^{\infty} J_1(x)\,J_1(y)\,J_2(c\sqrt{xy})\,\frac{dx\,dy}{e^{x+y}}$$

which, if it cannot be put to explicit form, can be proved by creative telescoping to satisfy the second-order linear ODE

$$\begin{aligned}(c^2-8)(c^2+8)(c^2+4c+8)(c^2-4c+8)c^2y''(c)+\\ (5c^8-32c^6-256c^4-2048c^2-4096)cy'(c)+\\ 16(c^6-16c^4-320c^2+1024)y(c)=0;\end{aligned}$$

- Extractions of coefficients, for instance by the Cauchy formula, like with the formula

$$\frac{1}{2\pi i}\oint\frac{(1+2xy+4y^2)\exp\left(\frac{4x^2y^2}{1+4y^2}\right)}{y^{n+1}(1+4y^2)^{\frac{3}{2}}}\,dy=\frac{H_n(x)}{\lfloor n/2\rfloor!},$$

  due to Doetsch (1930) and related to the Hermite orthogonal polynomials;

- Verifying identities in q-sums that appear in the combinatorial theory of partitions, like

$$\sum_{k=0}^{n}\frac{q^{k^2}}{(q;q)_k(q;q)_{n-k}}=\sum_{k=-n}^{n}\frac{(-1)^kq^{(5k^2-k)/2}}{(q;q)_{n-k}(q;q)_{n+k}},$$

$$\sum_{j=0}^{n}\sum_{i=0}^{n-j}\frac{q^{(i+j)^2+j^2}}{(q;q)_{n-i-j}(q;q)_i(q;q)_j}=\sum_{k=-n}^{n}\frac{(-1)^kq^{7/2k^2+1/2k}}{(q;q)_{n+k}(q;q)_{n-k}},$$

  which are finite forms of the Rogers–Ramanujan identities and of a generalisation and were respectively obtained by Andrews (1974) and Paule (1985);

- Computing explicit forms for scalar products w.r.t. various exponential/algebraic weights and in relation to families of orthogonal polynomials or of other parametrised families of functions, like the identities

$$\int_{-1}^{+1}\frac{e^{-px}T_n(x)}{\sqrt{1-x^2}}\,dx=(-1)^n\pi I_n(p),$$

$$\int_0^{+\infty}xe^{-px^2}J_n(bx)I_n(cx)\,dx=\frac{1}{2p}\exp\left(\frac{c^2-b^2}{4p}\right)J_n\left(\frac{bc}{2p}\right),$$

  which involve Chebyshev orthogonal polynomials and Bessel functions;

- Scalar products that appear in the theory of symmetric functions, like

$$\left\langle\exp\left((p_1^2-p_2)/2-p_2^2/4\right)\,\middle|\,\exp\left(t\,(p_1^2+p_2)/2\right)\right\rangle=\frac{e^{-\frac{1}{4}t\,(t+2)}}{\sqrt{1-t}},$$

  where p_1 and p_2 respectively denote the infinite symmetric power sums $x_1+x_2+\cdots$ and $x_1^2+x_2^2+\cdots$ and for the scalar product induced by the formula $\langle m_\lambda\,|\,h_\mu\rangle=\delta_{\lambda,\mu}$, in terms of the other classical monomial and homogeneous bases $(m_\lambda)_\lambda$ and $(h_\mu)_\mu$.

In all previous examples, the sequences and functions under consideration possess as many independent linear equations, whether differential, difference, or more general functional, as their number of variables. That is to say, they can be described as ∂-finite functions, by a set of linear functional equations and *finitely-many* initial conditions. A recent extension of the approach (Chyzak, Kauers, and Salvy, 2009) allows to deal as well with functions that possess fewer independent equations than variables. Given a linear functional system of this type, the description in explicit form of the general solutions necessarily involves an *arbitrary function* of at least one variable. Examples of applications of this extension include:

- Combinatorial identities involving: the graph-counting sequence k^{k-1}, like

$$\sum_{k=0}^{n} \binom{n}{k} i\,(k+i)^{k-1} (n-k+j)^{n-k} = (n+i+j)^n,$$

which is attributed to Abel; or Stirling numbers of the second kind and Eulerian numbers, like

$$\sum_{k=0}^{n} (-1)^{m-k} k! \binom{n-k}{m-k} \left\{ {n+1 \atop k+1} \right\} = \left\langle {n \atop m} \right\rangle,$$

attributed to Frobenius; or Bernoulli numbers, like

$$\sum_{k=0}^{m} \binom{m}{k} B_{n+k} = (-1)^{m+n} \sum_{k=0}^{n} \binom{n}{k} B_{m+k}$$

to be found in (Gessel, 2003);

- Identities in more special functions, like Hurwitz's zeta function, the beta function, polylogarithms, and the (upper) incomplete Gamma function, which appear in the following evaluations:

$$\int_0^\infty x^{k-1}\,\zeta(n, \alpha+\beta x)\,dx = \beta^{-k} B(k, n-k)\,\zeta(n-k, \alpha),$$

$$\int_0^\infty x^{\alpha-1}\,\mathrm{Li}_n(-xy)\,dx = \frac{\pi\,(-\alpha)^n y^{-\alpha}}{\sin(\alpha\pi)},$$

$$\int_0^\infty x^{s-1} \exp(xy)\,\Gamma(a, xy)\,dx = \frac{\pi y^{-s}}{\sin\big((a+s)\,\pi\big)} \frac{\Gamma(s)}{\Gamma(1-a)}.$$

Let us also mention that a very great deal of identities amenable to creative telescoping can be found in the series of books (Prudnikov, Brychkov, and Marichev, 1992) on identities involving elementary and special functions, and on Laplace transforms.

Different kinds of identities do not provide closed-form evaluations but are established by creative telescoping as a crucial step in the proof of a mathematical result. I refer readers to Koutschan's recent survey (2013) for a list.

1.6 Notation

When discussing bounds, whether on various mathematical parameters (degrees, number of terms, ...) or on the complexity of algorithms, I shall most often consider asymptotic upper bounds, which will be denoted using the big-O notation: for example, $u_n \in O(n^3)$ means that the sequence $u = (u_n)$ does not grow faster than a constant times the cubic function when n goes to infinity. When stating an upper bound by a big-O notation would involve logarithmic factors that are inessential to understanding the described phenomenon, I shall use the soft-O notation: for example, $u_n \in \tilde{O}(n)$ instead of $u_n \in O(n \log^2 n)$. On the other hand, I shall at times use the big-Θ notation to denote that a sequence grows in proportion to another: for example, $v_n \in \Theta(n^4)$ means that v_n is asymptotically equivalent to κn^4 for some fixed non-zero κ, as n goes to infinity. This stronger notion is crucial in two cases: to express that an upper bound is tight and to express that, asymptotically, a sequence becomes strictly greater than another.

As is usual in combinatorics, for a non-negative integer ℓ, the *falling factorial* $n^{\underline{\ell}}$ denotes the polynomial $n\,(n-1)\cdots(n-\ell+1)$.

Chapter 2

Early History of Creative Telescoping

Before the algorithmic results in the next chapters, I provide a brief history of the research that led to creative telescoping. My goal here is to highlight the flow of ideas, while providing context and motivation to some of the creative-telescoping techniques and to my research orientations.

2.1 From Zeilberger's Early Attempt to his "Holonomic-Systems Approach"

Zeilberger (1982) made the first attempt in the literature at giving generality to Cohen and Zagier's derivation (1.2)–(1.4) of Apéry's recurrence, by exploiting a technique of Fasenmyer (1945, 1949). Although this work of Zeilberger's makes good observations that have been used in later literature, it is flawed in several ways that make its main claims wrong. On the positive side, Zeilberger's observation is that, given a hypergeometric summand $h_{n,k}$, his variant of Fasenmyer's technique provides, if it succeeds, a relation of the form

$$\sum_{i=0}^{r}\sum_{j=0}^{s} c_{i,j}(n)\, h_{n+i,k+j} = 0 \tag{2.1}$$

from which an equation playing the role of (1.3) can be derived. As this equation does not involve the variable k intended for summation, it is called a k-*free recurrence*. In fact, Fasenmyer elaborated her original technique for hypergeometric series described in the form $h_n(x) = \sum_k h_{n,k} x^k$. She used an analogue of (2.1) that involves shifts in n only, together with powers of x:

$$\sum_{i=0}^{r}\sum_{j=0}^{s} c_{i,j}(n)\, x^j h_{n+i}(x) = 0. \tag{2.2}$$

Fasenmyer's technique was later described by Rainville (1960), who also slightly generalised it to sums of the form $h_n(x) = \sum_k h_{n,k} T_k(x)$, involving the kth Chebyshev polynomial $T_k(x)$. It was also used by Verbaeten, a name that will appear again in Section 3.3, for the quadrature of an integral problem parametrised by a Chebyshev series (Piessens and Verbaeten, 1973). In his paper, Zeilberger additionally observes that the approach generalises to sums of q-analogues and integrals of functions satisfying systems of first-order equations, and to all possible cases mixing these forms of operators.

It took Zeilberger a few more years before developing his seminal paper "A holonomic systems approach to special functions identities" (1990b), in which he introduced the proper definitions to prevent degenerate cases from occurring and ensure the existence of non-trivial (that is, non-zero) equations (1.3) and (2.1). This important paper bases on results of the theory of *holonomic D-modules* that had been developed in the 1970s (Bernšteĭn, 1971, 1972; Kashiwara, 1978). What is decisive here is to consider the set of *all* linear differential/difference equations that a given summand or integrand satisfies. Viewing these equations as linear differential/difference operators results in (non-commutative) polynomials that constitute a (left) ideal in a non-commutative polynomial ring $A_p = \mathbb{Q}[x, y, \dots, m, n, \dots]\langle D_x, D_y, \dots, S_m, S_n, \dots\rangle$ (in finitely many generators). This representation is now amenable to a theory of polynomial elimination, which was studied in depth in D-module theory.

Conceptually, however, a subtle distinction has to be done between the ideal of operators with rational-function coefficients in $A_r = \mathbb{Q}(x, y, \dots, m, n, \dots)\langle D_x, D_y, \dots, S_m, S_n, \dots\rangle$ and the ideal of operators with polynomial coefficients in A_p: the polynomial-elimination theory takes place in A_p, not in A_r. However, while the ideal of all operators in A_r that annihilate a given function is usually easily presented by finitely many explicit generators in applications involving special functions, the related ideal in A_p is not so easily described explicitly; in particular, generators in A_r cannot be used directly in A_p, even after renormalisation to remove denominators. A simple illustration of the problem is given by the polynomial $f = x^3$ in the ordinary differential case. It is annihilated by $xD_x - 3$, and any annihilator in A_r is a left multiple of the form $L(xD_x - 3)$, with L from A_r. But just restricting the cofactor L to A_p does not generate all annihilators from A_p: for example, D_x^4 cancels f, but D_x^4 factors as $(x^{-1}D_x^3)(xD_x - 3)$, requiring a denominator x in the cofactor. This phenomenon is part of the cause for the problems in (Zeilberger, 1982), and was not completely clarified even in (Zeilberger, 1990b). In the differential case, algorithms for obtaining generators of the ideal with polynomial coefficients from generators of the ideal with rational coefficients were given by Tsai, first in the ordinary differential case (2000), then in the partial differential case (2002). This process was named *Weyl closure*. On the other hand, no "Ore closure" algorithm is known yet for other types of operators in the multivariate case.

2.2 Other Early Elimination Approaches: Constant Terms and Diagonals

It is worth noting that the polynomial notation for operators had already been used for combinatorial matters and in connection to a polynomial elimination problem. Let me mention two works.

First, Zeilberger had already studied in (1980) means to derive difference equations satisfied by the constant term of products of powers with symbolic exponents of multivariate Laurent polynomials. An example (simple, but of pedagogical nature) is the constant term w.r.t. x_1, x_2, and x_3 and viewed as a function of a, b, and c, of

$$F(a, b, c, x_1, x_2, x_3) = \left(\left(1 - \frac{x_1}{x_2}\right)\left(1 - \frac{x_2}{x_1}\right)\right)^a \left(\left(1 - \frac{x_1}{x_3}\right)\left(1 - \frac{x_3}{x_1}\right)\right)^b \left(\left(1 - \frac{x_2}{x_3}\right)\left(1 - \frac{x_3}{x_2}\right)\right)^c,$$

which turns out to be

$$\frac{(2a)!\,(2b)!\,(2c)!\,(a+b+c)!}{a!\,b!\,c!\,(a+b)!\,(a+c)!\,(b+c)!}. \tag{2.3}$$

Zeilberger's approach was to consider the two-term first-order difference equation

$$F(a+1, b, c, x_1, x_2, x_3) = \left(1 - \frac{x_1}{x_2}\right)\left(1 - \frac{x_2}{x_1}\right) F(a, b, c, x_1, x_2, x_3),$$

or rather its normalised polynomial representation $x_1x_2S_a + (x_1 - x_2)^2$, together with its siblings obtained by shifting b or c instead of a. Then, eliminating x_1, x_2, and x_3 by successive runs of the fraction-free Euclidean algorithm in $\mathbb{Q}[S_a, S_b, S_c][x_1, x_2, x_3]$ yields the annihilator

$$S_aS_bS_c + S_a^2 + S_b^2 + S_c^2 - 2(S_aS_b + S_bS_c + S_aS_c)$$

of the constant term, from which *checking* that (2.3) is the constant term is easy. It is of interest that Zeilberger observed that the classical elimination theory for commutative polynomials can be applied to the case of partial difference operators with coefficients independent of the variables (a, b, and c above), but dependent of extra parameters (x_1, x_2, and x_3 above). In contrast, algorithms by creative telescoping would represent the constant term by the Cauchy integral

$$\frac{1}{(2\pi i)^3} \oint\oint\oint F(a, b, c, x_1, x_2, x_3)\, \frac{dx_1}{x_1} \frac{dx_2}{x_2} \frac{dx_3}{x_3},$$

then consider as well equations that are differential in x_1, x_2, and x_3, and perform an algorithmic elimination of x_1, x_2, and x_3 in $\mathbb{Q}(a, b, c)[x_1, x_2, x_3]\langle S_a, S_b, S_c, D_1, D_2, D_3\rangle$ (where D_i denotes derivation w.r.t. x_i). The approach of the present section prefigures Zeilberger's "slow algorithm" discussed in the next chapter.

A second work is the proof by Lipshitz (1988) that the diagonal of a D-finite power series is D-finite. These notions require definitions: Given a multivariate formal power series

$$f = \sum_{(n_1,\dots,n_r)\in\mathbb{N}^r} c_{n_1,\dots,n_r} x_1^{n_1} \cdots x_r^{n_r} \in \mathbb{Q}[[x_1, \dots, x_r]],$$

its *diagonal* is defined as the univariate series

$$\Delta f = \sum_{n\in\mathbb{N}} c_{n,\dots,n} x^n.$$

The case studied by Lipshitz is that of a *differentiably finite* series, in short *D-finite* series, that is, of a series f whose derivatives $D_1^{n_1} \cdots D_r^{n_r} f$ at all order $(n_1, \dots, n_r \geq 0)$ generate a finite-dimensional vector space over $\mathbb{Q}(x_1, \dots, x_r)$. (The power-series ring $\mathbb{Q}[[x_1, \dots, x_r]]$ can be embedded into a $\mathbb{Q}(x_1, \dots, x_r)$-vector space.) An equivalent definition is that such an f possesses for each i between 1 and r an ordinary non-zero annihilator L_i from $\mathbb{Q}(x_1, \dots, x_r)\langle D_i\rangle$. Lipshitz's approach does not appeal to any tool of D-modules theory, and remains on a very elementary level, but it bases on a counting argument that is at the heart of Bernstein's dimension theory for D-modules.

The derivation is as follows, after specialising to $r = 2$ for the sake of simplicity: The first idea is to express the diagonal as a residue of a suitable transform of f, namely by

$$\Delta f = \operatorname{res}_s g \quad \text{where} \quad g = \frac{1}{s} f\left(s, \frac{x}{s}\right).$$

(Here, the *residue* $\operatorname{res}_s \phi$ of a function ϕ that can be expressed for some $m \in \mathbb{Z}$ as the sum $\sum_{(p,q)\in\mathbb{Z}^2,\ p+q\geq m} \phi_{p,q} s^p x^q$ is defined as the univariate series $\sum_{q=m+1}^{\infty} \phi_{-1,q} x^q$.) Note that the diagonal could be expressed as a constant term or a Cauchy integral, as well. Then, Lipshitz introduces the ordinary annihilators L_i associated with g and he proves that s can be eliminated from the family of the L_i's. To this end, he considers the expressions

$$x^m D_s^n D_x^o g \quad \text{subject to} \quad m + n + o \leq N. \tag{2.4}$$

Then, he determines suitable integers a, b, and $h \geq 1$, all three independent of N, and a suitable polynomial p to show that the expressions (2.4) all rewrite as linear combinations of terms of the form $(q/p^N) D_s^i D_x^j g$, where $0 \leq i < a$, $0 \leq j < b$, and $q(s, x)$ is a polynomial of total degree bounded by Nh. Now, the number of initial expressions is a multinomial number, growing in proportion to N^3, while the simplified expressions (2.4) live in a vector space of dimension $O(N^2)$ over $\mathbb{Q}$. Therefore, for large enough N there must be a $\mathbb{Q}$-linear combination $Z(x, D_x, D_s)$ of the $x^m D_s^n D_x^o$'s in (2.4) that rewrites to 0 (as the rewriting is a linear map). He finally extracts the coefficient $Z'(x, D_x)$ of Z of lowest exponent w.r.t. D_s and proves that Z' cancels the diagonal.

Chapter 3

Creative-Telescoping Algorithms for Equations of the First Order

3.1 Zeilberger's Fast Algorithm for Hypergeometric Sums and its Variants

Zeilberger made explicit with (1990b) what was implied by the earlier works: designing algorithms based on the creative-telescoping approach for operations on integrals and series requires algorithmic means for non-commutative polynomial elimination in operator algebras. Earlier, Zeilberger (1980) had appealed to successive gcd computations by a sort of fraction-free Euclidean algorithm, which is likely to introduce spurious factors in the coefficients. The output (2.1) from Zeilberger's modification (1982) of Fasenmyer's technique is essentially an operator from $\mathbb{Q}(n)\langle S_n, S_k\rangle$, and could be obtained by an elimination of k from annihilators of $h_{n,k}$, provided the Ore closure problem was solved algorithmically. A searching algorithm by linear algebra is implied by Lipshitz's proof (1988). For his part, Zeilberger (1990b) based on a classical process to compute Sylvester's resultant (1840), which he names "Sylvester's dialytic elimination". This method was originally designed for univariate polynomials, but, despite non-commutativity, Zeilberger generalises it to an elimination method for operators viewed as bivariate polynomials in n and k with coefficients in $\mathbb{Q}[S_n, S_k]$. However, he dodges the question of Ore closure, which, in practice, becomes a large weakness of the method: one does not know if it will terminate if one starts with indiscriminate annihilators.

In addition, all these methods are merely existence proofs turned into algorithms and tend to be very slow in practice. Although no sufficient study of lower bounds for their complexity is available yet, an explanation is that elimination constrains the output to be a polynomial in high degrees, and in a form like (2.1) that is more restrictive than what is minimally required for creative telescoping to work. Indeed, by way of comparison, representing Cohen and Zagier's telescoping relation (1.3) as an operator results in an operator that involves k. Inspect the last term in the left-hand side of the following identity and how it is transformed by the commutation rule:

$$\begin{aligned}(n+1)^3 S_n - (34n^3+51n^2+27n+5) + n^3 S_n^{-1} - \big(1-S_k^{-1}\big)4(2n+1)\big(k(2k+1)-(2n+1)^2\big) = \\ (n+1)^3 S_n - (34n^3+51n^2+27n+5) + n^3 S_n^{-1} - 4(2n+1)\big(k(2k+1)-(2n+1)^2\big) \\ + 4(2n+1)\big((k-1)(2k-1)-(2n+1)^2\big)S_k^{-1}.\end{aligned}$$

A further consequence is that the order of the outputs from these methods cannot be expected to be minimal: when searching a fixed set of annihilators for elements of the form $P(n, S_n) + (S_k-1)Q(n,k,S_n,S_k)$, the less constrained Q is, the more pairs (P, Q) will exist, and the lower the minimal order of a possible output P is.

In view of this, Zeilberger called his algorithm in (1990b) his "slow algorithm", and turned his attention back to the more restricted class of inputs he had studied in (1982): the special class of hypergeometric sequences, that is sequences $(h_{n,k})$ for which the two ratios

$$\frac{h_{n+1,k}}{h_{n,k}} \quad \text{and} \quad \frac{h_{n,k+1}}{h_{n,k}}$$

are given by two fixed rational functions in n and k. For them, he designed a "fast algorithm" (Zeilberger, 1990a, 1991), which finally popularised the method of creative telescoping as an algorithm. Zeilberger based his fast approach for definite summation on Gosper's *decision algorithm* for the indefinite summation of hypergeometric sequences (1978). Given a (univariate) hypergeometric sequence (u_k), Gosper's interest is to find an *indefinite sum* (U_k) of (u_k), that is, a sequence satisfying $U_{k+1} - U_k = u_k$. He observed that any *hypergeometric* indefinite sum (U_k) must be a multiple of the summand by a fixed rational function R of k: necessarily,

$$U_k = R(k)\, u_k. \tag{3.1}$$

This leads to an auxiliary linear recurrence equation on R, for which he developed a decision procedure for solving. As a consequence, Gosper obtained an algorithm that *decides* whether a given hypergeometric sequence possesses a hypergeometric indefinite sum, or whether its sum is a non-trivial extension. Zeilberger realised that if the telescoper output from creative telescoping was known beforehand, like (1.4) in the example about $\zeta(3)$, then the certificate (1.2) could be obtained just by calling Gosper's algorithm on the right-hand side of (1.3). To make this into an algorithm, it was sufficient for Zeilberger to describe how to search at the same time for the coefficients of the output and for the rational function implied by Gosper's algorithm: this amounts to a parametrised variant of Gosper's calculation. Therefore, Zeilberger's algorithm proceeds by increasing a tentative order r of maximal shifts w.r.t. n for the right-hand side of (1.3), trying to solve for R in

$$\eta_r(n)\, h_{n+r,k} + \cdots + \eta_0(n)\, h_{n,k} = R(k+1)\, h_{n,k+1} - R(k)\, h_{n,k} \tag{3.2}$$

at each order. If there exists a non-zero family $\{\eta_i(n)\}_{i=0}^r$ of univariate rational functions and a bivariate rational function $R(n,k)$ such that (3.2) holds, then the algorithm terminates. If so, it obviously produces such an identity with least possible order r.

At about the same time, Almkvist and Zeilberger (1990) gave a differential analogue of Zeilberger's fast algorithm, which applies to functions called *hyperexponential functions*, that is, functions h of two variables x and y for which the two ratios

$$\frac{\frac{dh}{dx}(x,y)}{h(x,y)} \quad \text{and} \quad \frac{\frac{dh}{dy}(x,y)}{h(x,y)}$$

are two fixed rational functions in x and y. To this end, they produced a differential analogue of Gosper's algorithm and replaced (3.2) with an ansatz of the form

$$\begin{aligned} \eta_r(x)\, D_x^r h(x,y) + \cdots + \eta_0(x)\, h(x,y) = \\ D_y\ (R(x,y)\, h(x,y)) = \big(D_y R(x,y)\big)\, h(x,y) + R(x,y)\, \big(D_y h(x,y)\big). \end{aligned} \tag{3.3}$$

By the mid-1990s, Zeilberger had done a great job in popularising his theory, his fast algorithm, and his Maple implementation of it. A great deal of application papers were published, by he, admirers, and more often than not his computer whom he had named Shalosh B. Ekhad. The goal was to demonstrate that hypergeometric summation had become *routine*. To list a few of such papers: (Zeilberger, 1994; Ekhad and Zeilberger, 1994b,a, 1996; Prodinger, 1996). In addition, Petkovšek, Wilf, and Zeilberger coauthored a whole book (1996) to popularise algorithms for hypergeometric summation. Still, there was some confusion around Zeilberger's articles, caused by some allusiveness in the presentation and in algorithmic descriptions, especially in view of the many generalisations that were announced but not formalised rigorously.

This motivated Koornwinder to work on a new Maple implementation of Zeilberger's fast algorithm as well as on a q-analogue; he wrote (1993) with the purpose to describe them in a very rigorous way, to ensure that the outputs produced could be trusted. In the same vein, Paule and Schorn (1995) provided a Mathematica implementation (for the classical case), with special emphasis on speed and robustness The similar work for the q-analogue algorithm was done by Paule and Riese (1997); it was continued into a work on indefinite bibasic hypergeometric summation (Riese, 1996), that is, for identities involving an operator B such that

$$(Bf)(x, y) = f(qx, py).$$

(Here, "bibasic" refers to p and q playing the role of two bases, to be compare with the single base q for usual q-analogues.) For a Maple counterpart, (Böing and Koepf, 1999) describes an analogue implementation of Zeilberger's q-analogue fast algorithm and of Riese's bibasic Gosper algorithm.

3.2 Wilf and Zeilberger's Approach to Multiple Sums and Integrals

After single hypergeometric/hyperexponential sums/integrals, there remained to understand on what inputs the method would terminate with certainty and if it could be extended to multiple sums and integrals. This was addressed to some extent by Wilf and Zeilberger (1992a); see also the result announcement in (Wilf and Zeilberger, 1992b). There, they introduced the notion of a *proper hypergeometric term*, a special kind of hypergeometric term given as

$$h_{n,k} = P(n, k)\, \zeta^n \xi^k \prod_{\ell=1}^{L} \Gamma(a_\ell n + b_\ell k + c_\ell)^{\epsilon_\ell}, \tag{3.4}$$

where:

1. the a_ℓ's and b_ℓ's are specific *integers*, and the ϵ_ℓ's are ± 1;
2. the c_ℓ's, ζ, and ξ are *constants* independent from n and k;
3. P is a *polynomial* in n and k.

(The original definition separates the factors with $\epsilon = +1$ from those with $\epsilon = -1$ and insists on the c_ℓ's being integers, but I prefer this more formal view for what follows. Also, the term ζ^n was really introduced by Wegschaider (1997) only, but it alters what follows in no essential way.) As a general hypergeometric term could be represented formally in the same way, but with a rational function in place of the polynomial P, the wording *proper* emphasises that the factor P is polynomial. As we shall see, the crucial consequence of the definition (3.4) is the behaviour of its components under shifts: a term $\Gamma(an + bk + c)$ is multiplied by a rational function of degree in k linear in the total number of unit shifts; the degree in k of P is unchanged under shifts; the exponentials ζ^n and ξ^k are multiplied by constants.

A first contribution of Wilf and Zeilberger's is to show the existence of a non-trivial k-free relation of the form (2.1) for any proper hypergeometric h, together with explicit bounds r and s in (2.1) to ensure existence:

$$r = B \quad \text{and} \quad s = (A - 1)\, B + \deg_k(P) + 1, \tag{3.5}$$

where they could set

$$A = \sum_{\ell=1}^{L} |a_\ell| \quad \text{and} \quad B = \sum_{\ell=1}^{L} |b_\ell|.$$

The proof can be sketched as follows: Each $h_{n+i,k+j}$ in (2.1) involves terms of the form $\Gamma(a_\ell n + b_\ell k + c_\ell + u)$ for a shift u bounded in absolute value by $\sigma_\ell = |a_\ell| r + |b_\ell| s$, that is, in a linear way.

Now, observe that both

$$\frac{\Gamma(a_\ell n + b_\ell k + c_\ell + u)}{\Gamma(a_\ell n + b_\ell k + c_\ell - \sigma_\ell)} \quad \text{and} \quad \frac{\Gamma(a_\ell n + b_\ell k + c_\ell + \sigma_\ell)}{\Gamma(a_\ell n + b_\ell k + c_\ell + u)} \tag{3.6}$$

are polynomials in k of degree at most $2\sigma_\ell$; the former will be used if $\epsilon_\ell = +1$, the latter if $\epsilon_\ell = -1$. Therefore, multiplying the k-free ansatz (2.1) by the term

$$H_{n,k} = \zeta^{-n}\xi^{-k}\left(\prod_{\ell=1,\ \epsilon_\ell=+1}^{L} \Gamma(a_\ell n + b_\ell k + c_\ell - \sigma_\ell)\right)^{-1}\left(\prod_{\ell=1,\ \epsilon_\ell=-1}^{L} \Gamma(a_\ell n + b_\ell k + c_\ell + \sigma_\ell)\right) \tag{3.7}$$

results in an equivalent *polynomial* relation

$$\sum_{i=0}^{r}\sum_{j=0}^{s} c_{i,j}(n)\,(H_{n,k}h_{n+i,k+j}) = 0, \tag{3.8}$$

where each product $H_{n,k}h_{n+i,k+j}$ is a polynomial of degree in k not more than a linear function of r and s. The full analysis gives the degree bound

$$\deg_k\bigl(H_{n,k}h_{n+i,k+j}\bigr) \le \deg_k(P) + \sum_{\ell=1}^{L}\sigma_\ell = \deg_k(P) + Ar + Bs. \tag{3.9}$$

Therefore, the left-hand side of (3.8) is a polynomial in k of degree $O(r+s)$ while it is the combination of $\Theta\bigl((r+s)^2\bigr)$ non-zero polynomials. For r and s large enough, there must be a non-trivial relation (3.8). More specifically, it is sufficient to ensure

$$(r+1)(s+1) > \deg_k(P) + Ar + Bs + 1, \tag{3.10}$$

and choosing $r = B$ immediately results in (3.5).

Of course, the formula should not be used blindly in an implementation. For instance, applying it to the simple binomial term $\binom{n+k}{k}^2\binom{n}{k}^2$ results in $A = 4$, $B = 8$, $r = 8$, $s = 25$, and one has to expand products of 108 terms of the form $(n \pm k + u)^2$ or $(k+u)^2$, leading to a linear-algebra system in dimension 217, involving integers greater than $3 \cdot 10^{234}$!

The second main contribution from (Wilf and Zeilberger, 1992a) is to show that the whole work extends to multiple sums and integrals, providing an algorithm to compute an analogue of (2.1) when several k's are involved. To this end, the notion of proper hypergeometric term extends to terms depending on several k's in a natural way:

$$h_{n,k_1,\dots,k_m} = P(n,k_1,\dots,k_m)\,\zeta^n\xi_1^{k_1}\cdots\xi_m^{k_m}\prod_{\ell=1}^{L}\Gamma(a_\ell n + b_{\ell,1}k_1 + \cdots + b_{\ell,m}k_m + c_\ell)^{\epsilon_\ell},$$

with the obvious generalisation of the constraints on (3.4). In case of an m-fold sum, (2.1) takes the form

$$\sum_{i=0}^{r}\sum_{j_1=0}^{s_1}\cdots\sum_{j_m=0}^{s_m} c_{i,j_1,\dots,j_m}(n)\,h_{n+i,k_1+j_1,\dots,k_m+j_m} = 0. \tag{3.11}$$

The quantity A is defined as above while a sum B_i is associated with each k_i. Inequality (3.10) is transformed by considering total degrees w.r.t. $k_1,\dots,k_m$, so as to compare the number of c's in the $(m+1)$-dimensional sum (3.11) with the m-dimensional combinatorics of monomials in $k_1,\dots,k_m$ of bounded total degree:

$$(r+1)(s_1+1)\cdots(s_m+1) > \binom{\deg_{k_i}(P) + Ar + B_1s_1 + \cdots + B_ms_m + m}{m} + 1.$$

Choosing for instance $r = s_1 = \cdots = s_m$ and comparing exponents shows the existence of a solution; Wilf and Zeilberger (1992a) give a formula, but it is much less explicit than in the case $m = 1$.

Wilf and Zeilberger's method also has a q-analogue, for simple and multiple sums. They adapted their definition to call proper q-hypergeometric a term of the form:

$$h_{n,k} = P(q^n, q^k)\, \zeta^n \xi^k q^{\alpha n^2+\beta nk+\gamma k^2+\lambda\binom{n}{2}+\mu\binom{k}{2}} \prod_{\ell=1}^{L} \left((q; c_\ell)_{a_\ell n+b_\ell k}\right)^{\epsilon_\ell}, \tag{3.12}$$

where $(q; x)_N$ denotes the q-Pochhammer symbol, defined by (1.7), and, in addition to the constraints of the classical case:

1. the constants c_ℓ's, ζ, and ξ, as well as the coefficients of P may now be rational functions of q;
2. α, β, γ, λ, and μ are all relative integers.

(Again, I have used here the generalised form by Riese (2003), with no essential change in what follows.) As for the classical case, Wilf and Zeilberger (1992a) (resp. Riese (2003)) showed that a non-trivial relation

$$\sum_{i=0}^{r} \sum_{j=0}^{s} c_{i,j}(q^n) h_{n+i,k+j} = 0 \tag{3.13}$$

always exist. Here, an additional difficulty over the classical case is that the analogues of the factors (3.6) are no longer polynomials, but in general Laurent polynomials in q^n and q^k; this adds a lot of technicalities. This all generalises to the case of multiple q-sums. Furthermore, under the assumption $m = P = \zeta = 1$ (single q-sums), Wilf and Zeilberger gave the bound

$$|\gamma| + \sum_{\ell=1}^{L} b_\ell^2 \tag{3.14}$$

on r for a relation (3.13) to exist.

Another analogue of Wilf and Zeilberger's approach was developed by Tefera (2000, 2002) for the case of multiple integrals of functions that are (essentially) proper hypergeometric w.r.t. one discrete variable n and hyperexponential w.r.t. several continuous variables $x_1, \dots, x_u$. Written in the case of a single continuous variable, the corresponding "proper terms" are of the form

$$h_n(x) = P(n, x) e^{\zeta_0(x)} \zeta_1(x)^n \zeta_2(x)^d \prod_{\ell=1}^{L} \Gamma(a_\ell n + b_\ell k + c_\ell)^{\epsilon_\ell},$$

where P is again a polynomial, and the ζ's are now univariate rational functions, the c's and d are constant w.r.t. both n and x, and the a's, b's and ϵ's are like before.

3.3 Verbaeten's Completion and Non-k-Free Recurrences

Practical experimentation with the WZ-method soon revealed two shortcomings in formula (2.1), both causing the need for too high upper bounds r and s of the double sum, and too high running times in implementations.

Firstly, the support of the double sum being a rectangular box, as opposed to a more flexible set of pairs (i, j), can become a non-intrinsic difficulty when solving, as it can unnecessarily involve too many terms. A solution had already been studied in the similar context of obtaining what Verbaeten called called "pure recurrence relation" (2.2) in his PhD thesis (1974; 1976), years before the works on Fasenmyer's technique by Zeilberger (1982) and Wilf and Zeilberger (1992a). The idea consists in enlarging the support of the sum (2.1) insofar as this does not increase the degree in k of (3.8), a process that is called *Verbaeten's completion*. This results in transforming (2.1) into the form

$$\sum_{(i,j)\in S} c_{i,j}(n)\, h_{n+i,k+j} = 0 \tag{3.15}$$

for some finite $S \subset \mathbb{N}^2$ that is not necessarily of the form $[0,r]\times[0,s]$. Incidentally, by considering special maximal sets of pairs, Verbaeten had already obtained an existence proof for a relation (2.1), for a special case of proper hypergeometric terms, later called *irreducible*: terms such that P in (3.4) is 1 and no two factors $\Gamma(an+bk+c)$ with opposite ϵ's have the same a's and b's, and c's that differ by an integer. Verbaeten's proof was later greatly simplified by Hornegger (1992) and a sketch of it is available in (Wegschaider, 1997). It involves a very fine analysis of the degree of polynomials in an equivalent of (3.8) where the rectangular support of the double sum is replaced with the interior some convex polygon. The approach proceeds by counting the points on the integer-lattice that lie in a convex polygon defined by extremal directions related to the a_ℓ's and b_ℓ's. In addition, by thoroughly studying Verbaeten's completion, Wegschaider was able in his master's thesis (1997) to fill a gap in the proof in (Wilf and Zeilberger, 1992a) of the existence theorem of a recurrence in n for the sum $\sum_k h_{n,k}$ in the case of a proper hypergeometric term that is not necessarily irreducible.

Secondly, the way (2.1) is used to get a recurrence on the sum over k suggests that banning k from the coefficients $c_{i,j}$ in an absolute way is not optimal, as is best explained by observing how the recurrence (2.1) on the term $h_{n,k}$ is transformed into a recurrence

$$\sum_{i=0}^{\rho} a_i(n)\, h_{n+i} = 0 \quad \text{for} \quad h_n = \sum_{k=\alpha}^{\beta} h_{n,k}. \tag{3.16}$$

This transformation proceeds by rewriting (2.1) by the relations

$$h_{n+i,k+j} = h_{n+i,k} - \left(h^{(j)}_{n+i,k} - h^{(j)}_{n+i,k+1}\right) \quad \text{for} \quad h^{(j)}_{n+i,k} = h_{n+i,k} + \cdots + h_{n+i,k+j-1}, \tag{3.17}$$

which results in a term $g_{n,k}$ satisfying

$$\sum_{i=0}^{r} c_i(n)\, h_{n+i,k} = g_{n,k+1} - g_{n,k} \quad \text{for} \quad c_i(n) = \sum_{j=0}^{s} c_{i,j}(n). \tag{3.18}$$

Now, summation over k yields

$$\sum_{i=0}^{r} c_i(n)\, h_{n+i} = g_{n,\beta+1} - g_{n,\alpha}. \tag{3.19}$$

In applications, either the right-hand side is zero by itself, or it can be cancelled by applying a linear recurrence operator. In the former case, the output recurrence (3.16) is just (3.19), with $\rho = r$ and $a_i = c_i$ for each i. In the latter case, applying the proper operator to both sides of (3.19) results in a new recurrence (of order ρ greater than r), satisfied by (h_n). In the derivation above, the term $g_{n,k}$ is of the form $Lf_{n,k}$ for $L \in \mathbb{Q}[n]\langle S_n, S_k\rangle$, but this limitation on g is inessential for the derivation. In particular, allowing L to be in the larger set $\mathbb{Q}[n,k]\langle S_n, S_k\rangle$ should allow "more" g's to be tested, and "more" relations (3.19) to be found, with the hope of lower orders for the final recurrences. Wilf and Zeilberger (1992a) attribute this observation to

Gerdt Almkvist. Wegschaider (1997, Sec. 3.5.1) developed a heuristic to allow L to involve k. To this end, he considers the equation

$$\sum_{i=0}^{r}\sum_{j=0}^{s}\sum_{\ell=0}^{t} c_{i,j,\ell}(n)\, k^\ell h_{n+i,k+j} = 0 \tag{3.20}$$

in place of equation (2.1) and the following analogue for the transformations (3.17):

$$k^\ell h_{n+i,k+j} = (k-j)^\ell h_{n+i,k} - \left(h^{(j,\ell)}_{n+i,k} - h^{(j,\ell)}_{n+i,k+1}\right)$$
$$\text{for}\quad h^{(j)}_{n+i,k} = (k-j)^\ell h_{n+i,k} + \cdots + (k-1)^\ell h_{n+i,k+j-1}.$$

Using this in (3.20) before any calculations involving the actual value of h results in linear constraints on the c's for the ansatz to lead to a recurrence of the form (3.19), best expressed by the $(r+1)\times t$ linear constraints over $\mathbb{Q}(n)$ that each polynomial

$$\sum_{j=0}^{s}\sum_{\ell=0}^{t} c_{i,j,\ell}(n)\,(k-j)^\ell, \quad \text{for} \quad 0 \le i \le r,$$

of potential degree t in k, should actually not involve k. As no study is available of what a suitable degree t should be, Wegschaider's implementation lets the user heuristically input a value for it; his manuscript shows tremendous speed-ups by using this improvement.

The case of q-sums is amenable to the same two improvements, Verbaeten's completion and the reintroduction of the summation variable in the ansatz. This has been worked out thoroughly by Riese (2003), both theoretically and in his implementation.

3.4 From Elimination to Equations on Sums and Integrals

In the recurrence case, creative telescoping crucially relies on transforming (2.1) to (3.18), and on an analogue transformation in the differential case. In operator notation, this transformation derives from an operator L free of the summation and integration variables an operator P that involves exclusively the parameters of the sum/integral and the corresponding shifts and derivatives. A cause of concern is that nothing guarantees a priori that (3.18) or its analogue does not exhibit a zero left-hand side, which would make (3.19) provide no information on (h_n). As a matter of fact, the reader will only find handwaving in (Zeilberger, 1990b; Almkvist and Zeilberger, 1990), and no proof attempt at all in either of (Zeilberger, 1991; Wilf and Zeilberger, 1992a).

To the best of my knowledge, this problem was fixed first by Wegschaider (1997) in the case of recurrences. A similar idea works in the differential case. This late stage in the creative-telescoping method is the topic of the current section.

The differential case is a bit less technical, so let us start with it. Creative telescoping for the integration of a function f of variables $x_1, \ldots, x_r$ w.r.t. $x_2, \ldots, x_r$ first obtains a non-zero skew polynomial $L \in \mathbb{C}[x_1]\langle D_1, \ldots, D_r\rangle$ that annihilates f. Then, to obtain P, it rewrites L by successive divisions by $D_2, \ldots, D_r$ on the left as

$$L = P(x_1, D_1) + D_2 Q_2(x_1, D_1, \ldots, D_r) + \cdots + D_r Q_r(x_1, D_1, \ldots, D_r). \tag{3.21}$$

(There, the family of Q's is not uniquely defined.) Upon application to f and integration over $(x_2, \ldots, x_r)$ in some domain Ω, and under the assumption that the boundary terms (induced by the integration of derivatives) vanish, we obtain the equation

$$P(x_1, D_1)F(x_1) = 0 \quad \text{where} \quad F(x_1) = \int_\Omega f(x_1, \ldots, x_r)\, dx_2 \cdots dx_r.$$

This is a meaningful relation on the integral F *provided the remainder* P *is not zero*. In general, a transformation is needed to ensure that P be non-zero. To this end, consider a monomial $D_2^{\nu_2} \cdots D_r^{\nu_r}$ that makes it possible to write L with a non-zero $\tilde{P}$ as

$$L = D_2^{\nu_2} \cdots D_r^{\nu_r} \tilde{P}(x_1, D_1) + D_2^{\nu_2+1} U_2(x_1, D_1, \dots, D_r) + \cdots + D_r^{\nu_r+1} U_r(x_1, D_1, \dots, D_r),$$

and such that $(\nu_2, \dots, \nu_r)$ is maximal with this property (that is, increasing any of the ν's would not result in a $\tilde{P}$ that is free of $D_2, \dots, D_r$). By repeated use of the relation $x_i D_i^\ell = D_i^\ell x_i - \ell D_i^{\ell-1}$ and using $(*)$ to denote expressions whose explicit values are not needed, we get:

$$x_2^{\nu_2} \cdots x_r^{\nu_r} L = (-1)^{\nu_2+\cdots+\nu_r} \nu_2! \cdots \nu_r! \, \tilde{P}(x_1, D_1) + D_2\,(*) + \cdots + D_r\,(*). \tag{3.22}$$

As this new operator also cancels f, $\tilde{P}$ is a non-zero operator that cancels the integral F (provided the suitable boundary terms vanish). As a final result, we conclude with the existence of an annihilator of the form

$$\tilde{L} = \tilde{P}(x_1, D_1) - D_2 \tilde{Q}_2(x_1, \dots, x_r, D_1, \dots, D_r) - \cdots - D_r \tilde{Q}_r(x_1, \dots, x_r, D_1, \dots, D_r), \tag{3.23}$$

with non-zero $\tilde{P}$.

The case of recurrences is more technical, although it is essentially the same idea, and it too has a q-analogue. To mimic the formula $x_i^\ell D_i^\ell = (-1)^\ell \ell! + D_i\,(*)$, Wegschaider (1997, Sec. 3.2) multiplied on the left by the falling factorial $(n-a)^{\underline{\ell}}$ (for a constant a), leading to the identities

$$\begin{aligned}(n-a)^{\underline{\ell}}(S_n - 1) = S_n (n-a-1)^{\underline{\ell}} - (n-a)^{\underline{\ell}} = \\ (S_n - 1)(n-a-1)^{\underline{\ell}} - \Big((n-a)^{\underline{\ell}} - (n-a-1)^{\underline{\ell}}\Big) = \\ (S_n - 1)(n-a-1)^{\underline{\ell}} - \ell (n-a-1)^{\underline{\ell-1}}.\end{aligned}$$

Their iterated use yields $n^{\underline{\ell}}(S_n - 1)^\ell = (-1)^\ell \ell! + (S_n - 1)\,(*)$. The proof in the recurrence case then proceeds in a way similar to the differential case, by using a left factor of the form $n_2^{\underline{\nu_2}} \cdots n_r^{\underline{\nu_r}}$ instead of $x_2^{\nu_2} \cdots x_r^{\nu_r}$.

3.5 Telescoper, Certificates, Natural Boundaries

The model of (3.2)–(3.3) is so pervasive in the work on creative telescoping that some terminology is welcome here: the skew polynomial P constructed on the η's, whether

$$P = \eta_r(n) S_n^r + \cdots + \eta_0(n)$$

in the recurrence case or

$$P = \eta_r(x) D_x^r + \cdots + \eta_0(x)$$

in the differential case, is called a *telescoper* or *telescoping operator* for the sequence or function h. The corresponding rational function R is the (rational-function) *certificate* for h and P. The motivation for this terminology is that, in certain kinds of applications that are named *over natural boundaries* in the literature, the telescoper P by itself encodes a linear differential/difference equation on the sum or integral of h. This was the case for both examples (1.1) and (1.5) when we discussed the wording "creative telescoping" in Section 1.1. In this situation, the certificate R "certifies" that the telescoper P is well associated with h, by enabling a cancellation property on h of the form

$$P(n, S_n)\, h = (S_n - 1)\big(R(n,k)\, h\big), \quad \text{respectively} \quad P(x, D_x)\, h = D_y\big(R(x,y)\, h\big), \tag{3.24}$$

or, in terms of annihilating ideals

$$P(n, S_n) - (S_n - 1)\, R(n,k) \in \operatorname{ann} h, \quad \text{respectively} \quad P(x, D_x) - D_y\, R(x,y) \in \operatorname{ann} h.$$

To be more explicit, the situation of natural boundaries is when it can be predicted that the right-hand sides in (3.24) go to 0 when summed w.r.t. k or integrated w.r.t. y. This is so for example when a summand $h_{n,k}$ and all its shifts have finite support for each n, as $b_{n,k}$ in (1.1), or when an integrand $h(x,y)$ and all its derivatives have exponential decrease at the boundaries of integration, as in the integral (1.5).

The annihilator $\tilde{L}$ of f in (3.23), too, can be interpreted in terms of telescoper and certificate by generalising (3.24). To this end, introduce $R_2 = f^{-1}\tilde{Q}_2 f, \dots, R_r = f^{-1}\tilde{Q}_r f$, so that

$$\tilde{P} - D_2 R_2 - \cdots - D_r R_r \in \operatorname{ann} f.$$

Here again, $\tilde{P}$ is called a telescoper for f and the rational functions $R_2, \dots, R_r$ are the (rational) certificates for f and $\tilde{P}$. For that matter, the operators $\tilde{Q}_2, \dots, \tilde{Q}_r$ are also at times called certificates.

Chapter 4

Termination Questions: Criteria and Bounds

Zeilberger's fast algorithm and its variants and extensions all perform an exhaustive search of an analogue of (3.19) in some suitable space of equations, in relation to the annihilator of the input summand or integrand. Thus described, when no guarantee of existence of a telescoper, the approach is only a heuristic, as no argument justifies its termination, especially in the non-purely differential situations. This has motivated a number of works to prove termination properties, which I shall separate in two main bodies.

First, a series of works endeavour to determine a criteria that is able to decide, before any complicated calculation, whether Zeilberger's approach will be successful. Such results give no hint as to the order of the outputs from the method.

Second, for certain classes of inputs, a bound on the output order has been developed, which depends on degrees and other arithmetic parameters of the input. The bounds (3.5) already mentioned for proper hypergeometric terms are of this type, as is any order bound for Wilf and Zeilberger's approach is a bound for Zeilberger's fast algorithm. When they exist, bounds can hopefully be reused in estimating the complexity of some summation or integration algorithm.

4.1 Consequence of Holonomy

In this section, I recall the notions of holonomic functions and sequences, and the sufficient condition of holonomy for the existence of (3.19) or its differential variant. These notions are adapted from the notion of holonomic module, itself borrowed from D-module theory.

A series f, possibly of Taylor kind or a formal power series, or more generally a function of variables $x_1, \dots, x_r$, or even a distribution, is called *holonomic* when the functions $x_1^{\alpha_1} \cdots x_r^{\alpha_r} D_1^{\beta_1} \cdots D_r^{\beta_r} f$ obtained by multiplying monomials in the variables and higher-order derivatives of f subject to the constraint $\alpha_1 + \cdots + \alpha_r + \beta_1 + \cdots + \beta_r \leq N$ span a vector space $V_N(f)$ whose dimension over $\mathbb{C}$ grows like $O(N^r)$. For comparison sake, note that the number of monomials under consideration to describe $V_N(f)$ grows like $\Theta(N^{2r})$. Even the vector space of elements from $\mathbb{C}[x_1]\langle D_1, \dots, D_r \rangle$ with total degree not more than N grows "faster" than the $V_N(f)$'s, with a dimension $\Theta(N^{r+1})$. As a consequence, there must exist for large enough N a non-zero skew polynomial L that maps f to 0. The implied identity $Lf = 0$ is a differential analogue to (2.1). This means that a differential analogue of Wilf and Zeilberger's approach will always terminate. After all, this was Lipshitz's argument (1988), which I summarised in Section 2.2.

Furthermore, L can be put in the form (3.21) and transformed into (3.23) for a non-zero $\tilde{P}(x_1, D_1)$. This implies that the differential analogue (Almkvist and Zeilberger, 1990) of Zeilberger's fast algorithm will always terminate.

A sequence $u = (u_{n_1,\dots,n_r})_{n_1,\dots,n_r \geq 0}$ is commonly called *holonomic* when its generating series

$$U(x_1, \dots, x_r) = \sum_{n_1,\dots,n_r \geq 0} u_{n_1,\dots,n_r} x_1^{n_1} \cdots x_r^{n_r}$$

is holonomic in the original sense. I shall show the existence of an equation of the form (3.11) for $h = u$, but, for the sake of presentation, I shall give the idea in the bivariate case, with x_1 and x_2 respectively denoted by x and y. By the same type of reasoning as above, there exists a non-zero $L \in \mathbb{C}[x, y]\langle D_x \rangle$ that annihilates U. We proceed to make the relation $LU = 0$ explicit on the coefficient level. To this end, remark that L rewrites as a (Laurent) polynomial $\Lambda(x, y, \theta_x) \in \mathbb{C}[x, x^{-1}, y]\langle \theta_x \rangle$, where θ_x is Euler's derivative xD_x. Next, for any sum V of the form

$$V(x, y) = \sum_{n \geq 0, k \geq 0} v_{n,k} x^n y^k,$$

we have the formulas

$$\theta_x^b V = \sum_{n \geq 0, k \geq 0} n^b v_{n,k} x^n y^k \quad \text{and} \quad x^a y^b V = \sum_{n \geq a, k \geq b} v_{n-a,k-b} x^n y^k,$$

for any $a \in \mathbb{Z}$ and $b \in \mathbb{N}$. Write Λ more explicitly as a sum

$$\Lambda = \sum_{(a,b) \in S} x^a y^b \lambda_{a,b}(\theta_x),$$

where S is a finite set of pairs $(a, b) \in \mathbb{Z} \times \mathbb{N}$. Applying Λ to U results in

$$\Lambda U = \sum_{(a,b) \in S} \sum_{n \geq a, k \geq b} \lambda_{a,b}(n - a)\, u_{n-a,k-b} x^n y^k.$$

Extracting the coefficient of $x^n y^k$ proves that the relation

$$\sum_{(a,b) \in S} \lambda_{a,b}(n - a)\, u_{n-a,k-b} = 0$$

holds for all $n \geq A$ and all $k \geq B$, where A and B denote the partial degrees in x and y, respectively. This is a non-trivial k-free recurrence.

As in the differential case, this means that Wilf and Zeilberger's approach and Zeilberger's fast algorithm both terminate on holonomic inputs.

4.2 Termination of Zeilberger's Fast Algorithm

In Section 3.2, proving the existence of bounds like (3.5) and (3.14) induces that Zeilberger's algorithm terminates on *proper* hypergeometric terms and on its q-analogues. Ten years lasted before Abramov obtained a sort of a converse result, in the form of a criterion (Abramov, 2002, 2003) for deciding whether Zeilberger's algorithm terminates on a given hypergeometric input. This criterion was a continuation of similar studies for the termination in the rational case (Abramov and Le, 2000, 2002) and it can be tested algorithmically by appealing to algorithms in (Abramov and Petkovšek, 2002; Abramov and Le, 2002). Abramov's criterion takes a suitable remainder $R_{n,k}$ modulo finite differences in what is called an *additive decomposition* of the given summand, then tests if the denominator in a non-standard rational normal form of the quotient $R_{n,k+1}/R_{n,k}$ involves only linear factors of the form $an + bk + c$ for integers a and b. A similar criterion was also elaborated for q-analogues in (Le, 2001; Chen, Hou, and Mu, 2005).

It is clear that Zeilberger's algorithm terminates on a hypergeometric term h if and only if a telescoper for h exists. Another line of study was to relate the termination of Zeilberger's algorithm and the nature of h to be holonomic or not. Here, a sequence is called holonomic when its generating series $H(x,y) = \sum_{n,k\geq 0} h_{n,k}x^n y^k$ is holonomic, which is equivalent to H being D-finite, or again to the fact that the sequence h satisfies a set of recurrences that relates it and a suitable family of specialisations of it, making it *P-recursive* in the sense of (Lipshitz, 1989).

Wilf and Zeilberger's conjecture is that every holonomic hypergeometric term is a proper hypergeometric term. It was proved for the bivariate case independently in (Abramov and Petkovšek, 2002) and in (Hou, 2004), modulo the identification of terms h that share the same quotients $h_{n+1,k}/h_{n,k}$ and $h_{n,k+1}/h_{n,k}$.

Technically, all those criteria commonly base on or partially redevelop a structural study that had been done seemingly independently by (Ore, 1929, 1930) and Sato (1990). (See also (Gel'fand, Graev, and Retakh, 1992).) The essential result is that two first-order recurrences, one w.r.t. n, another w.r.t. k, must satisfy some compatibility property to define a bivariate hypergeometric sequence $h_{n,k}$: the two rational functions $h_{n+1,k}/h_{n,k}$ and $h_{n,k+1}/h_{n,k}$ must admit a very specific factorisation form involving integer coefficients.

A further natural step is to prolong this research to mixed-type (q-)hypergeometric-hyperexponential terms. A work was initiated during Chen's PhD thesis and will be presented in the forthcoming paper (Chen, Chyzak, Feng, and Li, 2013) (under revision at the time of writing): this establishes a computational test for the existence of telescopers, by developing suitable generalisations of additive decompositions and rational normal forms.

4.3 Proving Identities by Numerical Evaluations

An application of creative telescoping is to decide—prove or disprove—a conjectured identity of the form

$$\sum_{k=a}^{b} u_{n,k} = U(n). \tag{4.1}$$

Performing creative telescoping, either in the form of Zeilberger's fast algorithm or of Wilf and Zeilberger's approach, produces a k-free recurrence from which a recurrence for the sum is derived. The identity is decided by:

- verifying that U satisfies the computed recurrence;
- specialising (4.1) on sufficiently many values of n and observing equality or mismatch.

Indeed, if the recurrence is of order r and expressed as

$$a_0(n)\, w_n + \cdots + a_r(n)\, w_{n-r} = 0,$$

and if n_0 is defined as the maximal integer root of $a_0(n)$ if it has one, or 0 if it has none, the values of w_n at $n_0, \dots, n_0 - (r-1)$ define the values at $n > n_0$ uniquely.

This is the starting point of a strategy, initiated by Yen (1993, 1996, 1997), for deciding identities of the form (4.1) by numerical evaluations. Yen's approach is to rewrite the relation (3.18) predicted by Wilf and Zeilberger's theory as

$$a_0(n)\, u_{n,k} + \cdots + a_r(n)\, u_{n-r,k} = R(n,k)\, u_{n,k} - R(n,k-1)\, u_{n,k-1}$$

for a rational function R. The approach in Section 3.2, and especially the bounds (3.5) and (3.9), allow to bound the degree β in k of the numerator $c_0(n) + \cdots + c_\beta(n)\, k^\beta$ of R, then to view the a_i's and c_i's as solutions of a linear system of size $(\delta+1) \times (r+\beta+2)$, where δ is the upper bound in (3.9). Then, expressing the unknown a's and c's by Cramer's rules and using Hadamard-type bounds allows to derive bounds on the degrees and heights of those polynomials. These bounds are polynomial functions in parameters of the input (degrees of P, heights of P, A, B, L, in (3.4)–(3.5)); they induce an exponential bound on the maximal integer root of $a_0(n)$.

The exponential nature of the bound makes it absolutely impractical: the simplest possible example of the sum $\sum_k \binom{n}{k} = 2^n$ would be proved by checking the identity on a number of consecutive integers that is much more than ... 10^{38000}, according to the formula in (Yen, 1993)!

Yen's evaluation was later refined by Zhang (2003) and (Guo, Hou, and Sun, 2008). There, no explicit bound is provided: rather, an algorithm to produce a bound is developed, which experimentally provides dramatically lower values: for the same simple sum of the binomial coefficients, the bound goes down to ... just 4 by the method in (Guo, Hou, and Sun, 2008)!

At the time of writing, I cannot say if the possibility of the simultaneous cancellation of all the c_i's in (3.18) has been integrated in either of the works mentioned above. However, Wegschaider's transformation presented in Section 3.4 does not increase degrees and heights by much, which should not drastically change the results.

4.4 Bounds related to First-Order Equations

A series of works has produced sharper and sharper bounds on the minimal order of a telescoper that can be obtained for a proper, respectively q-proper, hypergeometric term by Wilf and Zeilberger's approach and by Zeilberger's fast algorithm. However, bound improvements seemingly require a genericity assumption of some kind.

Wilf and Zeilberger (1992a) formulated the linear bound (3.5), that is

$$r \leq \sum_{\ell=1}^{L} |b_\ell|,$$

for a proper hypergeometric term, together with a quadratic bound (3.14) for q-analogues. This was refined by Yen (1993), who produced the bound

$$r \leq \sum_{\ell=1,\epsilon=+1}^{L} b_\ell^+ + \sum_{\ell=1,\epsilon=-1}^{L} (-b_\ell)^+ + \left(- \sum_{\ell=1,\epsilon=\pm 1}^{L} \epsilon b_\ell \right)^+,$$

where the notation x^+ denotes $\max(0, x)$; this bound relies on distinguishing terms that are factorials and terms that are inverses of factorials. An even sharper bound can be obtained by further collecting the terms according to the signs of the b's in (3.4), respectively in (3.12). Under presentation and genericity hypotheses, specifically that all a's are non-negative and that the polynomial part $P(n, k)$ has maximal degree, Mohammed and Zeilberger (2005) derived the better bound

$$\max \left(\sum_{\epsilon=+1, b_\ell \geq 0} b_\ell - \sum_{\epsilon=-1, b_\ell \leq 0} b_\ell, - \sum_{\epsilon=+1, b_\ell \leq 0} b_\ell + \sum_{\epsilon=-1, b_\ell \geq 0} b_\ell \right).$$

Apagodu (after Mohammed changed his name to this) and Zeilberger (Apagodu, 2006; Apagodu and Zeilberger, 2006) obtained similar bounds for various classes of hyperexponential functions, for mixed hypergeometric-hyperexponential functions, for q-analogues, and for multiple summations and integrations. For example, for the class of non-rational hyperexponential functions of the form

$$H(x, y) = p(x, y) \exp\left(\frac{a(x, y)}{b(x, y)} \right) \prod_{s \in S} s(x, y)^{\alpha_s}, \tag{4.2}$$

where

- p, a, and b are polynomial such that a/b is a non-constant function of y,
- the s's are polynomials with no non-trivial content w.r.t. y,

- the α's are transcendental constants,

the bound is given as

$$\deg_y(b) + \max\bigl(\deg_y(a), \deg_y(b)\bigr) + \left(\sum_{s\in S} \deg_y(s)\right) - 1. \tag{4.3}$$

(The choice of hypotheses in (Apagodu and Zeilberger, 2006) for their proof to be valid is not cautious, so I chose here a combination that makes the proof work with no additional idea.)

An improvement over this bound can be found in my recent joint work (Bostan, Chen, Chyzak, Li, and Xin, 2013). There, we developed a bound for *any* bivariate hyperexponential function, whether presented in the form (4.2) or not. In particular, the proofs in our work does not require the α's to be transcendental constants. The bouding process starts by a decomposition of the logarithmic derivative $D_y H/H$ in the form

$$\frac{D_y H}{H} = K + \frac{D_y S}{S}$$

with some technical conditions: the *kernel* K and *shell* S need to have coprime denominators and, after writing the kernel in the form $K = k_1/k_2$, the polynomials k_2 and $k_1 - \ell D_y k_2$ must be coprime for any $\ell \in \mathbb{Z}$. Kernel and shell can be obtained by a simple algorithm. After setting β to the squarefree part of the denominator of the shell, our bound reads

$$\deg_y \beta + \max\left(\deg_y k_1, \deg_y k_2 - 1\right). \tag{4.4}$$

In addition, we proved in (Bostan, Chen, Chyzak, Li, and Xin, 2013) that our bound (4.4) is at most Apagodu and Zeilberger's bound (4.3) on their class of more special functions.

Beside bounds on the output order, (Wilf and Zeilberger, 1992a; Yen, 1993) provide bounds on the degree of the output that are, informally speaking, quadratic in the quantities that appear in the order bound. In contrast, no degree bound can be found in (Mohammed and Zeilberger, 2005). However, this discussion means nothing as to the degree of the telescoper *of minimal order*. It may well be that (Apagodu and Zeilberger, 2006) describes the generic case, with polynomial order and degree for the minimal-order telescoper, but that degenerate cases require non-polynomial degrees, as suggested by the encoding of these degrees as roots of a resultant in (Almkvist and Zeilberger, 1990).

Chapter 5

Creative Telescoping Algorithms for Equations of Arbitrary Order

The algorithms and considerations of Chapters 3 and 4 all discuss how to obtain a k-free relation like (2.1), a differential variant, or a non-k-free generalisation like (3.20). In each case, the calculation can be viewed as some sort of skew-polynomial elimination of the summation index or integration variable, possibly modulo derivatives like in (3.23) or modulo finite differences like in (3.19). In each case, too, the application of operators to a function f can be expressed very explicitly, owing to the assumption on the hypergeometric-hyperexponential nature of f, as an explicit rational function times f.

Both aspects lose their simplicity in presence of higher-order equations, and a common understanding lies in an algorithmic theory for skew-polynomial elimination, which was modelled after the classical commutative theory of *Gröbner bases*. This is why the present chapter begins with an account on a non-commutative analogue for the theory of Gröbner bases that is well adapted to the skew algebras under consideration. In relation to (non-commutative) ideal theory, many termination arguments or arguments that some calculation returns a non-trivial output more often than not rely on a non-commutative analogue of the *dimension theory* of algebraic geometry. Dimension is a quantity that, on an intuitive level, distinguishes between the infinite vector-space dimensions over $\mathbb{Q}$ of $\mathbb{Q}[x]$, $\mathbb{Q}[x, y]$, $\mathbb{Q}[x, y, z]$, etc., and is able to capture such notions as ∂-finiteness and holonomy.

Algorithms in the later sections all rely on this Gröbner-basis theory in a way or another, although they not necessarily base on the direct computation of a Gröbner basis.

5.1 Skew Gröbner Bases and a Dimension Theory

As Galligo (1985) and Takayama (1989) noticed, respectively in the differential and in the differential-difference cases, and as was developed by Kandri-Rody and Weispfenning in the more general setting of polynomial rings of solvable type (1990), Buchberger's algorithm for Gröbner bases can be adapted to our non-commutative context: whether with rational-function coefficients as in $\mathbb{Q}(x, y, \dots, m, n, \dots)\langle D_x, D_y, \dots, S_m, S_n, \dots\rangle$, or with polynomial coefficients as in $\mathbb{Q}[x, y, \dots, m, n, \dots]\langle D_x, D_y, \dots, S_m, S_n, \dots\rangle$, or in any intermediate situation. Denoting any such non-commutative ring by A, this theory provides:

- a procedure for putting the presentation of an ideal of A in normal form: two ideals given by sets of generators can be compared for equality by testing equality of the normalised sets, and additionally, inclusion of ideals can be tested easily;
- a procedure for division of an element of A by an ideal $I \subset A$ with unique remainder, or, equivalently, for normal forms in the quotient module A/I;

- a procedure for (skew-)polynomial elimination: for a sub-algebra B of A given by a subset of the generators of A (the $D_x, D_y, \dots, S_m, S_n, \dots$, and possibly the $x, y, \dots, m, n, \dots$ in the variant with polynomial coefficients), a Gröbner-basis calculation results in a presentation of the intersection ideal $I \cap B$.

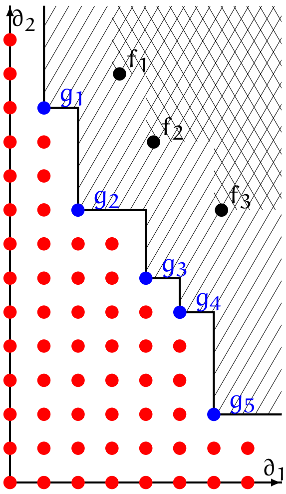


Figure 5.1: The stairs of a typical ideal

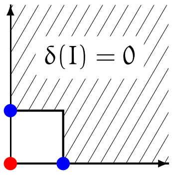


Figure 5.2: Binomial coefficients $\binom{n}{k}$ w.r.t. S_n, S_k; hypergeometric sequences

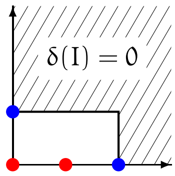


Figure 5.3: Bessel $J_\nu(x)$ w.r.t. S_ν, D_x; orthogonal polynomials w.r.t. S_n, D_x

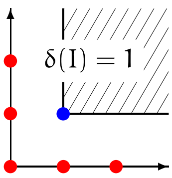


Figure 5.4: Stirling numbers w.r.t. S_n, S_k

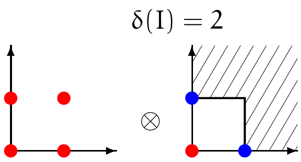


Figure 5.5: Abel-type sequences w.r.t. S_m, S_k, S_r, S_s

As in the commutative case, these notions are parametrised by a monomial ordering on the variables of the problem: $D_x, D_y, \dots, S_m, S_n, \dots$ in the case of rational-function coefficients, $x, y, \dots, m, n, \dots, D_x, D_y, \dots, S_m, S_n, \dots$ in the case of polynomial-function coefficients. In each situation, this designates a leading exponent of each non-zero (skew) polynomial: the maximal one with non-zero coefficient, w.r.t. the ordering. Given an ideal I generated in A by a collection of generators $f_1, f_2, f_3, \dots$ as in Figure 5.1, the combinatorial problem solved by Gröbner bases is to describe the set $S(I)$ of monomials obtained as leading monomials of elements of I. Such monomials are usually depicted as points on a lattice as in the figure. The data of the f_i's directly induces that the doubly-hatched zone is part of $S(I)$. These are the monomial that are multiples of the leading monomials of the f_i's. But it generally happens that combinations of the f_i's with coefficients in A can produce more leading monomials, as depicted by the simply hatched zone for $S(I)$. The corners of this stair-shaped hatched zone correspond to elements g_i of the ideal I, which collectively constitute a Gröbner basis. Buchberger's algorithm really is an algorithm to compute the (finite) collection of the g_i's from the (finite) collection of the f_i's. As a consequence, only the monomials under the stairs can be involved in the remainder of a polynomial under division by I; they constitute a vector basis of the classes in A/I.

Functions that are ∂-finite (in the sense of Section 1.3) correspond to quotient modules A/I that are finite-dimensional vector spaces over $\mathbb{Q}(x, y, \dots, m, n, \dots)$. This situation corresponds to special shapes of stairs that touch all axes, like on Figures 5.2 and 5.3. Given ∂-finite functions $f_1, \dots, f_s$, described by (Gröbner bases for) respective ideals $I_1, \dots, I_s$ w.r.t. the same algebra A, the normal-form procedures in the quotient modules A/I_k provide a means to normalise any polynomial expression $h = P(f_1, \dots, f_s, \dots)$ in the f_j's and their shifts and derivatives.

This leads to an algorithm to compute a system of operators for h and a proof that h is ∂-finite too. The computation is as follows: Each of the successive derivatives $D_x^i h$ of the composed function are normalised as linear combinations of a fixed, finite set F (independent of i) of monomials in the f_j's and their shifts and derivatives. (More specifically, for each j, these monomials involve only those shifts and derivatives of f_j that cannot be simplified by the normal-form procedure w.r.t. I_j; these are the monomials under the stairs associated to I_j.) This normalisation can be written $D_x^i h = \sum_{f \in F} c_{i,f} f$ for rational functions $c_{i,f}$. By the finiteness of F, for large enough N, the family $\{h, \dots, D_x^N h\}$ is linearly dependent, and a linear-algebra solver can be used to find the kernel of the matrix $(c_{i,f})$. Any non-trivial kernel element encodes a differential equation satisfied by h. This process is then reproduced for other derivations D_y, ..., as well as for shifts S_m, S_n, ..., resulting in a system that describes h as ∂-finite. A natural refinement of the algorithm is to consider families that involve partial shifts and derivatives of h (Chyzak and Salvy, 1998).

Often in applications, like in the case of the examples in Section 1.5, a sum or integral to be evaluated involves an expression like h above. As the summation/integration method of the next section inputs a system for h, the algorithmic closure of the present section has to be used to prepare the encoding of h.

The discussion of non-∂-finite functions requires the introduction of a notion of *ideal dimension* $\delta(I)$, not to be confused with the vector-space dimension discussed above for the quotient A/I viewed as a vector space. The finite dimension of A/I for ∂-finite functions (Figures 5.2 and 5.3) corresponds to ideals of dimension zero: $\delta(I) = 0$. Higher values of the ideal dimension $\delta(I)$ refine the classification by distinguishing between different ideals I for which A/I is of infinite vectorial dimension. Figures 5.1 and 5.4 depict ideals of dimension 1; Figure 5.5 depicts an ideal of dimension 2, for sequences of the form $h(m,k)r(k+r)^{k-1}(m-k+s)^{m-k}$ involving a hypergeometric sequence h (and called *of Abel type* in the literature).

Very importantly, the notion of ideal dimension is related to (skew-)polynomial elimination: the lower the dimension, the more variables can be eliminated. In terms of the algebras A and B in the discussion at the beginning of the section, the lower the dimension of the ideal, the fewer generators B needs to have in order that there exists a non-trivial polynomial in $I \cap B$: in fact, as few as $\delta(I) + 1$. Under suitable choices of A, this relates to the possibility to perform multiple summations and integrations w.r.t. more variables when the dimension is low. For an explicit formula, let us denote by g the number of generators $x, y, \dots, m, n, \dots, D_x, D_y, \dots, S_m, S_n, \dots$ of A. Then, skew-polynomial elimination can eliminate $e = g - \big(\delta(I) + 1\big)$ generators simultaneously. Thus, creative telescoping can be used to perform up to e simultaneous summations and integrations.

5.2 Elimination Based on Gröbner Bases

The relation (2.1) can be rephrased by saying that the skew polynomial

$$\sum_{i=0}^{r} \sum_{j=0}^{s} c_{i,j}(n)\, S_n^i S_k^j$$

is in the annihilating ideal w.r.t. $\mathbb{Q}(n)[k]\langle S_n, S_k\rangle$ of the hypergeometric term h. Similarly in the differential case, the $\mathbb{Q}$-linear combination $Z(x, D_x, D_s)$ of the $x^m D_s^n D_x^o$'s obtained from (2.4) is also a skew polynomial in an annihilating ideal: that w.r.t. $\mathbb{Q}[x,s]\langle D_x, D_s\rangle$ of the D-finite power series g. Likewise, the skew polynomial L in Section 3.4 is in the annihilating ideal w.r.t. $\mathbb{C}[x_1, \dots, x_r]\langle D_1, \dots, D_r\rangle$ of f. In each case, (at least) one generator of the skew-polynomial algebra is not involved in the annihilator under consideration, respectively: *(i)* k; *(ii)* s; *(iii)* $x_2, \dots, x_r$. These examples motivate the study of creative telescoping from the point of view of skew-polynomial elimination via Gröbner-bases computations. Algorithms that follow can be viewed as an algorithmic continuation of the research on D-modules by Bernstein and Kashiwara in the 1970s.

To the best of my knowledge, the first published study in this spirit was on integration by Takayama (1990b,a), a topic continued in (Takayama, 1992). More specifically, Takayama

(1990b) considers elimination monomial orderings in $\mathbb{C}[x_1, x_2, \dots, x_r]\langle D_1, \dots, D_r\rangle$: to perform integration w.r.t. $x_2, \dots, x_r$, a monomial ordering is chosen such that any of $x_2, \dots, x_r$ is lexicographically higher than any of $x_1, D_1, \dots, D_r$ (and ties are broken in some way). As, intuitively speaking, a Gröbner-basis calculation looks for (skew) polynomials with small monomials, for that ordering it will try to find polynomials free of $x_2, \dots, x_r$. For a given function f of variables x and y, to be integrated w.r.t. y, after choosing an ordering that eliminates y, it will compute generators for

$$\operatorname{ann}_{W_{x,y}} f \cap W_x[D_y], \tag{5.1}$$

where, for this discussion, we write $W_{x,y}$ for $\mathbb{C}[x, y]\langle D_x, D_y\rangle$ and W_x for $\mathbb{C}[x]\langle D_x\rangle$. Telescopers are then obtained by setting $D_y = 0$ in the result, or by more cautiously following the procedure in Section 3.4.

As in the case of usual (commutative) polynomial rings, this type of elimination procedure is too slow in practice. (And algorithms for efficient computation via change of orderings (Faugère, Gianni, Lazard, and Mora, 1993; Collart, Kalkbrener, and Mall, 1997) had not been discovered yet, even in the simpler commutative case.) So Takayama suggested another approach (1990a), which computes telescopers from the (polynomial) *telescoper ideal* defined as

$$\left(\operatorname{ann}_{W_{x,y}} f + D_y W_{x,y}\right) \cap W_x. \tag{5.2}$$

Here, note that the parenthesis is the sum of a left ideal and a right ideal. It does not have any ideal structure: it is stable under multiplications on the left by elements of $W_x[D_y]$, but not by multiplications by y, because y and D_y do not commute. In fact, Takayama's algorithm computes only an approximation of the telescoper ideal: it views $\operatorname{ann}_{W_{x,y}} f$ as a left $W_x[D_y]$-module, which, as such, has an infinite Gröbner basis, and then works by truncation. Given $m \in \mathbb{N}$ and a (finite) Gröbner basis $\{g_1, \dots, g_s\}$ of $\operatorname{ann}_{W_{x,y}} f$, Takayama's algorithm will introduce all products $y^i g_j$ and $D_y y^i$ of degree in y bounded by m, before running a module Gröbner-basis calculation to eliminate y and provide

$$\left(\operatorname{ann}_{W_{x,y}} f + D_y W_{x,y}\right)_{\le m} \cap W_x[D_y], \tag{5.3}$$

where the subscript "$\le m$" denotes truncation. Telescopers are then obtained by setting $D_y = 0$ in the result, or again by the procedure in Section 3.4. Takayama's method is faster in practice than the previous plain elimination, which motivated Salvy and I to generalise it to non-differential operators and to optimise it by discarding D_y earlier in the calculations (Chyzak and Salvy, 1998).

A specific weakness of Takayama's approach is that it stops as soon as it gets a non-zero element of W_x, but nothing excludes that a lower-order operator could not be found for higher m. This question of a bound on m to get all telescopers was solved by Oaku's algorithm (1997b) for the computation of b-functions. In its simplest variant, the b-function represents the obstruction to expressing a backward shift operator as a differential operator: the *b-function of a polynomial* $\lambda(x_1, \dots, x_r)$ is the monic polynomial $b(s) \in \mathbb{C}[s]$ of minimal degree for which there exists an operator $P(s, x_1, \dots, x_r, D_1, \dots, D_r) \in \mathbb{C}[s, x_1, x_2, \dots, x_r]\langle D_1, \dots, D_r\rangle$ satisfying

$$P(s, x_1, \dots, x_r, D_1, \dots, D_r)\, \lambda^{s+1} = b(s)\, \lambda^s.$$

The notion extends to a notion of the b-function of a D-module. It is sometimes called *indicial polynomial* of a D-module, as it plays the role of a multivariate indicial polynomial with respect to multivariate series solutions of a given differential system, in the sense that it provides with constraints on the minimal (weighted) degrees that appear in the solutions (Saito, Sturmfels, and Takayama, 2000). The article (Oaku, 1997b) essentially gave an algorithm for computing all telescopers by determining the largest integer root of a suitable b-function, before using it as an appropriate bound for m in (5.3) to call Takayama's algorithm. This was later extended in (Oaku and Takayama, 2001).

Two points may obfuscate the discussion of this matter in the literature. First, Oaku's algorithm was not intended for integration, but rather for its dual operation of *restriction*, that is, the computation of annihilating operators for the specialisation of the function f at $y = 0$. This is taken into account in the algorithms by exchanging the roles of y and D_y by means of a transformation named *Fourier transform*. Second, several sub-tasks of the algorithms require computing Gröbner bases in $W_{x,y}$ for orderings that are not well-orders, that is, for which some variable is less than the monomial 1. This situation would lead to non-terminating procedures without amendment of the Gröbner-basis theory. To this end, two options are available, which both use homogenisation to ensure termination of reductions during Gröbner-basis calculations:

- a homogenisation of each generator of the annihilator by an additional variable x_0 so as to work with homogeneous polynomials in $W_{x,y}[x_0]$; this was introduced by (Oaku, 1997a; Oaku and Takayama, 2001) in algorithms for the computation of b-functions and restrictions;

- a homogenisation of the whole algebra $W_{x,y}$ by introducing a slack variable h and an algebra with new commutation rules: $D_x x = xD_x + h^2$ and $D_y y = yD_y + h^2$; this was introduced by Sturmfels and Takayama (1998); Saito, Sturmfels, and Takayama (2000) for the computation of the *integral of a module*, that is, the W_x-module $M/D_y M$ obtained from a $W_{x,y}$-module M.

In addition, it has to be noted that all methods require a sufficient description of $\operatorname{ann}_{W_{x,y}} f$ for (5.1) or (5.3) to be certainly non-trivial. What is at sake here is that a generating family for $\operatorname{ann}_{W_{x,y}} f$ need not be easy to find in practice: applications usually provide with annihilators w.r.t. the algebra $R_{x,y}$ of differential operators with rational-function coefficients, that is, $\mathbb{C}(x, y)\langle D_x, D_y\rangle$, not w.r.t. $W_{x,y}$. And a generating set for $\operatorname{ann}_{R_{x,y}} f$ is not a generating set for $\operatorname{ann}_{W_{x,y}} f$, even after clearing denominators. This is the problem of Weyl closure already mentioned in Section 2.1. In other words, one easily has generators of a sub-ideal of $\operatorname{ann}_{W_{x,y}} f$ only. Although an algorithm for Weyl closure exists (Tsai, 2002), the only implementation of it and the related algorithms that is available in a mainstream computer-algebra system is for Singular, by Andres, Levandovskyy, and Martín-Morales. This may be caused by their apparent intricateness. Indeed, at least in its first presentation, the algorithm for Weyl closure is based on a localisation algorithm for D-modules (Oaku, Takayama, and Walther, 2000), which in turn uses the two different homogenised Gröbner-basis theories mentioned above for sub-tasks.

Recently, Oaku's algorithm has been extended to compute certificates after obtaining telescopers (Nakayama and Nishiyama, 2010). Another recent article (Oaku, 2013) develops similar algorithms for integrals over a domain defined by polynomial inequalities by restricting the domain of the function via the Heaviside function. In the context of recurrences, Kauers, Koutschan, and Zeilberger (2009) have used Takayama's approach in dual form, to compute recurrences for the specialisation of some sequence $f_{n,i,j}$ at $i = j = 0$.

5.3 Summation and Integration of ∂-Finite Functions

Zeilberger's fast algorithm for definite hypergeometric sums of the form

$$U_n = \sum_{k=a}^{b} u_{n,k}$$

and the differential analogue by Almkvist and Zeilberger for definite hyperexponential integrals of the form

$$U(x) = \int_a^b u(x, y)\, dy$$

are dedicated to hypergeometric/hyperexponential terms by their choice of an ansatz,

$$P(n, S_n)\,u = (S_k - 1)\big(R(n,k)\,u\big) \quad \text{for} \quad P \in \mathbb{C}(n)\langle S_n\rangle \text{ and } R \in \mathbb{C}(n,k)$$

in the discrete case and

$$P(x, D_x)\,u = D_y\,\big(R(x,y)\,u\big) \quad \text{for} \quad P \in \mathbb{C}(x)\langle D_x\rangle \text{ and } R \in \mathbb{C}(x,y) \tag{5.4}$$

in the continuous case. In each case, the rationale to ask for a term of the form Ru, for a rational function R, is that the evaluation of Qu for a skew polynomial $Q \in \mathbb{C}(n,k)\langle S_n, S_k\rangle$, respectively $Q \in \mathbb{C}(x,y)\langle D_x, D_y\rangle$, simply leads to a rational multiple of u when u is hypergeometric, respectively hyperexponential.

But more general classes of functions u require more general terms to take the role of Ru. With the motivation of section (3.4), which, in the (differential) holonomic case, guarantees the existence of a non-zero $P(x, D_x)$ and of some $Q(x, y, D_x, D_y) \in \mathbb{C}[x,y]\langle D_x, D_y\rangle$ such that

$$P(x, D_x)\,u = D_y v \quad \text{for} \quad v = Q(x, y, D_x, D_y)\,u, \tag{5.5}$$

it is just natural to replace Ru with an expression that can represent all the possible Qu's. I realised in (Chyzak, 2000) that a nice solution is available for a D-finite u, which is the topic of the present section.

5.3.1 Chyzak's algorithm in basic form

Indeed, given that there exists a finite basis $\{v_i\}$, indexed by $1 \le i \le d$, for the vector space V over $\mathbb{C}(x,y)$ generated by all the derivatives $D_x^a D_y^b u$ at any orders, the ansatz (5.5) in the unknown operator Q can be replaced with the ansatz

$$P(x, D_x)\,u = D_y v \quad \text{for} \quad v = \sum_{i=1}^{d} \phi_i v_i \tag{5.6}$$

in terms of unknown bivariate rational functions ϕ_i's from $\mathbb{C}(x,y)$. Now, expanding $D_y v$ results in an equation that is linear in the v_i and $D_y v_i$, on the one hand, and linear in the ϕ_i and the $D_y \phi_i$, on the other hand. As the $D_y v_i$'s are also in V, the derivative $D_y v$ can be rewritten in the form

$$D_y v = \sum_{j=1}^{d} (D_y \phi_j)\,v_j + \sum_{i,j=1}^{d} \phi_i\, a_{i,j} v_j \tag{5.7}$$

for explicit rational functions $a_{i,j} \in \mathbb{C}(x,y)$ that depend only on the choice of the basis $\{v_i\}$. As in the case of (3.3) for hyperexponential functions (that is, when $d = 1$), an ansatz $P = \eta_r(x)\,D_x^r + \cdots + \eta_0(x)$ is made, and leads to writing Pu as a linear combination of the v_j with coefficients that are linear in the η's:

$$Pu = \sum_{j=1}^{d} \sum_{i=0}^{r} \eta_i b_{i,j} v_j \tag{5.8}$$

for explicit rational functions $b_{i,j} \in \mathbb{C}(x,y)$. For each j between 1 and d, extracting from (5.6) the coefficients w.r.t. the basis element v_j results in a non-homogeneous linear differential relation between $D_y \phi_j$ and the ϕ_i's, with non-homogeneous part involving the η_i's:

$$D_y \phi_j + \sum_{i=1}^{d} a_{i,j} \phi_i = \sum_{i=0}^{r} b_{i,j} \eta_i \qquad (1 \le j \le d). \tag{5.9}$$

This system is solved by eliminating all ϕ_j's but one, say, $\psi = \phi_d$, which results in a non-homogeneous higher-order linear differential equation in $\psi(x,y)$, with derivations w.r.t. y only and a non-homogeneous part that is linear in the $\eta_i(x)$'s. This can be solved by a non-homogeneous variant of Abramov's decision algorithm for rational solutions of a linear ODE (1991). Then, if Abramov's algorithm proves the absence of solutions, there is provably no solution to the ansatz (5.6) for the current value of r. Else, putting the solution ψ back into the system (5.9) results in a similar system in fewer unknown ϕ's, which can in turn be examined for solutions.

The algorithm I formulated in (2000) applies to general operators in place of just the derivations D_x and D_y, as long as the same kind of finiteness as with D-finite functions is preserved. This is why I presented my algorithm for ∂-finite functions. This includes sequences defined by recurrences or q-recurrences, functions defined by mixed differential-difference equations. What varies with the nature of operators is how $D_y v$ is changed in (5.6) and the exact form it takes in the analogue of (5.7). Still, the induced system that plays the role of (5.9) can each time be solved by resorting to a variant of Abramov's algorithms for rational solutions (1991; 1995).

In practice, one takes for the v_i's a family of derivatives $\{D_x^{a_i} D_y^{b_i} u\}_{1\le i\le d}$ with good properties w.r.t. derivation, and the matrix $(a_{i,j})$ is rather sparse. Such a family is obtained naturally when manipulating the ∂-finite function in an algorithmic way. A ∂-finite function f is given by a family operators $P_1, \dots, P_s$ that generate the annihilating ideal $\operatorname{ann} f$ w.r.t. a skew-polynomial algebra $A = \mathbb{Q}(x, y, \dots, m, n, \dots)\langle D_x, D_y, \dots, S_m, S_n, \dots\rangle$. But most often, as the result of a preceding calculation, the P's constitute moreover a Gröbner basis of $\operatorname{ann} f$ w.r.t. some monomial order. So there is a natural family of derivatives that are reduced w.r.t. the P's, that is, that are equal to their remainder after division by the P's.

5.3.2 Iterated integrals and sums

Multiple summations and integrations can also be computed by the same approach as in the previous section, as I explained in the case of natural boundaries in (2000) and as we later extended to non-natural boundaries in (Bostan, Chyzak, van Hoeij, and Pech, 2011). For presentation sake, I shall only present the case of double integration w.r.t. y and z of a hyperexponential function u of variables x, y, and z.

The case of double integrals leads to generalising (5.4) into a form

$$P(x, D_x)\, u = D_y\big(R_1(x,y,z)\, u\big) + D_z\big(R_2(x,y,z)\, u\big) \tag{5.10}$$

for P in $\mathbb{C}(x)\langle D_x\rangle$ and R_1 and R_2 in $\mathbb{C}(x,y,z)$, but the solving for R_1 and R_2 does not generalise so smoothly: attempting reduces to a linear *partial* differential equation relating R_1 and R_2 with $D_y R_1$ and $D_z R_2$. To the best of our knowledge, although this overdetermined linear partial differential equation has a very specific form, no algorithm is available to solve it for its rational solutions. (For comparison sake, the algorithmic problem of recognising whether a general linear partial differential equations with polynomial coefficients has a rational-function solution has recently been proved to be undecidable (Paramonov, 2013).)

Therefore, instead of a direct approach, I developed a cascading approach (2000) which I shall now summarise. Noting that the dependency of P on a *single* derivation D_x in (5.4) is inessential, the same approach is possible for the creative telescoping w.r.t. the (single) variable z of a *trivariate* hyperexponential function u from $\mathbb{Q}(x,y,z)$. Indeed, setting P to the undetermined form

$$P = \sum_{0\le i+j\le r} \eta_{i,j}(x,y)\, D_x^i D_y^j$$

for some tentative total order r and unknown rational functions $\eta_{i,j}$ from $\mathbb{Q}(x,y)$, then performing the same solving as previously, now relying on linear algebra over $\mathbb{Q}(x,y)$, leads to a basis of $P^{(\alpha)}$'s of total order at most r for which there exists a rational function $\phi^{(\alpha)}(x,y,z)$ satisfying

$$P^{(\alpha)}\, u = D_z\left(\phi^{(\alpha)}\, u\right).$$

The theory (as developed, e.g., by Zeilberger (1990b)) guarantees that the set of $P^{(\alpha)}$'s obtained for sufficiently large r describes a D-finite function $\hat{u}$ of x and y, and can therefore be used to determine the finite set needed as an input to the algorithm in (Chyzak, 2000) and in (5.6).

Finally, a double integration algorithm is obtained by continuing the approach used for natural boundaries in (Chyzak, 2000) (Stages A and B below) by a suitable recombination of the outputs (Stage C below). The resulting treatment of multiple integrals over non-natural boundaries is an extension over (Chyzak, 2000), and the corresponding algorithm is as follows:

- *Stage A: First iteration of creative telescoping.* Using the univariate algorithm for trivariate hyperexponential functions in variables (x, y, z) delivers identities

$$P^{(\alpha)}(x, y, D_x, D_y)\, u = D_z\big(\phi^{(\alpha)}(x, y, z)\, u\big). \tag{5.11}$$

- *Stage B: Second iteration of creative telescoping.* Considering a function $\hat{u}$ of (x, y) that is annihilated by all $P^{(\alpha)}$ and using the univariate algorithm for D-finite functions in variables (x, y) delivers an identity

$$P(x, D_x)\, \hat{u} = D_y\big(Q(x, y, D_x, D_y)\, \hat{u}\big). \tag{5.12}$$

- *Stage C: Recombination.* By the theory of linear-differential-operators ideals, the calculations of the algorithm can be interpreted as a proof of existence of operators $L^{(\alpha)}(x, y, D_x, D_y)$ satisfying

$$P(x, D_x) - D_y Q(x, y, D_x, D_y) = \sum_\alpha L^{(\alpha)}(x, y, D_x, D_y)\, P^{(\alpha)}(x, y, D_x, D_y). \tag{5.13}$$

  These $L^{(\alpha)}$ can effectively be obtained either by following the calculations step by step or (less efficiently) by a postprocessing (non-commutative multivariate division). Hence, defining

$$R_1 = u^{-1}\,\big(Q(x, y, D_x, D_y)\, u\big) \quad \text{and} \quad R_2 = u^{-1} \sum_\alpha L^{(\alpha)}(x, y, D_x, D_y)\,\big(\phi^{(\alpha)}(x, y, z)\, u\big)$$

  delivers a solution (P, R_1, R_2) of (5.10).

Note that this two-stage process inherently introduces a dissymmetry in the treatment of the variables y and z: the output from the first iteration tends to be larger than its input; in turn, the output from the second is larger than the output from the first. As a consequence, the order we deal with the variables may have an impact on the running time.

5.3.3 Koutschan's heuristics

Mainly two aspects of the algorithms for ∂-finite functions make them slow in practice. Firstly, solving of (5.9) by uncoupling is sub-optimal. Although algorithms for direct solving of a system exist in the ordinary differential/difference case, they have never been tested in creative-telescoping implementations. Instead, Chyzak's algorithm proceeds by uncoupling, which turns out to be a bottleneck. Secondly, even if no algorithm is known to solve (5.10) as an overdetermined linear partial differential equation, patterns in the orders of poles emerge by experimentation. Having no algorithm remains an obstruction to the direct multiple summation/integration by creative telescoping in Chyzak's algorithm. This has motivated Koutschan (2010) to develop heuristics to guess the exponents in the denominators, which have allowed to solve difficult problems in sizes that can so far not be attacked by the complete algorithmic approaches (Koutschan, Kauers, and Zeilberger, 2011).

Koutschan proceeds with a refined ansatz to solve (5.6), where, owing to the remark at the end of Section 5.3.1, the ϕ_i's have to be thought of as low-order derivatives $D_x^{a_i} D_y^{b_i} u$. Once the (tentative) order r of P is fixed, the left-hand side Pu rewrites according to the differential/recurrence equations satisfied by u by introducing denominators involving y, as introduced by the $b_{i,j}$'s in (5.8). Koutschan adjusts heuristic denominators of the ϕ_i's to match those obtained for the left-hand side via the formula (5.7) for the right-hand side. Then, he fixes heuristic (possibly large) partial degrees for the numerators of the ϕ_i's. To test quickly whether (5.6) is solvable for the current choices of r, denominators, and supports of the numerators, he then proceeds modulo specialisation of x and modulo a large prime. This is more subtle than it first looks like: one has to work around non-commutativity, but performing all non-commutative multiplications before specialisations. Once a modular image proves that the right order r has been found, the same modular process is used to minimise the orders of poles in the denominators and the numerator supports before proceeding with the final non-modular (exact) solving.

It is of interest to compare the way Koutschan's algorithm optimises the shape of numerators and denominators by fast homomorphic computations with Verbaeten's approach for hypergeometric terms in Section 3.3. There, a theoretical study describes the set S of shifts of interest for a relation (3.15) to exist and, starting from a possibly too small S, a completed S is obtained, which does not enlarge denominators and numerators. In contrast, Koutschan's strategy tends to first overestimate the denominators and numerators, then uses fast heuristics to reduce and fit them exactly to the minimal needed.

5.4 Beyond Holonomy

Another direction of extension concerns functions or sequences that cannot be defined by a ∂-finite ideal. Majewicz (1996, 1997) has given an algorithm that is able to produce Abel's summation identity

$$\sum_{k=0}^{n} \binom{n}{k} i(k+i)^{k-1}(n-k+j)^{n-k} = (n+i+j)^n$$

automatically and to find similar new identities. (Here, the reader will easily believe that a term like k^k makes such summand unable to enjoy a recurrence with polynomial coefficients in k.) Kauers (2007) has suggested a summation algorithm applicable to sums involving Stirling numbers and similar sequences defined by triangular recurrence equations (and infinitely many initial conditions). This algorithm finds, for instance, the identity

$$\sum_{k=0}^{n} (-1)^{m-k} k! \binom{n-k}{m-k} S_2(n+1, k+1) = E_1(n, m),$$

where S_2 and E_1 refer to the Stirling numbers of second kind and the Eulerian numbers of first kind, respectively. A summation algorithm of Chen and Sun (2009) is able to discover certain summation identities involving Bernoulli numbers B_n or similar quantities, for example

$$\sum_{k=0}^{m} \binom{m}{k} B_{n+k} = (-1)^{m+n} \sum_{k=0}^{n} \binom{n}{k} B_{m+k}.$$

None of the quantities covered by these algorithms admits a definition via a ∂-finite ideal, but all three algorithms are based on principles that resemble those employed for holonomic systems and ∂-finite ideals. In each case, it turns out that the differential/difference equations defining the integrand/summand are of a form that permits to prove the existence of at least one non-trivial differential/difference equation for the integral/sum by a counting argument that can be compared with Lipshitz's proof for diagonals in Section 2.2.

In (Chyzak, Kauers, and Salvy, 2009), we have given algorithms dealing with functions described by ideals of linear functional operators that are *not* ∂-finite. They generalise the algorithms known for the ∂-finite case and cover the extensions to non-holonomic functions discussed in the beginning of the present section. Holonomy being lost, it is not always the case that creative telescoping can succeed—whatever the algorithm. However, holonomy being only a sufficient condition, it is shown that by considering more generally the *dimension* of the ideals (as discussed in Section 5.1) and another quantity that Kauers, Salvy, and I have called *polynomial growth*, it is possible to predict termination of a generalisation of Chyzak's generalisation of Zeilberger's fast algorithm.

To state it in a nutshell, the lower the dimension of the annihilating ideal of a function, the more variables can be summed and integrated by creative telescoping. (See the formula at the end of Section 5.1.) It is therefore natural to try and bound the dimension related to a multiple sum/integral in terms of the dimension of the summand/integrand. In doing this, the bound we could find is parametrised by the new notion of possible growth.

The notion of polynomial growth originates in observing how the "common denominator" $H_{n,k}$ could be chosen in Wilf and Zeilberger's treatment of proper hypergeometric sums, in contrast to the behaviour of the same approach if confronted with the non-proper input $1/(n^2 + k^2)$. In the former case, the common denominator has a number of factors that is linear w.r.t. $r + s$; in the latter case, it has to be chosen as

$$\prod_{i=0}^{r}\prod_{j=0}^{s}\left((n+i)^2+(k+j)^2\right)$$

and thus has a quadratic number of factors. The same difference in behaviours—linear versus quadratic—occurs for the numerators. Intuitively speaking, the exponent in this polynomial growth of the degree is our notion of polynomial growth.

To make this formal, let us distinguish between variables $x_1, \dots, x_\xi$ that are parameters of the integral/sum and variables $t_1, \dots, t_\tau$ that are integration/summation variables. That is, we consider a multiple integral/sum of the form

$$F(x_1, \dots, x_\xi) = \int f(x_1, \dots, x_\xi, t_1, \dots, t_\tau)\, dt_1 \cdots dt_\tau,$$

or

$$F(x_1, \dots, x_\xi) = \sum_{(t_1,\dots,t_\tau)\in I} f(x_1, \dots, x_\xi, t_1, \dots, t_\tau),$$

viewed as a sequence with indices $x_1, \dots, x_\xi$, or some mixed case of integrations and summations. In the work (Chyzak, Kauers, and Salvy, 2009), we managed to have a complete description in terms of operators with rational-function coefficients. Therefore, let us introduce $A_{x,t} = \mathbb{Q}(x_1, \dots, x_\xi, t_1, \dots, t_\tau)\langle \partial_{x_1}, \dots, \partial_{x_\xi}, \partial_{t_1}, \dots, \partial_{t_\tau}\rangle$ and $A_x = \mathbb{Q}(x_1, \dots, x_\xi)\langle \partial_{x_1}, \dots, \partial_{x_\xi}\rangle$, where ∂_{x_i} denotes either D_{x_i} or S_{x_i}, according to the case, and similarly for ∂_{t_i}.

As creative telescoping has to do with the elimination of the t's, we consider how reduction modulo (a fixed Gröbner basis for) the ideal $I = \operatorname{ann}_{A_{x,t}} f$ lets the degrees in the t's grow. For any given integer $s \geq 0$, there exists a polynomial $P_s(x_1, \dots, x_\xi, t_1, \dots, t_\tau)$ such that each of the $P_s(x_1, \dots, x_\xi, t_1, \dots, t_\tau)\, \partial_{x_1}^{\alpha_1} \cdots \partial_{x_\xi}^{\alpha_\xi} \partial_{t_1}^{\beta_1} \cdots \partial_{t_\tau}^{\beta_\tau}$ has a remainder modulo I that is a linear combination of monomials in the ∂'s with coefficients in $\mathbb{Q}(x_1, \dots, x_\xi)[t_1, \dots, t_\tau]$: such a polynomial P_s can be found as a common denominator, for instance. When, additionally, there is an integer $\pi \in \mathbb{N}$ such that the coefficients in $\mathbb{Q}(x_1, \dots, x_\xi)[t_1, \dots, t_\tau]$ have total degree in the t's *bounded by* $O(s^\pi)$, then the annihilating ideal of f is said *to have polynomial growth* π.

The main result of (Chyzak, Kauers, and Salvy, 2009) is to bound the dimension of the ideal of annihilators that can be obtained by creative telescoping, respectively

$$\left(I + \sum_{i=1}^{\tau} D_{t_i} A_{x,t}\right) \cap A_x \quad \text{and} \quad \left(I + \sum_{i=1}^{\tau} (S_{t_i} - 1) A_{x,t}\right) \cap A_x \tag{5.14}$$

in the integration and in the summation cases, where we have set $A_{x,t} = \mathbb{Q}(x)[t]\langle \partial_x, \partial_t\rangle$ and $A_x = \mathbb{Q}(x)\langle \partial_x\rangle$. These (rational) *telescoping ideals* play a role similar to (5.2) in Takayama's and Oaku's algorithms in Section 5.2. Under natural technical assumptions, this bound can be expressed in terms of the dimension $\delta(I)$ and polynomial growth $\pi(I)$: the dimension (w.r.t. A_x) of the output (5.14) is bounded by

$$\delta(I) + \big(\pi(I) - 1\big)\,\tau.$$

In particular, when this bound is smaller than ξ, then non-trivial annihilators exist in A_x for the sum/integral, as a consequence of the remarks at the end of Section 5.1. This explains why the case $\pi(I) = 1$ is particularly pleasant in applications.

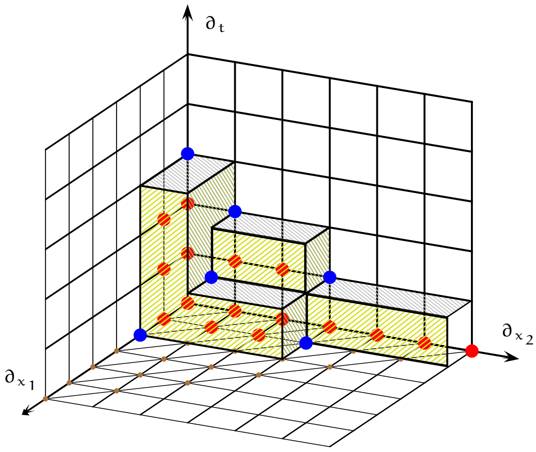


Figure 5.6: Non-∂-finite algorithm: relation to the stairs of the annihilating ideal

Several nice cases correspond to polynomial growth $\pi(I) = 1$:

- *Functions in the purely differential situation*. Intuitively speaking, taking a derivative increases the order of finite poles by 1, while not increasing the order at infinity, and this enables the polynomial growth to be 1. The situation that is described here is very analogous to and should be compared with the proofs on diagonals in Section 2.2.

  The set of the singularities t of the elements of $A_{x,t}f$ (with the x's viewed as parameters) is given as the zero set of a single polynomial $p \in \mathbb{Q}[x,t]$. A suitable polynomial p can be obtained as the lcm of the coefficients of the leading monomials of the elements of Gröbner basis for I, written without denominators. Even more, this polynomial $p = p(x_1, \dots, x_\xi, t_1, \dots, t_\tau)$ is such that, for any function $g = Lf$ given by $L \in A_{x,t}$, each of the derivatives $D_{x_i}g$, for $1 \le i \le \xi$, and $D_{t_j}g$, for $1 \le j \le \tau$, can be written in the form $p^{-1}L'f$ for $L' \in A_{x,t}$, where the total degree of L' w.r.t. the t's is not more than that of L augmented by the degree of p. Therefore, P_s can be taken to be p^s, with $\pi = 1$.

  In this case, starting with a ∂-finite function f results in an integral that is ∂-finite as well.

- *Proper hypergeometric terms*. The ideas have been provided already in Section 3.2. The main difference is that we now want to bound the total shift order, instead of each partial shift order as in (2.1). We provide the idea in the bivariate case for simplicity.

  Indeed, for h as in Section 3.2, each $h_{n+i,k+j}$ for $0 \le i + j \le s$ involves terms of the form $\Gamma(a_\ell n + b_\ell k + c_\ell + u)$ for a shift u bounded in absolute value by $\sigma'_\ell = \max(|a_\ell|, |b_\ell|)\,s$, that is, in a linear way w.r.t. s. Now, observe that $\big(\Gamma(a_\ell n + b_\ell k + c_\ell + u)/\Gamma(a_\ell n + b_\ell k +$

$c_\ell))^{\epsilon_\ell}$ has a denominator bounded by

$$B_\ell = \begin{cases} \dfrac{\Gamma(a_\ell n + b_\ell k + c_\ell)}{\Gamma(a_\ell n + b_\ell k + c_\ell - \sigma'_\ell)} & \text{when } \epsilon_\ell = +1, \\ \dfrac{\Gamma(a_\ell n + b_\ell k + c_\ell + \sigma'_\ell)}{\Gamma(a_\ell n + b_\ell k + c_\ell)} & \text{when } \epsilon_\ell = -1. \end{cases}$$

Both B_ℓ and the polynomial

$$B_\ell \left(\frac{\Gamma(a_\ell n + b_\ell k + c_\ell + u)}{\Gamma(a_\ell n + b_\ell k + c_\ell)} \right)^{\epsilon_\ell}$$

have a degree at most $2\sigma'_\ell$ in k. For a common denominator, we therefore take $P_s = P(n,k)\, B_1 \cdots B_s$, and we find $\pi = 1$.

- *Abel-type sequences*. An Abel-type sequence is a sequence of the form

$$h_{n,k}(k+r)^{k-1}(n-k+s)^{n-k}r$$

where $h_{n,k}$ is a proper hypergeometric term. In (Chyzak, Kauers, and Salvy, 2009), it was shown that a suitable common denominator is

$$P_s = P(n,k)\left(\prod_{\ell=1}^{L} B_\ell\right)\left(\prod_{j=-s}^{s} (k+r+j)(n-k+j)\right),$$

leading to $\pi = 1$ again.

- More examples considered in (Chyzak, Kauers, and Salvy, 2009) include certain products of Stirling numbers or Bernoulli numbers with hypergeometric terms, or more types of special functions, as listed at the end of Section 1.5.

Chapter 6

Efficiency and Complexity Issues

Until recently, little was known about the theoretical complexity of Zeilberger's fast algorithm and about that of creative telescoping in general. Indeed, most of the studies that investigated efficiency issues in creative-telescoping algorithms were more concerned with heuristics to recast a summation problem in to another sum with same value, either by using a symmetry of the summand w.r.t. the summation range ("creative symmetrising", (Paule, 1994)), or by other changes of the form of the sum ("creative subtituting", (Riese, 2002)). An exception seems to be the optimisation in (Lisoněk, Paule, and Strehl, 1993) for rational summation, which presents a very detailed analysis of the degree of a polynomial to be solved for in Zeilberger's fast algorithm, namely the numerator of the rational function R in (3.1).

Yet, experimenting had soon indicated that the arithmetic size of the rational certificate, whether $R(k)$ in (3.2) or the collection of the ϕ_i's in (5.6), is a limiting factor, as it is in practice much larger than the size of the corresponding telescoper, whether $\eta_r(n)\, S_n^r + \cdots + \eta_0(n)$ or $P(x, D_x)$. Complexity study has to take the certificate size into account. In fact, several phenomena occur:

- Algorithms that operate on linear differential or recurrence equations with polynomial equations are sensitive to a quantity N that governs the arithmetic size of their solutions and that is in the worst case exponential in the binary size of the input equation. In particular, the degrees in the rational certificates are governed by such an N, which is often much larger than the order of the telescoper.
- The total binary size of the coefficients of rational certificates tends to grow quadratically in their degrees, making the binary complexity of any algorithm that writes the ϕ's in expanded form at least $\Theta(N^2)$.
- Additionally, rational certificates typically involve more variables than the corresponding telescopers, entailing a combinatorial blow-up of the number of monomials in their dense representation.

To solve both questions of the theoretical complexity of creative telescoping and of the practical efficiency of its algorithms, two directions of research have emerged in the last years:

- First, it was observed that the rational certificates can be represented in condensed form, by keeping denominators factored and viewing numerator coefficients as produced by recurrences. This led to two works: (Bostan, Cluzeau, and Salvy, 2005) for the differential case, then (Bostan, Chyzak, Cluzeau, and Salvy, 2006), in which I participated, for the difference case. The main contribution there is to explain how non-homogeneous linear differential or difference equations can be solved for rational-function solutions with lesser size explosion of intermediate computations, so as to achieve algorithms with complexity $\tilde{O}(N)$.

- Second, it has to be noticed that rational certificates are typically not needed in expanded normal form in applications: often, they are not needed at all (integration over a closed contour or other type of "natural boundaries"); even when they are needed, they are first specialised at end points of the summation or integration range, resulting in much smaller expressions that can more easily be normalised than the original rational certificates. This point of view led to the design of parametrised extensions of Hermite reduction in another series of articles, mostly for the integration of rational functions (Bostan, Chen, Chyzak, and Li, 2010; Bostan, Lairez, and Salvy, 2013), but also for the integration of hyperexponential functions (Bostan, Chen, Chyzak, Li, and Xin, 2013).

6.1 Compact Forms

Section 5.3.1 has shown how Chyzak's creative-telescoping algorithm reduces to solving linear ordinary differential or difference equations for their rational-function solutions, which constitute the rational certificates of the telescopers obtained by the algorithm. In my introductory discussion, I suggested that computing these rational functions—in normal form as is done in Abramov's algorithms—is the bottleneck in Chyzak's algorithm. The situation was similar for Zeilberger's and Almkvist and Zeilberger's algorithms in Section 3.1, although a more direct approach is classically used to find the denominator, then the numerator, of the rational certificates.

Very simple examples show that the degrees appearing in rational solutions can be exponential in the binary size of the equation: for any explicit integer value of the parameter N,

$$(x+1)(x+2)\,u' + Nu = 0 \quad \text{and} \quad n(n+1/2-N)\,u(n+1) - (n-N)(n+1/2)\,u(n) = 0 \tag{6.1}$$

have respective solutions $(x+2)^N/(x+1)^N$ and $\binom{n-1/2}{N}/\binom{n-1}{N}$, which are rational functions with numerators and denominators of degree N. In addition, expanding the numerators and denominators of these rational functions in the monomial bases introduces integers with a total binary size $\Theta(N^2)$. For instance, $\binom{N}{i}$ appears in the denominator of the first example, whose binary size is proportional to i for $0 \le i \le N/2$: summing over i leads to the quadratic bound.

So, any way of accelerating these solving algorithms to a complexity less than quadratic has to base on a way to avoid representing numerators and denominators in expanded normal form. For denominators, this is obtained by keeping them factored. For numerators, this is obtained by a more drastic change of representations: the coefficients of the numerators in a suitable basis are governed by a (compact) recurrence, which permits to represent the numerator by a recurrence and initial conditions.

The goal in (Bostan, Cluzeau, and Salvy, 2005; Bostan, Chyzak, Cluzeau, and Salvy, 2006) is to reduce the binary complexity from $\tilde{O}(N^2)$ for naive algorithms to $\tilde{O}(N)$. Still, if the certificates were needed in expanded form, expanding would cost $\tilde{\Theta}(N^2)$. In the articles, related probabilistic algorithms are introduced for testing existence of rational solutions in binary complexity $\tilde{O}(N^{1/2})$. The precise complexity estimates are derived in the articles. For simplicity, I more informally use here the wording "huge", respectively "small", to indicate quantities whose (arithmetic) size or value are exponential, respectively polynomial, in the binary size of the input.

6.1.1 Abramov algorithms for rational solutions

To understand how to control sizes in and the complexity of rational solving, let us examine how Abramov's algorithms work. A problem of integration w.r.t. y, parametrised by x, leads to a differential equation

$$a_r(x,y)\,\frac{d^r\phi}{dy^r}(x,y) + \cdots + a_0(x,y)\,\phi(x,y) = b(x,y), \tag{6.2}$$

while a problem of summation w.r.t. k, parametrised by n, leads to a recurrence

$$a_r(n,k)\,\phi(n,k+r)+\cdots+a_0(n,k)\,\phi(n,k)=b(n,k), \tag{6.3}$$

in both cases for polynomials a_i's and b in $\mathbb{C}(x)[y]$ or $\mathbb{C}(n)[k]$. Owing to the role of x and n as parameters only in what follows, we now drop the dependency in them in the notation. A first step of the algorithm is to find a *denominator bound*, that is, a polynomial D that is a multiple of the denominator of any rational solution. Then, after a change of unknowns by $\phi=\psi/D$ and clearing denominators, a new equation of the same form is obtained, whose polynomial solutions parametrise the solutions of the original problem. To solve the new problem, one simply determines a degree bound δ on polynomial solutions before proceeding by undetermined coefficients, which amounts to solving a linear system. For fast solving, it was observed in (Abramov, Bronstein, and Petkovšek, 1995) that expressing polynomials in the usual monomial basis $(y^m)_{m\in\mathbb{N}}$ is sufficient to obtain a sparse system in the differential case, while obtaining a sparse system in the recurrence case requires expanding polynomials in the binomial basis $(\binom{y}{m})_{m\in\mathbb{N}}$.

For its part, the polynomial D is obtained by a local analysis of the possible singularities of solutions at finite singular points, first studying their potential location before bounding pole orders in a way that is similar in spirit to the degree bounding for polynomial solutions: One first observes that potential singular points are given by zeroes of the leading coefficient $a_r(y)$ in the differential case and by the gcd $g(k)$ of suitable shifts of the leading and trailing coefficients, namely $p_r(k-r)$ and $p_0(k)$, in the recurrence case. After this, the denominator bound is expected in the form

$$D(y)=\prod_{\ell=1}^{\lambda} d_\ell(y)^{m_\ell} \tag{6.4}$$

for factors $d_\ell(y)$ of $a_r(y)$ in the differential case, and in the form

$$D(k)=\prod_{\ell=1}^{\lambda} d(k)\cdots d(k+m_\ell-1)=\prod_{\ell=1}^{\lambda} d_\ell(k+m_\ell-1)^{\underline{m_\ell}} \tag{6.5}$$

for factors $d_\ell(k)$ of $g(k)$ in the recurrence case.

In both types of equations, the m_ℓ's and the degree bound δ are obtained as maximal integer zeroes of polynomial equations called *indicial equations*. Indicial equations for δ, as well as for the m_ℓ's in the differential case, are constructed by reading off their coefficients after simple transformations of the original equation (6.2). In the recurrence case, the indicial equation for the m_ℓ's is obtained from (6.3) as the resultant

$$R(h)=\operatorname{Res}_h\big(p_0(k+h),p_r(k-r)\big).$$

This provides an exponential upper bound on the m_ℓ's, and thus on the degrees that appear in rational solutions. The examples (6.1) show that these bounds are met. The quantity N of the introduction is the maximum of δ and the m_ℓ's.

6.1.2 Keeping the denominator bound compact

A scrutiny of Abramov's classical algorithms shows that, even if the denominator bound is obtained in factored form (6.4) or (6.5) in a first step of the algorithm, it is immediately expanded by the next step when realising the change of unknown $\phi=\psi/D$ in the equation.

To explain how to improve on this situation, I shall focus on the recurrence case that we dealt with in (Bostan, Chyzak, Cluzeau, and Salvy, 2006). The key observation is that, even if D in (6.5) is "huge", the quotient $\rho(k):=D(k)/D(k+1)$ is "small". This makes it possible to transform the recurrence for ϕ into one for ψ *without expanding* D. Indeed, the quotient is

$$\rho(k)=\frac{D(k)}{D(k+1)}=\prod_{\ell=1}^{\lambda}\frac{d_\ell(0)}{d_\ell(k+m_\ell)}$$

and has degrees not more than that of a_r. After observing, for $j \geq 0$,

$$\phi(k+j) = \rho(k+j-1)\cdots\rho(k)D(k)^{-1}\,\psi(k+j),$$

changing unknowns in (6.3) and multiplying by $D(k)$ lead to the equivalent formulation

$$a_r(k)\rho(k+r-1)\cdots\rho(k)\,\phi(k+r)+\cdots+a_1(k)\rho(k)\,\phi(k+1)+a_0(k)\,\phi(k) = D(k)\,b(k).$$

Here, observe that the left-hand side is "small", as it involves only "small" products of "small" rational functions. However, the right-hand side has grown to a "huge" degree. To avoid this, we first prepare (6.3) by forcing $b = 0$: recombining the equation with its shift, with cofactors $-b(k+1)$ and $b(k)$, results in a homogeneous recurrence. This requires to perform polynomial operations in "small" degrees only, that is, in not more than linear in the original degrees.

After these transformations, (6.3) is turned into a homogeneous analogue,

$$\tilde{a}_{r+1}(k)\,\psi(k+r+1)+\cdots+\tilde{a}_0(k)\,\psi(k) = 0, \tag{6.6}$$

with new polynomial coefficients $\tilde{a}_i$'s of degrees not more than polynomial in the original degrees and order, and with $b = 0$. A similar technique for obtaining

$$\tilde{a}_{r+1}(y)\,\frac{d^{r+1}\psi}{dy^{r+1}}(y)+\cdots+\tilde{a}_0(y)\,\psi(y) = 0, \tag{6.7}$$

with same constraint on the a_i's, is available for (6.2) in the differential case.

6.1.3 Numerators as recurrences

It is classical that, if a solution $\psi(y)$ of (6.7) can be expanded as a Taylor or formal series $\psi(y) = \sum_{j=0}^{\infty} c_j y^j$, its coefficients satisfy a linear recurrence

$$\bar{a}_s(j)c_{j+s}+\cdots+\bar{a}_0(j)c_j = 0 \tag{6.8}$$

with polynomial coefficients $\bar{a}_i$'s. Then, a *polynomial solution* $\psi(y)$ is reflected as a sequence (c_j) *with finite support*, that is, such that c_j is 0 for all sufficiently large j's. Obtaining (6.8) from (6.7) can be done by substituting $\sum_{j=0}^{\infty} c_j y^j$ for $\psi(y)$ in the latter, before extracting the coefficient of y^j. However, the formal expression for this coefficient involves backward shifts of c, which are not defined for too small values of j. This explains why j has to be restricted to be larger than some $j_0 \geq 0$. Alternatively, write $P(y, D_y)$ for the operator underlying the left-hand side of (6.7) and, for and ℓ and m in $\mathbb{N}$, note the formula

$$y^\ell D_y^m \sum_{j=0}^{\infty} c_j y^j = \sum_{j=\max(0,\ell-m)}^{\infty} (j+1-\ell)\cdots(j+m-\ell)\,c_{j+m-\ell}y^j = \sum_{j=\max(0,\ell-m)}^{\infty} c'_j y^j, \tag{6.9}$$

where $c' = (j+1-\ell)\cdots(j+m-\ell)S_j^{m-\ell}c$. This can be viewed as the result of composing simpler formulas for ψ' and $y\psi$. By linearity, (6.7) results in the relation $(Rc)(j) = 0$ for the result R of the non-commutative evaluation of $P(S_j^{-1}, (j+1)S_j)$. This relation is valid provided j is not less than zero and the maximum of the $\ell - m$ over monomials $x^\ell D_x^m$ with non-zero coefficients in P. As it is written, R may involve the shift S_j with negative exponent. Multiplying R with a suitable power S_j^α of S_j to force all exponents to be positive results in an operator $S_j^\alpha R$ that expresses (6.6) and in the corresponding lower bound j_0.

It is less known that polynomial solutions $\psi(k)$ of a recurrence (6.6) can be dealt with in essentially the same manner, after representing $\psi(k)$ in the binomial basis, in the form $\psi(k) = \sum_{j=0}^{\infty} c_j \binom{k}{j}$. Indeed, a (less explicit) analogue of formula (6.9) can be obtained by composing

$$S_k\psi(k) = \sum_{j=0}^{\infty}(c_j+c_{j+1})\binom{k}{j} \quad\text{and}\quad k\psi(k) = \sum_{j=1}^{\infty} j\,(c_j+c_{j-1})\binom{k}{j}.$$

Again, an alternative is to introduce the operator $P(k, S_k)$ underlying the left-hand side of (6.6) and to perform the non-commutative evaluation of $P\big(j\,(1+S_j^{-1}), S_j+1\big)$, which requires to be multiplied by S_j^{α}, where α is the degree of P in k, so as to force only non-negative exponents in powers of S_j. This also makes the recurrence valid from some j_0 on only.

In both the differential and recurrence case, polynomial solutions of the original equation, whether (6.7) or (6.6), are in bijection after one of the previous encodings with the finitely supported sequences that are solutions of a recurrence of the form (6.8). Denote δ the maximal index j such that a finitely supported solution c satisfies $c_\delta \neq 0$ and $c_j = 0$ for $j > \delta$; setting $j = \delta$ in (6.8) results in the indicial equation $\bar{a}_0(\delta) = 0$. Of course, not all initial conditions $C = (c_0, \dots, c_{s-1})$ lead to a finitely supported solution, but, as $C' = (c_{\delta+1}, \dots, c_{\delta+s})$ depends linearly on C, determining all finitely supported solutions reduces to linear algebra, after one has computed a matrix M such that $C' = CM$.

To make the matrix M explicit, it is classical to introduce a matrix equivalent of (6.8): setting $C_j = (c_j, \dots, c_{j-1})$ results in a companion matrix M_j with coefficients in $\mathbb{C}(j)$ such that $C_{j+1} = C_j M_j$; after setting $\nu = \bar{a}_s$, we find a polynomial matrix $U(j)$ to write $M_j = \nu(j)^{-1}U(j)$. In view of $C = C_0$ and $C' = C_{\delta+1}$, we have

$$C' = CM \quad \text{for} \quad M = \frac{1}{\nu(0)\cdots\nu(\delta)} U(0)\cdots U(\delta).$$

At this stage, evaluating the product $U(0)\cdots U(\delta)$ consists in δ matrix products in size s. The arithmetic complexity of this operation is clearly polynomial in s and δ, and, owing to the companion structure of U, in $O(\delta s^2)$. Evaluating the matrices for j between 0 and δ can be done simply in $O(d\delta s^2)$ operations, where d denotes the maximal degree in (6.8).

Here, the potentially “huge” parameter is δ: it contributes to N. Ensuring that the binary complexity is quasi-linear in N, and not quadratic, requires a final algorithmic idea. Intermediate products $U(0)\cdots U(i)$ are a kind of generalisation of the factorial, and like $i!$, they have a binary size in $\tilde{O}(i)$, a bound that is most often matched in practice. So, computing all these products iteratively results in data of size $\Theta(N^2)$, which is too large for the desired linear complexity. Instead, it is now classical to perform *binary splitting*: after introducing $F(i,j) = U(i)\cdots U(j)$, the relevant matrix $M = F(0,\delta)$ is computed by the recursion:

$$F(a,b) = \begin{cases} F(a,m)\,F(m+1,b) & \text{for } m = \lfloor \frac{a+b}{2} \rfloor \text{ if } b \geq a+2, \\ U(a)\,U(a+1) & \text{if } b = a+1, \\ U(a) & \text{if } b = a. \end{cases}$$

This scheme balances binary sizes in multiplications, as $F(a,b)$ can be proved to have a binary size in $\tilde{O}(b-a)$; this also results in a total memory used for the whole calculation bounded by $\tilde{O}(N)$, and the complexity study shows the wanted quasi-linear overall binary complexity.

6.2 The Bivariate Rational Case (and Beyond?)

As it proved too involved to try and tackle the problem of the complexity of creative telescoping as a whole, a line of research endeavours to address simpler classes of integrals. The case of rational-function integration was initiated by the PhD of Shaoshi Chen, which led to the joint work (Bostan, Chen, Chyzak, and Li, 2010). This has continued with trivariate rational functions and a class of bivariate algebraic functions in (Chen, Kauers, and Singer, 2012) leading to faster implementations but no complexity study. Then, in (Bostan, Lairez, and Salvy, 2013), an algorithm was given for a class of rational functions in any number of indeterminates, with a complexity result, but with practical interest only under a regularity condition. In parallel, similar ideas have led to a faster algorithm for the integration of bivariate hyperexponential functions (Bostan, Chen, Chyzak, Li, and Xin, 2013), with no complexity result yet. In the present text, I shall focus on the initial work (Bostan, Chen, Chyzak, and Li, 2010).

In a way or another, creative telescoping boils down to obtaining the intersection in (5.14), whether solely, represented as a set of telescopers P in (5.6), or augmented by the corresponding ϕ_i's. This can be viewed as a sort of skew-polynomial elimination modulo D_y on the left, respectively modulo the D_{t_i}'s or $(S_{t_i} - 1)$'s on the left. In the context of rational-function integration, Hermite introduced a reduction that was named by his name: given a (univariate) rational function $R(t) \in \mathbb{Q}(t)$, Hermite reduction returns another rational function $H(t) \in \mathbb{Q}(t)$ such that H is a normal form for R in a certain sense and $R - H$ is the derivative of a rational function. The similarity of situations suggested that a bivariate extension of Hermite reduction could be used for creative telescoping.

Given a rational function $f = P/Q \in \mathbb{Q}(x, y)$, the bivariate Hermite reduction, that is, Hermite reduction w.r.t. y for coefficients taken in $\mathbb{Q}(x)$, computes a reduced form $[f]$ such that $f - [f]$ is a derivative w.r.t. y of a rational function. A crucial point is now that taking the reduced form commutes with multiplication by any rational function in x only. So, we compute in turn the normal forms $[f]$, $[D_x f]$, $[D_x^2 f]$, ..., and stop when there is a linear dependency

$$\alpha_0(x)\,[f] + \cdots + \alpha_r(x)\,[D_x^r f] = 0.$$

This implies that

$$\alpha_0(x)\,f + \cdots + \alpha_r(x)\,D_x^r f$$

has zero as its normal form, hence is the derivative w.r.t. y of some rational function. This makes $\sum_{i=0}^{r} \alpha_i D_x^i$ a telescoper for f. As we can prove that any telescoper can be obtained in this way, the same procedure as in Zeilberger's algorithm, that is, dealing with increasing values of r, ensures to discover the telescopers of minimal order if starting with $r = 0$. This procedure is the algorithm we called HermiteCT in (Bostan, Chen, Chyzak, and Li, 2010).

With regard to complexity, both creative-telescoping algorithms based on finding rational solutions and the algorithm to be described in this section ultimately resort to linear-algebra solving. But Hermite reduction leads to more manageable dimensions, in relation to the fact that the normal forms obtained by the method lie in a vector space of small dimension.

6.2.1 Hermite reduction

In the classical case, Hermite reduction rewrites a rational function $f = p/q \in \mathbb{Q}(y)$ in the form

$$f = \frac{p}{q} = D_y g + \frac{a}{b}$$

where: *(i)* g is another rational function; *(ii)* a and b are polynomials in $\mathbb{Q}[y]$ such that $\deg a < \deg b$; *(iii)* and b is squarefree, that is, b can be divided (exactly) by the square of no non-constant polynomial. When n is the maximum of the degree of p and q, it is classical that the decomposition (6.2.1) can be obtained in $\tilde{O}(n)$ operations in $\mathbb{Q}$. The same applies to other computational fields of characteristic 0 in place of $\mathbb{Q}$. This is so for example for $\mathbb{Q}(x)$, but then the complexity does not reflect the reality on the computer, owing to a growth of the degrees in x in the calculations.

Considering polynomials P and Q from $\mathbb{Q}[x, y]$ and viewing them in $\mathbb{Q}(x)[y]$, Hermite reduction results in

$$f = \frac{P}{Q} = D_y \frac{A}{Q^-} + \frac{a}{Q^*}, \tag{6.10}$$

where A, a, Q^-, and Q^* are in $\mathbb{Q}(x)[y]$, and, more specifically, Q^* is the squarefree part of Q and Q^- is Q/Q^*. In other words, if an irreducible factor u appears with positive exponent exactly e in f, then it appears with exponent 1 in Q^* and $e - 1$ in Q^-. Coefficient extraction w.r.t. y results in a linear system with coefficients in $\mathbb{Q}[x]$ for the coefficients of A and a. Determining bounds on the degrees in x in (6.10) is more easily done by avoiding coefficients in $\mathbb{Q}(x)$: to this

Method	ord L	$\deg_x$ L	size L	$\deg_x$ g	$\deg_y$ g	size g	Complexity	
RatAZ	$O(n)$	$O(n^3)$	$O(n^4)$	$O(n^3)$	$O(n^2)$	$O(n^5)$	$\tilde{O}(n^{2\omega+3})$	$\tilde{O}(n^9)$
HermiteCT	$O(n)$	$O(n^3)$	$O(n^4)$	$O(n^3)$	$O(n^2)$	$O(n^5)$	$\tilde{O}(n^{\omega+4})$	$\tilde{O}(n^7)$
Lipshitz	$O(n^2)$	$O(d^2)$	$O(n^4)$	$O(n^3)$	$O(n^3)$	$O(n^6)$	$O(n^{6\omega})$	$O(n^{18})$
CubicSize	$O(n^2)$	$O(n)$	$O(n^3)$	$O(n^2)$	$O(n^2)$	$O(n^4)$	$O(n^{4\omega})$	$O(n^{12})$

Figure 6.1: Degrees and sizes of telescopers L and certificates g, and complexity of the algorithms HermiteCT and RatAZ for minimal order, Lipshitz and CubicSize for non-minimal order

end, Cramer rules provide a determinant $\delta \in \mathbb{Q}[x]$ that is guaranteed to be a multiple of the denominators in A and a, so as to rewrite (6.10) as

$$f = \frac{P}{Q} = D_y \frac{B}{\delta Q^-} + \frac{b}{\delta Q^*}, \tag{6.11}$$

where $B = \delta A$ and $b = \delta a$. Let n denote a bound on the partial degrees w.r.t. x and y of P and Q. Interpreting Cramer rules provides degree bounds on δ and the coefficients of B and b: the system has size $O(n)$ and involves polynomial coefficients of degrees $O(n)$, with the result that the solutions of the system have degrees $O(n^2)$ w.r.t. x. This bounds the arithmetic size of B and b by $O(n^3)$.

After getting a cubic bound, a fast algorithm is obtained by the now classical technique of evaluation and interpolation, w.r.t. the variable x. The idea is to compute sufficiently many specialisations of (6.11), in the form

$$f(x_0, y) = \frac{P(x_0, y)}{Q(x_0, y)} = D_y \frac{B(x_0, y)}{\delta(x_0)\, Q^-(x_0, y)} + \frac{b(x_0, y)}{\delta(x_0)\, Q^*(x_0, y)}, \tag{6.12}$$

so as to be able to reconstruct (6.11). This approach bases on the possibility to evaluate a polynomial of degree n at $N \gg n$ consecutive integers in complexity $\tilde{O}(N)$, and to reconstruct a polynomial of degree less than N from evaluations of it at N consecutive integers in the same complexity $\tilde{O}(N)$. The problem is that, for specific x_0's, (6.12) need not be well defined or have the same monomial structure as for generic x. Technically, such unlucky points are defined to be those x_0 for which the degree w.r.t. y of $Q(x_0, y)$ is less than the degree of $Q(x, y)$, or for which the degree w.r.t. y of $\gcd\big(Q(x_0, y), D_y Q(x_0, y)\big)$ is less than the degree of $Q^- = \gcd\big(Q(x, y), D_y Q(x, y)\big)$. This can be tested from the specialisations in complexity $\tilde{O}(n)$. A technical lemma shows that there are not more than $O(n^2)$ unlucky points, so that it is sufficient to test $O(n^2)$ points to get $O(n^2)$ consecutive lucky points, for a total complexity of $\tilde{O}(n^3)$. As computing the Hermite reduction of $P(x_0, y)/Q(x_0, y)$ has complexity $\tilde{O}(n)$, getting all the quadratically many specialisations (6.12) is done in complexity $\tilde{O}(n^3)$. Finally, reconstructing each of $\delta(x)$ and the $O(n)$ coefficients w.r.t. y of B and b in degree $O(n^2)$ w.r.t. x fits in the same complexity, leading to an algorithm for bivariate Hermite reduction in complexity $\tilde{O}(n^3)$.

As a remark, the case when $Q = (Q^*)^2$ permits some level of optimisation. This motivates an improvement of the algorithm algorithm of creative telescoping by Hermite reduction to compute each $\big[D_x^{i+1} f\big]$ as $\big[D_x \big[D_x^i f\big]\big]$, keeping denominators with exponents two after the initial reduction $[f]$.

6.2.2 Minimal-order telescopers

Almkvist and Zeilberger's now classical algorithm (1990) was described for general hyperexponential terms. Specialising it to rational inputs results in an algorithm that we called RatAZ in (Bostan, Chen, Chyzak, and Li, 2010). Interestingly enough, the complexity analysis that resulted in the estimates in Figure 6.1 shows a subtle interplay between the analysis of both algorithms, based on the fact that both algorithms search for minimal-order telescopers and thus return essentially the same outputs. First, the analysis of HermiteCT provides $\deg_y Q^* \leq n$ as an order bound, resulting in quartic bounds for the degrees w.r.t. x of both the obtained telescoper L and the corresponding certificate, and a quadratic bound for the degree w.r.t. y of the certificate. A technical step in Almkvist and Zeilberger's algorithm is the determination of a specific normal form for the rational function $(D_y Lf)/(Lf)$, which requires the computation of a resultant. This is frightening from the complexity viewpoint, owing to the generic size of such resultants. However, the rational situation is sufficiently explicit that the algorithm RatAZ can be simplified so as to predict the form of $(D_y Lf)/(Lf)$ at smaller cost. Importing the tight order bound on the telescoper from the analysis of HermiteCT makes it possible to show that RatAZ finds a telescoper and a certificate of degrees w.r.t. x that are cubic in n. These bound therefore apply to HermiteCT as well, leading to the first columns in Figure 6.1.

The final difference in the complexity analysis is due to the size of the linear systems solved in each algorithms. As RatAZ primarily looks for the certificate g, with the telescoper L obtained as a side effect, it sets up a linear problem to find the coefficients w.r.t. y of g. This system has size given by $1 + \deg_y g \in O(n^2)$, leading to the factor 2 in front of ω in the exponent. For its part, HermiteCT focuses on obtaining L without computing g in expanded form. To this end, it writes the derivatives of f as

$$D_x^i f = D_y \frac{B_i}{\delta^{i+1}\delta'^i Q^{*i} Q^-} + \frac{b_i}{\delta^{i+1}\delta'^i Q^*},$$

where δ' is a technical variant of δ. It then sets up a linear problem to find the coefficients w.r.t. D_x of L, which is obtained by equating to 0 a proper linear combination of the b_i's. This leads to a linear problem in size bounded by $n+1$, explaining the factor 1 in front of ω in the exponent of complexity. Even setting $\omega = 3$ results in HermiteCT beating RatAZ both with regard to theoretical complexity and in practice.

6.2.3 Towards minimal-size telescopers

A line of research initiated in (Bostan, Chyzak, Lecerf, Salvy, and Schost, 2007) for the search of a recurrence satisfied by the coefficients of an algebraic series relaxes the constraint of order minimality on the wanted recurrence operator in order to look for an operator of smaller total arithmetic size. In that article, one option was to encode the algebraic series as an integral and to apply the creative-telescoping approach. This was continued for general integrals of bivariate rational functions in (Bostan, Chen, Chyzak, and Li, 2010), proving the existence of a telescoper of cubic size, as opposed to the quartic size obtained by either RatAZ or HermiteCT. The implied algorithms are however not practical.

The method in (Bostan, Chyzak, Lecerf, Salvy, and Schost, 2007) is reminiscent of the approach by filtration in (Lipshitz, 1988), sketched in Section 2.2. Given a rational integrand $f = P/Q$, it consists in studying the action of the derivations on rational functions of prescribed form H/Q^ℓ so as to obtain bounds on the degrees w.r.t. x and y of the numerators $H_{i,j,k}$ in

$$x^i D_x^j D_y^k f = \frac{H_{i,j,k}}{Q^{j+k+1}}.$$

(Cf. (2.4) in Lipshitz's treatment.) Next, these degree bounds are exploited to determine constraints on i, j, and k that are sufficient to impose a linear dependency between the $x^i D_x^j D_y^k f$'s. A first choice, leading to the algorithm Lipshitz, is to follow Lipshitz's work and use (Bernstein's) total-degree filtration, and set $i + j + k \leq \nu$; a better choice, leading to the algorithm CubicSize, is to enforce $i \leq \kappa$ and $j + k \leq \nu$. Both choices select a cubic number of (i, j, k) w.r.t. ν, and, after normalisation over a common denominator $Q^{\nu+1}$, results in bivariate numerators that live in a vector space of dimension $O(\nu^2)$. This approach provides the existence of smaller telescopers, but is unfortunately only able to compute them as the kernel of a linear system over $\mathbb{Q}$ in size in $O(n^6)$, respectively $O(n^4)$, leading to the coefficients 6 and 4 in front of ω in the complexity estimates.

Chapter 7

Implementations

7.1 Existing Software

Several implementations of Zeilberger's algorithm and of other creative-telescoping algorithms started to be developed in the 1990s. After the first implementations of Zeilberger's algorithm by D. Zeilberger (EKHAD and qEKHAD for Maple) and T. Koornwinder (zeilb and qzeilb for Maple), the Algorithmic Combinatorics group RISC, led by P. Paule, devoted much effort in packages (fastZeil by P. Paule, M. Schorn, and A. Riese, for Mathematica; qZeil by A. Riese, for Mathematica; Zeilberger by F. Caruso, for Maxima). Wilf and Zeilberger's approach was also implemented in RISC (MultiSum by K. Wegschaider and A. Riese, for Mathematica; qMultiSum by A. Riese, for Mathematica). Since several versions, Maple has been shipped with its own standard packages (DEtools, SumTools, QDifferenceEquations), which to the best of my knowledge, stem from initial works by W. Koepf and H. Le. For the list of Mathematica software maintained by the group at RISC, readers are referred to `http://www.risc.jku.at/research/combinat/software/`.

For higher-order creative telescoping, only two implementations exist. My Maple package Mgfun was the first to propose calculations of integration and summation of special functions, as well as in in mixed differential-difference settings. It is still maintained and distributed as a component of Algolib from the URL `http://algo.inria.fr/libraries/`. The part of it that implements my algorithm in (Chyzak, 2000) was reimplemented by L. Pech (2009). It also contains the recent developments on bivariate rational functions described in Section 6.2, implemented by S. Chen and Z. Li. For Mathematica, HolonomicFunctions is a homologue package written by C. Koutschan, which also contains the fast heuristics discussed in Section 5.3.3. It is available from `http://www.risc.jku.at/research/combinat/software/HolonomicFunctions/`.

Specifically for the algorithms based on Gröbner-bases calculations in the Weyl algebras and discussed in Section 5.2, two dedicated softwares have been developed: a package Dmodules is available for Macaulay2 (`http://www.math.uiuc.edu/Macaulay2/`), but is probably outdated; Risa/Asir (`http://www.math.kobe-u.ac.jp/Asir/`) is a computer-algebra system with a Gröbner engine that is targeted to Weyl algebras. Packages implementing algorithms by Oaku, Takayama, and Tsai are part of the distribution of Singular (`http://www.singular.uni-kl.de/`).

7.2 Working Out an Example on the Computer

Let me proceed to exemplify how the algorithms of Chapter 5 can be used on the computer in practice. Here, I use my package Mgfun for Maple.

The following call loads the package and returns the list of exported functions:

```
> with(Mgfun);
```

[MG_Internals, creative_telescoping, dfinite_expr_to_diffeq, dfinite_expr_to_rec,
dfinite_expr_to_sys, diag_of_sys, int_of_sys, pol_to_sys,
rational_creative_telescoping, sum_of_sys, sys*sys, sys+sys]

The main function that I shall exemplify here is creative_telescoping, the function to compute equations like (1.3), (1.6), (3.18), and more generally (5.6) or (5.10). The functions dfinite_expr_to_sys and related take a Maple expression and return a system (resp. a differential equation or a recurrence) satisfied by it. The functions 'sys+sys', 'sys*sys', and related input systems defining several functions and compute a system for the composed function by the algorithms of Section 5.1.

Our interest is now to prove that the integrals

$$I_{m,n} := \int_{-1}^{+1} C_m^{(\mu)}(x) C_n^{(\nu)}(x) (1-x^2)^{\nu-1/2}\, dx,$$

where $C_k^{(\alpha)}(z)$ denotes the kth Gegenbauer polynomial, satisfy the recurrences

$$(n+1)(m-2\nu+2\mu-1-n) I_{m,n+1} = (2\nu+n)(m+1-n) I_{m+1,n},$$
$$(m+2-n)(2+n+2\nu+m) I_{m+2,n} = (2\mu+m+n)(2\mu+m-n-2\nu) I_{m,n},$$

when $\nu > 1/2$. Running the commands:

```
> f := GegenbauerC(m, mu, x) * GegenbauerC(n, nu, x) * (1 - x^2) ^ (nu - 1/2);
> ct := creative_telescoping(f, [m::shift, n::shift], x::diff);
```

performs the approach of Section 5.3.2 on the integrand $f_{m,n}(x)$, looking for a generalisation of (5.6) (cf. (5.11)–(5.12)) in the form

$$P(m, n, S_m, S_n)\, f = D_x\big(Q(m, n, x, S_m, S_n, D_x)\, f\big). \tag{7.1}$$

The third argument of creative_telescoping encodes the summation/integration variable; here, for an integration, it is qualified by "`::diff`". Using "`::shift`" instead would apply to summation problems and would look for relations involving $S_x - 1$ in place of D_x. The second argument to the call encodes the parameters that remain after summation/integration, and the kind of dependency in those parameters; here, using "`::shift`" twice orders the function to look for recurrences w.r.t. m and n (encoded by P). Using "`::diff`" would look for differential relations. Not providing a qualification for a parameter is also possible and keeps the parameter as a constant of the problem, without trying to relate derivatives and shifts w.r.t. it. This is mostly useful to avoid complicated computations for large problems, at the cost of returning less information.

In our example, the output from creative_telescoping is a pair of pairs, encoded in Maple as a list of lists. After some renormalisation, the first (inner) list is of the form:

```
> collect(ct, {_F, _f}, factor);
```

$$\big[\, (n+1)(m-2\nu+2\mu-1-n)\, \_F(m, n+1) - (2\nu+n)(m+1-n)\, \_F(m+1, n),$$
$$(2\nu+n)\, \_f(m, n, x) - x(n+1)\, \_f(m, n+1, x) - (2\nu+n)x\, \_f(m+1, n, x) +$$
$$(n+1)\, \_f(m+1, n+1, x)\, \big]$$

This can be understood as follows (releases of Mgfun before release 15 of Maple had a different notation). Upon integration between -1 and $+1$, (7.1) takes the form

$$P(m, n, S_m, S_n) \int_{-1}^{+1} f\, dx = \big[Q(m, n, x, S_m, S_n, D_x)\, f\big]_{-1}^{+1}. \tag{7.2}$$

The expression on the left-hand side is represented by the first component of the pair, with _F denoting the integral of f; the inside of the brackets on the right-hand side is represented by the second component of the pair, with _f denoting f.

Under our assumption $\nu > 1/2$, the function f is continuous on $[-1,+1]$ and is zero at both ends. This implies that the right-hand side of (7.2) is zero, therefore, that $I_{m,n}$ satisfies the first announced recurrence. The reasoning is similar for the second recurrence, with an additional consideration because the second component of the corresponding pair involves denominators. By chance, the only denominator involved is $m+2$ and we do not need to consider the relations when $m=-2$. Thus, the second recurrence on $I_{m,n}$ can be proved as well.

We fed the algorithm with a closed-form expression for f, but internally it had to represent it by a linear functional system. To this end, Mgfun contains a database of systems for the basic functions, together with implementations of algorithms to combine them. For example, one can ask the package about its knowledge on the Gegenbauer polynomials in the following way:

```
> dfinite_expr_to_sys(GegenbauerC(k, alpha, z),
    c(k::shift, alpha::shift, z::diff));
```

$$\begin{gathered}\{\,(-2\alpha-k)\,c(k,\alpha,z)+(zk+z)\,c(k+1,\alpha,z)+(-2\alpha z^2+2\alpha)\,c(k,\alpha+1,z),\\ (2\alpha+k)\,c(k,\alpha,z)+z\frac{dc}{dz}(k,\alpha,z)-2\alpha c(k,\alpha+1,z),\\ (-2\alpha-4k\alpha-4\alpha^2-k-k^2)\,c(k,\alpha,z)+\\ (4k\alpha-4kz^2\alpha-4\alpha z^2+6\alpha-4\alpha^2z^2+8\alpha^2)\,c(k,\alpha+1,z)+\\ (4\alpha^2z^2-4\alpha^2+4\alpha z^2-4\alpha)\,c(k,\alpha+2,z)\,\}\end{gathered}$$

Here, each element of the set has to be viewed as equal to zero. Combining specialisations of this system (more or less by the function 'sys*sys'), creative_telescoping starts by essentially performing the following command to get a system for f:

```
> dfinite_expr_to_sys(f, c(m::shift, n::shift, x::diff));
```

$$\begin{gathered}\{\,(*)\,c(m,n,x)+(*)\,c(m,n+1,x)+(*)\,c(m,n+2,x),\\ (*)\,c(m,n,x)+(*)\,c(m,n+1,x)+(*)\,c(m+1,n,x)+(*)\,\frac{dc}{dx}(m,n,x),\\ (*)\,c(m,n,x)+(*)\,c(m,n+1,x)+(*)\,\frac{dc}{dx}(m,n,x)+(*)\,\frac{d^2c}{dx^2}(m,n,x)+(*)\,\frac{d^3c}{dx^3}(m,n,x),\\ (*)\,c(m,n,x)+(*)\,c(m,n+1,x)+(*)\,\frac{dc}{dx}(m,n+1,x)+(*)\,\frac{dc}{dx}(m,n,x)+(*)\,\frac{d^2c}{dx^2}(m,n,x)\,\}\end{gathered}$$

where we have abridged by $(*)$ coefficients that are polynomials of total degree up to 6 in m, n, and x.

Chapter 8

Conclusions

8.1 Another Classification of Creative-Telescoping Algorithms

In the previous chapters, I have followed a more or less chronological ordering in my presentation, which at the same time corresponds to the development of algorithms for larger and larger classes of inputs. Questions of efficiency also came after questions of feasibility for larger classes of inputs, and led to other algorithmic paradigms (compact forms and Hermite reduction). But at the present stage, I would like to propose another classification of algorithms, based on the approach they follow:

- *Algorithms based on computations of Gröbner bases.* They have been developed mostly by researchers in D-module theory, in relation with dedicated computer-algebra systems. A limitation of this approach is that functions need to be described by means of ideals in an algebra of operators with polynomial coefficients. On the positive side, multiple integrations can be computed directly. It seems that research by this approach has focused on hypergeometric series that are solutions of A-hypergeometric systems.
- *Algorithms based on an ansatz (undetermined coefficients), whether constant, polynomial or rational.* They have been applied to combinatorics in the summation case and to special functions in the integration case. Their current limitation is that either heuristics have to be used for multiple summations/integrations, or incremental algorithms, with worse behaviour, have to be used.
- *Algorithms based on some generalised Hermite reduction.* This is the most recent family of algorithms and the scope of application is so far more limited, as only the cases of rational and algebraic functions have been studied. Additionally, at the time of writing, algorithms will only terminate under regularity conditions. When applicable, the approach is much faster by avoiding the computation of certificates, but whether an analogue of Hermite reduction can be developed for larger (non-algebraic) classes of inputs is unclear.

At the time of writing, no complete algorithm has been described yet for difference analogues of the Gröbner-based and Hermite-based approaches.

8.2 Perspectives

Several seemingly promising approaches remain untouched or not enough explored after a few decades of algorithmic study on creative telescoping:

- *Verbaeten's completion for non-hypergeometric sequences.* In the hypergeometric situation, Verbaeten's approach has led to faster implementations. Although technical, the theoretical structural study of the degrees that occur in shifts of a hypergeometric term is simplified by the hypergeometric nature of the sequence. It would be of interest to get a similar gain in non-hypergeometric situations. A possible simple starting case for such a study is that of ∂-finite functions with a (vector) dimension 2, like for example families of classical orthogonal polynomials.

- *Creative telescoping based on guessing.* A practical method that has not been mentioned in this memoir to obtain recurrences and differential equations for special functions in applications are *guessing* heuristics: from sufficiently many evaluations of a sequence or sufficiently large series expansion of a functions, equations can often be reconstructed. The question becomes how to bound the number of evaluations or truncation order to get a proof out of a guess. This is related to the research on proving identities by numerical evaluations.

- *Giving up the minimal order.* Creative-telescoping algorithms tend to search for minimal-order telescopers, as this seems to be the right mathematical object. But it is not clear that searching directly for minimal-order telescopers is computationally favorable. Another possible approach is to look for telescopers with larger order but smaller total arithmetic size, which recombine to minimal-order telescopers. The approach was suggested by (Bostan, Chyzak, Lecerf, Salvy, and Schost, 2007; Bostan, Chen, Chyzak, and Li, 2010) and is now a very vivid line of research (Chen and Kauers, 2012b,a; Chen, Jaroschek, Kauers, and Singer, 2013).

- *Signature-based computations of skew Gröbner bases.* It struck me while writing this memoir that there has been only limited interaction (after the 1990s) between the lines of research on Gröbner-based creative telescoping and ansatz-based creative telescoping. Possible explanations can be the nature of inputs expected by the D-module algorithms (operators with polynomial coefficients) and the difficulty to control the speed of Gröbner-bases computations with skew polynomials, both in practice and in theoretical complexity studies. However, recent results (Oaku, 2013) suggest that the Gröbner-based approach can have advantages also for non-natural boundaries. Also, for sequences, it is after all detrimental to work with ∂-finite representations, as they cannot distinguish for example, the sequence $(\delta_{m,n})_{m,n\in\mathbb{N}}$ (ones on the diagonal) from the zero sequence. Such a situation is however frequent in presence of orthogonal polynomials, for example. With the current advent of (faster) signature-based algorithms for computing Gröbner bases, including in algebras of skew polynomials, we can expect that creative telescoping will have to be reconsidered in this light.

- *Ansatz-based algorithms with no uncoupling.* It seems that creative-telescoping algorithms rely on uncoupling only for bad reasons: direct rational-solving algorithms exist for systems and should be tried. In particular, Barkatou (1999) gave an algorithm that includes the case of parametrised non-homogeneous equations, which should do for the differential case. Beside the practical questions of getting implementations and their efficiency, a comparison of the uncoupling and direct approaches from the point of view of complexity would be of interest. In this direction, our recent result (Bostan, Chyzak, and de Panafieu, 2013) gives the complexity of uncoupling in the generic case. A more refined study is thus needed, as the systems to be solved for creative telescoping may well be structured.

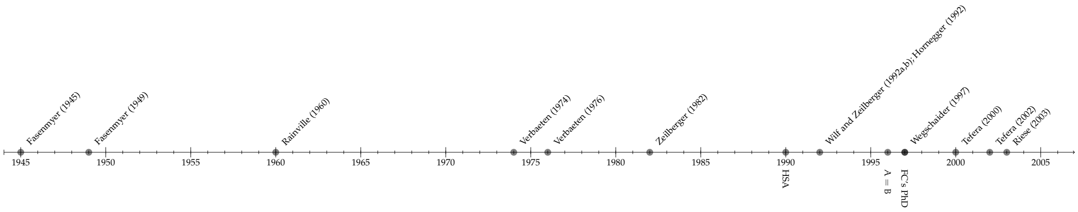
- *Telescopers may be non-minimal annihilators.* A weakness of creative telescoping has long be observed: it need not produce the minimal-order annihilator for a sum or integral, even if it produces a minimal-order telescoper. A classical example (Paule and Schorn, 1995) is given by the sums $F_n = \sum_{k=0}^{n}(-1)^k f_{n,k}$ for $f_{n,k} = \binom{n}{k}\binom{dk}{n}$, which evaluate to $F_n = (-d)^n$. Therefore, a minimal-order annihilator for F is $S_n + d$ and has order 1, but *any* algorithms that bases on the current definition of telescopers or telescoper ideals must return a telescoper of order $d+1$. This is also reflected in the fact that the sum of $(S_n + d)\,f_{n,k}$, which is zero and thus admits the identity operator as an annihilator, can only be proved to have telescopers of order d by current algorithms. A theoretical explanation is still missing and would be welcome in order to design algorithms for minimal-order annihilators.

- *Formal proofs by/of creative telescoping*. It has long been declared in the literature that computer proofs of identities involving special functions or hypergeometric sequences could be considered *routine*. However, a blind appeal of creative-telescoping algorithms can be misleading, as the use of computed telescoper and certificate to produce an equation for a sum or integral often involves *human reasoning* and not just computation. For instance, in the discrete case, it is not clear that the algebraic normalisation to check that a rational certificate certifies a telescoper is valid for the whole range of summation; dealing with summations over non-natural boundaries is very technical and error-prone; in the continuous case, examples show that even if an integral to be computed is well defined, creative telescoping can lead to auxiliary integrals that are divergent; another difficulty lies in creative telescoping producing certificates that have singularities on the integration interval. But we naturally want to assert that the results of our computer-algebra calculations have proved mathematical results. And in this vein in several application domains (combinatorics, physics), there are examples where the only known proof are computer proofs, using creative telescoping. This situation calls for a proper interaction between computer algebra and formal-proofs theory, for a combination of computation and reasoning on the computer. In the context of special functions and creative telescoping, this is a direction in which I engaged recently. Besides producing computer proofs beyond reasonable doubt of identities like those in the list of examples in Section 1.5, a goal is to endow proof assistants with computer-algebra facilities providing them with reliable tools that they will be inclined to reuse in other applications.

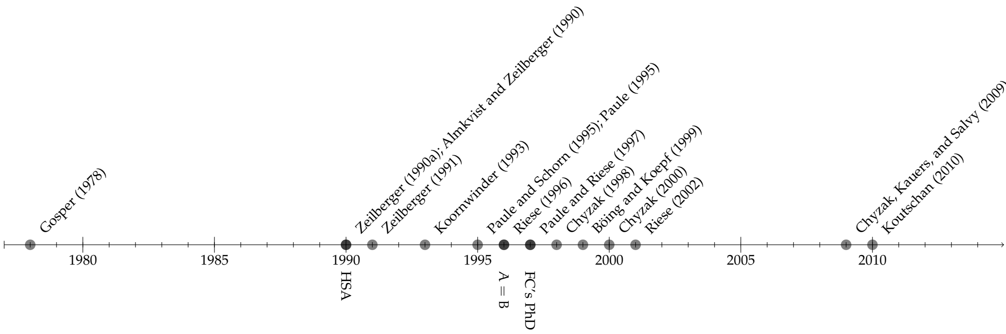
Chapter 9

Timelines

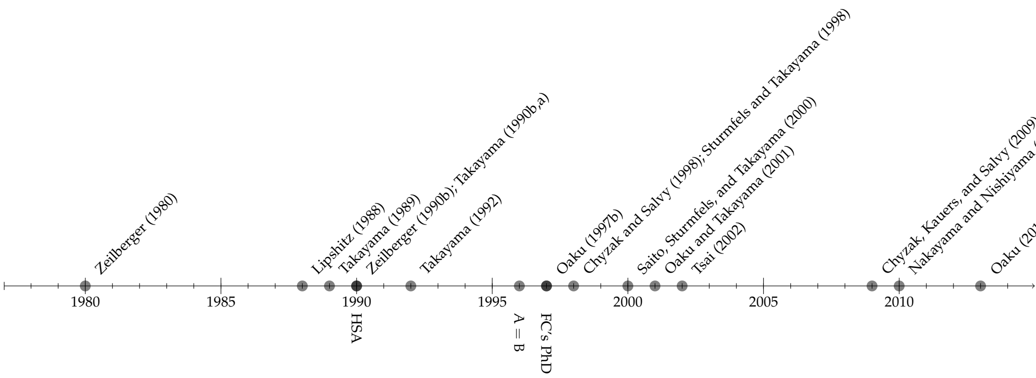
Timeline for the Fasenmyer/Verbaeten/Wilf–Zeilberger approach:



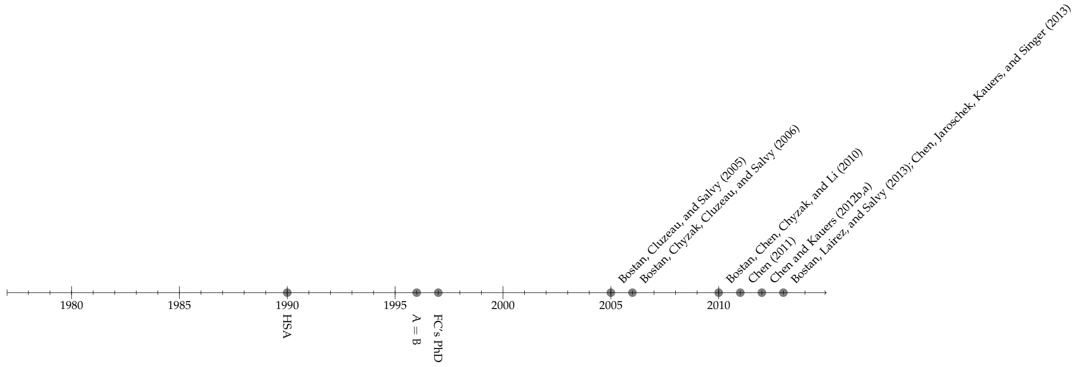
Timeline for the Zeilberger/Chyzak/Koutschan approach:



Timeline for various skew-polynomial-elimination approaches:



Timeline for the complexity study and bounds on creative telescoping:



Bibliography

Abramov, S. A. (1991). Rational solutions of linear differential and difference equations with polynomial coefficients. *U.S.S.R. Comput. Math. and Math. Phys. 29*(6), 7–12. Translation from *Zh. Vychisl. Mat. i Mat. Fiz. 29*(11), 1611–1620 (1989).

Abramov, S. A. (1995). Rational solutions of linear difference and q-difference equations with polynomial coefficients. *Program. Comput. Software 21*(6), 273–278. Translation from *Programmirovanie 6*, 3–11 (1995).

Abramov, S. A. (2002). Applicability of Zeilberger's algorithm to hypergeometric terms. In *Proceedings of the 2002 International Symposium on Symbolic and Algebraic Computation*, New York, pp. 1–7. ACM.

Abramov, S. A. (2003). When does Zeilberger's algorithm succeed? *Adv. in Appl. Math. 30*(3), 424–441.

Abramov, S. A., M. Bronstein, and M. Petkovšek (1995). On polynomial solutions of linear operator equations. In A. H. M. Levelt (Ed.), *Symbolic and Algebraic Computation*, New York, pp. 290–296. ACM Press. Proceedings of ISSAC'95, July 1995, Montreal, Canada.

Abramov, S. A. and H. Q. Le (2000). Applicability of Zeilberger's algorithm to rational functions. In *Formal power series and algebraic combinatorics (Moscow, 2000)*, pp. 91–102. Berlin: Springer.

Abramov, S. A. and H. Q. Le (2002). A criterion for the applicability of Zeilberger's algorithm to rational functions. *Discrete Math. 259*(1-3), 1–17.

Abramov, S. A. and M. Petkovšek (2002). On the structure of multivariate hypergeometric terms. *Adv. in Appl. Math. 29*(3), 386–411.

Almkvist, G. and D. Zeilberger (1990). The method of differentiating under the integral sign. *J. Symbolic Comput. 10*(6), 571–591.

Andrews, G. E. (1974). A general theory of identities of the Rogers-Ramanujan type. *Bull. Amer. Math. Soc. 80*, 1033–1052.

Apagodu, M. (2006). *The sharpening of Wilf-Zeilberger theory*. ProQuest LLC, Ann Arbor, MI. Ph. D. thesis. Rutgers, The State University of New Jersey, New Brunswick.

Apagodu, M. and D. Zeilberger (2006). Multi-variable Zeilberger and Almkvist-Zeilberger algorithms and the sharpening of Wilf-Zeilberger theory. *Adv. in Appl. Math. 37*(2), 139–152.

Askey, R. (1992). The world of q. *CWI Quarterly 5*(4), 251–269.

Barkatou, M. A. (1999). On rational solutions of systems of linear differential equations. *J. Symbolic Comput. 28*(4-5), 547–567. Differential algebra and differential equations.

Bernšteĭn, I. N. (1971). Modules over a ring of differential operators. An investigation of the fundamental solutions of equations with constant coefficients. *Funkcional. Anal. i Priložen. 5*(2), 1–16.

Bernšteĭn, I. N. (1972). Analytic continuation of generalized functions with respect to a parameter. *Funkcional. Anal. i Priložen. 6*(4), 26–40.

Blodgett, R. J., G. E. Andrews, P. Paule, and G. W. Peck (1990, March). Problem E3376. *Amer. Math. Monthly*.

Böing, H. and W. Koepf (1999). Algorithms for q-hypergeometric summation in computer algebra. *J. Symbolic Comput. 28*(6), 777–799. Special issue on Orthogonal Polynomials and Computer Algebra.

Bostan, A., S. Chen, F. Chyzak, and Z. Li (2010). Complexity of creative telescoping for bivariate rational functions. In *ISSAC'10: Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation*, New York, NY, USA, pp. 203–210. ACM.

Bostan, A., S. Chen, F. Chyzak, Z. Li, and G. Xin (2013). Hermite reduction and creative telescoping for hyperexponential functions. In *ISSAC 2013*, pp. 77–84. New York: ACM.

Bostan, A., F. Chyzak, T. Cluzeau, and B. Salvy (2006). Low complexity algorithms for linear recurrences. In *ISSAC 2006*, pp. 31–38. New York: ACM.

Bostan, A., F. Chyzak, and É. de Panafieu (2013). Complexity estimates for two uncoupling algorithms. In *ISSAC 2013*, pp. 85–92. New York: ACM.

Bostan, A., F. Chyzak, G. Lecerf, B. Salvy, and E. Schost (2007). Differential equations for algebraic functions. In *ISSAC 2007*, pp. 25–32. New York: ACM.

Bostan, A., F. Chyzak, M. van Hoeij, and L. Pech (2011). Explicit formula for the generating series of diagonal 3D rook paths. *Sém. Loth. Comb. B66a*. 27 pp.

Bostan, A., T. Cluzeau, and B. Salvy (2005). Fast algorithms for polynomial solutions of linear differential equations. In *ISSAC'05*, pp. 45–52. New York: ACM.

Bostan, A., P. Lairez, and B. Salvy (2013). Creative telescoping for rational functions using the Griffiths–Dwork method. In *ISSAC'13*, pp. 93–100. New York: ACM.

Bronstein, M. and M. Petkovšek (1994). Ore rings, linear operators and factorization. *Programmirovanie* (1), 27–44.

Bronstein, M. and M. Petkovšek (1996). An introduction to pseudo-linear algebra. *Theoret. Comput. Sci. 157*(1), 3–33. Algorithmic complexity of algebraic and geometric models (Creteil, 1994).

Chen, S. (2011). *Some applications of differential-difference algebra to creative telescoping*. Ph. D. thesis, École polytechnique. French original title: Quelques applications de l'algèbre différentielle et aux différences pour le télescopage créatif.

Chen, S., F. Chyzak, R. Feng, and Z. Li (2013). On the existence of telescopers for hyperexponential-hypergeometric sequences. 21 pp. Submitted.

Chen, S., M. Jaroschek, M. Kauers, and M. F. Singer (2013). Desingularization explains order-degree curves for Ore operators. In *ISSAC'13*, pp. 157–164. New York: ACM.

Chen, S. and M. Kauers (2012a). Order-degree curves for hypergeometric creative telescoping. In *ISSAC 2012*, pp. 122–129. New York: ACM.

Chen, S. and M. Kauers (2012b). Trading order for degree in creative telescoping. *J. Symbolic Comput. 47*(8), 968–995.

Chen, S., M. Kauers, and M. Singer (2012). Telescopers for rational and algebraic functions via residues. In *ISSAC 2012*, pp. 130–137. New York: ACM.

Chen, W. Y. C., Q.-H. Hou, and Y.-P. Mu (2005). Applicability of the q-analogue of Zeilberger's algorithm. *J. Symbolic Comput. 39*(2), 155–170.

Chen, W. Y. C. and L. H. Sun (2009). Extended Zeilberger's algorithm for identities on Bernoulli and Euler polynomials. *J. Number Theory 129*(9), 2111–2132.

Chyzak, F. (1998). *Fonctions holonomes en calcul formel*. Ph. D. thesis, École polytechnique. INRIA, TU 0531. 227 pp.

Chyzak, F. (2000). An extension of Zeilberger's fast algorithm to general holonomic functions. *Discrete Math. 217*(1-3), 115–134.

Chyzak, F., M. Kauers, and B. Salvy (2009). A non-holonomic systems approach to special function identities. In *ISSAC'09: Proceedings of the 2009 International Symposium on Symbolic and Algebraic Computation*, pp. 111–118. New York: ACM.

Chyzak, F. and B. Salvy (1998). Non-commutative elimination in Ore algebras proves multivariate identities. *J. Symbolic Comput. 26*(2), 187–227.

Collart, S., M. Kalkbrener, and D. Mall (1997). Converting bases with the Gröbner walk. *J. Symbolic Comput. 24*(3-4), 465–469. Special issue on Computational Algebra and Number Theory (London, 1993).

Doetsch, G. (1930). Integraleigenschaften der Hermiteschen Polynome. *Math. Z. 32*(1), 587–599.

Ekhad, S. B. and D. Zeilberger (1994a). A high-school algebra, "formal calculus", proof of the Bieberbach conjecture [after L. Weinstein]. In *Jerusalem Combinatorics'93*, Volume 178 of *Contemp. Math.*, pp. 113–115. Providence, RI: Amer. Math. Soc. Conference proceedings.

Ekhad, S. B. and D. Zeilberger (1994b). A WZ proof of Ramanujan's formula for π. In J. M. Rassias (Ed.), *Geometry, Analysis, and Mechanics*, pp. 107–108. Singapore: World Scientific.

Ekhad, S. B. and D. Zeilberger (1996). The number of solutions of $X^2 = 0$ in triangular matrices over $\mathrm{GF}(q)$. *Electron. J. Combin. 3*(1), Research Paper 2, 2 pp.

Fasenmyer, M. C. (1945). *Some generalized hypergeometric polynomials*. Ph. D. thesis, University of Michigan.

Fasenmyer, M. C. (1949). A note on pure recurrence relations. *Amer. Math. Monthly 56*, 14–17.

Faugère, J. C., P. Gianni, D. Lazard, and T. Mora (1993). Efficient computation of zero-dimensional Gröbner bases by change of ordering. *J. Symbolic Comput. 16*(4), 329–344.

Galligo, A. (1985). Some algorithmic questions on ideals of differential operators. In *Eurocal'85, Vol. 2 (Linz, 1985)*, Volume 204 of *Lecture Notes in Comput. Sci.*, pp. 413–421. Berlin: Springer.

Gel'fand, I. M., M. I. Graev, and V. S. Retakh (1992). General hypergeometric systems of equations and series of hypergeometric type. *Russian Math. Surveys 47*(4), 1–88. Translation from *Uspekhi Matematicheskikh Nauk 47*(4(286)), 3–82.

Gessel, I. M. (2003). Applications of the classical umbral calculus. *Algebra Universalis 49*(4), 397–434. Dedicated to the memory of Gian-Carlo Rota.

Glasser, M. L. and E. Montaldi (1994). Some integrals involving Bessel functions. *J. Math. Anal. Appl. 183*(3), 577–590.

Gosper, Jr., R. W. (1978). Decision procedure for indefinite hypergeometric summation. *Proc. Nat. Acad. Sci. U.S.A. 75*(1), 40–42.

Guo, Q.-H., Q.-H. Hou, and L. H. Sun (2008). Proving hypergeometric identities by numerical verifications. *J. Symbolic Comput. 43*, 895–907.

Hornegger, J. (1992). Hypergeometrische Summation und polynomiale Rekursion. Diplomarbeit, Universität Erlangen–Nürnberg.

Hou, Q.-H. (2004). k-Free recurrences of double hypergeometric terms. *Adv. in Appl. Math. 32*(3), 468–484.

Jacobson, N. (1937). Pseudo-linear transformations. *Ann. of Math. (2) 38*(2), 484–507.

Kandri-Rody, A. and V. Weispfenning (1990). Noncommutative Gröbner bases in algebras of solvable type. *J. Symbolic Comput. 9*(1), 1–26.

Karr, M. (1981). Summation in finite terms. *J. Assoc. Comput. Mach. 28*(2), 305–350.

Karr, M. (1985). Theory of summation in finite terms. *J. Symbolic Comput. 1*(3), 303–315.

Kashiwara, M. (1978). On the holonomic systems of linear differential equations. II. *Invent. Math. 49*(2), 121–135.

Kauers, M. (2007). Summation algorithms for Stirling number identities. *J. Symbolic Comput. 42*(10), 948–970.

Kauers, M., C. Koutschan, and D. Zeilberger (2009). Proof of Ira Gessel's lattice path conjecture. *Proc. Natl. Acad. Sci. USA 106*(28), 11502–11505.

Koornwinder, T. H. (1993). On Zeilberger's algorithm and its q-analogue. In *Proceedings of the Seventh Spanish Symposium on Orthogonal Polynomials and Applications (VII SPOA) (Granada, 1991)*, Volume 48, pp. 91–111.

Koutschan, C. (2010). A fast approach to creative telescoping. *Math. Comput. Sci. 4*(2-3), 259–266.

Koutschan, C. (2013). Creative telescoping for holonomic functions. In C. Schneider and J. Blümlein (Eds.), *Computer Algebra in Quantum Field Theory: Integration, Summation and Special Functions*, Texts & Monographs in Symbolic Computation, pp. 171–194. Wien: Springer.

Koutschan, C., M. Kauers, and D. Zeilberger (2011). Proof of George Andrews's and David Robbins's q-TSPP conjecture. *Proc. Natl. Acad. Sci. USA 108*(6), 2196–2199.

Le, K. (2001). On the q-analogue of Zeilberger's algorithm for rational functions. *Programmirovanie* (1), 49–58.

Lipshitz, L. (1988). The diagonal of a D-finite power series is D-finite. *J. Algebra 113*(2), 373–378.

Lipshitz, L. (1989). D-finite power series. *J. Algebra 122*(2), 353–373.

Lisoněk, P., P. Paule, and V. Strehl (1993). Improvement of the degree setting in Gosper's algorithm. *J. Symbolic Comput. 16*(3), 243–258.

Majewicz, J. E. (1996). WZ-style certification and Sister Celine's technique for Abel-type sums. *J. Differ. Equations Appl. 2*(1), 55–65.

Majewicz, J. E. (1997). *WZ certification of Abel-type identities and Askey's positivity conjecture*. ProQuest LLC, Ann Arbor, MI. Ph. D. thesis. Temple University.

Mohammed, M. and D. Zeilberger (2005). Sharp upper bounds for the orders of the recurrences output by the Zeilberger and q-Zeilberger algorithms. *J. Symbolic Comput. 39*(2), 201–207.

Nakayama, H. and K. Nishiyama (2010). An algorithm of computing inhomogeneous differential equations for definite integrals. In *Proceedings of the Third international congress conference on Mathematical software*, ICMS'10, Berlin, Heidelberg, pp. 221–232. Springer-Verlag.

Oaku, T. (1997a). An algorithm of computing b-functions. *Duke Math. J. 87*(1), 115–132.

Oaku, T. (1997b). Algorithms for b-functions, restrictions, and algebraic local cohomology groups of D-modules. *Adv. in Appl. Math. 19*(1), 61–105.

Oaku, T. (2013). Algorithms for integrals of holonomic functions over domains defined by polynomial inequalities. *J. Symbolic Comput. 50*, 1–27.

Oaku, T. and N. Takayama (2001). Algorithms for D-modules: restriction, tensor product, localization, and local cohomology groups. *J. Pure Appl. Algebra 156*(2-3), 267–308.

Oaku, T., N. Takayama, and U. Walther (2000). A localization algorithm for D-modules. *J. Symbolic Comput. 29*(4-5), 721–728. Special issue on Symbolic Computation in Algebra, Analysis, and Geometry (Berkeley, CA, 1998).

Ore, O. (1929). Sur les fonctions hypergéométriques de plusieurs variables. *Comptes Rendus des Séances de l'Académie des Sciences 189*, 1238–1241.

Ore, O. (1930). Sur la forme des fonctions hypergéométriques de plusieurs variables. *Journal de Mathématiques Pures et Appliquées 9*(9), 311–326.

Ore, O. (1931). Linear equations in non-commutative fields. *Ann. of Math. (2) 32*(3), 463–477.

Ore, O. (1933). Theory of non-commutative polynomials. *Ann. of Math. (2) 34*(3), 480–508.

Paramonov, S. V. (2013). On rational solutions of linear partial differential or difference equations. *Programming and Computer Software 39*(2), 57–60.

Paule, P. (1985). On identities of the Rogers-Ramanujan type. *J. Math. Anal. Appl. 107*(1), 255–284.

Paule, P. (1994). Short and easy computer proofs of the Rogers-Ramanujan identities and of identities of similar type. *Electron. J. Combin. 1*, Research Paper 10, 9 pp.

Paule, P. (1995). Greatest factorial factorization and symbolic summation. *J. Symbolic Comput. 20*(3), 235–268.

Paule, P. and A. Riese (1997). A Mathematica q-analogue of Zeilberger's algorithm based on an algebraically motivated approach to q-hypergeometric telescoping. In *Special functions, q-series and related topics (Toronto, ON, 1995)*, Volume 14 of *Fields Inst. Commun.*, pp. 179–210. Providence, RI: Amer. Math. Soc.

Paule, P. and M. Schorn (1995). A Mathematica version of Zeilberger's algorithm for proving binomial coefficient identities. *J. Symbolic Comput. 20*(5-6), 673–698. Special issue for Symbolic Computation in Combinatorics Δ_1 (Ithaca, NY, 1993).

Pech, L. (2009, novembre). Algorithmes pour la sommation et l'intégration symboliques. Master's thesis, École Normale Supérieure. 22 pp.

Petkovšek, M., H. S. Wilf, and D. Zeilberger (1996). $A = B$. Wellesley, MA: A K Peters Ltd.

Piessens, R. and P. Verbaeten (1973). Numerical solution of the Abel integral equation. *Nordisk Tidskr. Informationsbehandling (BIT) 13*, 451–457.

Prodinger, H. (1996). Descendants in heap ordered trees, or, A triumph of computer algebra. *Electron. J. Combin. 3*(1), Research Paper 29, 9 pp.

Prudnikov, A. P., Y. A. Brychkov, and O. I. Marichev (1986–1992). *Integrals and series. Vol. 1 to 5*. New York: Gordon and Breach Science Publishers.

Raab, C. (2013). Generalization of Risch's algorithm to special functions. In C. Schneider and J. Blümlein (Eds.), *Computer Algebra in Quantum Field Theory: Integration, Summation and Special Functions*, Texts & Monographs in Symbolic Computation, pp. 285–304. Wien: Springer.

Rainville, E. D. (1960). *Special functions*. New York: The Macmillan Co.

Riese, A. (1996). A generalization of Gosper's algorithm to bibasic hypergeometric summation. *Electron. J. Combin. 3*(1), Research Paper 19, 16 pp.

Riese, A. (2002). Fine-tuning Zeilberger's algorithm. The methods of automatic filtering and creative substituting. In *Symbolic computation, number theory, special functions, physics and combinatorics (Gainesville, FL, 1999)*, Volume 4 of *Developments in Mathematics*, pp. 243–254. Dordrecht: Kluwer Acad. Publ.

Riese, A. (2003). qMultiSum: a package for proving q-hypergeometric multiple summation identities. *J. Symbolic Comput. 35*(3), 349–376.

Risch, R. H. (1969). The problem of integration in finite terms. *Trans. Amer. Math. Soc. 139*, 167–189.

Risch, R. H. (1970). The solution of the problem of integration in finite terms. *Bull. Amer. Math. Soc. 76*, 605–608.

Saito, M., B. Sturmfels, and N. Takayama (2000). *Gröbner deformations of hypergeometric differential equations*, Volume 6 of *Algorithms and Computation in Mathematics*. Berlin: Springer-Verlag.

Sato, M. (1990). Theory of prehomogeneous vector spaces (algebraic part). *Nagoya Math. J. 120*, 1–34. Notes by Takuro Shintani of a lecture by Mikio Sato. Translated from the Japanese by Masakazu Muro.

Schneider, C. (2013). Simplifying multiple sums in difference fields. In C. Schneider and J. Blümlein (Eds.), *Computer Algebra in Quantum Field Theory: Integration, Summation and Special Functions*, Texts & Monographs in Symbolic Computation, pp. 325–360. Wien: Springer.

Stanley, R. P. (1980). Differentiably finite power series. *European J. Combin. 1*(2), 175–188.

Stanley, R. P. (1999). *Enumerative combinatorics. Vol. 2*, Volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge: Cambridge University Press.

Strehl, V. (1994). Binomial identities: combinatorial and algorithmic aspects. *Discrete Math. 136*(1-3), 309–346. Special issue for Trends in Discrete Mathematics.

Sturmfels, B. and N. Takayama (1998). Gröbner bases and hypergeometric functions. In *Gröbner bases and applications (Linz, 1998)*, Volume 251 of *London Math. Soc. Lecture Note Ser.*, pp. 246–258. Cambridge: Cambridge Univ. Press.

Sylvester, J. J. (1840). A method of determining by mere inspection the derivatives from two equations of any degree. *Philosophical Magazine XVI*, 132–135. Available from his Collected Mathematical Papers (`http://www.archive.org/details/collectedmathema00sylv`).

Takayama, N. (1989). Gröbner basis and the problem of contiguous relations. *Japan J. Appl. Math. 6*(1), 147–160.

Takayama, N. (1990a). An algorithm of constructing the integral of a module: an infinite dimensional analog of Gröbner basis. In *Symbolic and Algebraic Computation*, pp. 206–211. ACM and Addison-Wesley. Proceedings of ISSAC'90 (Kyoto, Japan).

Takayama, N. (1990b). Gröbner basis, integration and transcendental functions. In *Symbolic and Algebraic Computation*, pp. 152–156. ACM and Addison-Wesley. Proceedings of ISSAC'90 (Kyoto, Japan).

Takayama, N. (1992). An approach to the zero recognition problem by Buchberger algorithm. *J. Symbolic Comput. 14*(2-3), 265–282.

Tefera, A. (2000). *Improved algorithms and implementations in the multi-WZ theory*. ProQuest LLC, Ann Arbor, MI. Ph. D thesis. Temple University.

Tefera, A. (2002). MultInt, a MAPLE package for multiple integration by the WZ method. *J. Symbolic Comput. 34*(5), 329–353.

Tsai, H. (2000). Weyl closure of a linear differential operator. *J. Symbolic Comput. 29*(4-5), 747–775. Special issue on Symbolic Computation in Algebra, Analysis, and Geometry (Berkeley, CA, 1998).

Tsai, H. (2002). Algorithms for associated primes, Weyl closure, and local cohomology of D-modules. In *Local cohomology and its applications (Guanajuato, 1999)*, Volume 226 of *Lecture Notes in Pure and Appl. Math.*, pp. 169–194. New York: Dekker.

van der Poorten, A. (1979). A proof that Euler missed: Apéry's proof of the irrationality of $\zeta(3)$. *Math. Intelligencer 1*(4), 195–203. An informal report.

Verbaeten, P. (1974). The automatic construction of pure recurrence relations. In *Proceedings of Eurosam 74, Sigsam Bulletin*, pp. 96–98.

Verbaeten, P. (1976). *Rekursiebetrekkingen voor lineaire hypergeometrische funkties*. Ph. D. thesis, Department of Computer Science, K. U. Leuven, Leuven, Belgium. 239 pp.

Wegschaider, K. (1997, May). Computer generated proofs of binomial multi-sum identities. Diplomarbeit, RISC, J. Kepler University. 99 pp.

Wilf, H. S. and D. Zeilberger (1992a). An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities. *Invent. Math. 108*(3), 575–633.

Wilf, H. S. and D. Zeilberger (1992b). Rational function certification of multisum/integral/"q" identities. *Bull. Amer. Math. Soc. (N.S.) 27*(1), 148–153.

Yen, L. (1993). *Contributions to the proof theory of hypergeometric identities*. Ph. D. thesis, University of Pennsylvania.

Yen, L. (1996). A two-line algorithm for proving terminating hypergeometric identities. *J. Math. Anal. Appl. 198*(3), 856–878.

Yen, L. (1997). A two-line algorithm for proving q-hypergeometric identities. *J. Math. Anal. Appl. 213*(1), 1–14.

Zeilberger, D. (1980). The algebra of linear partial difference operators and its applications. *SIAM J. Math. Anal. 11*(6), 919–932.

Zeilberger, D. (1982). Sister Celine's technique and its generalizations. *J. Math. Anal. Appl. 85*(1), 114–145.

Zeilberger, D. (1990a). A fast algorithm for proving terminating hypergeometric identities. *Discrete Math. 80*(2), 207–211.

Zeilberger, D. (1990b). A holonomic systems approach to special functions identities. *J. Comput. Appl. Math. 32*(3), 321–368.

Zeilberger, D. (1991). The method of creative telescoping. *J. Symbolic Comput. 11*(3), 195–204.

Zeilberger, D. (1994). Towards a WZ evaluation of the Mehta integral. *SIAM J. Math. Anal. 25*(2), 812–814.

Zhang, B.-Y. (2003). A new elementary algorithm for proving q-hypergeometric identities. *J. Symbolic Comput. 35*(3), 293–303.

Frédéric Chyzak • INRIA (France) • frederic.chyzak@inria.fr •
`http://specfun.inria.fr/chyzak/`