JOHANNES KEPLER
UNIVERSITÄT LINZ

Eingereicht von
Hui Huang

Angefertigt am
Institut für Algebra

Betreuer und
Erstbeurteiler
**Univ.-Prof. Dr.
Manuel Kauers**

Zweitbeurteiler
Prof. Dr. Ziming Li

Mitbetreuung
Prof. Dr. Ziming Li

Januar 2017

Definite Sums of Hypergeometric Terms and Limits of P-Recursive Sequences

Dissertation

zur Erlangung des akademischen Grades

Doktorin der Naturwissenschaften

im Doktoratsstudium

Naturwissenschaften

**JOHANNES KEPLER
UNIVERSITÄT LINZ**
Altenbergerstraße 69
4040 Linz, Österreich
www.jku.at
DVR 0093696

Definite Sums of Hypergeometric Terms and Limits of P-Recursive Sequences

Hui Huang

Doctoral Thesis

Institute for Algebra
Johannes Kepler University Linz

advised by

Univ.-Prof. Dr. Manuel Kauers

Prof. Dr. Ziming Li

examined by

Univ.-Prof. Dr. Manuel Kauers

Prof. Dr. Ziming Li

The research was partially funded by the Austrian Science Fund (FWF): W1214-N15, project DK13.

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.
Die vorliegende Dissertation ist mit dem elektronisch übermittelten Textdokument identisch.

Linz, im Januar 2017 Hui Huang

Abstract

The ubiquity of the class of D-finite functions and P-recursive sequences in symbolic computation is widely recognized. This class is defined in terms of linear differential and difference equations with polynomial coefficients. In this thesis, the presented work consists of two parts related to this class.

In the first part, we generalize the reduction-based creative telescoping algorithms to the hypergeometric setting. This generalization allows to deal with definite sums of hypergeometric terms more quickly.

The Abramov-Petkovšek reduction computes an additive decomposition of a given hypergeometric term, which extends the functionality of Gosper's algorithm for indefinite hypergeometric summation. We modify this reduction so as to decompose a hypergeometric term as the sum of a summable term and a non-summable one. Properties satisfied by the output of the original reduction carry over to our modified version. Moreover, the modified reduction does not solve any auxiliary linear difference equation explicitly.

Based on the modified reduction, we design a new algorithm to compute minimal telescopers for bivariate hypergeometric terms. This new algorithm can avoid the costly computation of certificates, and outperforms the classical Zeilberger algorithm no matter whether certificates are computed or not according to the computational experiments.

We further employ a new argument for the termination of the above new algorithm, which enables us to derive order bounds for minimal telescopers. Compared to the known bounds in the literature, our bounds are sometimes better, and never worse than the known ones.

In the second part of the thesis, we study the class of D-finite numbers, which is closely related to D-finite functions and P-recursive sequences. It consists of the limits of convergent P-recursive sequences. Typically, this class contains many well-known mathematical constants in addition to the algebraic numbers. Our definition of the class of D-finite numbers depends on two subrings of the field of complex numbers. We investigate how different choices of these two subrings affect the class. Moreover, we show that D-finite numbers over the Gaussian rational field are essentially the same as the values of D-finite functions at non-singular algebraic number arguments (so-called the regular holonomic constants). This result makes it easier to recognize certain numbers as belonging to this class.

Zusammenfassung

Die Allgegenwart der Klasse der D-finiten Funktionen und der P-rekursiven Folgen im Gebiet des Symbolischen Rechnens ist allgemein bekannt. Diese Klasse ist definiert durch lineare Differential- und Differenzengleichungen mit polynomiellen Koeffizienten. Die Ergebnisse dieser Arbeit bestehen aus Teilen, die mit dieser Klasse zu tun haben.

Im ersten Teil verallgemeinern wir die reduktions-basierten Algorithmen für *creative telescoping* auf den hypergeometrischen Fall. Diese Verallgemeinerung erlaubt eine effizientere Behandlung von definiten Summen hypergeometrischer Terme.

Die Abramov-Petkovšek-Reduktion berechnet eine additive Zerlegung eines gegebenen hypergeometrischen Terms, durch die die Funktionalität des Gosper-Algorithmus für indefinite hypergeometrische Summen erweitert. Wir adaptieren diese Reduktion so, dass sie einen hypergeometrischen Term in einen summierbaren und einen nichtsummierbaren Term zerlegt. Eigenschaften des Outputs der ursprünglichen Zerlegung bleiben für unsere modifizierte Version erhalten. Darüber hinaus braucht man bei der modifizierten Reduktion keine lineare Hilfsrekurrenz explizit zu lösen.

Ausgehend von der modifizierten Reduktion entwickeln wir einen neuen Algorithmus zur Berechnung minimaler Telescoper für bivariate hypergeometrische Terme. Dieser neue Algorithmus can die teure Berechnung von Zertifikaten vermeiden, und gemäß unserer Experimente läuft er schneller als der klassische Zeilberger-Algorithmus, egal ob man Zertifikate mitberechnet oder nicht.

Wir verwenden außerdem ein neues Argument für die Terminierung der genannten neuen Algorithmen, das es uns erlaubt, Schranken für die Ordnung des minimalen Telescopers herzuleiten. Verglichen mit den bekannten Schranken in der Literatur sind unsere Schranken manchmal besser und nie schlechter als die bekannten.

Im zweiten Teil der Arbeit untersuchen wir die Klasse der D-finiten Zahlen, die eng verwandt mit D-finiten Funktionen und P-rekursiven Folgen ist. Sie besteht aus den Grenzwerten der konvergenten P-rekursiven Folgen. Typischerweise enthält diese Klasse neben den algebraischen Zahlen viele weitere bekannte mathematische Konstanten. Unsere Definition der Klasse der D-finiten Zahlen hängt von zwei Unterringen des Körpers der komplexen Zahlen ab. Wir untersuchen, wie die Klasse von der Wahl dieser zwei Unterringe abhängt. Außerdem zeigen wir, dass die D-finiten Zahlen über dem Körper der Gaußschen rationalen Zahlen im wesentlichen dieselben Zahlen sind, die auch als Werte von D-finiten Funktionen an nicht-singulären algebraischen Argumenten auftreten (die sogenannten regulären holonomen Konstanten). Dieses Resultat erleichtert es, gewisse Zahlen als Elemente der Klasse zu erkennen.

Acknowledgments

I would like to express my deepest gratitude to my two co-supervisors: Manuel Kauers and Ziming Li, for their academic guidance, constant support and sincere advices. I thank Manuel, for giving me the opportunity to benefit from his immense knowledge, excellent programming skills, amazing scientific insights and high enthusiasm for math. I thank Ziming, for sharing his rigorous scientific attitude, assisting with mathematical and other matters, training my speaking skills over and over again with great patience, and also providing countless valuable suggestions.

My special thanks go to Shaoshi Chen, from whom I profited a lot. I thank him very much for his useful suggestions and constructive comments, which subsequently improved my work considerably. Besides, I am impressed with his obsession with books and high enthusiasm for math.

I was very lucky to be a student of the two lectures "Computer Algebra for Concrete Mathematics" and "Algorithmic combinatorics" given by Peter Paule. I thank him for his enlightening lessons and valuable encouragement. I also learned much from discussions with Hao Du, Ruyong Feng, Christoph Koutschan and Stephen Melczer. I would especially like to thank Hao Du for improving my code.

I really appreciate many members of the Key Laboratory of Mathematics Mechanization for their help, particularly, Wen-tsun Wu for creating this beautiful subject and Xiaoshan-Gao for making our lab a magnificent place to work in. I also appreciate all of my colleagues at the Institute for Algebra as well as my former colleagues at RISC. Special thanks are due to all secretaries for their assistance, and to my friends: Zijia, Miriam, Peng, Ronghua, Rika, Liangjie, for making my life in Linz awesome.

Moreover, I wish to thank Mark Giesbrecht, George Labahn and Éric Schost, for accepting me as a Postdoctoral Fellow in the Symbolic Computation Group at the University of Waterloo.

Most importantly, I would like to dedicate this work to my beloved parents, who make all these possible. Thank you for your unbounded love and care. I thank all my family, for their selfless support and for their understanding and appreciation of my work.

This work was supported by the Austrian Science Fund (FWF) grant W1214-N15 (project DK13), two NSFC grants (91118001, 60821002/F02) and a 973 project (2011CB302401).

Contents

Chapter 1

Introduction

1.1 Background and motivation

Using computer instead of human thought is one of the main themes in the study of symbolic computation for the past century. In particular, finding algorithmic solutions for problems about special functions is one of the very popular topics nowadays.

As an especially attractive class of special functions, D-finite functions have been recognized long ago [59, 45, 70, 57, 46, 60]. They are interesting on the one hand because each of them can be easily described by a finite amount of data, and efficient algorithms are available to do exact as well as approximate computations with them. On the other hand, the class is interesting because it covers a lot of special functions which naturally appear in various different context, both within mathematics as well as in applications.

The defining property of a *D-finite* function is that it satisfies a linear differential equation with polynomial coefficients. This differential equation, together with an appropriate number of initial terms, uniquely determines the function at hand. Similarly, a sequence is called *P-recursive* (or rarely, *D-finite*) if it satisfies a linear recurrence equation with polynomial coefficients. Also in this case, the equation together with an appropriate number of initial terms uniquely determine the object.

The set of P-recursive sequences covers a lot of important combinatorial sequences, including C-finite sequences, hypergeometric terms and sequences whose generating functions are algebraic (called algebraic sequences in this thesis). Rather than talking about sequences themselves, our main interest focus on their definite sums and limits. This thesis is divided into two components.

⋆ ⋆ ⋆ ⋆ ⋆

Part I. Hypergeometric terms. The set of hypergeometric terms is a basic and powerful class of P-recursive sequences. It is defined to be the nonzero solutions of first-order (partial) difference equations with polynomial coefficients.

Many familiar functions are hypergeometric terms, for instance, nonzero rational functions, exponential functions, factorial terms, binomial coefficients, etc. In the study of symbolic summation, there are mainly two kinds of problems related to hypergeometric terms.

Problem 1.1 (Hypergeometric summation)**.** Investigate whether or not the following sum is expressible in simple "closed form",

$$\sum_{k=a}^{b} f(n,k), \quad f(n,k) \text{ is a bivariate hypergeometric term in } n,k, \tag{1.1}$$

where a, b are fixed constants independent of all variables. By a closed form, we mean a linear combination of a fixed number of hypergeometric terms, where the fixed number must be a constant independent of all variables.

Problem 1.2 (Hypergeometric identities)**.** Prove the following identity

$$\sum_{k=a}^{b} f(n,k) = h(n), \quad f(n,k) \text{ is a bivariate hypergeometric term in } n,k, \tag{1.2}$$

where a, b are fixed constants independent of all variables, and $h(n)$ is a known univariate function.

Analogous to the first fundamental theorem of calculus, Problem 1.1 could be solved in terms of indefinite summation provided that there exists a so-called "anti-difference". More precisely, we compute a hypergeometric term $g(n,k)$ such that

$$f(n,k) = g(n,k+1) - g(n,k),$$

and then Problem 1.1 easily follows by the telescoping sum technique. To our knowledge, the first complete algorithm for indefinite summation was designed by Gosper [36] in 1978, namely the famous Gosper's algorithm. To address the case when Gosper's algorithm is not applicable, i.e., there exists no such g, Wilf and Zeilberger developed a constructive theory in a series of articles [65, 66, 67, 68, 69, 70, 71] in early 1990s. This theory came to be known as Wilf-Zeilberger's theory, whose main idea is to construct a so-called telescoper for f to derive a difference equation with polynomial coefficients satisfied by (1.1), and then applying Petkovšek's algorithm [53], which detects the existence of the hypergeometric terms solutions, to this equation gives the final answer for Problem 1.1.

On the other hand, Wilf-Zeilberger's theory also works for Problem 1.2. To be precise, after deriving a difference equation satisfied by the left-hand side of (1.2) from a telescoper as for Problem 1.1, we verify that h satisfies the same equation and then (1.2) easily follows by checking the initial values.

Wilf-Zeilberger's theory not only provides an algorithmic method to solve the problems about hypergeometric summations or identities, but also gives a constructive way to find new combinatorial identities. In terms of algorithms, Wilf-Zeilberger's theory is a strong fundamental tool for combinatorics and also the theory of special functions.

From the above discussion, one sees that the key step of Wilf-Zeilberger's theory is to construct a telescoper. This process is referred to as *creative telescoping*. To be more specific, for a bivariate hypergeometric term $f(n, k)$, the task consists in finding some nonzero recurrence operator L and another hypergeometric term g such that

$$L \cdot f(n, k) = g(n, k + 1) - g(n, k). \tag{1.3}$$

It is required that the operator L does not contain k or the shift operator σ_k, i.e., it must have the form $L = e_0 + e_1\sigma_n + \cdots + e_\rho\sigma_n^\rho$ for some $e_0, \ldots, e_\rho$ that only depend on n. If L and $g(n, k)$ are as above, we say that L is a *telescoper* for $f(n, k)$, and $g(n, k)$ is a *certificate* for L.

As outlined in the introduction of [19], we can distinguish four generations of creative telescoping algorithms.

The first generation [29, 70, 54, 27] dates back to the 1940s, and the algorithms were based on elimination techniques. **The second generation** [69, 11, 71, 54] started with what is now known as Zeilberger's (fast) algorithm. The algorithms of this generation use the idea of augmenting Gosper's algorithm for indefinite summation (or integration) by additional parameters $e_0, \ldots, e_\rho$ that are carried along during the calculation and are finally instantiated, if at all possible, such as to ensure the existence of a certificate g in (1.3). These algorithms have been implemented in many computer algebra programs, for example MAPLE [5] and MATHEMATICA [52]. See [54] for details about the first two generations.

The third generation [49, 12] was initiated by Apagodu and Zeilberger. In a sense, they applied a second-generation algorithm by hand to a generic input and worked out the resulting linear system of equations for the parameters $e_0, \ldots, e_\rho$ and the coefficients inside the certificate g. Their algorithm then merely consists in solving this system. This approach is interesting not only because it is easier to implement and tends to run faster than earlier algorithms, but also because it is easy to analyze. In fact, the analysis of algorithms from this family gives rise to the best output size estimates for creative telescoping known so far [20, 21, 22]. A disadvantage is that these algorithms may not always find the smallest possible output.

The fourth generation of the creative telescoping algorithms, so-called reduction-based algorithms, originates from [14]. The basic idea behind these algorithms is to bring each term $\sigma_n^i f$ of the left-hand side of (1.3) into some kind of normal form modulo all terms that are differences of other terms. Then to find $e_0, \ldots, e_\rho$ amounts to finding a linear dependence among these normal forms. The key advantage of this approach is that it separates the computation of the e_i from the computation of g. This is interesting because a certificate is not always needed, and it is typically much larger (and thus computationally more expensive) than the telescoper, so we may not want to compute it if we don't have to. With previous algorithms there was no way to obtain telescopers without also computing the corresponding certificates, but with fourth generation algorithms there is. So far this approach has only been worked out for several instances in the differential case [14, 16, 15]. The goal of the first part of the present thesis is to give a fourth-generation algorithm for the shift case, namely for the classical setting of hypergeometric telescoping.

⋆ ⋆ ⋆ ⋆ ⋆

Part II. D-finite numbers. In a sense, the theory of D-finite functions generalizes the theory of algebraic functions. Many concepts that have first been introduced for the latter have later been formulated also for the former. In particular, every algebraic function is D-finite (Abel's theorem), and many properties the class of algebraic function enjoys carry over to the class of D-finite functions.

The theory of algebraic functions in turn may be considered as a generalization of the classical and well-understood class of algebraic numbers. The class of algebraic numbers suffers from being relatively small. There are many important numbers, most prominently the numbers e and π, which are not algebraic.

Many larger classes of numbers have been proposed, let us just mention three examples. The first is the class of periods (in the sense of Kontsevich and Zagier [43]). These numbers are defined as the values of multivariate definite integrals of algebraic functions over a semi-algebraic set. In addition to all the algebraic numbers, this class contains important numbers such as π, all zeta constants (the Riemann zeta function evaluated at an integer) and multiple zeta values, but it is so far not known whether for example e, $1/\pi$ or Euler's constant γ are periods (conjecturally they are not). The second example is the class of all numbers that appear as values of so-called G-functions (in the sense of Siegel [58]) at algebraic number arguments [30, 31]. The class of G-functions is a subclass of the class of D-finite functions, and it inherits some useful properties of that class. Among the values that G-functions can assume are π, $1/\pi$, values of elliptic integrals and multiple zeta values, but it is so far not known whether for example e, Euler's constant γ or a Liouville number are such a value (conjecturally not).

Another class of numbers is the class of holonomic constants, studied by Flajolet and Vallée [35, §4]. (We thank Marc Mezzarobba for pointing us to this reference.) A number is *holonomic* if it is equal to the (finite) value of a D-finite function at an algebraic point. The number is further called a *regular holonomic constant* if the evaluation point is an ordinary point of the defining differential equation of the given D-finite function; otherwise it is called a *singular holonomic constant*. Typical examples of the regular holonomic constants are π, $\log(2)$, e and the polylogarithmic value $\mathrm{Li}_4(1/2)$; while several famous constants like Apéry's constant $\zeta(3)$, Catalan's constant G are of singular type.

It is tempting to believe that there is a strong relation between holonomic constants and limits of convergent P-recursive sequences. To make this relation precise, we introduce the class of D-finite numbers in this thesis. Let R be a subring of $\mathbb{C}$ and $\mathbb{F}$ be a subfield of $\mathbb{C}$. A complex number ξ is called *D-finite* (w.r.t. R and $\mathbb{F}$) if it is the limit of a convergent sequence in $R^{\mathbb{N}}$ which is P-recursive over $\mathbb{F}$. We denote by $\mathcal{D}_{R,\mathbb{F}}$ the set of all D-finite numbers with respect to R and $\mathbb{F}$.

It is clear that $\mathcal{D}_{R,\mathbb{F}}$ contains all the elements of R, but it typically contains many further elements. For example, let i be the imaginary unit, then $\mathcal{D}_{\mathbb{Q}(i)}$ contains many (if not all) the periods and, as we will see below, many (if not all) the values of G-functions. In addition, it is not hard to see that e and $1/\pi$ are D-finite numbers. According to Fischler and Rivoal's work [31], also Euler's constant γ and any value of the Gamma function at a rational number are D-finite. (We thank Alin Bostan for pointing us to this reference.)

The definition of D-finite numbers given above involves two subrings of $\mathbb{C}$ as parameters: the ring to which the sequence terms of the convergent sequences are supposed to belong, and the field to which the coefficients of the polynomials in the recurrence equations should belong. Obviously, these choices matter, because we have, for example, $\mathcal{D}_{\mathbb{R},\mathbb{R}} = \mathbb{R} \neq \mathbb{C} = \mathcal{D}_{\mathbb{C},\mathbb{C}}$. Also, since $\mathcal{D}_{\mathbb{Q},\mathbb{Q}}$ is a countable set, we have $\mathcal{D}_{\mathbb{Q},\mathbb{Q}} \neq \mathcal{D}_{\mathbb{R},\mathbb{R}}$. On the other hand, different choices of R and $\mathbb{F}$ may lead to the same classes. For example, we would not get more numbers by allowing $\mathbb{F}$ to be a subring of $\mathbb{C}$ rather than a field, because we can always clear denominators in a defining recurrence. One of our goals is to investigate how R and $\mathbb{F}$ can be modified without changing the resulting class of D-finite numbers.

As a long-term goal, we hope to establish the notion of D-finite numbers as a class that naturally relates to the class of D-finite functions in the same way as the classical class of algebraic numbers relates to the class of algebraic functions.

1.2 Main results and outline

This section is intended to provide an outline of the thesis and the main results.

In Chapter 2, we recall basic notions and facts about hypergeometric terms.

In Chapter 3, our starting point is the Abramov-Petkovšek reduction for hypergeometric terms introduced in [7, 10]. Unfortunately the reduced forms obtained by this reduction are not sufficiently "normal" for our purpose. Therefore, we present a modified version of the reduction process, which does not solve any auxiliary linear difference equation explicitly like the original one and totally separates the summable and non-summable parts of a given hypergeometric term. The outputs of the Abramov-Petkovšek reduction and our modified version share the same required properties. According to the experimental comparison, the modified reduction is also more efficient than the original one.

Chapter 4 is mainly used to connect univariate hypergeometric terms with bivariate ones for later use. We explore some important properties of discrete residual forms by means of rational normal forms [10]. Furthermore, we show that the residual forms are well-behaved with respect to taking linear combinations.

We translate terminology concerning univariate hypergeometric terms to bivariate ones in Chapter 5. Based on the modified version of Abramov-Petkovšek reduction in Chapter 3, we present a new algorithm to compute minimal telescopers for bivariate hypergeometric terms. This new algorithm keeps the key feature of the fourth generation, that is, it separates the computations of telescopers and certificates. Experimental results illustrate that the new algorithm is faster than the classical Zeilberger's algorithm if it returns a normalized certificate; and the new algorithm is much more efficient if it omits certificates.

In Chapter 6, we present a new argument for the termination of the new algorithm in Chapter 5. This new argument provides an independent proof of the existence of telescopers and even enables us to obtain upper and lower bounds for the order of minimal telescopers for hypergeometric terms. Compared to the known bounds in the literature, our bounds are sometimes better and never worse than the known ones. Moreover, we present a variant of the new algorithm by combining our bounds, which improves the new algorithm in some special cases.

In Chapter 7, we review basic notions and useful properties of the class of D-finite functions and P-recursive sequences mainly from [34, 41].

In Chapter 8, we study the class of D-finite numbers, defined as the limits of convergent P-recursive sequences. In general, this class is much larger than the class of algebraic numbers. The definition of the class depends on two subrings of the field of complex numbers. We investigate the possible choices of these two subrings that keep the class unchanged. Moreover, we connect this class with the class of holonomic constants [35] and show that D-finite numbers over the Gaussian rational field are essentially the same as the regular holonomic constants. With this result, certain numbers are easily recognized as belonging to this class, including many periods as well as many values of G-functions.

1.3 Remarks

The main results in Chapters 3 – 5 are joint work with S. Chen, M. Kauers and Z. Li, which have been published in [19]. The main results in Chapter 6 were published in [38]. The main results in Chapter 8 are joint work with M. Kauers, and are in preparation [39].

Part I

Definite Sums of Hypergeometric Terms

Chapter 2

Hypergeometric Terms

In this chapter, we recall basic notions and facts on difference rings (fields) and hypergeometric terms. In addition, we review the context of summability and multiplicative decomposition for hypergeometric terms. These topics are well-known and more details can be found in [50, 28].

2.1 Basic concepts

Let $\mathbb{F}$ be a field of characteristic zero, and $\mathbb{F}(k)$ be the field of rational functions in k over $\mathbb{F}$. Let σ_k be the automorphism that maps $r(k)$ to $r(k+1)$ for every rational function $r \in \mathbb{F}(k)$. The pair $(\mathbb{F}(k), \sigma_k)$ is called a *difference field*. A *difference ring extension* of $(\mathbb{F}(k), \sigma_k)$ is a ring $\mathbb{D}$ containing $\mathbb{F}(k)$ together with a distinguished endomorphism $\sigma_k : \mathbb{D} \to \mathbb{D}$ whose restriction to $\mathbb{F}(k)$ agrees with the automorphism defined before. An element $c \in \mathbb{D}$ is called a constant if $\sigma_k(c) = c$. It is readily seen that all constants in $\mathbb{D}$ form a subring of $\mathbb{D}$, denoted by $C_{\sigma_k,\mathbb{D}}$. In particular, $C_{\sigma_k,\mathbb{D}}$ is a field whenever $\mathbb{D}$ is one. Moreover, we have $C_{\sigma_k,\mathbb{F}(k)} = \mathbb{F}$ according to [9, Theorem 2]. In other words, the set of all constants in $\mathbb{F}(k)$ w.r.t. σ_k is exactly the field $\mathbb{F}$.

Throughout the thesis, for a polynomial $p \in \mathbb{F}[k]$, its degree and leading coefficient are denoted by $\deg_k(p)$ and $\mathrm{lc}_k(p)$, respectively. For convenience, we define the degree of zero to be $-\infty$.

Definition 2.1. *Let $\mathbb{D}$ be a difference ring extension of $\mathbb{F}(k)$. A nonzero element $T \in \mathbb{D}$ is called a* hypergeometric term *over $\mathbb{F}(k)$ if it is invertible and $\sigma_k(T) = rT$ for some $r \in \mathbb{F}(k)$. We call r the* shift-quotient *of T w.r.t. k.*

In the following two chapters, whenever we mention hypergeometric terms, they always belong to some difference ring extension $\mathbb{D}$ of $\mathbb{F}(k)$, unless specified otherwise.

Example 2.2. All nonzero rational functions are hypergeometric. Moreover, the following two classes of combinatorial functions are also hypergeometric.

1. (Exponential functions). $T = c^k$ where $c \in \mathbb{F} \setminus \{0\}$. The shift-quotient of T is $\sigma_k(T)/T = c$.
2. (Factorial terms). $T = (ak)!$ with $a \in \mathbb{N}$ and $a > 0$. The shift-quotient of T is $\sigma_k(T)/T = (ak+a)(ak+a-1)\cdots(ak+1)$.

One can easily show that the product of hypergeometric terms and the reciprocal of a hypergeometric term are again hypergeometric. However, the sum of hypergeometric terms is not necessarily hypergeometric. For example, $2^k + 1$ is not a hypergeometric term although 2^k and 1 both are; otherwise we would have $(2^{k+1}+1)/(2^k+1) \in \mathbb{F}(k)$, and then a straightforward calculation would yield that $2^k \in \mathbb{F}(k)$, a contradiction.

Recall [50, 54] that two hypergeometric terms T_1, T_2 over $\mathbb{F}(k)$ are called *similar* if there exists a rational function $r \in \mathbb{F}(k)$ such that $T_1 = rT_2$. This is an equivalence relation and all rational functions form one equivalence class. By Proposition 5.6.2 in [54], the sum of similar hypergeometric terms is either hypergeometric or zero.

2.2 Hypergeometric summability

Analogous to indefinite integrals of elementary functions in calculus, we consider indefinite sums of hypergeometric terms in shift case. More precisely, given a hypergeometric term $T(k)$, we compute another hypergeometric term $G(k)$ such that

$$T(k) = G(k+1) - G(k).$$

This motivates the notion of hypergeometric summability.

Definition 2.3. *A univariate hypergeometric term T over $\mathbb{F}(k)$ is called* hypergeometric summable*, if there exists another hypergeometric term G such that*

$$T = \Delta_k(G), \quad \textit{where } \Delta_k \textit{ denotes the difference of } \sigma_k \textit{ and the identity map.}$$

We call G an indefinite summation *(or* anti-difference*) of T. If T and G are both rational functions, we also say T is* rational summable*.*

We abbreviate "hypergeometric summable" as "summable" in this thesis.

Example 2.4. All polynomials are summable. Moreover, we see that $k \cdot k!$ is summable since $k \cdot k! = \Delta_k(k!)$, but $k!$ is not which will be shown in Example 3.7.

To solve the problem of indefinite summation, Gosper [36] developed a first complete algorithm which is known as Gosper's algorithm. This is a deterministic procedure. It determines whether or not the input hypergeometric term is summable, and then returns an indefinite summation if the answer is yes. The basic idea is to reduce the summation problem to finding polynomial solutions of a first-order difference equation with polynomial coefficients.

2.3 Multiplicative decomposition

By [7, 10], every hypergeometric term admits a multiplicative decomposition. This enables us to analyze a hypergeometric term by rational functions. To recall it, let us first review the notion of shift-free polynomials and shift-reduced rational functions [7, §1].

Definition 2.5. *A nonzero polynomial $p \in \mathbb{F}[k]$ is said to be* shift-free *if for any nonzero integer i, we have $\gcd(p, \sigma_k^i(p)) = 1$.*

Consequently, no two distinct roots of a shift-free polynomial differ by an integer. The following lemma indicates the relation between shift-freeness and rational summability, whose proof can be found in [1, Proposition 1].

Lemma 2.6. *Let $f = p/q$ be a rational function in $\mathbb{F}(k)$, where $p, q \in \mathbb{F}[k]$ are coprime and $\deg_k(p) < \deg_k(q)$. Further assume that q is shift-free. If there exists a rational function $r \in \mathbb{F}(k)$ such that $f = \Delta_k(r)$, then $f = 0$.*

Definition 2.7. *A nonzero rational function $f = p/q \in \mathbb{F}(k)$ with $p, q \in \mathbb{F}[k]$ coprime, is said to be* shift-reduced *if for any integer i, we have $\gcd(p, \sigma_k^i(q)) = 1$.*

Some basic properties of shift-reduced rational functions are given below.

Lemma 2.8. *Let $f \in \mathbb{F}(k)$ be shift-reduced.*

(i) If there exists a nonzero rational function $r \in \mathbb{F}(k)$ such that $f = \sigma_k(r)/r$, then $r \in \mathbb{F}$ and thus $f = 1$.

(ii) If $f \neq 1$ and there exists $r \in \mathbb{F}[k]$ such that $f\sigma_k(r) - r = 0$, then $r = 0$.

Proof. (i) Suppose that $r = s/t \in \mathbb{F}(k) \setminus \mathbb{F}$, where s, t are coprime and at least one of them does not belong to $\mathbb{F}$. W.l.o.g., we assume that $s \notin \mathbb{F}$. Then there exists a nontrivial factor $p \in \mathbb{F}[k]$ of s such that $\deg_k(p) > 0$. Let

$$\ell = \min\{k \in \mathbb{Z} : \sigma_k^k(p) \mid s\} \quad \text{and} \quad m = \max\{k \in \mathbb{Z} : \sigma_k^k(p) \mid s\}.$$

It follows that $m, \ell \geq 0$ and

- $\sigma_k^{-\ell}(p) \mid s$ but $\sigma_k^{-\ell}(p) \nmid \sigma_k(s)$;
- $\sigma_k^{m+1}(p) \mid \sigma_k(s)$ but $\sigma_k^{m+1}(p) \nmid s$.

Since s and t are coprime, so are $\sigma_k(s)$ and $\sigma_k(t)$. Note that

$$f = \frac{\sigma_k(r)}{r} = \frac{\sigma_k(s)t}{s\sigma_k(t)}.$$

Hence $\sigma_k^{m+1}(p)$ is a nontrivial factor of the numerator of f and $\sigma_k^{-\ell}(p)$ is a nontrivial factor of the denominator of f, a contradiction as f is shift-reduced.

(ii) Suppose that $r \neq 0$. Then

$$f = \frac{r}{\sigma_k(r)} = \frac{\sigma_k(1/r)}{1/r}.$$

Since f is unequal to one, $1/r$ does not belong to $\mathbb{F}$. It follows from (i) that f is not shift-reduced, a contradiction.

□

According to [7, 10], every hypergeometric term T admits a *multiplicative decomposition* SH, where S is in $\mathbb{F}(k)$ and H is another hypergeometric term whose shift-quotient is shift-reduced. We call the shift-quotient $K := \sigma_k(H)/H$ a *kernel* of T w.r.t. k and S a corresponding *shell*. By Lemma 2.8 (i), we know that $K = 1$ if and only if T is a rational function, which is then equal to cS for some constant $c \in C_{\sigma_k,\mathbb{D}}$. Here $\mathbb{D}$ is a difference ring extension of $\mathbb{F}(k)$.

Let $T = SH$ be a multiplicative decomposition, where S is a rational function and H a hypergeometric term with a kernel K. Assume that $T = \Delta_k(G)$ for some hypergeometric term G. A straightforward calculation shows that G is similar to T. So there exists $r \in \mathbb{F}(k)$ such that $G = rH$. One can easily verify that

$$SH = \Delta_k(rH) \iff S = K\sigma_k(r) - r. \tag{2.1}$$

Chapter 3

Additive Decomposition for Hypergeometric Terms [1]

Computing an indefinite summation of a given hypergeometric term is one of the basic problems in the theory of difference equations. In terms of algorithms, Gosper's algorithm [36] is the first complete algorithm for solving this problem. However, when there exist no indefinite summations, Gosper's algorithm is not applicable any more, but we still desire more information so as to handle definite summations. As far as we know, the first description of the non-summable case was given by Abramov. In 1975, Abramov [2] developed a reduction algorithm to compute an additive decomposition of a given rational function, which was improved later by Pirastu and Strehl [55], Paule [51], and by Abramov himself [3], etc. These algorithms decompose a rational function into a summable part and a proper fractional part whose denominator is shift-free and of minimal degree. We refer to it as a *minimal additive decomposition* of the given rational function. According to Lemma 2.6, the fractional part is in fact non-summable. Hence a rational function is summable if and only if the fractional part of a minimal decomposition is zero. In 2001, Abramov and Petkovšek [7, 10] generalized these ideas to the hypergeometric case. We call it the Abramov-Petkovšek reduction. It preserves the minimality of additive decompositions. It loses, however, the separation of summable and non-summable parts. More precisely, given a hypergeometric term T, Abramov-Petkovšek reduction computes two hypergeometric terms T_1, T_2 such that

$$T = \underbrace{\Delta_k(T_1)}_{\text{summable}} + \underbrace{T_2}_{\text{possibly summable}},$$

where T_2 is minimal in some sense. To determine the summability of T, one needs to further solve an auxiliary difference equation [10, §4]. The discrepancy in the reductions for the rational case and the hypergeometric case is unpleasant.

[1] The main results in this chapter are joint work with S. Chen, M. Kauers, Z. Li, published in [19].

In this chapter, in order to obtain the consistency, we modify the Abramov-Petkovšek reduction by a shift variant of the method developed by Bostan et al. [15]. The modified Abramov-Petkovšek reduction not only preserves the minimality of the output additive decomposition, but also decomposes a hypergeometric term as a sum of a summable part and a non-summable part. It laid a solid foundation for the new reduction-based creative telescoping algorithm in Chapter 5. Moreover, we implement the modified reduction in Maple 18 and compare it with the built-in Maple procedure SumDecomposition, which is based on the Abramov-Petkovšek reduction. The experimental results illustrate that the modified Abramov-Petkovšek reduction is more efficient than the original one.

3.1 Abramov-Petkovšek reduction

In the shift case, reduction algorithms for computing minimal additive decompositions of rational functions have been well-developed. More details can be found in [1, 2, 3, 51, 55]. For this reason, we will mainly focus on irrational hypergeometric terms.

The Abramov-Petkovšek reduction [7, 10] is fundamental for the first part of this thesis, which computes a minimal additive decomposition of a given hypergeometric term. It can not only be used to determine hypergeometric summability, but also provide some description of the non-summable part when the given hypergeometric term is not summable. In this sense, the Abramov-Petkovšek reduction is more useful than Gosper's algorithm in some cases, as illustrated by the following example.

Example 3.1. [2] Consider a definite sum

$$\sum_{k=0}^{\infty} T(k), \quad \text{where } T(k) = \frac{1}{(k^4+k^2+1)k!}.$$

Applying Gosper's algorithm shows that T is not summable, and thus we cannot evaluate the sum in terms of indefinite summations. Applying the Abramov-Petkovšek reduction to T, however, yields

$$T(k) = \Delta_k\left(\frac{k^2}{2(k^2-k+1)k!}\right) + \frac{1}{2k!}.$$

Summing over k from zero to infinity and using the telescoping sum technique leads to a "closed form" of the summation,

$$\sum_{k=0}^{\infty} T(k) = \lim_{k\to\infty}\left(\frac{k^2}{2(k^2-k+1)k!}\right) - 0 + \sum_{k=0}^{\infty}\frac{1}{2k!} = \frac{1}{2}e.$$

Thus the given sum in fact admits a simple form.

[2] We thank Yijun Chen for providing this example.

To describe the Abramov-Petkovšek reduction concisely, we need a notational convention and a technical definition.

Convention 3.2. *Let T be a hypergeometric term over $\mathbb{F}(k)$ with a kernel K and a corresponding shell S. Then $T = SH$, where H is a hypergeometric term whose shift-quotient is K. Further write $K = u/v$, where u, v are nonzero polynomials in $\mathbb{F}[k]$ with $\gcd(u, v) = 1$.*

Moreover, we let $\mathbb{U}_T$ be the union of $\{0\}$ and the set of summable hypergeometric terms that are similar to T, and $\mathbb{V}_K = \{K\sigma_k(r) - r \mid r \in \mathbb{F}(k)\}$.

With the above convention, it is clear that $\mathbb{U}_T$ and $\mathbb{V}_K$ are both $\mathbb{F}$-linear vector spaces and $\mathbb{U}_T = \mathbb{U}_H$ since H is similar to T. Then (2.1) translates into

$$SH \equiv_k 0 \mod \mathbb{U}_H \iff S \equiv_k 0 \mod \mathbb{V}_K. \tag{3.1}$$

These congruences enable us to shorten expressions.

Definition 3.3. *With Convention 3.2, a nonzero polynomial p in $\mathbb{F}[k]$ is said to be* strongly coprime *with K if $\gcd(p, \sigma_k^{-i}(u)) = \gcd(p, \sigma_k^{i}(v)) = 1$ for all $i \geq 0$.*

The proof of Lemma 3 in [7] contains a reduction algorithm whose inputs and outputs are given below.

Algorithm 3.4 (Abramov-Petkovšek Reduction)**.**
Input: Two rational functions $K, S \in \mathbb{F}(k)$ as defined in Convention 3.2.
Output: A rational function $S_1 \in \mathbb{F}(k)$ and two polynomials $b, w \in \mathbb{F}[k]$ such that b is shift-free and strongly coprime with K, and the following equation holds:

$$S = K\sigma_k(S_1) - S_1 + \frac{w}{b \cdot \sigma_k^{-1}(u) \cdot v}. \tag{3.2}$$

The algorithm contained in the proof of Lemma 3 in [7] is described as pseudo code on page 4 of the same paper, in which the last ten lines make the denominator of the rational function V in its output minimal in some technical sense. We shall not execute these lines. Then the algorithm will compute two rational functions U_1 and U_2. They correspond to S_1 and $w/(b\,\sigma_k^{-1}(u)\,v)$ in (3.2), respectively.

We slightly modify the output of the Abramov-Petkovšek reduction so that we can analyze it more easily in the next section. Note that K is shift-reduced and b is strongly coprime with K. Thus, b, $\sigma_k^{-1}(u)$ and v are pairwise coprime. By partial fraction decomposition, (3.2) can be rewritten as

$$S = K\sigma_k(S_1) - S_1 + \left(\frac{a}{b} + \frac{p_1}{\sigma_k^{-1}(u)} + \frac{p_2}{v}\right),$$

where $a, p_1, p_2 \in \mathbb{F}[k]$. Furthermore, set $r = p_1/\sigma_k^{-1}(u)$ and a direct calculation yields

$$r = K\sigma_k(-r) - (-r) + \frac{\sigma_k(p_1)}{v}.$$

Update S_1 to be $S_1 - r$ and set p to be $\sigma_k(p_1) + p_2$. Then

$$S = K\sigma_k(S_1) - S_1 + \left(\frac{a}{b} + \frac{p}{v}\right). \tag{3.3}$$

This modification leads to shell reduction specified below.

Algorithm 3.5 (Shell Reduction)**.**
Input: Two rational functions $K, S \in \mathbb{F}(k)$ as defined in Convention 3.2.
Output: A rational function $S_1 \in \mathbb{F}(k)$ and three polynomials $a, b, p \in \mathbb{F}[k]$ such that b is shift-free and strongly coprime with K, and that (3.3) holds.

Shell reduction provides us with a necessary condition on summability.

Proposition 3.6. *With Convention 3.2, let a, b, p be polynomials in $\mathbb{F}[k]$ where b is shift-free and strongly coprime with K. Assume further that (3.3) holds. If T is summable, then a/b belongs to $\mathbb{F}[k]$.*

Proof. Recall that $T = SH$ by Convention 3.2 and it has a kernel K and a corresponding shell S. It follows from (3.1) and (3.3) that

$$T \equiv_k \left(\frac{a}{b} + \frac{p}{v}\right) H \mod \mathbb{U}_H.$$

Thus, T is summable if and only if $(a/b + p/v)\, H$ is summable.

Set $H' = (1/v)H$, which has a kernel $K' = u/\sigma_k(v)$. Note that since b is strongly coprime with K, so is K'. Applying [10, Theorem 11] to $(av/b + p)\, H'$, which is equal to $(a/b + p/v)\, H$, yields that $(av/b + p)$ is a polynomial. Thus, a/b is a polynomial because b is coprime with v. □

The above proposition enables us to determine hypergeometric summability directly in some instances.

Example 3.7. Let $T = k^2 k!/(k+1)$. Then it has a kernel $K = k+1$ and the shell $S = k^2/(k+1)$. Shell reduction yields

$$S \equiv_k -\frac{1}{k+2} + \frac{k}{v} \mod \mathbb{V}_K,$$

where $v = 1$. By Proposition 3.6, T is not summable. By a similar argument as before, one sees that $k!$ is indeed not summable as mentioned in Example 2.4.

Note that $a/b + p/v$ in (3.3) can be nonzero for a summable T.

Example 3.8. Let $T = k \cdot k!$ whose kernel is $K = k+1$ and shell is $S = k$. Then

$$S \equiv_k \frac{k}{v} \mod \mathbb{V}_K,$$

where $v = 1$. But T is summable as it is equal to $\Delta_k\,(k!)$.

The above example illustrates that neither shell reduction nor the Abramov-Petkovšek reduction can decide summability directly when $a/b \in \mathbb{F}[k]$ in (3.3). One way to proceed is, according to [10], to find a polynomial solution of the auxiliary first-order linear difference equation $u\sigma_k(z) - vz = av/b + p$, under the hypotheses of Algorithm 3.5. If there is a polynomial solution, say $f \in \mathbb{F}[k]$, then $T = \Delta_k((S_1 + f)H)$; otherwise T is not summable. This method reduces the summability problem to solving a linear system over $\mathbb{F}$. We show in the next section how this can be avoided so as to read out summability directly from a minimal decomposition.

3.2 Modified Abramov-Petkovšek reduction

After the shell reduction described in (3.3), it remains to check the summability of the hypergeometric term $(a/b + p/v)\,H$. In the rational case, i.e., when the kernel K is one, the rational function $a/b+p/v$ in (3.3) can be further reduced to a/b with $\deg_k(a) < \deg_k(b)$, because all polynomials are rational summable. However, a hypergeometric term with a polynomial shell is not necessarily summable, for example, $k!$ has a polynomial shell but it is not summable.

In this section, we define the notion of discrete residual forms for rational functions, and present a discrete variant of the polynomial reduction for hyperexponential functions given in [15]. This variant not only leads to a direct way to decide summability, but also reduces the number of terms of p in (3.3).

3.2.1 Discrete residual forms

With Convention 3.2, we define an $\mathbb{F}$-linear map

$$\begin{array}{rccc} \phi_K: & \mathbb{F}[k] & \to & \mathbb{F}[k] \\ & p & \mapsto & u\sigma_k(p) - vp, \end{array}$$

for all $p \in \mathbb{F}[k]$. We call ϕ_K the *map for polynomial reduction w.r.t. K*.

Lemma 3.9. *Let*

$$\mathbb{W}_K = \operatorname{span}_{\mathbb{F}} \left\{ k^{\ell} \mid \ell \in \mathbb{N}, \ell \neq \deg_k(p) \text{ for all nonzero } p \in \operatorname{im}(\phi_K) \right\}.$$

Then $\mathbb{F}[k] = \operatorname{im}(\phi_K) \oplus \mathbb{W}_K$.

Proof. By the definition of $\mathbb{W}_K$, $\operatorname{im}(\phi_K) \cap \mathbb{W}_K = \{0\}$. The same definition also implies that, for every nonnegative integer m, there exists a polynomial f_m in $\operatorname{im}(\phi_K) \cup \mathbb{W}_K$ such that the degree of f_m is equal to m. The set $\{f_0, f_1, f_2, \ldots\}$ forms an $\mathbb{F}$-basis of $\mathbb{F}[k]$. Thus $\mathbb{F}[k] = \operatorname{im}(\phi_K) \oplus \mathbb{W}_K$. □

In view of the above lemma, we call $\mathbb{W}_K$ the *standard complement of* $\operatorname{im}(\phi_K)$. Note that if $K = 1$, then $\phi_K = \Delta_k$ and $\mathbb{W}_K = \{0\}$ since all polynomials are rational summable. According to Lemma 3.9, every polynomial $p \in \mathbb{F}$ can be uniquely decomposed as $p = p_1 + p_2$ where $p_1 \in \operatorname{im}(\phi_K)$ and $p_2 \in \mathbb{W}_K$.

Lemma 3.10. *With Convention 3.2, let p be a polynomial in $\mathbb{F}[k]$. Then there exists a polynomial $q \in \mathbb{W}_K$ such that $p/v \equiv_k q/v \mod \mathbb{V}_K$.*

Proof. Let $q \in \mathbb{F}[k]$ be the projection of p on $\mathbb{W}_K$. Then there exists f in $\mathbb{F}[k]$ such that $p = \phi_K(f) + q$, that is, $p = u\sigma_k(f) - vf + q$. So $p/v = K\sigma_k(f) - f + q/v$, which is equivalent to $p/v \equiv_k q/v \mod \mathbb{V}_K$. □

Remark 3.11. Replacing the polynomial p in the above lemma by vp, we see that, for every polynomial $p \in \mathbb{F}[k]$, there exists $q \in \mathbb{W}_K$ such that $p \equiv_k q/v \mod \mathbb{V}_K$.

By Lemma 3.10 and Remark 3.11, (3.3) implies that

$$S \equiv_k \frac{a}{b} + \frac{q}{v} \mod \mathbb{V}_K, \tag{3.4}$$

where a, b, q are polynomials in $\mathbb{F}[k]$, $\deg_k(a) < \deg_k(b)$, b is shift-free and strongly coprime with K, and $q \in \mathbb{W}_K$. The congruence (3.4) motivates us to translate the notion of (continuous) residual forms [15] into the discrete setting.

Definition 3.12. *With Convention 3.2, we further let f be a rational function in $\mathbb{F}(k)$. Another rational function r in $\mathbb{F}(k)$ is called a* (discrete) residual form *of f w.r.t. K if there exist a, b, q in $\mathbb{F}[k]$ such that*

$$f \equiv_k r \mod \mathbb{V}_K \quad \text{and} \quad r = \frac{a}{b} + \frac{q}{v},$$

where $\deg_k(a) < \deg_k(b)$, b is shift-free and strongly coprime with K, and q belongs to $\mathbb{W}_K$. For brevity, we just say that r is a residual form *w.r.t. K if f is clear from the context. Moreover, we call b the* significant denominator *of r if $\gcd(a, b) = 1$ and b is monic, i.e., $\operatorname{lc}_k(b) = 1$.*

Residual forms help us to decide summability, as shown below.

Proposition 3.13. *With Convention 3.2, we further assume that r is a nonzero residual form w.r.t. K. Then the hypergeometric term rH is not summable.*

Proof. Suppose that rH is summable. Let $r = a/b + q/v$, where $a, b, q \in \mathbb{F}[k]$, $\deg_k(a) < \deg_k(b)$, b is shift-free and strongly coprime with K, and $q \in \mathbb{W}_K$. By Proposition 3.6, a/b is a polynomial. Since $\deg_k(a) < \deg_k(b)$, we have $a = 0$ and thus the term $(q/v)H$ is summable. It follows from (2.1) that there exists a rational function $w \in \mathbb{F}(k)$ such that $u\sigma_k(w) - vw = q$. Thus, $w \in \mathbb{F}[k]$ by Theorem 5.2.1 in [54, page 76], which implies that q belongs to $\operatorname{im}(\phi_K)$. But q also belongs to $\mathbb{W}_K$. By Lemma 3.9, $q = 0$ and then $r = 0$, a contradiction. □

With Convention 3.2, let r be a residual form of the shell S w.r.t. K. Then

$$SH \equiv_k rH \mod \mathbb{U}_H$$

according to (3.1) and (3.4). By Proposition 3.13, SH is summable if and only if $r = 0$. Thus, determining the summability of a hypergeometric term T amounts to computing a residual form of a corresponding shell with respect to a kernel of T, which is studied below.

3.2.2 Polynomial reduction

With Convention 3.2, to compute a residual form of a rational function, we project a polynomial on $\operatorname{im}(\phi_K)$ and also its standard complement $\mathbb{W}_K$, both defined in the previous subsection. If the given term T is a rational function, i.e., $K = 1$, then this projection is trivial because $\operatorname{im}(\phi) = \operatorname{im}(\Delta_k) = \mathbb{F}[k]$ and $\mathbb{W}_K = \{0\}$.

Now we assume $K \neq 1$ and let $\mathbb{B}_K = \{\phi_K(k^i) \mid i \in \mathbb{N}\}$. Since $K \neq 1$, the $\mathbb{F}$-linear map ϕ_K is injective by Lemma 2.8 (ii). So $\mathbb{B}_K$ is an $\mathbb{F}$-basis of $\operatorname{im}(\phi_K)$, which allows us to construct an echelon basis of $\operatorname{im}(\phi_K)$. By an echelon basis, we mean an $\mathbb{F}$-basis in which distinct elements have distinct degrees. We can easily project a polynomial using an echelon basis and linear elimination.

To construct an echelon basis, we rewrite $\operatorname{im}(\phi_K)$ as

$$\operatorname{im}(\phi_K) = \{u\Delta_k(p) - (v-u)p \mid p \in \mathbb{F}[k]\}\,.$$

Set $\alpha_1 = \deg_k(u)$, $\alpha_2 = \deg_k(v)$, and $\beta = \deg_k(v-u)$. Moreover, set

$$\tau_K = \frac{\operatorname{lc}_k(v-u)}{\operatorname{lc}_k(u)},$$

which is nonzero since $K \neq 1$ and let p be a nonzero polynomial in $\mathbb{F}[k]$.

We make the following case distinction.

Case 1. $\beta > \alpha_1$. Then $\beta = \alpha_2$, and

$$\phi_K(p) = -\operatorname{lc}_k(v-u)\operatorname{lc}_k(p)k^{\alpha_2+\deg_k(p)} + \text{ lower terms.}$$

So $\mathbb{B}_K$ is an echelon basis of $\operatorname{im}(\phi_K)$, in which $\deg_k(\phi_K(k^i))$ is equal to $\alpha_2 + i$ for all $i \in \mathbb{N}$. Accordingly, $\mathbb{W}_K$ has an echelon basis $\{1, k, \ldots, k^{\alpha_2-1}\}$ and has dimension α_2.

Case 2. $\beta = \alpha_1$. Then

$$\phi_K(p) = -\operatorname{lc}_k(v-u)\operatorname{lc}_k(p)k^{\alpha_1+\deg_k(p)} + \text{ lower terms.}$$

So $\mathbb{B}_K$ is an echelon basis of $\operatorname{im}(\phi_K)$, in which $\deg_k(\phi_K(k^i))$ is equal to $\alpha_1 + i$ for all $i \in \mathbb{N}$. Accordingly, $\mathbb{W}_K$ has an echelon basis $\{1, k, \ldots, k^{\alpha_1-1}\}$ and has dimension α_1.

Case 3. $\beta < \alpha_1 - 1$. If $\deg_k(p) = 0$, then $\phi_K(p) = (u - v)p$. Otherwise, we have

$$\phi_K(p) = \deg_k(p)\,\mathrm{lc}_k(u)\,\mathrm{lc}_k(p)k^{\alpha_1+\deg_k(p)-1} + \text{ lower terms.}$$

It follows that $\mathbb{B}_K$ is an echelon basis of $\mathrm{im}(\phi_K)$, in which $\deg_k(\phi_K(1)) = \beta$ and

$$\deg_k(\phi_K(k^i)) = \alpha_1 + i - 1 \quad \text{for all } i \geq 1.$$

Accordingly, $\mathbb{W}_K$ has an echelon basis $\{1, \ldots, k^{\beta-1}, k^{\beta+1}, \ldots, k^{\alpha_1-1}\}$ and has dimension $\alpha_1 - 1$.

Case 4. $\beta = \alpha_1 - 1$ and τ_K is not a positive integer. Then

$$\phi_K(p) = (\deg_k(p)\,\mathrm{lc}_k(u) - \mathrm{lc}_k(v - u))\,\mathrm{lc}_k(p)k^{\alpha_1+\deg_k(p)-1} + \text{ lower terms.} \quad (3.5)$$

Accordingly, $\mathbb{B}_K$ is an echelon basis of $\mathrm{im}(\phi_K)$, in which $\deg_k(\phi_K(k^i)) = \alpha_1+i-1$ for all $i \in \mathbb{N}$. Accordingly, $\mathbb{W}_K$ has an echelon basis $\{1, k, \ldots, k^{\alpha_1-2}\}$ and has dimension $\alpha_1 - 1$.

Case 5. $\beta = \alpha_1 - 1$ and τ_K is a positive integer. It follows from (3.5) that for $i \neq \tau_K$, we have $\deg_k(\phi_K(k^i)) = \alpha_1 + i - 1$. Moreover, for every polynomial p of degree τ_K, $\phi_K(p)$ is of degree less than $\alpha_1 + \tau_K - 1$. So any echelon basis of $\mathrm{im}(\phi_K)$ does not contain a polynomial of degree $\alpha_1 + \tau_K - 1$. Set

$$\mathbb{B}'_K = \left\{\phi_K(k^i) \mid i \in \mathbb{N}, i \neq \tau_K\right\}.$$

Reducing $\phi_K(k^{\tau_K})$ by the polynomials in $\mathbb{B}'_K$, we obtain a polynomial p' with degree less than $\alpha_1 - 1$. Since $\mathbb{B}_K$ is an $\mathbb{F}$-basis and $\mathbb{B}'_K \subset \mathbb{B}_K$, $p' \neq 0$. Hence $\mathbb{B}'_K \cup \{p'\}$ is an echelon basis of $\mathrm{im}(\phi_K)$. Consequently, $\mathbb{W}_K$ has an echelon basis $\{1, k, \ldots, k^{\deg_k(p')-1}, k^{\deg_k(p')+1}, \ldots, k^{\alpha_1-2}, k^{\alpha_1+\tau_K-1}\}$. The dimension of $\mathbb{W}_K$ is equal to $\alpha_1 - 1$.

Example 3.14. Let $K = (k^4+1)/(k+1)^4$, which is shift-reduced. Then $\tau_K = 4$. According to Case 5, $\mathrm{im}(\phi_K)$ has an echelon basis

$$\{\phi_K(p)\} \cup \{\phi_K(k^m) \mid m \in \mathbb{N}, m \neq 4\},$$

where $p = k^4 + k/3 + 1/2$, $\phi_K(p) = (5/3)k^2 + 2k + 4/3$, and

$$\phi_K(k^m) = (m - 4)k^{m+3} + \text{lower terms.}$$

Therefore, $\mathbb{W}_K$ has a basis $\{1, k, k^7\}$.

From the above case distinction and example one observes that, although the degree of a polynomial in the standard complement depends on τ_K, which may be arbitrarily high, the number of its terms depends merely on the degrees of u and v. We record this observation in the next proposition.

Proposition 3.15. *With Convention 3.2, further let* $\alpha_1 = \deg_k(u), \alpha_2 = \deg_k(v)$ *and* $\beta = \max\{0, \deg_k(v-u)\}$*. Then there exists a set* $\mathcal{P} \subset \{k^i \mid i \in \mathbb{N}\}$ *with*

$$|\mathcal{P}| \leq \max\{\alpha_1, \alpha_2\} - [\![\beta \leq \alpha_1 - 1]\!]$$

such that every polynomial in $\mathbb{F}[k]$ *can be reduced modulo* $\mathrm{im}(\phi_K)$ *to an* $\mathbb{F}$*-linear combination of the elements in* $\mathcal{P}$*. Note that here the expression* $[\![\beta \leq \alpha_1 - 1]\!]$ *equals* 1 *if* $\beta \leq \alpha_1 - 1$*, otherwise it is* 0*.*

Proof. If $K = 1$, then $\mathrm{im}(\phi_K) = \mathrm{im}(\Delta_k) = \mathbb{F}[k]$ and $\alpha_1 = \alpha_2 = \beta = 0$. Taking $\mathcal{P} = \emptyset$ completes the proof. Otherwise $K \neq 1$. By the above case distinction, the dimension of $\mathbb{W}_K$ over $\mathbb{F}$ is no more than $\max\{\alpha_1, \alpha_2\} - [\![\beta \leq \alpha_1 - 1]\!]$. The lemma follows. □

When $K \neq 1$, the above case distinction enables one to find an infinite sequence $p_0, p_1, \ldots$ in $\mathbb{F}[k]$ such that

$$\mathbb{E}_K = \{\phi_K(p_i) | i \in \mathbb{N}\} \text{ with } \deg_k \phi_K(p_i) < \deg_k \phi_K(p_{i+1}),$$

is an echelon basis of $\mathrm{im}(\phi_K)$. This basis allows one to project a polynomial on $\mathrm{im}(\phi_K)$ and $\mathbb{W}_K$, respectively. In the first four cases, the p_i's can be chosen as powers of k. But in the last case, one of the p_i's is not necessarily a monomial as shown in Example 3.14.

Based on the above discussion, we have the following algorithm.

Algorithm 3.16 (Polynomial Reduction)**.**
Input: A polynomial $p \in \mathbb{F}[k]$ and a shift-reduced rational function $K \in \mathbb{F}(k)$.
Output: Two polynomials $f, q \in \mathbb{F}[k]$ such that $q \in \mathbb{W}_K$ and $p = \phi_K(f) + q$.

1 If $p = 0$, then set $f = 0$ and $q = 0$; and return.

2 If $K = 1$, then set $f = \Delta_k^{-1}(p)$ and $q = 0$; and return.

2 Set $d = \deg_k(p)$.
Find the subset $\mathbb{P} = \{p_{i_1}, \ldots, p_{i_s}\}$ consisting of the preimages of all polynomials in the echelon basis $\mathbb{E}_K$ whose degrees are at most d.

3 For $j = s, s-1, \ldots, 1$, perform linear elimination to
find $c_s, c_{s-1}, \ldots, c_1 \in \mathbb{F}$ such that $p - \sum_{j=1}^{s} c_j \phi_K(p_{i_j}) \in \mathbb{W}_K$.

4 Set $f = \sum_{j=1}^{s} c_j p_{i_j}$ and $q = p - \phi_K(f)$; and return.

Together with Algorithms 3.5 and 3.16, we are ready to present a modified version of the Abramov-Petkovšek reduction, which is summarized as Algorithm 3.17. This modified reduction determines summability without solving any auxiliary difference equations explicitly.

Algorithm 3.17 (Modified Abramov-Petkovšek Reduction)**.**
Input: A hypergeometric term T over $\mathbb{F}(k)$.
Output: A hypergeometric term H with a kernel K and two rational functions $f, r \in \mathbb{F}(k)$ such that r is a residual form w.r.t. K and

$$T = \Delta_k(fH) + rH. \tag{3.6}$$

1 Find a kernel K and a corresponding shell S of T.

2 Apply Algorithm 3.5, namely the shell reduction, to S w.r.t. K to find three polynomials $b, s, t \in \mathbb{F}[k]$ and a rational function $g \in \mathbb{F}(k)$ such that b is shift-free and strongly coprime with K, and

$$T = \Delta_k\left(gH\right) + \left(\frac{s}{b} + \frac{t}{v}\right)H, \tag{3.7}$$

where $\sigma_k(H)/H = K$ and v is the denominator of K.

3 Set p and a to be the quotient and remainder of s and b, respectively.

4 Apply Algorithm 3.16, namely the polynomial reduction, to $vp + t$ to find $h \in \mathbb{F}[k]$ and $q \in \mathbb{W}_K$ such that $vp + t = \phi_K(h) + q$.

5 Set $f = g + h$ and $r = a/b + q/v$; and return H, f and r.

Theorem 3.18. *With Convention 3.2, Algorithm 3.17 computes a rational function f in $\mathbb{F}(k)$ and a residual form r w.r.t. K such that (3.6) holds. Moreover, T is summable if and only if $r = 0$.*

Proof. Recall that $T = SH$, where H has a kernel K and S is a rational function. Applying shell reduction to S w.r.t. K yields (3.7), which can be rewritten as

$$T = \Delta_k\left(gH\right) + \left(\frac{a}{b} + \frac{vp + t}{v}\right)H,$$

where a and p are given in step 3 of Algorithm 3.17. The polynomial reduction in step 4 yields that $vp + t = u\sigma_k(h) - vh + q$. Substituting this into (3.7) gives

$$\begin{aligned} T &= \Delta_k(gH) + \left(K\sigma_k(h) - h\right)H + \left(\frac{a}{b} + \frac{q}{v}\right)H \\ &= \Delta_k((g+h)H) + rH, \end{aligned}$$

where $r = a/b + q/v$. Thus, (3.6) holds. By Proposition 3.13, T is summable if and only if r is equal to zero. $\square$

Example 3.19. Let T be the same hypergeometric term as in Example 3.7. Then we know $K = k+1$ and $S = k^2/(k+1)$. Set $H = k!$. By the shell reduction in Example 3.7,

$$T = \Delta_k\left(\frac{-1}{k+1}H\right) + \left(\frac{-1}{k+2} + \frac{k}{v}\right)H \quad \text{with } v = 1.$$

Applying the polynomial reduction to $(k/v)H$ yields $(k/v)H = \Delta_k(1 \cdot H)$. Combining the above steps, we decompose T as

$$T = \Delta_k\left(\frac{k}{k+1}H\right) - \frac{1}{k+2}H.$$

So the input term T is not summable, which is consistent with Example 3.7.

Example 3.20. Let T be the same hypergeometric term as in Example 3.8. Then we know $K = k+1$ and $S = k$. Set $H = k!$. The shell reduction in Example 3.8 gives

$$T = \Delta_k(0) + \frac{k}{v}H \quad \text{with } v = 1.$$

By the polynomial reduction, $(k/v)H = \Delta_k(1 \cdot H)$, and hence $T = \Delta_k(k!)$, implying that T is summable.

Remark 3.21. With the notation given in step 5 of Algorithm 3.17, we can rewrite rH as $(s_1/s_2)G$, where $s_1 = av + bq$, $s_2 = b$, and $G = H/v$. It follows from the case distinction in this subsection that the degree of s_1 is bounded by λ given in [7, Theorem 8]. The polynomial s_2 is equal to b in (3.2) whose degree is minimal by [7, Theorem 3]. Moreover, $\sigma_k(G)/G$ is shift-reduced because $\sigma_k(H)/H$ is. These are exactly the same required properties of the output of the Abramov-Petkovšek reduction [7]. In summary, the modified reduction preserves all required conditions for the outputs of the original reduction, namely, it also returns a minimal additive decomposition of a given hypergeometric term.

It is remarkable that the modified Abramov-Petkovšek reduction also applies to Example 3.1. Moreover, compared to the original reduction, the modified reduction not only further decomposes a hypergeometric term into a summable part and a non-summable part, but also provides a new method for proving identities in several examples.

Example 3.22. Consider the following two famous combinatorial identities

$$\sum_{k=0}^{\infty}\binom{n}{k} = 2^n \quad \text{and} \quad \sum_{k=0}^{\infty}\binom{n}{k}^2 = \binom{2n}{n}.$$

Many methods can be used to prove the above identities. In this example, we use the modified Abramov-Petkovšek reduction.

Referring to the first identity, we apply the modified reduction to the summand and get

$$\binom{n}{k} = \Delta_k\left(-\frac{1}{2}\binom{n}{k}\right) + \frac{n+1}{2(k+1)}\binom{n}{k}.$$

Summing over k from zero to infinity and using the telescoping sum technique yields

$$\begin{aligned}\sum_{k=0}^{\infty}\binom{n}{k} &= \lim_{k\to\infty}\left(-\frac{1}{2}\binom{n}{k}\right) - \left(-\frac{1}{2}\right) + \sum_{k=0}^{\infty}\frac{n+1}{2(k+1)}\binom{n}{k}\\ &= \frac{1}{2} + \frac{1}{2}\sum_{k=0}^{\infty}\binom{n+1}{k+1} = \frac{1}{2}\sum_{k=0}^{\infty}\binom{n+1}{k}.\end{aligned}$$

Let $F(n) = \sum_{k=0}^{\infty}\binom{n}{k}$. Then the above equation can be rewritten as a first-order difference equation about $F(n)$,

$$F(n+1) - 2F(n) = 0.$$

It is readily seen that 2^n is a solution. Since $2^0 = 1 = F(0)$, we have $F(n) = 2^n$, which proves the first identity.

For the second identity, applying the modified reduction to the summand yields

$$\binom{n}{k}^2 = \Delta_k\left(-\frac{1}{2}\frac{n+2k+1}{2n+1}\binom{n}{k}^2\right) + \frac{1}{2}\frac{(n+1)^3}{(2n+1)(k+1)^2}\binom{n}{k}^2.$$

Along entirely similar lines as the first identity, we get a first-order difference equation

$$(n+1)F(n+1) - 2(2n+1)F(n) = 0,$$

where $F(n) = \sum_{k=0}^{\infty}\binom{n}{k}^2$. The second identity follows since $\binom{2n}{n}$ satisfies the same difference equation and has the same initial value at zero as $F(n)$.

However, the Abramov-Petkovšek reduction applies to neither the first identity nor the second one.

3.3 Implementation and timings

We have implemented Algorithms 3.5 – 3.17 in Maple 18. The procedures are included in our Maple package **ShiftReductionCT**. A detailed description of this package is given in Appendix A.

In order to get an idea about the efficiency of our new procedures, we compared their runtime and memory requirements to the performance of known algorithms. Since the comparisons of runtime and memory requirements almost have the same indication, we only show that of runtime in this section. One can refer to Appendix B for the memory requirements. All timings are measured in seconds on a Linux computer with 388Gb RAM and twelve 2.80GHz Dual core processors. The computations for this experiment did not use any parallelism. For brevity, we denote

- G: the procedure Gosper in **SumTools[Hypergeometric]**, which is based on Gosper's algorithm;
- AP: the procedure SumDecomposition in **SumTools[Hypergeometric]**, which is based on the Abramov-Petkovšek reduction;
- S: the procedure IsSummable in **ShiftReductionCT**, which determines hypergeometric summability in a similar way as Gosper's algorithm;
- MAP: the procedure ModifiedAbramovPetkovsekReduction in **ShiftReductionCT**, which is based on the modified reduction.

We make the following two comparisons. One is for random hypergeometric terms, while the other is for summable hypergeometric terms.

Example 3.23 (Random hypergeometric terms)**.** Consider hypergeometric terms of the form

$$T(k) = \frac{f(k)}{g_1(k)g_2(k)} \prod_{\ell=m_0}^{k} \frac{u(\ell)}{v(\ell)}, \tag{3.8}$$

where $f \in \mathbb{Z}[k]$ of degree 20, $m_0 \in \mathbb{F}$ is fixed, u, v are both the product of two polynomials in $\mathbb{Z}[k]$ of degree one, $g_i = p_i\sigma_k^{\lambda}(p_i)\sigma_k^{\mu}(p_i)$ with $p_i \in \mathbb{Z}[k]$ of degree 10, $\lambda, \mu \in \mathbb{N}$, and $\alpha, \beta \in \mathbb{Z}$. For a selection of random terms of this type for different choices of λ and μ, Table 3.1 compares the timings of the four procedures described above.

(λ,μ)	G	AP	S	MAP
$(0,0)$	0.09	0.16	0.12	0.12
$(5,5)$	0.36	3.99	0.37	0.45
$(10,10)$	0.66	13.70	0.65	0.86
$(10,20)$	4.05	40.82	1.41	2.53
$(10,30)$	12.13	294.52	2.22	6.26
$(10,40)$	19.09	564.71	3.31	14.11
$(10,50)$	34.89	865.01	4.76	26.02

Table 3.1: Timing comparison of Gosper's algorithm, the Abramov-Petkovšek reduction and the modified version for random hypergeometric terms (in seconds)

Example 3.24 (Summable hypergeometric terms)**.** Consider the summable terms $\sigma_k(T) - T$, where T is of the form (3.8). Similarly, for the same choices of λ and μ as the previous example, Table 3.2 compares the timings of the four procedures.

(λ, μ)	G	AP	S	MAP
$(0, 0)$	1.13	2.34	1.27	1.26
$(5, 5)$	1.86	6.44	1.59	1.59
$(10, 10)$	2.22	13.78	1.63	1.63
$(10, 20)$	7.09	29.76	2.09	2.10
$(10, 30)$	19.61	57.63	2.34	2.33
$(10, 40)$	30.83	95.31	2.49	2.49
$(10, 50)$	64.69	168.72	2.69	2.69

Table 3.2: Timing comparison of Gosper's algorithm, the Abramov-Petkovšek reduction and the modified version for summable hypergeometric terms (in seconds)

Notice that μ is the dispersion of g_i and itself in (3.8) (see Definition 4.13). From Table 3.1 and Table 3.2, we observe that for different procedures, the effect of dispersion is quite different. Figure 3.1 describes the effect of dispersion on the above four procedures in Example 3.23 and Example 3.24.

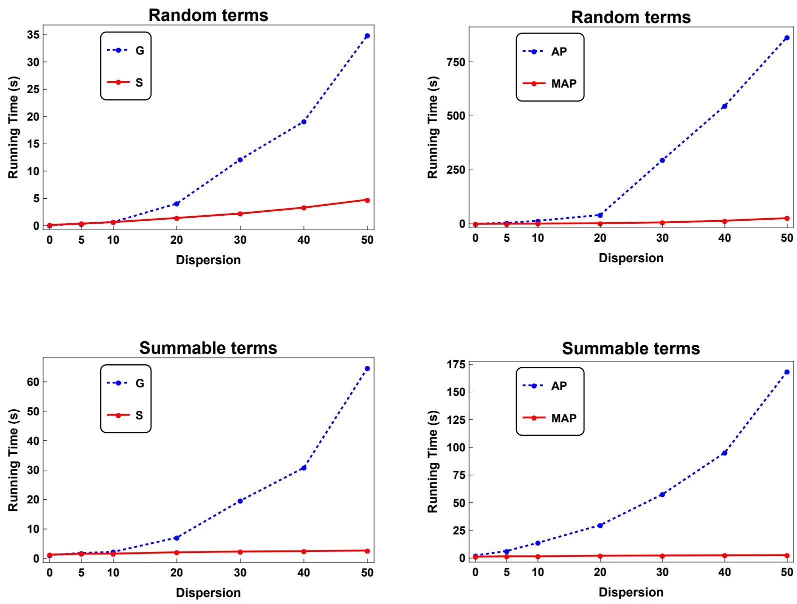

Figure 3.1: Comparison of the effect of dispersion on Gosper's algorithm, the Abramov-Petkovšek reduction and the modified version for Examples 3.23 and 3.24

Chapter 4

Further Properties of Residual Forms [1]

In Chapter 3, we presented a modified version of the Abramov-Petkovšek reduction, which decomposes a univariate hypergeometric term into a summable part and a non-summable part. Moreover, the non-summable part is described by a residual form. In [15], the authors used the Hermite reduction for univariate hyperexponential functions to compute telescopers for bivariate hyperexponential functions. It allows one to separate the computation of telescopers from that of certificates. We try to translate their idea into the hypergeometric setting.

We call a bivariate nonzero term *hypergeometric* if its shift-quotients with respect to the two variables are both rational functions. Given a hypergeometric term $T(n,k)$. Let σ_n and σ_k be the shift operators w.r.t. n and k, respectively. Applying the modified Abramov-Petkovšek reduction to T as well as its shifts $\sigma_n(T),\dots,\sigma_n^i(T)$ w.r.t. k, where i is a nonnegative integer, we obtain

$$\sigma_n^j(T) \equiv_k r_j H \mod \mathbb{U}_K \quad \text{for } j = 0,\dots,i,$$

where H is another bivariate hypergeometric term whose shift-quotient K w.r.t. k is shift-reduced w.r.t. k, and r_j is a residual form w.r.t. K. For univariate rational functions $c_0(n), c_1(n),\dots,c_i(n)$, not all zero, we have

$$\sum_{j=0}^{i} c_j \sigma_n^j(T) \equiv_k \sum_{j=0}^{i} c_j r_j H \mod \mathbb{U}_K.$$

It is readily seen that $\sum_{j=0}^{i} c_j \sigma_n^j$ is a telescoper for T w.r.t. k if $\sum_{j=0}^{i} c_j r_j = 0$. Unfortunately, the converse is false. This is because $\sum_{j=0}^{i} c_j r_j$ is not necessarily a residual form, although all the r_j's are. Thus Theorem 3.18 is not applicable.

[1] The main results in this chapter are joint work with S. Chen, M. Kauers, Z. Li, published in [19].

This situation does not occur in the differential case [15]. To make Theorem 3.18 applicable, we need to find a way to make $\sum_{j=1}^{i} c_j r_j$ a residual form.

This chapter aims at connecting univariate hypergeometric terms with bivariate ones for the next two chapters. In this chapter, we present further properties of residual forms so as to estimate the order bounds of telescopers in Chapter 6. To make the modified reduction applicable to compute telescopers for hypergeometric terms in Chapter 5, we also show that the linear combination of residual forms is well-behaved in terms of congruences.

4.1 Rational normal forms

In this section, we recall the notion of rational normal forms from [10] and review the relation between distinct rational normal forms of a rational function.

Definition 4.1. *Two polynomials $p, q \in \mathbb{F}[k]$ are called* shift-equivalent *w.r.t. k if there exists an integer m such that $p = \sigma_k^m(q)$. We denote it by $p \sim_k q$.*

It is readily seen that $\sim_k$ is an equivalence relation. We call a polynomial in $\mathbb{F}[k]$ *monic* if its leading coefficient w.r.t. k is 1.

Definition 4.2. *Let f be a rational function in $\mathbb{F}(k)$. A rational function pair (K, S) with $K, S \in \mathbb{F}(k)$ is called a* rational normal form *of f if*

$$f = K \cdot \frac{\sigma_k(S)}{S}$$

and K is shift-reduced.

By Theorem 1 in [10], every rational function has a rational normal form. It is not hard to see that there is a one-to-one correspondence between multiplicative decompositions for a given hypergeometric term and rational normal forms for the corresponding shift-quotient. More precisely, for a hypergeometric term T over $\mathbb{F}(k)$, a rational function pair (K, S) is a rational normal form of $\sigma_k(T)/T$ if and only if K is a kernel of T and S a corresponding shell, if and only if T has a multiplicative decomposition $T = SH$ with H a hypergeometric term whose shift-quotient is K.

In fact, a rational function can have more than one rational normal form, as illustrated by the following example.

Example 4.3 (Example 1 in [10])**.** Consider a rational function

$$f = \frac{k(k+2)}{(k-1)(k+1)^2(k+3)}.$$

It can be verified that the following rational function pairs

$$\left(\frac{1}{(k+1)(k+3)}, (k-1)(k+1)\right), \quad \left(\frac{1}{(k+1)^2}, \frac{k-1}{k+2}\right),$$
$$\left(\frac{1}{(k-1)(k-3)}, \frac{k+1}{k}\right), \quad \left(\frac{1}{(k-1)(k+1)}, \frac{1}{k(k+2)}\right).$$

are all rational normal forms of f.

The next theorem describes a relation between two distinct rational normal forms of a rational function.

Theorem 4.4 (Theorem 2 in [10]). *Assume that* $(K, S), (K', S') \in \mathbb{F}(k)^2$ *are distinct rational normal forms of a rational function in* $\mathbb{F}(k)$. *Write*

$$K = c\,\frac{u}{v} \quad \textit{and} \quad K' = c'\,\frac{u'}{v'},$$

where $c, c' \in \mathbb{F}$, $u, u', v, v' \in \mathbb{F}[k]$ *are all monic, and* $\gcd(u, v) = \gcd(u', v') = 1$. *Then*

(*i*) $c = c'$;

(*ii*) $\deg_k(u) = \deg_k(u')$ *and* $\deg_k(v) = \deg_k(v')$;

(*iii*) *there is a one-to-one correspondence* ϕ *between the multi-sets of nontrivial monic irreducible factors of* u *and* u' *such that* $p \sim_k \phi(p)$ *for any nontrivial monic irreducible factor* p *of* u.

(*iv*) *there is a one-to-one correspondence* ψ *between the multi-sets of nontrivial monic irreducible factors of* v *and* v' *such that* $p \sim_k \phi(p)$ *for any nontrivial monic irreducible factor* p *of* v.

4.2 Uniqueness and relatedness of residual forms

In this section, we will present two useful properties of residual forms, which enables us to derive order bounds in Chapter 6. For the notion of residual forms, one can refer to Definition 3.12.

Unlike the differential case, a rational function may have more than one residual form in the shift case. These residual forms, however, are related to each other in some way. To describe it precisely, we introduce the notion of shift-relatedness.

Definition 4.5. *Two shift-free polynomials* $p, q \in \mathbb{F}[k]$ *are called* shift-related*, denoted by* $p \approx_k q$*, if for any nontrivial monic irreducible factor* f *of* p*, there exists a unique monic irreducible factor* g *of* q *with the same multiplicity as* f *in* p *such that* $f \sim_k g$*, and vice versa.*

It is readily seen that $\approx_k$ is an equivalence relation. The following theorem describes the uniqueness of residual forms.

Theorem 4.6. *Let $K \in \mathbb{F}(k)$ be a shift-reduced rational function. Assume that r_1, r_2 are both residual forms of a same rational function in $\mathbb{F}(k)$ w.r.t. K. Then the significant denominators of r_1 and r_2 are shift-related to each other.*

Proof. Assume that r_1, r_2 are of the forms

$$r_1 = \frac{a_1}{b_1} + \frac{q_1}{v} \quad \text{and} \quad r_2 = \frac{a_2}{b_2} + \frac{q_2}{v},$$

where for $i = 1, 2$, $a_i, b_i \in \mathbb{F}[k]$, $\deg_k(a_i) < \deg_k(b_i)$, $\gcd(a_i, b_i) = 1$, b_i is monic, shift-free and strongly coprime with K, $q_i \in \mathbb{W}_K$, and v is the denominator of K. Since r_1, r_2 are both residual forms of the same rational function, $r_1 \equiv_k r_2 \mod \mathbb{V}_K$, which is equivalent to

$$\frac{a_1}{b_1} \equiv_k \frac{a_2}{b_2} + \frac{q_2 - q_1}{v} \mod \mathbb{V}_K.$$

By (2.1), there exists $w \in \mathbb{F}(k)$ so that

$$\frac{a_1 v}{b_1} = u\sigma_k(w) - vw + \frac{a_2 v}{b_2} + (q_2 - q_1). \tag{4.1}$$

Let $f \in \mathbb{F}[k]$ be a nontrivial monic irreducible factor of b_1 with multiplicity $\alpha > 0$. If f^α divides b_2, then we are done. Otherwise, let $\text{den}(w)$ be the denominator of w. Since b_1 is strongly coprime with K, we have $\gcd(f^\alpha, v) = 1$. By (4.1) and partial fraction decomposition, f^α either divides $\text{den}(w)$ or $\sigma_k(\text{den}(w))$. If f^α divides $\text{den}(w)$, let

$$m = \max\{k \in \mathbb{Z} \mid \sigma_k^k(f)^\alpha \text{ divides } \text{den}(w)\},$$

and then $m \geq 0$. Since b_1 is strongly coprime with K, $\gcd(\sigma_k^{m+1}(f)^\alpha, u) = 1$. Apparently, $\sigma_k^{m+1}(f)^\alpha$ divides $\sigma_k(\text{den}(w))$ but doesn't divide $\text{den}(w)$ as m is maximal. Note that b_1 is shift-free and $f \mid b_1$, thus b_1 is not divisible by $\sigma_k^{m+1}(f)^\alpha$. Hence (4.1) implies $\sigma_k^{m+1}(f)^\alpha$ is the required factor of b_2. Similarly, we can show that $\sigma_k^\ell(f)^\alpha$ with

$$\ell = \min\{k \in \mathbb{Z} \mid \sigma_k^k(f)^\alpha \text{ divides } \text{den}(w)\} \leq -1,$$

is the required factor of b_2, if f^α divides $\sigma_k(\text{den}(w))$.

In summary, there always exists a monic irreducible factor of b_2 with multiplicity at least α such that it is shift-equivalent to f. Due to the shift-freeness of b_2, this factor is unique. The same conclusion holds when we switch the roles of b_1 and b_2. Therefore, $b_1 \approx_k b_2$ by definition. $\square$

For a given hypergeometric term, the above theorem reveals the relation between two residual forms of the shell with respect to a same kernel. To study the case with different kernels, we need the following two lemmas.

Lemma 4.7. *Let* (K, S) *be a rational normal form of* $f \in \mathbb{F}(k)$ *and* r *a residual form of* S *w.r.t.* K*. Write* $K = u/v$ *with* $u, v \in \mathbb{F}[k]$ *and* $\gcd(u, v) = 1$*. Assume that* p *is a nontrivial monic irreducible factor of* v *with multiplicity* $\alpha > 0$*. Then the pair*

$$(K', S') = \left(\frac{u}{v'\sigma_k(p)^\alpha}, p^\alpha S\right)$$

is a rational normal form of f*, in which* $v' = v/p^\alpha$*. Moreover, there exists a residual form* r' *of* S' *w.r.t.* K' *whose significant denominator equals that of* r*.*

Proof. Since K is shift-reduced, so is K'. The first assertion follows by noticing

$$K\frac{\sigma_k(S)}{S} = \frac{u}{v'p^\alpha}\frac{\sigma_k(S)}{S} = \frac{u}{v'\sigma_k(p)^\alpha}\frac{\sigma_k(p^\alpha S)}{p^\alpha S} = K'\frac{\sigma_k(S')}{S'}.$$

Let r be of the form $r = a/b + q/v$, where $a, b, q \in \mathbb{F}[k]$, $\deg_k(a) < \deg_k(b)$, $\gcd(a, b) = 1$, b is monic, shift-free and strongly coprime with K, and $q \in \mathbb{W}_K$. Then there exists a rational function $g \in \mathbb{F}(k)$ such that

$$S = K\sigma_k(g) - g + \frac{a}{b} + \frac{q}{v'p^\alpha},$$

which implies

$$\begin{aligned}
S' = p^\alpha S &= p^\alpha K\sigma_k(g) - p^\alpha g + \frac{ap^\alpha}{b} + \frac{q}{v'} \\
&= \frac{u}{v'\sigma_k(p)^\alpha}\sigma_k(p^\alpha g) - p^\alpha g + \frac{ap^\alpha}{b} + \frac{q\sigma_k(p)^\alpha}{v'\sigma_k(p)^\alpha} \\
&= K'\sigma_k(p^\alpha g) - p^\alpha g + \frac{ap^\alpha}{b} + \frac{q\sigma_k(p)^\alpha}{v'\sigma_k(p)^\alpha}
\end{aligned}$$

Since b is strongly coprime with K and $\gcd(a, b) = 1$, we have $\gcd(ap^\alpha, b) = 1$. Using step 3 and step 4 in Algorithm 3.17 computes polynomials $a', q' \in \mathbb{F}[k]$ with $\deg_k(a') < \deg_k(b)$, $\gcd(a', b) = 1$ and $q' \in \mathbb{W}_{K'}$ so that

$$S' \equiv_k \frac{a'}{b} + \frac{q'}{v'\sigma_k(p)^\alpha} \mod \mathbb{V}_{K'}.$$

Note that b is strongly coprime with K, so b is also strongly coprime with K'. Since b is shift-free, $a'/b + q'/(v'\sigma_k(p)^\alpha)$ is a residual form of S' w.r.t. K'. □

Lemma 4.8. *Let (K, S) be a rational normal form of $f \in \mathbb{F}(k)$ and r a residual form of S w.r.t. K. Write $K = u/v$ with $u, v \in \mathbb{F}[k]$ and $\gcd(u, v) = 1$. Assume that p is a nontrivial monic irreducible factor of u with multiplicity $\alpha > 0$. Then the pair*

$$(K', S') = \left(\frac{u' \sigma_k^{-1}(p)^\alpha}{v}, \sigma_k^{-1}(p)^\alpha S \right)$$

is a rational normal form of f, in which $u' = u/p^\alpha$. Moreover, there exists a residual form r' of S' w.r.t. K' whose significant denominator equals that of r.

Proof. Similar to Lemma 4.7. □

Proposition 4.9. *Let (K, S) be a rational normal form of $f \in \mathbb{F}(k)$ and r a residual form of S w.r.t. K. Then there exists a rational normal form $(\tilde{K}, \tilde{S})$ of f such that*

1. *$\tilde{K}$ has shift-free numerator and shift-free denominator;*
2. *there exists a residual form $\tilde{r}$ of $\tilde{S}$ w.r.t. $\tilde{K}$ whose significant denominator is equal to that of r.*

Proof. Let $K = u/v$ with $u, v \in \mathbb{F}[k]$ and $\gcd(u, v) = 1$, and b be the significant denominator of r.

Assume that v is not shift-free. Then there exist two nontrivial monic irreducible factors p and $\sigma_k^m(p)$ $(m > 0)$ of v with multiplicity $\alpha > 0$ and $\beta > 0$, respectively. W.l.o.g., assume further that $\sigma_k^\ell(p)$ is not a factor of v for all $\ell < 0$ and $\ell > m$. By Lemma 4.7, f has a rational normal form (K', S'), in which K' has a denominator of the form $\mathrm{den}(K') = v' \sigma_k(p)^\alpha$, where $v' = v/p^\alpha$, and the numerator remains to be u. Moreover, there exists a residual form of S' w.r.t. K' whose significant denominator is b. If $m = 1$, then $\sigma_k(p)$ is an irreducible factor of $\mathrm{den}(K')$ with multiplicity $\alpha + \beta$. Otherwise, it is an irreducible factor of $\mathrm{den}(K')$ with multiplicity α. More importantly, $\sigma_k^\ell(p)$ is not a factor of $\mathrm{den}(K')$ for all $\ell < 1$. Iteratively using the argument, we arrive at a rational normal form of f such that $\sigma_k^m(p)$ divides the denominator of the new kernel with certain multiplicity but $\sigma_k^i(p)$ does not whenever $i \neq m$, and the numerator remains to be u. Moreover, there exists a residual form of the new shell with respect to the new kernel whose significant denominator is equal to b. Applying the same argument to each irreducible factor, we can obtain a rational normal form of f whose kernel has the numerator u and a shift-free denominator, and whose shell has a residual form with significant denominator b.

With Lemma 4.8, one can obtain a rational normal form of f whose kernel has a shift-free numerator and whose shell has a residual form with significant denominator b. □

A nonzero rational function is said to be *shift-free* if it is shift-reduced and its denominator and numerator are both shift-free. The relatedness of residual forms with respect to different kernels is given below.

Theorem 4.10. *Let $(K, S), (K', S')$ be two rational normal forms of $f \in \mathbb{F}(k)$, and r, r' residual forms of S (w.r.t. K) and S' (w.r.t. K'), respectively. Then the significant denominators of r and r' are shift-related.*

Proof. Let b and b' be the significant denominators of r and r', respectively. By the above proposition, there exist two rational normal forms $(\tilde{K}, \tilde{S})$ and $(\tilde{K}', \tilde{S}')$ of f such that their kernels are shift-free and their shells have residual forms whose significant denominators are b and b', respectively.

According to Theorem 4.4, the respective denominators $\tilde{v}$ and $\tilde{v}'$ of $\tilde{K}$ and $\tilde{K}'$ are shift-related. It follows that for a nontrivial monic irreducible factor p of $\tilde{v}$ with multiplicity $\alpha > 0$, there exists a unique factor $\sigma_k^\ell(p)$ with $\ell \in \mathbb{Z}$ of $\tilde{v}'$ with the same multiplicity. W.l.o.g., we may assume $\ell \leq 0$. Otherwise, we can switch the roles of two pairs $(\tilde{K}, \tilde{S})$ and $(\tilde{K}', \tilde{S}')$. If $\ell < 0$, a repeated use of Lemma 4.7 leads to a new rational normal form $(\tilde{K}'', \tilde{S}'')$ from $(\tilde{K}', \tilde{S}')$, such that $\tilde{K}''$ is shift-free with the same numerator as $\tilde{K}'$, and p is a factor of the denominator of $\tilde{K}''$ with the same multiplicity α. Moreover, $\tilde{S}''$ has a residual form w.r.t. $\tilde{K}''$ with significant denominator b'.

Applying the above argument to each irreducible factor and using Lemma 4.8 for numerators in the same fashion, we can obtain two new rational normal forms whose kernels are equal and whose shells have respective residual forms with significant denominators b and b'. It follows from Theorem 4.6 that b and b' are shift-related. □

4.3 Sum of two residual forms

To compute a telescoper for a given bivariate hypergeometric terms by the modified Abramov-Petkovšek reduction, we are confronted with the difficulty that the sum of two residual forms is not necessarily a residual form, as mentioned at the beginning of this chapter. This is because the least common multiple of two shift-free polynomials is not necessarily shift-free.

The goal of this section is to show that the sum of two residual forms is congruent to a residual form modulo $\mathbb{V}_K$.

Example 4.11. Let $K = 1/k$, $r = 1/(2k+1)$ and $s = 1/(2k+3)$. Then both r and s are residual forms w.r.t. K, but their sum is not, because the denominator $(2k+1)(2k+3)$ is not shift-free. However, we can still find an equivalent residual form. For example, we have

$$r + s \equiv_k -\frac{1}{2(2k+1)} + \frac{1}{2k} \mod \mathbb{V}_K.$$

Note that the residual form is not unique. Another possible choice is

$$r + s \equiv_k \frac{1}{3(2k+3)} + \frac{1}{3k} \mod \mathbb{V}_K.$$

Lemma 4.12. *With Convention 3.2, let $r, s \in \mathbb{F}(k)$ be two residual forms w.r.t. K, i.e., r and s can be written as*

$$r = \frac{a}{f} + \frac{p}{v} \quad \text{and} \quad s = \frac{b}{g} + \frac{q}{v},$$

where $a, f, b, g \in \mathbb{F}[k]$, $\deg_k(a) < \deg_k(f)$, $\deg_k(b) < \deg_k(g)$, $p, q \in \mathbb{W}_K$, and f, g are shift-free and strongly coprime with K. Assume that $\gcd(a, f) = \gcd(b, g) = 1$. Then for all $\lambda, \mu \in \mathbb{F}$, $\lambda r + \mu s$ is a residual form w.r.t. K if and only if the least common multiple of f and g is shift-free.

Proof. Let h be the least common multiple of f and g. Then

$$\lambda r + \mu s = \frac{\lambda a(h/f) + \mu b(h/g)}{h} + \frac{\lambda p + \mu q}{v}. \tag{4.2}$$

We first show the sufficiency. Assume that h is shift-free. It is clear that

$$\deg_k(\lambda a(h/f) + \mu b(h/g)) < \deg_k(h).$$

Since $\mathbb{W}_K$ is a $\mathbb{F}$-vector space, we have $\lambda p + \mu q \in \mathbb{W}_K$. Note that f and g are strongly coprime with K, so is h. By definition, $\lambda r + \mu s$ is a residual form w.r.t. K.

To show the necessity, we suppose otherwise that h is not shift-free. Since $\lambda r + \mu s$ is a residual form w.r.t. K, there exist $b^*, h^* \in \mathbb{F}[k]$ and $q^* \in \mathbb{W}_K$ with $\deg_k(b^*) < \deg_k(h^*)$, and h^* shift-free and strongly coprime with K, such that

$$\lambda r + \mu s = \frac{b^*}{h^*} + \frac{q^*}{v}.$$

It follows from (4.2) that

$$\frac{(\lambda a(h/f) + \mu b(h/g))v}{h} = \frac{b^* v}{h^*} + q^* - \lambda p - \mu q. \tag{4.3}$$

Since h is not shift-free and f, g are shift-free, there exist nontrivial monic irreducible factors p' and $\sigma_k^\ell(p')$ of h such that $p' \mid f$ and $\sigma_k^\ell(p') \mid g$, where ℓ is a nonzero integer. Because $\gcd(a, f) = \gcd(b, g) = 1$ and $h \mid fg$, so

- $p' \nmid (h/f)$ and $p' \nmid a$, but $p' \mid (h/g)$;
- $\sigma_k^\ell(p') \nmid (h/g)$ and $\sigma_k^\ell(p) \nmid b$, but $\sigma_k^\ell(p') \mid (h/f)$.

Since h is also strongly coprime with K, p' and $\sigma_k^\ell(p')$ are coprime with v. Thus they both divide the denominator of the left-hand side of (4.3). By partial fraction decomposition, p' and $\sigma_k^\ell(p')$ both divide h^*, a contradiction as h^* is shift-free. $\square$

To describe the shift-freeness of the least common multiple of two polynomials, we introduce the following notions.

Definition 4.13. *Let f and g be two nonzero polynomials in $\mathbb{F}[k]$. According to [10, §3], the* dispersion *of f and g is defined to be the largest nonnegative integer ℓ such that f and $\sigma_k^\ell(g)$ have a nontrivial common divisor, or -1 if no such ℓ exists. Moreover, we say that f and g are* shift-coprime *if $\gcd(f, \sigma_k^\ell(g)) = 1$ for all nonzero integer ℓ.*

It is clear that the least common multiple of two shift-free polynomials is shift-free if and only if these two polynomials are shift-coprime. Let f and g be two nonzero shift-free polynomials in $\mathbb{F}[k]$. By polynomial factorization and dispersion computation (see [10]), one can uniquely decompose

$$g = \tilde{g}\sigma_k^{\ell_1}\left(p_1^{m_1}\right)\cdots\sigma_k^{\ell_\rho}\left(p_\rho^{m_\rho}\right), \tag{4.4}$$

where $\tilde{g}$ is shift-coprime with f, $p_1, \ldots, p_\rho$ are pairwise distinct and monic irreducible factors of f, $\ell_1, \ldots, \ell_\rho$ are nonzero integers, $m_1, \ldots, m_\rho$ are multiplicities of the factors $\sigma_k^{\ell_1}(p_1)$, …, $\sigma_k^{\ell_\rho}(p_\rho)$ in g, respectively. We refer to (4.4) as the *shift-coprime decomposition* of g w.r.t. f.

Remark 4.14. The factors $\tilde{g}, \sigma_k^{\ell_1}\left(p_1^{m_1}\right)$, …, $\sigma_k^{\ell_\rho}\left(p_\rho^{m_\rho}\right)$ in (4.4) are pairwise coprime, since f and g are shift-free.

To construct a residual form congruent to the sum of two given residual ones, we need three technical lemmas. The first one corresponds to the kernel reduction in [15].

Lemma 4.15. *With Convention 3.2, assume that p_1, p_2 are in $\mathbb{F}[k]$ and m in $\mathbb{N}$. Then there exist q_1, q_2 in $\mathbb{W}_K$ such that*

$$\frac{p_1}{\prod_{i=0}^m \sigma_k^i(v)} \equiv_k \frac{q_1}{v} \mod \mathbb{V}_K \quad \textit{and} \quad \frac{p_2}{\prod_{j=1}^m \sigma_k^{-j}(u)} \equiv_k \frac{q_2}{v} \mod \mathbb{V}_K.$$

Proof. To prove the first congruence, let $w_m = \prod_{i=0}^m \sigma_k^i(v)$.

We proceed by induction on m. If $m = 0$, then the conclusion holds by Lemma 3.10. Assume that the lemma holds for $m-1$ with $m > 0$. Consider the equality

$$\frac{p_1}{w_m} = K\sigma_k\left(\frac{s}{w_{m-1}}\right) - \frac{s}{w_{m-1}} + \frac{t}{w_{m-1}},$$

where $s, t \in \mathbb{F}[k]$ are to be determined. This equality holds if and only if

$$\sigma_k(s)u + (t-s)\sigma_k^m(v) = p_1.$$

Since u and $\sigma_k^m(v)$ are coprime, such s and t can be computed by the extended Euclidean algorithm. Thus, $p_1/w_m \equiv_k t/w_{m-1} \mod \mathbb{V}_K$. Consequently, p_1/w_m has a required residual form by the induction hypothesis.

To prove the second congruence, we use the identity

$$\frac{p_2}{\sigma_k^{-1}(u)} = K\sigma_k\left(-\frac{p_2}{\sigma_k^{-1}(u)}\right) - \left(-\frac{p_2}{\sigma_k^{-1}(u)}\right) + \frac{\sigma_k\left(p_2\right)}{v},$$

which implies that $p_2/\sigma_k^{-1}(u) \equiv_k \sigma_k\left(p_2\right)/v \bmod \mathbb{V}_K$. By Lemma 3.10, there exists a polynomial $q_2 \in \mathbb{W}_K$ such that q_2/v is a residual form of $p_2/\sigma_k^{-1}(u)$ w.r.t. K. Thus the conclusion holds for $m = 0$. Assume that the congruence holds for $m-1$ with $m > 0$. The induction can be completed as in the proof for p_1/w_m. □

The next lemma provides us with flexibility to rewrite a rational function modulo $\mathbb{V}_K$.

Lemma 4.16. *Let $K \in \mathbb{F}(k)$ be nonzero and shift-reduced. Then for every rational function $f \in \mathbb{F}(k)$ and every positive integer ℓ,*

$$f \equiv_k \sigma_k^{\ell}(f)\prod_{i=0}^{\ell-1}\sigma_k^i(K) \equiv_k \sigma_k^{-\ell}(f)\prod_{i=1}^{\ell}\sigma_k^{-i}\left(\frac{1}{K}\right) \bmod \mathbb{V}_K.$$

Proof. Let's show the first congruence by induction on ℓ. For $\ell = 1$, the identity

$$f = K\sigma_k(-f) - (-f) + \sigma_k(f)K$$

implies that f is congruent to $\sigma_k(f)K$ modulo $\mathbb{V}_K$. Assume that it holds for $\ell - 1$ with $\ell > 1$. Set $w_\ell = \prod_{i=0}^{\ell-1}\sigma_k^i(K)$. Then by the induction hypothesis,

$$f \equiv_k \sigma_k^{\ell-1}(f)w_{\ell-1} \mod \mathbb{V}_K.$$

Moreover, $\sigma_k^{\ell-1}(f)w_{\ell-1} \equiv_k \sigma_k^{\ell}(f)w_\ell \mod \mathbb{V}_K$ by the induction base, in which f is replaced with $\sigma_k^{\ell-1}(f)w_{\ell-1}$. Hence, f is congruent to $\sigma_k^{\ell}(f)w_\ell$ modulo $\mathbb{V}_K$.

The second congruence can be shown similarly. For the base case $\ell = 1$, let $r = \sigma_k^{-1}(f)\sigma_k^{-1}(1/K)$. Then the identity $f = K\sigma_k(r) - r + r$ implies that f is congruent to r modulo $\mathbb{V}_K$. We can then proceed as in the proof of the first congruence. □

Lemma 4.17. *With Convention 3.2, let $a, b \in \mathbb{F}[k]$ with $b \neq 0$. Assume that b is shift-free and strongly coprime with K. Assume further that $\sigma_k^{\ell}(b)$ is strongly coprime with K for some integer ℓ, then a/b has a residual form $c/\sigma_k^{\ell}(b) + q/v$ w.r.t. K, where $c \in \mathbb{F}[k]$ with $\deg_k(c) < \deg_k(b)$ and $q \in \mathbb{W}_K$.*

Proof. First, consider the case in which $\ell \geq 0$. If $\ell = 0$, then there exist two polynomials $c, p \in \mathbb{F}[k]$ with $\deg_k(c) < \deg_k(b)$ such that $a/b = c/b + p$. The lemma follows from Remark 3.11. Assume that $\ell > 0$. By the first congruence of Lemma 4.16,

$$\frac{a}{b} \equiv_k \sigma_k^\ell\left(\frac{a}{b}\right)\left(\prod_{i=0}^{\ell-1}\sigma_k^i(K)\right) = \frac{\sigma_k^\ell(a)\prod_{i=0}^{\ell-1}\sigma_k^i(u)}{\sigma_k^\ell(b)\prod_{i=0}^{\ell-1}\sigma_k^i(v)} \mod \mathbb{V}_K.$$

Note that $\sigma_k^\ell(b)$ is strongly coprime with v by assumption. Then it is coprime with the product $v\sigma_k(v)\cdots\sigma_k^{\ell-1}(v)$. By partial fraction decomposition, we get

$$\frac{a}{b} \equiv_k \frac{\tilde{a}}{\sigma_k^\ell(b)} + \frac{\tilde{q}}{\prod_{i=0}^{\ell-1}\sigma_k^i(v)} \mod \mathbb{V}_K,$$

where $\tilde{a}, \tilde{q} \in \mathbb{F}[k]$ and $\deg_k \tilde{a} < \deg_k(b)$. By the first congruence of Lemma 4.15, the second summand in the right-hand side of the above congruence can be replaced by a residual form whose denominator is equal to v. The first assertion holds.

The case $\ell < 0$ can be handled in the same way, in which the second congruences of Lemmas 4.16 and 4.15 will be used instead of the first ones. □

Remark 4.18. With the assumptions of the above lemma, let p be a nontrivial factor of b with $\gcd(b', p) = 1$ where $b' = b/p$. Assume that $\sigma_k^\ell(p)$ is also strongly coprime with K. Then by partial fraction decomposition and Lemma 4.17, there exist $c, q \in \mathbb{F}[k]$ with $\deg_k(c) < \deg_k(b)$ and $q \in \mathbb{W}_K$ such that $c/(b'\sigma_k^\ell(p)) + q/v$ is a residual form of a/b w.r.t. K.

We will refer to Lemma 4.17 and Remark 4.18 as *the shifting property of significant denominators*. Now we are ready to present the main result of this section.

Theorem 4.19. *With Convection 3.2, let r and s be two residual forms w.r.t. K. Then there exists a residual form t congruent to s modulo $\mathbb{V}_K$ so that for all constants $\lambda, \mu \in \mathbb{F}$, the sum $\lambda r + \mu t$ is a residual form w.r.t. K congruent to $\lambda r + \mu s$ modulo $\mathbb{V}_K$.*

Proof. Since r and s are two residual forms w.r.t. K, they can be written as

$$r = \frac{a}{f} + \frac{p}{v} \quad \text{and} \quad s = \frac{b}{g} + \frac{q}{v}, \tag{4.5}$$

where $a, f, b, g \in \mathbb{F}[k]$, $\deg_k(a) < \deg_k(f)$, $\deg_k(b) < \deg_k(g)$, $p, q \in \mathbb{W}_K$, and f, g are shift-free and strongly coprime with K.

Assume that (4.4) is the shift-coprime decomposition of g w.r.t. f. Define $P_i = \sigma_k^{\ell_i}(p_i)$ for $i = 1, \ldots, \rho$. By Remark 4.14 and partial fraction decomposition,

$$\frac{b}{g} = \frac{b_0}{\tilde{g}} + \sum_{i=1}^{\rho} \frac{b_i}{P_i^{m_i}}, \tag{4.6}$$

where $b_0, b_1, \ldots, b_\rho \in \mathbb{F}[k]$, $\deg_k(b_0) < \deg_k(\tilde{g})$ and $\deg_k(b_i) < m_i \deg_k(p_i)$. Note that $p_i = \sigma_k^{-\ell_i}(P_i)$, which is a factor of f. Thus it is strongly coprime with K. So we can apply Lemma 4.17 to each fraction $b_i/P_i^{m_i}$ in (4.6) to get

$$\frac{b}{g} \equiv_k \frac{b_0}{\tilde{g}} + \sum_{i=1}^{\rho} \frac{b_i'}{p_i^{m_i}} + \frac{q'}{v} \mod \mathbb{V}_K, \tag{4.7}$$

where $b_1', \ldots, b_\rho' \in \mathbb{F}[k]$, $\deg_k(b_i') < m_i \deg_k(p_i)$ and $q' \in \mathbb{W}_K$.

Let $h = \tilde{g} \prod_{i=1}^{\rho} p_i^{m_i}$. Then h is shift-free and strongly coprime with K as both f and g are. Since f is shift-free, all its factors are shift-coprime with f, so are the p_i's, and so is h. Let t be the sum of q/v and the rational function in the right-hand side of (4.7). Then there exist $b^* \in \mathbb{F}[k]$ with $\deg_k(b^*) < \deg_k(h)$ and $q^* \in \mathbb{W}_K$ such that

$$t = \frac{b^*}{h} + \frac{q^*}{v}.$$

Since f and h are shift-coprime, their least common multiple is shift-free. Therefore, $\lambda r + \mu t$ is a residual form w.r.t. K by Lemma 4.12, and $\lambda r + \mu t$ is congruent to $\lambda r + \mu s$ modulo $\mathbb{V}_K$. □

The above proof contains an algorithm, which can translate a residual form properly according to a given one, so that the resulting sum is again a residual form. We outline this algorithm as follows.

Algorithm 4.20 (Translation of Discrete Residual Forms)**.**

Input: A shift-reduced rational function $K \in \mathbb{F}(k)$, a polynomial $f \in \mathbb{F}[k]$ which is shift-free and strongly coprime with K, and a residual form s w.r.t. K of the form (4.5).

Output: A rational function $w \in \mathbb{F}(k)$ and a residual form t w.r.t. K such that

$$s = K\sigma_k(w) - w + t,$$

and the least common multiple of the given polynomial f and the significant denominator of t is shift-free.

1 Compute the shift-coprime decomposition, say (4.4), of g w.r.t. f.

2 Set $P_i = \sigma_k^{\ell_i}(p_i)$ for $i = 1, \ldots, \rho$.

3 Compute the partial fraction decomposition (4.6) of b/g.

4 Apply Lemma 4.17 to each $b_i/P_i^{m_i}$ to find $w_i \in \mathbb{F}(k)$ and $b_i', q_i' \in \mathbb{F}[k]$ with $\deg_k(b_i') < m_i \deg_k(p_i)$ and $q_i' \in \mathbb{W}_K$ such that

$$\frac{b_i}{P_i^{m_i}} = K\sigma_k(w_i) - w_i + \frac{b_i'}{p_i^{m_i}} + \frac{q_i'}{v}.$$

5 Set $w = \sum_{i=1}^{\rho} w_i$ and

$$t = \frac{b_0}{\tilde{g}} + \sum_{i=1}^{\rho} \frac{b_i'}{p_i^{m_i}} + \frac{\sum_{i=1}^{\rho} q_i' + q}{v};$$

and return.

Chapter 5

Creative Telescoping for Hypergeometric Terms [1]

In the study of combinatorics, we often encounter problems about evaluating definite sums or proving identities of hypergeometric terms. These terms are exactly nonzero solutions of first-order (partial) difference equations with polynomial coefficients. Traditionally [56], such problems were solved case by case using methods that do not give rise to general algorithms. Based on a series of work [65, 66, 67, 68, 69, 70, 71] in early 1990s, Wilf and Zeilberger developed a constructive theory, which is now known as Wilf-Zeilberger's theory. This theory provides a systematic solution to a large class of problems concerning hypergeometric summations and identities, and has wide application in the areas of combinatorics and physics. The key step of Wilf-Zeilberger's theory is to compute a telescoper for a given hypergeometric term. The efficiency of the computation determines the utility of this theory. During the past 26 years, numerous algorithms have been developed for computing telescopers. In early 1990s, Zeilberger [70] first came up with an algorithm based on elimination techniques. This algorithm was improved later by Takayama [61] and Chyzak, Salvy [27], respectively. In 1990, Zeilberger [69] developed another algorithm, known as Zeilberger's (fast) algorithm, based on a parametrization of Gosper's algorithm. 15 years later, Apagodu and Zeilberger designed a new algorithm which reduced the problem to solving a linear system. The common feature of the above algorithms is that there was no way to obtain a telescoper without also computing a certificate. In many applications, however, certificates are not needed, and they typically require more storage space than telescopers do. It would be more efficient to avoid computing certificates if we don't need them. To achieve this goal, Bostan et al. [14] presented a new algorithm for bivariate rational functions in the differential case, based on the Hermite reduction. This algorithm separates the computation of telescopers and the corresponding certificates. So far, this approach has been generalized to

[1] The main results in this chapter are joint work with S. Chen, M. Kauers, Z. Li, published in [19].

several instances including rational functions in three variables [24], multivariate rational functions [16], bivariate hyperexponential functions [15] and bivariate algebraic functions [23]. These algorithms turn out to be more efficient than the classical algorithms in practice. However, all these algorithms only work for the differential case.

In this chapter, we discuss how to translate their ideas into the hypergeometric setting. Using the modified Abramov-Petkovšek reduction, we develop a new creative telescoping algorithm. This new algorithm separates the computation of telescopers from that of certificates. We have implemented the new algorithm in MAPLE 18 and compare it to the built-in Maple procedure Zeilberger in the package **SumTools[Hypergeometric]**, which is based on Zeilberger's algorithm. The experimental results indicate that the new algorithm is faster than the Maple procedure if it returns a normalized certificate, and the new algorithm is much more efficient if it omits the computation of certificates.

5.1 Bivariate hypergeometric terms

In this section, we translate terminology concerning univariate hypergeometric terms to bivariate ones and introduce the notions of telescopers as well as certificates for bivariate hypergeometric terms. Moreover, we recall [67, 4] an existence criterion for telescopers.

Let $\mathbb{K}$ be a field of characteristic zero, and $\mathbb{K}(n,k)$ be the field of rational functions in n and k over $\mathbb{K}$. Let σ_n and σ_k be the shift operators w.r.t. n and k, respectively, defined by

$$\sigma_n(f(n,k)) = f(n+1,k) \quad \text{and} \quad \sigma_k(f(n,k)) = f(n,k+1),$$

for any rational function $f \in \mathbb{K}(n,k)$. Clearly, σ_n and σ_k are both automorphisms of $\mathbb{K}$. The pair $(\mathbb{K}(n,k), \{\sigma_n, \sigma_k\})$ forms a *partial difference field.* A *partial difference ring extension* of $(\mathbb{K}(n,k), \{\sigma_n, \sigma_k\})$ is a ring $\mathbb{D}$ containing $\mathbb{K}(n,k)$ together with two distinguished endomorphism σ_n and σ_k from $\mathbb{D}$ to itself, whose restrictions to $\mathbb{K}(n,k)$ agree with the two automorphisms defined before, respectively.

Analogous to the univariate case in Chapter 2, an element $c \in \mathbb{D}$ is called a *constant* if it is invariant under the applications of σ_n and σ_k. It is readily seen that all constants in $\mathbb{D}$ form a subring of $\mathbb{D}$. Moreover, Theorem 2 in [9] yields that the set of all constants in $\mathbb{K}(n,k)$ w.r.t. σ_n and σ_k is exactly the field $\mathbb{K}$.

Definition 5.1. *Let $\mathbb{D}$ be a partial difference ring extension of $\mathbb{K}(n,k)$. A nonzero element $T \in \mathbb{D}$ is called a* hypergeometric term *over $\mathbb{K}(n,k)$ if it is invertible and there exist $f, g \in \mathbb{K}(n,k)$ such that $\sigma_n(T) = fT$ and $\sigma_k(T) = gT$. We call f and g the* shift-quotients *of T w.r.t. n and k, respectively.*

In the rest of this chapter and also the next chapter, whenever we mention hypergeometric terms, they always belong to some difference ring extension $\mathbb{D}$ of $\mathbb{K}(n,k)$, unless specified otherwise.

Let $\mathbb{F}$ be the field $\mathbb{K}(n)$, and $\mathbb{F}\langle S_n\rangle$ be the ring of linear recurrence operators in n, in which the commutation rule is that $S_n r = \sigma_n(r) S_n$ for all $r \in \mathbb{F}$. The application of an operator $L = \sum_{i=0}^{\rho} \ell_i S_n^i \in \mathbb{F}\langle S_n\rangle$ to a hypergeometric term T is defined as

$$L(T) = \sum_{i=0}^{\rho} \ell_i \sigma_n^i(T).$$

Definition 5.2. *Let T be a hypergeometric term over $\mathbb{F}(k)$. A nonzero recurrence operator $L \in \mathbb{F}\langle S_n\rangle$ is called a* telescoper *for T w.r.t. k if there exists a hypergeometric term G such that*

$$L(T) = \Delta_k(G).$$

We call G a certificate *of L.*

In contrast to the differential case, telescopers for hypergeometric terms do not always exist. To describe the existence of telescopers concisely, we recall [4] the definitions of integer-linear polynomials and proper terms.

Definition 5.3. *An irreducible polynomial $p \in \mathbb{K}[n, k]$ is called* integer-linear *over $\mathbb{K}$ if there exists $P \in \mathbb{K}[z]$ and $\lambda, \mu \in \mathbb{Z}$ such that $p = P(\lambda n + \mu k)$. A polynomial in $\mathbb{K}[n, k]$ is called* integer-linear *over $\mathbb{K}$ if all of its irreducible factors are integer-linear. A rational function in $\mathbb{K}(n, k)$ is called* integer-linear *over $\mathbb{K}$ if its denominator and numerator are both integer-linear.*

Definition 5.4. *A hypergeometric term T over $\mathbb{K}(n, k)$ is called a* factorial term *if the shift-quotients $\sigma_n(T)/T$ and $\sigma_k(T)/T$ are integer-linear over $\mathbb{K}$. A* proper term *over $\mathbb{K}(n, k)$ is the product of a factorial term and a polynomial in $\mathbb{K}[n, k]$.*

We have the following existence criterion for telescopers according to [67, 4].

Theorem 5.5 (Existence criterion)**.** *Let T be a hypergeometric term over $\mathbb{F}(k)$ and let $K = u/v$ with $u, v \in \mathbb{F}[k]$, $\gcd(u, v) = 1$ be a kernel of T w.r.t. k and S a corresponding shell of T. Assume that applying Algorithm 3.17, i.e., the modified Abramov-Petkovšek reduction w.r.t. k to T yields*

$$T = \Delta_k(gH) + \left(\frac{a}{b} + \frac{q}{v}\right) H, \tag{5.1}$$

where $g \in \mathbb{F}(k)$, $H = T/S$, and $a/b + q/v$ is a residual form of S w.r.t. K, that is, $a, b \in \mathbb{F}[k]$ with $\deg_k(a) < \deg_k(b)$, b is shift-free and strongly coprime with K w.r.t. k, and $q \in \mathbb{W}_K$. Then T has a telescoper w.r.t. k if and only if b is integer-linear over $\mathbb{K}$.

Proof. Since the kernel $K = \sigma_k(H)/H$ is shift-reduced w.r.t. k, it follows from [8, Theorem 8] that H is a factorial term over $\mathbb{F}(k)$. Thus K is integer-linear over $\mathbb{K}$, and then so are the numerator u and the denominator v.

We first show the sufficiency. Since b is integer-linear over $\mathbb{K}$, the term $H/(bv)$ is again a factorial term. Hence

$$\left(\frac{a}{b} + \frac{q}{v}\right) H = (av + bq)\frac{H}{bv}$$

is a proper term, whose telescopers exist according to the fundamental lemma in [67]. By (5.1), T has a telescoper w.r.t. k.

To show the necessity, assume that T has a telescoper w.r.t. k. Then the term $(a/b+q/v)H$ is proper by [4, Theorem 10]. Thus $H/(bv)$ is a factorial term. Note that

$$\frac{\sigma_k(H/(bv))}{H/(bv)} = \frac{u}{\sigma_k(v)}\frac{b}{\sigma_k(b)}.$$

Hence, $b/\sigma_k(b)$ is integer-linear over $\mathbb{K}$ as u, v are integer-linear. Because b is shift-free w.r.t. k, so $\gcd(b, \sigma_k(b)) \in \mathbb{F}$. The assertion follows by noticing that all elements in $\mathbb{F}$ are integer-linear. □

5.2 Telescoping via reductions

Let T be a hypergeometric term over $\mathbb{F}(k)$. If there exists a telescoper for T w.r.t. k by Theorem 5.5, then all telescopers for T w.r.t. k together with the zero operator form a left ideal of the principal ideal ring $\mathbb{F}\langle S_n\rangle$. We refer to a generator of this ideal as a *minimal telescoper* for T w.r.t. k. Roughly speaking, a minimal telescoper is a telescoper of the minimal order.

Since 1990, various algorithms [69, 70, 71, 44, 6] have been designed to compute a minimal telescoper for a given hypergeometric term, typically the classical Zeilberger's algorithm [69]. When telescopers exist, Zeilberger's algorithm constructs a telescoper for a given hypergeometric term T by iteratively using Gosper's algorithm to detect the summability of $L(T)$ for an ansatz

$$L = \sum_{i=0}^{\rho} \ell_i S_n^i \in \mathbb{F}\langle S_n\rangle,$$

where ℓ_i are indeterminates. In order to get a telescoper, one needs to solve a linear system with unknowns ℓ_i and also unknowns from the certificate. Any nontrivial solution gives rise to a telescoper and a corresponding certificate simultaneously. There is no obvious way to avoid the computation of certificates in Zeilberger's algorithm.

In order to separate the computations of telescopers and certificates, we follow the ideas in the continuous case [14, 18, 16, 15], and use the modified Abramov-Petkovšek reduction to develop a creative telescoping algorithm. The algorithm is outlined below.

Algorithm 5.6 (Reduction-based creative telescoping)**.**
Input: A hypergeometric term T over $\mathbb{F}(k)$.
Output: A minimal telescoper for T w.r.t. k and a corresponding certificate if telescopers exist; "No telescoper exists!", otherwise.

1 Find a kernel K and shell S of T w.r.t. k such that $T = SH$ with $K = \sigma_k(H)/H$.

2 Apply the modified Abramov-Petkovšek reduction to T to get

$$T = \Delta_k(g_0 H) + r_0 H. \tag{5.2}$$

If $r_0 = 0$, then return $(1, g_0 H)$.

3 If the denominator of r_0 is not integer-linear, return "No telescoper exists!".

4 Set $N = \sigma_n(H)/H$ and $R = \ell_0 r_0$, where ℓ_0 is an indeterminate.
For $i = 1, 2, \ldots$ do

4.1 View $\sigma_n(r_{i-1})NH$ as a hypergeometric term with kernel K and shell $\sigma_n(r_{i-1})N$. Using Algorithm 3.5 and Algorithm 3.16 w.r.t. K, find $g_i' \in \mathbb{F}$ and a residual form $\tilde{r}_i$ w.r.t. K such that

$$\sigma_n(r_{i-1})NH = \Delta_k(g_i' H) + \tilde{r}_i H.$$

4.2 Set $\tilde{g}_i = \sigma_n(g_{i-1})N + g_i'$, so that

$$\sigma_n^i(T) = \Delta_k(\tilde{g}_i H) + \tilde{r}_i H. \tag{5.3}$$

4.3 Apply Algorithm 4.20 to $\tilde{r}_i$ w.r.t. K and R, to find $g_i, r_i \in \mathbb{F}(k)$ such that $r_i \equiv_k \tilde{r}_i \mod \mathbb{V}_K$,

$$\sigma_n^i(T) = \Delta_k(g_i H) + r_i H, \tag{5.4}$$

and $R + \ell_i r_i$ is a residual form w.r.t. K, where ℓ_i is an indeterminate.

4.4 Update R to $R + \ell_i r_i$.
Find $\ell_j \in \mathbb{F}$ such that $R = 0$ by solving a linear system in $\ell_0, \ldots, \ell_i$ over $\mathbb{F}$. If there is a nontrivial solution, return

$$\left(\sum_{j=0}^{i} \ell_j S_n^j, \ \sum_{j=0}^{i} \ell_j g_j H \right).$$

Theorem 5.7. *Let T be a hypergeometric term over $\mathbb{F}(k)$. If T has a telescoper, then Algorithm 5.6 terminates and returns a telescoper of minimal order for T w.r.t. k.*

Proof. By Theorem 3.18, $r_0 = 0$ in step 2 implies that T is summable, and thus 1 is a minimal telescoper for T w.r.t. k. Now let r_0 obtained from step 2 be of the form $r_0 = a_0/b_0 + q_0/v$, where $a_0, b_0, v \in \mathbb{F}[k]$, $\deg_k(a_0) < \deg_k(b_0)$, b_0 is strongly coprime with K, $q_0 \in \mathbb{W}_K$, and v is the denominator of K. According to [8, Theorem 8], K is integer-linear and so is v. It follows that b_0 is integer-linear if and only if $b_0 v$ is. By Theorem 5.5, T has a telescoper if and only if the denominator of r_0 is integer-linear. Thus, steps 2 and 3 are correct.

It follows from (5.2) and $\sigma_n(r_0 H) = \sigma_n(r_0)NH$ that (5.3) holds for $i = 1$. By Algorithm 4.20, there exists a rational function $u_1 \in \mathbb{F}(k)$ and a residual form r_1 w.r.t. K such that

$$\tilde{r}_1 = K\sigma_k(u_1) - u_1 + r_1, \quad \text{i.e.,} \quad r_1 \equiv_k \tilde{r}_1 \mod \mathbb{V}_K,$$

and $R + \ell_1 r_1$ is again a residual form for all $\ell_0, \ell_1 \in \mathbb{F}$. Setting $g_1 = \tilde{g}_1 + u_1$, we see that (5.4) holds for $i = 1$. By a direct induction on i, (5.4) holds in the loop of step 4.

Assume that $L = \sum_{i=0}^{\rho} c_i S_n^i$ is a minimal telescoper for T with $\rho \in \mathbb{N}$, $c_i \in \mathbb{F}$ and $c_\rho \neq 0$. Then $L(T)$ is summable w.r.t. k. By Theorem 3.18, $\sum_{i=0}^{\rho} c_i r_i$ is equal to zero. Thus, the linear homogeneous system (over $\mathbb{F}$) obtained by equating $\sum_{i=0}^{\rho} \ell_i r_i$ to zero has a nontrivial solution, which yields a minimal telescoper. □

Remark 5.8. Algorithm 5.6 indeed separates the computation of minimal telescopers from that of certificates. In applications where certificates are irrelevant, we can drop step 4.2, and in step 4.3 we compute g_i and r_i with

$$r_i \equiv_k \tilde{r}_i \mod \mathbb{V}_K, \quad \sigma_n^i(r_{i-1})NH = \Delta_k(g_i H) + r_i H$$

and $R + \ell_i r_i$ is a residual form w.r.t. K, where ℓ_i is an indeterminate. Moreover, the rational function g_i can be discarded, and we do not need to calculate $\sum_{j=0}^{i} \ell_j g_j H$ in the end.

Remark 5.9. Instead of applying the modified reduction to $\sigma_n(r_{i-1})NH$ in step 4.1, it is also possible to apply the reduction to $\sigma_n^i(T)$ directly, but our experiments suggest that this variant takes considerably more time. This observation agrees with the advices given in [6, Example 6].

Since Algorithm 5.6 performs the same function as Zeilberger's algorithm, it is also applicable to the examples and applications indicated in [54]. In other words, it can be used to evaluate definite sums and prove identities of hypergeometric terms efficiently.

Example 5.10. Consider the hypergeometric term $T = \binom{n}{k}^3$. Then the respective shift-quotients of T with respect to n and k are

$$f = \frac{\sigma_n(T)}{T} = \frac{(n+1)^3}{(n+1-k)^3} \quad \text{and} \quad g = \frac{\sigma_k(T)}{T} = \frac{(n-k)^3}{(k+1)^3}.$$

Since g is shift-reduced w.r.t. k, its kernel is equal to g itself, and the corresponding shell is 1, implying that $H = T$ in step 1 of Algorithm 5.6. In step 4, applying the modified Abramov-Petkovšek reduction to $T, \sigma_n(T), \sigma_n^2(T)$, incrementally, yields

$$\sigma_k^i(T) = \Delta_k(g_i H) + \frac{q_i}{v} H,$$

where $i = 0, 1, 2$, $v = (k+1)^3$,

$$q_0 = \frac{1}{2}(n+1)(n^2 - n + 3k(k-n+1)+1), \quad q_1 = (n+1)^3,$$

$$q_2 = \frac{(n+1)^3}{(n+2)^2}\left(11n^2 - 12nk + 17n + 20 + 12k + 12k^2\right),$$

and $g_0, g_1, g_2 \in \mathbb{F}(k)$ which are too complicated to be reproduced here. By finding an $\mathbb{F}$-linear dependency among q_0, q_1, q_2, we see that

$$L = (n+2)^2 S_n^2 - (7n^2 + 21n + 16) S_n - 8(n+1)^2$$

is a minimal telescoper for T w.r.t. k. For a corresponding certificate G, one can choose to leave it as an unnormalized term

$$G = (n+2)^2 g_2 - (7n^2 + 21n + 16) g_1 - 8(n+1)^2 g_0,$$

or normalize it as one rational function according to the specific requirements.

5.3 Implementation and timings

We have implemented Algorithm 5.6 in Maple 18. The procedure is named as ReductionCT in the Maple package **ShiftReductionCT**. See Appendix A for more details.

In this section, we compare the runtime of the new procedure to the performance of Zeilberger's algorithm. All timings are measured in seconds on a Linux computer with 388Gb RAM and twelve 2.80GHz Dual core processors. No parallelism was used in this experiment. In addition, we also compare the memory requirements of all procedures, which is shown in Appendix B. For brevity, we denote

- Z: the procedure **SumTools[Hypergeometric]**[Zeilberger], which is based on Zeilberger's algorithm;
- RCT_{tc}: the procedure ReductionCT in **ShiftReductionCT**, which computes a minimal telescoper and a corresponding normalized certificate;
- RCT_t: the procedure ReductionCT in **ShiftReductionCT**, which computes a minimal telescoper without constructing a certificate.
- order: the order of the resulting minimal telescoper.

Example 5.11. Consider bivariate hypergeometric terms of the form

$$T = \frac{f(n,k)}{g_1(n+k)g_2(2n+k)} \frac{\Gamma(2\alpha n + k)}{\Gamma(n+\alpha k)}$$

where $f \in \mathbb{Z}[n,k]$ of degree d_2, and for $i = 1, 2$, $g_i = p_i \sigma_z^\lambda(p_i)\sigma_z^\mu(p_i)$ with $p_i \in \mathbb{Z}[z]$ of degree d_1 and $\alpha, \lambda, \mu \in \mathbb{N}$. For different choices of $d_1, d_2, \alpha, \mu, \lambda$, Table 5.1 compares the timings of the four procedures.

$(d_1, d_2, \alpha, \lambda, \mu)$	Z	RCT_{tc}	RCT_t	order
$(1, 0, 1, 5, 5)$	17.12	5.00	1.80	4
$(1, 0, 2, 5, 5)$	74.91	26.18	5.87	6
$(1, 0, 3, 5, 5)$	445.41	92.74	17.34	7
$(1, 8, 3, 5, 5)$	649.57	120.88	23.59	7
$(2, 0, 1, 5, 10)$	354.46	58.01	4.93	4
$(2, 0, 2, 5, 10)$	576.31	363.25	53.15	6
$(2, 0, 3, 5, 10)$	2989.18	1076.50	197.75	7
$(2, 3, 3, 5, 10)$	3074.08	1119.26	223.41	7
$(2, 0, 1, 10, 15)$	2148.10	245.07	11.22	4
$(2, 0, 2, 10, 15)$	2036.96	1153.38	153.21	6
$(2, 0, 3, 10, 15)$	11240.90	3932.26	881.12	7
$(2, 5, 3, 10, 15)$	10163.30	3954.47	990.60	7
$(3, 0, 1, 5, 10)$	18946.80	407.06	43.01	6
$(3, 0, 2, 5, 10)$	46681.30	2040.21	465.88	8
$(3, 0, 3, 5, 10)$	172939.00	5970.10	1949.71	9

Table 5.1: Timing comparison of Zeilberger's algorithm to reduction-based creative telescoping with and without construction of a certificate (in seconds)

Remark 5.12. The difference between RCT_{tc} and RCT_t mainly comes from the time needed to bring the rational function g in the certificate gH on a common denominator. When it is acceptable to keep the certificate as an unnormalized linear combination of rational functions, their timings are virtually the same.

Chapter 6

Order Bounds for Minimal Telescopers [1]

In the previous chapter, we have presented a reduction-based creative telescoping algorithm for bivariate hypergeometric terms, namely Algorithm 5.6. Roughly speaking, its basic idea is as follows. Using the modified Abramov-Petkovšek reduction from Chapter 3, we first reduce a given hypergeometric term and its shifts to some required "standard forms" (called *remainders* in the sequel), such that the difference between the original function and its remainder is summable. Then computing a telescoper amounts to finding a linear dependence among these remainders. In order to show that this algorithm terminates, we show that for every summable term, its remainder is zero. This ensures that the algorithm terminates by the existence criterion given in Theorem 5.5, and in fact it will find the smallest possible telescoper, but it does not provide a bound on its order. Another possible approach is to show that the vector space spanned by the remainders has a finite dimension. Then, as soon as the number of remainders exceeds this dimension, we can be sure that a telescoper will be found. This approach was taken in [15, 16, 23]. As a nice side result, this approach provides an independent proof of the existence of telescopers, and even a bound on the order of minimal telescopers.

In this chapter, we show that the approach for the differential case also works for the shift case, i.e., the remainders in the shift case also form a finite-dimensional vector space, so as to eliminate the discrepancy. As a result, we obtain a new argument for the termination of Algorithm 5.6, and also get new bounds for the order of minimal telescopers for hypergeometric terms. We do not find exactly the same bounds that are already in the literature [49, 6]. Comparing our bounds to the known bounds in the literature, it appears that for "generic" input (see Subsection 6.4.1 for a definition), the values often agree (of course, because the known bounds are already generically sharp). However, there are some

[1] The main results in this chapter are published in [38].

special examples in which our bounds are better than the known bounds. On the other hand, our bounds are never worse than the known ones. In addition, we give a variant of Algorithm 5.6 based on the new bounds. An experimental comparison is presented in the final section.

6.1 Shift-homogeneous decompositions

In this section, we generalize the notion of shift-equivalence in Chapter 4 to the bivariate case, and then derive a useful decomposition for an integer-linear polynomial.

Using the same notations as the previous chapter, $\mathbb{K}$ is a field of characteristic zero, and $\mathbb{K}(n,k)$ is the field of rational functions in n and k over $\mathbb{K}$. Let σ_n and σ_k be the shift operators w.r.t. n and k, respectively.

Definition 6.1. *Two polynomials $p, q \in \mathbb{K}[n,k]$ are called* shift-equivalent *w.r.t. n and k if there exist integers ℓ, m such that $q = \sigma_n^{\ell}\sigma_k^{m}(p)$. We denote it by $p \sim_{n,k} q$.*

Clearly $\sim_{n,k}$ is an equivalence relation. In particular, when $\ell = 0$ or $m = 0$, the above definition degenerates to Definition 4.1. Thus $\sim_n$ or $\sim_k$ implies $\sim_{n,k}$. Choosing the pure lexicographic order $n \prec k$, we say a polynomial is *monic* if its highest term has coefficient 1. A rational function is said to be *shift-homogeneous* if all non-constant monic irreducible factors of its denominator and numerator belong to the same shift-equivalence class.

By grouping together the factors in the same shift-equivalence class, every rational function $r \in \mathbb{K}(n,k)$ can be decomposed into the form

$$r = c\, r_1 \ldots r_s, \tag{6.1}$$

where $c \in \mathbb{K}$, $s \in \mathbb{N}$, each r_i is a shift-homogeneous rational function, and any two non-constant monic irreducible factors of r_i and r_j are pairwise shift-inequivalent whenever $i \neq j$. We call each r_i a *shift-homogeneous component* of r and (6.1) a *shift-homogeneous decomposition* of r. Noticing that the field $\mathbb{K}(n,k)$ is a unique factorization domain, one can easily show that the shift-homogeneous decomposition is unique up to the order of the factors and multiplication by nonzero constants.

Let $p \in \mathbb{K}[n,k]$ be an irreducible integer-linear polynomial. Then it is of the form $p = P(\lambda n + \mu k)$ for some $P \in \mathbb{K}[z]$ and $\lambda, \mu \in \mathbb{Z}$, not both zero. W.l.o.g., we further assume that $\mu \geq 0$ and $\gcd(\lambda, \mu) = 1$. Under this assumption, making ansatz and comparing coefficients yield the uniqueness of P since $\mathbb{Z}$ is a unique factorization domain. In view of this, we call the pair $(P, \{\lambda, \mu\})$ the *univariate representation* of the integer-linear polynomial p. By Bézout's relation, there exist unique integers α, β with $|\alpha| < |\mu|$ and $|\beta| < |\lambda|$ such that $\alpha\lambda + \beta\mu = 1$. Define $\delta^{(\lambda,\mu)}$ to be $\sigma_n^{\alpha}\sigma_k^{\beta}$. For brevity, we just write δ if (λ, μ) is clear from the context.

Note that $\delta(P(z)) = P(z+1)$ with $z = \lambda n + \mu k$, which allows us to treat integer-linear polynomials as univariate ones. For a Laurent polynomial $\xi = \sum_{i=\ell}^{\rho} m_i \delta^i$ in $\mathbb{Z}[\delta, \delta^{-1}]$ with $\ell, \rho, m_i \in \mathbb{Z}$ and $\ell \leq \rho$, define

$$p^{\xi} = \delta^{\ell}(p^{m_\ell})\delta^{\ell+1}(p^{m_{\ell+1}}) \cdots \delta^{\rho}(p^{m_\rho}).$$

Let $p, q \in \mathbb{K}[n,k]$ be two irreducible integer-linear polynomials of the forms

$$p = P(\lambda_1 n + \mu_1 k) \quad \text{and} \quad q = Q(\lambda_2 n + \mu_2 k),$$

where $(P, \{\lambda_1, \mu_1\})$ and $(Q, \{\lambda_2, \mu_2\})$ are the univariate representations of p and q, respectively. Namely, $P, Q \in \mathbb{K}[z]$, $\lambda_1, \mu_1, \lambda_2, \mu_2 \in \mathbb{Z}$, $\mu_1, \mu_2 \geq 0$ and $\gcd(\lambda_1, \mu_1) = \gcd(\lambda_2, \mu_2) = 1$. It is readily seen that $p \sim_{n,k} q$ if and only if $\lambda_1 = \lambda_2$, $\mu_1 = \mu_2$ and $q = p^{\delta^{\ell}}$ for some integer ℓ, in which $\delta = \delta^{(\lambda_1,\mu_1)} = \delta^{(\lambda_2,\mu_2)}$.

Given a shift-homogeneous and integer-linear rational function $r \in \mathbb{K}(n,k)$, let h be a monic, irreducible and integer-linear polynomial in $\mathbb{K}[n,k]$ with the property that all monic irreducible factors of the numerator and denominator of r are equal to some shift of h w.r.t. n and k. Assume that the univariate representation of h is the pair $(P_h, \{\lambda_h, \mu_h\})$. Then r can be written as $c\,h^{\xi_h}$ for some $c \in \mathbb{K}$ and $\xi_h \in \mathbb{Z}[\delta^{-1}, \delta]$ with $\delta = \delta^{(\lambda_h,\mu_h)}$. We call $(P_h, \{\lambda_h, \mu_h\}, \xi_h)$ a *univariate representation* of r. Assume that $(P_g, \{\lambda_g, \mu_g\}, \xi_g)$ is another univariate representation of r with $g \in \mathbb{K}[n,k]$. By the conclusion made in the preceding paragraph, we find that $g = h^{\delta^{\ell}}$ for some $\ell \in \mathbb{Z}$, or, equivalently, $P_g(z) = P_h(z + \ell)$. Moreover, $(\lambda_g, \mu_g) = (\lambda_h, \mu_h)$. It follows that $\xi_g = \delta^{\ell}\xi_h$. In particular, $\deg_z(P_h)$ is equal to $\deg_z(P_g)$ and the nonzero coefficients of ξ_h are exactly the same as those of ξ_g. When the choice of h and g is insignificant, we say that a tuple $(P, \{\lambda, \mu\}, \xi)$ is a *univariate representation* of r if the polynomial $P \in \mathbb{K}[z]$ is irreducible and $r(n,k) = cP(\lambda n + \mu k)^{\xi}$ for some $c \in \mathbb{K}$. Note that the coefficients of ξ are all nonnegative if r is a polynomial.

Let $r \in \mathbb{K}(n,k)$ be integer-linear with the shift-homogeneous decomposition

$$r = c\,r_1 \cdots r_s.$$

For $i = 1, \ldots, s$, assume that $U_i = (P_i, (\lambda_i, \mu_i), \xi_i)$ is a univariate representation of r_i. Then we call the tuple

$$(c,\ (U_1, \ldots, U_s))$$

a *univariate representation* of r.

To avoid unnecessary duplication, we make a notational convention.

Convention 6.2. *Let T be a hypergeometric term over $\mathbb{K}(n,k)$ with a multiplicative decomposition SH, where $S \in \mathbb{K}(n,k)$ and H is a hypergeometric term whose shift-quotient K w.r.t. k is shift-reduced w.r.t. k. By [8, Theorem 8], we know K is integer-linear over $\mathbb{K}$. Write $K = u/v$ where $u, v \in \mathbb{K}(n)[k]$ and $\gcd(u,v) = 1$.*

6.2 Shift-relation of residual forms

In this section, we describe a relation among residual forms of a given hypergeometric term and its shifts. This relation enables us to derive a shift-free common multiple of significant denominators of those residual forms, provided that telescopers exist. The existence of this common multiple implies that the residual forms span a finite-dimensional vector space over $\mathbb{K}(n)$, and then lead to order bounds for the minimal telescopers presented in the next section.

Lemma 6.3. *With Convention 6.2, let r be a residual form of S w.r.t. K. Then $\sigma_n(K)$ and $\sigma_n(S)$ are a kernel and a corresponding shell of $\sigma_n(T)$ w.r.t. k, respectively. Moreover, $\sigma_n(r)$ is a residual form of $\sigma_n(S)$ w.r.t. $\sigma_n(K)$.*

Proof. By Convention 6.2, $\sigma_n(T) = \sigma_n(S)\sigma_n(H)$ and $\sigma_n(K)$ is the shift-quotient of $\sigma_n(H)$ w.r.t. k. To prove the first assertion, one needs to show that $\sigma_n(K)$ is shift-reduced w.r.t. k. This can be proven by observing that, for any two polynomials $p_1, p_2 \in \mathbb{K}(n)[k]$, we have $\gcd(\sigma_n(p_1), \sigma_n(p_2)) = 1$ if and only if $\gcd(p_1, p_2) = 1$.

For the second assertion, since r is a residual form w.r.t. K, we write

$$r = \frac{a}{b} + \frac{q}{v},$$

where $a, b, q \in \mathbb{K}(n)[k]$, $\deg_k(a) < \deg_k(b)$, $\gcd(a, b) = 1$, b is shift-free and strongly coprime with K, and $q \in \mathbb{W}_K$. It is clear that $\deg_k(\sigma_n(a)) < \deg_k(\sigma_n(b))$ and $\gcd(\sigma_n(a), \sigma_n(b)) = 1$. By the above observation, $\sigma_n(b)$ is shift-free and strongly coprime with $\sigma_n(K)$.

Note that $\sigma_n \circ \deg_k = \deg_k \circ \sigma_n$ and $\sigma_n \circ \mathrm{lc}_k = \mathrm{lc}_k \circ \sigma_n$, where $\mathrm{lc}_k(p)$ is the leading coefficient of $p \in \mathbb{K}(n)[k]$ w.r.t. k. So the standard complements $\mathbb{W}_K$ and $\mathbb{W}_{\sigma_n(K)}$ for polynomial reduction have the same echelon basis according to the case study in Subsection 3.2.1. It follows from $q \in \mathbb{W}_K$ that $\sigma_n(q) \in \mathbb{W}_{\sigma_n(K)}$. Accordingly, $\sigma_n(r)$ is a residual form of $\sigma_n(S)$ w.r.t. $\sigma_n(K)$. □

Theorem 6.4. *With Convention 6.2, for every nonnegative integer i assume*

$$\sigma_n^i(T) = \Delta_k(g_i H) + \left(\frac{a_i}{b_i} + \frac{q_i}{v}\right) H, \tag{6.2}$$

where $g_i \in \mathbb{K}(n, k)$, $a_i, b_i \in \mathbb{K}(n)[k]$ with $\deg_k(a_i) < \deg_k(b_i)$, $\gcd(a_i, b_i) = 1$, b_i is shift-free w.r.t. k and strongly coprime with K, and q_i belongs to $\mathbb{W}_K$. Then b_i is shift-related to $\sigma_n^i(b_0)$, i.e., $b_i \approx_k \sigma_n^i(b_0)$.

Proof. We proceed by induction on i. For $i = 0$, the reflexivity of the relation $\approx_k$ implies that $b_0 \approx_k b_0$.

Assume that $b_{i-1} \approx_k \sigma_n^{i-1}(b_0)$ for $i \geq 1$. Note that K is also a kernel of $\sigma_n^{i-1}(T)$ and $\sigma_n^i(T)$ w.r.t. k. Let S_{i-1} and S_i be the corresponding shells, respectively. Consider the equality

$$\sigma_n^{i-1}(T) = \Delta_k(g_{i-1}H) + \left(\frac{a_{i-1}}{b_{i-1}} + \frac{q_{i-1}}{v}\right) H,$$

where $g_{i-1} \in \mathbb{K}(n,k)$ and $a_{i-1}/b_{i-1} + q_{i-1}/v$ is a residual form of S_{i-1} w.r.t. K. Applying σ_n to both sides yields

$$\begin{aligned}\sigma_n^i(T) &= \sigma_n(\Delta_k(g_{i-1}H)) + \sigma_n\left(\frac{a_{i-1}}{b_{i-1}} + \frac{q_{i-1}}{v}\right)\sigma_n(H) \\ &= \Delta_k(\sigma_n(g_{i-1}H)) + \left(\frac{\sigma_n(a_{i-1})}{\sigma_n(b_{i-1})} + \frac{\sigma_n(q_{i-1})}{\sigma_n(v)}\right)\sigma_n(H)\end{aligned}$$

It follows from Lemma 6.3 that $\sigma_n(K)$ and $\sigma_n(S_{i-1})$ are a kernel and the corresponding shell $\sigma_n^i(T)$ w.r.t. k, and $\sigma_n(a_{i-1})/\sigma_n(b_{i-1})+\sigma_n(q_{i-1})/\sigma_n(v)$ is a residual form of S_i w.r.t. $\sigma_n(K)$. By (6.2) with $i = 1$, we know that $a_i/b_i + q_i/v$ is a residual form of S_i w.r.t. K. By Theorem 4.10, $b_i \approx_k \sigma_n(b_{i-1})$. Thus $b_i \approx_k \sigma_n^i(b_0)$ by the induction hypothesis. □

Using the relation revealed in the above theorem, we can derive a common multiple as promised at the beginning of this section. To this end, we need two simple lemmas.

The first lemma says that, with Convention 6.2, for any $f \in \mathbb{K}(n)[k]$, there always exists $g \in \mathbb{K}(n)[k]$ such that $f \approx_k g$ and g is strongly coprime with K.

Lemma 6.5. *With Convention 6.2, assume that p is an irreducible polynomial in $\mathbb{K}(n)[k]$. Then there exists an integer m such that $\sigma_k^m(p)$ is strongly coprime with K.*

Proof. According to the definition of strong coprimeness, there is one and only one of the following three cases true.

Case 1. p is strongly coprime with K. Then the lemma follows by letting $m = 0$.

Case 2. There exists an integer $k \geq 0$ such that $\sigma_k^k(p) \mid u$. Then for every integer ℓ, we have $\gcd(\sigma_k^\ell(p), v) = 1$, since K is shift-reduced w.r.t. k. Let

$$m = \max\{i \in \mathbb{N} \mid \sigma_k^i(p) \mid u\} + 1.$$

One can see that $\sigma_k^m(p)$ is strongly coprime with K.

Case 3. There exists an integer $k \leq 0$ such that $\sigma_k^k(p) \mid v$. Then for every integer ℓ, we have $\gcd(\sigma_k^\ell(p), u) = 1$, since K is shift-reduced w.r.t. k. Letting

$$m = \min\{i \in \mathbb{N} \mid \sigma_k^i(p) \mid v\} - 1$$

yields that $\sigma_k^m(p)$ is strongly coprime with K. □

The next lemma shows that for any integer-linear polynomial in $\mathbb{K}[n,k]$, the number of shift-equivalence classes w.r.t. k produced by shifting the polynomial as a univariate one is finite.

Lemma 6.6. *Let $q \in \mathbb{K}[n,k]$ be integer-linear, and then $q = P(\lambda n + \mu k)$ for $P \in \mathbb{K}[z]$ and $\lambda, \mu \in \mathbb{Z}$ not both zero. Then any shift of q w.r.t. n or $z = \lambda n + \mu k$ is shift-equivalent to $\delta^j(q)$ w.r.t. k for $\delta = \delta^{(\lambda,\mu)}$ and $0 \le j \le \mu - 1$. More precisely, let*

$$S = \{\delta^j(q) \mid j = 0, \ldots, \mu - 1\},\ S_1 = \{\sigma_n^i(q) \mid i \in \mathbb{N}\} \text{ and } S_2 = \{\delta^j(q) \mid j \in \mathbb{N}\}.$$

Then for any element f in $S_1 \cup S_2$, there exists $g \in S$ such that $f \sim_k g$.

Proof. Assume that $f \in S_1 \cup S_2$. Since $\sigma_n = \delta^\lambda$, there exists a nonnegative integer i such that

$$f = \delta^i(q) = P(\lambda n + \mu k + i).$$

By Euclidean division, there exist unique integers j, ℓ with $0 \le j \le \mu - 1$, such that $i = \ell\mu + j$. It follows that

$$f = \sigma_k^\ell(P(\lambda n + \mu k + j)) = \sigma_k^\ell(\delta^j(q)).$$

Letting $g = \delta^j(q)$ completes the proof. □

Now we are ready to compute a common multiple as mentioned before.

Theorem 6.7. *With Convention 6.2, assume that*

$$T = \Delta_k(gH) + \left(\frac{a}{b} + \frac{q}{v}\right) H, \tag{6.3}$$

where $g \in \mathbb{K}(n,k)$, $a, b, q \in \mathbb{K}(n)[k]$, $\deg_k(a) < \deg_k(b)$, $\gcd(a,b) = 1$, b is shift-free w.r.t. k and strongly coprime with K, and $q \in \mathbb{W}_K$. Further assume that b is integer-linear and has a univariate representation

$$(c, (U_1, \ldots, U_s)), \quad \text{where } U_j = (P_j, (\lambda_j, \mu_j), \xi_j),\ j = 1, \ldots, s.$$

Then there exists $B \in \mathbb{K}(n)[k]$ such that $b \mid B$ and for all $i \in \mathbb{N}$,

$$\sigma_n^i(T) = \Delta_k(g_i H) + \left(\frac{a_i}{B} + \frac{q_i}{v}\right) H \tag{6.4}$$

for some $g_i \in \mathbb{K}(n,k), a_i \in \mathbb{K}(n)[k]$ with $\deg_k(a_i) < \deg_k(B)$, and $q_i \in \mathbb{W}_K$. Moreover,

(i) B is shift-free w.r.t. k and strongly coprime with K;

(ii) $\deg_k(B) = \sum_{j=1}^s \mu_j m_j \deg_k(P_j)$, where m_j is the maximum of the coefficients of ξ_j.

Proof. Since the shift-homogeneous components of b are coprime to each other, it suffices to consider the case when b is shift-homogeneous. W.l.o.g., assume that b is shift-homogeneous and has a univariate representation $(P, \{\lambda, \mu\}, \xi)$ such that

$$b = P(\lambda n + \mu k)^{\xi}.$$

Write $\xi = \sum_{i=0}^{d} m_i' \delta^i$ where $d \in \mathbb{N}$, $m_i' \in \mathbb{Z}$ and $\delta = \delta^{(\lambda,\mu)}$.

If $\mu = 0$ then $b \in \mathbb{K}(n)$. By the modified Abramov-Petkovšek reduction we can assume that (6.2) holds for every $i > 0$ and thus $b_i \in \mathbb{K}(n)$ by Theorem 6.4. The assertion follows by letting $B = 1$.

Otherwise we have $\mu > 0$. By Lemma 6.6, for every $i \in \mathbb{N}$ there are unique integers j and ℓ_j with $0 \leq j \leq \mu - 1$ such that

$$P(\lambda n + \mu k)^{\delta^i} = P(\lambda n + \mu k + j)^{\sigma_k^{\ell_j}},$$

which is equivalent to

$$P(\lambda n + \mu k + i) = P(\lambda n + \mu k + \mu \ell_j + j).$$

Since P is irreducible, we have $i = \mu \ell_j + j$. Let $m_j'' = m_{\mu \ell_j + j}'$. Since b is shift-free w.r.t. k,

$$b = \prod_{j=0}^{\mu-1} P(\lambda n + \mu k + j)^{m_j'' \sigma_k^{\ell_j}}.$$

For each j, if $m_j'' \neq 0$ then set $m_j = \ell_j$; otherwise by Lemma 6.5, let m_j be an integer so that $P(\lambda n + \mu k + j)^{\sigma_k^{m_j}}$ is strongly coprime with K. Let $m = \max_{0 \leq j \leq \mu - 1}\{m_j''\}$ and

$$B = \prod_{j=0}^{\mu-1} P(\lambda n + \mu k + j)^{m \sigma_k^{m_j}}. \tag{6.5}$$

Then $\deg_k(B) = \mu m \deg_k(P)$. Since $m_j = \ell_j$ when $m_j'' \neq 0$, every irreducible factor of b divides B and thus $b \mid B$ by the maximum of m. Because $0 \leq j \leq \mu - 1$, so B is shift-free w.r.t. k. Moreover, B is strongly coprime with K by the choice of m_j.

It remains to show that (6.4) holds for every nonnegative integer i. To prove this, we first show $\sigma_n(B) \approx_k B$. By (6.5), we have

$$B \approx_k \prod_{j=0}^{\mu-1} P(\lambda n + \mu k + j)^m,$$

which yields

$$\sigma_n(B) \approx_k \prod_{j=0}^{\mu-1} P(\lambda n + \mu k + j + \lambda)^m.$$

By Lemma 6.6, there exists a unique integer ℓ with $0 \leq \ell \leq \mu - 1$ such that

$$P(\lambda n + \mu k + j + \lambda) \sim_k P(\lambda n + \mu k + \ell).$$

Conversely, for any $0 \leq \ell \leq \mu - 1$, there exists a unique integer $0 \leq j \leq \mu - 1$ such that the above equivalence holds. Thus

$$\sigma_n(B) \approx_k \prod_{\ell=0}^{\mu-1} P(\lambda n + \mu k + \ell)^m \approx_k B.$$

For $i = 0$, letting $g_0 = g$, $a_0 = aB/b$ and $q_0 = q$ gives (6.4). Since $\sigma_n(B) \approx_k B$, we have $\sigma_n^i(B) \approx_k \sigma_n^{i-1}(B)$ for every positive integer i, and then $\sigma_n^i(B) \approx_k B$.

On the other hand, by the modified Abramov-Petkovšek reduction (6.2) holds for every $i \geq 0$, in which $b_0 = b$. According to Theorem 6.4, $b_i \approx_k \sigma_n^i(b_0)$. It follows from $b \mid B$ that $\sigma_n^i(b) \mid \sigma_n^i(B)$. Consequently, we have

$$b_i \approx_k \sigma_n^i(b) \mid \sigma_n^i(B) \approx_k B.$$

Thus there is $\tilde{b}_i \in \mathbb{K}(n)[k]$ dividing B so that $\tilde{b}_i \approx_k b_i$. Moreover, $\tilde{b}_i$ is strongly coprime with K as B is. By the shifting property of significant denominators (i.e., Lemma 4.17 and Remark 4.18), there exist $\tilde{g}_i \in \mathbb{K}(n,k)$, $\tilde{a}_i, \tilde{q}_i \in \mathbb{K}(n)[k]$ with $\deg_k(\tilde{a}_i) < \deg_k(\tilde{b}_i)$, and $\tilde{q}_i \in \mathbb{W}_K$ such that $\sigma_n^i(T) = \Delta_k(\tilde{g}_i H) + (\tilde{a}_i/\tilde{b}_i + \tilde{q}_i/v)H$. The assertion follows by noticing

$$\sigma_n^i(T) = \Delta_k(\tilde{g}_i H) + \left(\frac{\tilde{a}_i B/\tilde{b}_i}{B} + \frac{\tilde{q}_i}{v}\right) H.$$

□

Under the assumptions of Theorem 6.7, applying Algorithm 3.17 to T w.r.t. k yields $T = \Delta_k(gH) + rH$, where $g \in \mathbb{K}(n,k)$ and r is a residual form w.r.t. K. By Theorems 4.6 and 4.10, b and the significant denominator r_d of r are shift-related w.r.t. k, and thus so are the respective shift-homogeneous components. W.l.o.g., assume that b is shift-homogeneous (then so is r_d). Let $(P_b, \{\lambda_b, \mu_b\}, \xi_b)$ be a univariate representation of b and $(P_{r_d}, \{\lambda_{r_d}, \mu_{r_d}\}, \xi_{r_d})$ be one of r_d. Definition 4.5 yields that $(\lambda_b, \mu_b) = (\lambda_{r_d}, \mu_{r_d})$ and for each integer i, there exists a unique integer j and another integer ℓ_{ij} such that

$$P_b(z)^{\delta^i} = \sigma_k^{\ell_{ij}}\left(P_{r_d}(z)^{\delta^j}\right) = P_{r_d}(z + \mu_{r_d}\ell_{ij})^{\delta^j} \quad \text{with } \delta = \delta^{(\lambda_b,\mu_b)}.$$

Moreover, the nonzero coefficients of ξ_b are exactly the same as those of ξ_{r_d}. In summary, we have the following remark.

Remark 6.8. Although the form of B in Theorem 6.7 depends on the choice of b, the shift-equivalence classes w.r.t. $\sim_{n,k}$ as well as the degree of B w.r.t. k depend only on the hypergeometric term T.

6.3 Upper and lower order bounds

In this section, we show that Theorem 6.7 implies that some residual forms $\{a_i/b_i+q_i/v\}_{i\geq 0}$ satisfying (6.2) form a finite-dimensional vector space over $\mathbb{K}(n)$, and then derive an upper bound for the order of minimal telescopers.

Theorem 6.9. *With the assumptions of Theorem 6.7, we have that the order of a minimal telescoper for T w.r.t. k is no more than*

$$\max\{\deg_k(u), \deg_k(v)\} - [\![\deg_k(v-u) \leq \deg_k(u) - 1]\!] + \sum_{j=1}^{s} \mu_j m_j \deg_k(P_j),$$

where $[\![\varphi]\!]$ equals 1 if φ is true, otherwise it is 0.

Proof. Let $L = \sum_{i=0}^{\rho} e_i S_n^i$ be a minimal telescoper for T w.r.t. k, where $\rho \in \mathbb{N}$ and $e_0, \ldots, e_\rho \in \mathbb{K}(n)$ not all zero. By Theorem 6.7, there exists $B \in \mathbb{K}(n)[k]$ such that (6.4) holds for every nonnegative integer i. Then by Theorem 5.7, the residual forms $\{a_i/B + q_i/v\}_{i=0}^{\rho}$ are linearly dependent over $\mathbb{K}(n)$; equivalently, the following linear system with unknowns $e_0, \ldots, e_\rho$

$$\begin{cases} A_\rho = e_0 a_0 \ + \ e_1 a_1 \ + \ \cdots \ + \ e_\rho a_\rho = 0 \\ Q_\rho = e_0 q_0 \ + \ e_1 q_1 \ + \ \cdots \ + \ e_\rho q_\rho = 0 \end{cases} \tag{6.6}$$

has a nontrivial solution in $\mathbb{K}(n)^{\rho+1}$. Since $\deg_k(a_i) < \deg_k(B)$,

$$\deg_k(A_\rho) < \deg_k(B) = \sum_{j=1}^{s} \mu_j m_j \deg_k(P_j). \tag{6.7}$$

Note that $\mathbb{W}_K$ is a vector space, so $Q_\rho \in \mathbb{W}_K$. By Proposition 3.15, the number of nonzero terms w.r.t. k in Q_ρ is no more than the dimension $\dim_{\mathbb{K}(n)}(\mathbb{W}_K)$, which is bounded by

$$\max\{\deg_k(u), \deg_k(v)\} - [\![\deg_k(v-u) \leq \deg_k(u) - 1]\!]. \tag{6.8}$$

Comparing coefficients of like powers of k of the linear system (6.6) yields at most

$$\deg_k(A_\rho) + \dim_{\mathbb{K}(n)}(\mathbb{W}_K) + 1 \tag{6.9}$$

equations. Hence this system has nontrivial solutions whenever the order ρ exceeds $\deg_k(A_\rho) + \dim_{\mathbb{K}(n)}(\mathbb{W}_K)$. It implies that the order of a minimal telescoper for T w.r.t. k is no more than the number (6.9). Therefore, the theorem follows by (6.7) and (6.8). □

In addition, we can further obtain a lower order bound for telescopers for T.

Theorem 6.10. *With the assumptions of Theorem 6.7, further assume that T is not summable w.r.t. k. Then the order of a telescoper for T w.r.t. k is at least*

$$\max_{\substack{p|b,\, \deg_k(p)>0 \\ \text{multiplicity } \alpha \\ \text{monic \& irred.}}} \min\left\{\rho \in \mathbb{N}\setminus\{0\} : \sigma_k^{\ell}(p)^{\alpha} \mid \sigma_n^{\rho}(b) \text{ for some } \ell \in \mathbb{Z}\right\}.$$

Proof. Let $L = \sum_{i=0}^{\rho} e_i S_n^i$ be a minimal telescoper for T w.r.t. k, where $\rho \in \mathbb{N}$ and $e_0, \ldots, e_\rho \in \mathbb{K}(n)$ not all zero. Since T is not summable w.r.t. k, we have $\rho \geq 1$. By the modified Abramov-Petkovšek reduction, (6.2) holds for $1 \leq i \leq \rho$. Since L is a minimal telescoper, $e_0 \neq 0$ and by Theorem 5.7,

$$e_0 \frac{a}{b} \;+\; e_1 \frac{a_1}{b_1} \;+\; \cdots \;+\; e_\rho \frac{a_\rho}{b_\rho} = 0.$$

By partial fraction decomposition, for any monic irreducible factor p of b with $\deg_k(p) > 0$ and multiplicity $\alpha > 0$, there exists an integer i with $1 \leq i \leq \rho$ so that p^α is also a factor of b_i. By Theorem 6.4, $b_i \approx_k \sigma_n^i(b)$. Thus there is a factor p' of $\sigma_n^i(b)$ with multiplicity at least α such that $p' \sim_k p$. Let i_p be the minimal one with this property. Then the assertion follows by the fact that for each factor p of b there exists no telescoper for T of order less than i_p. □

Together with the bounds given above, we can further develop a variant of Algorithm 5.6 by omitting step 4.5 for each loop until the loop index i reaches and exceeds the lower bound.

Algorithm 6.11 (Bound and Reduction-based creative telescoping)**.**
Input: A hypergeometric term T over $\mathbb{F}(k)$.
Output: A minimal telescoper for T w.r.t. k and a corresponding certificate if telescopers exist; "No telescoper exists!", otherwise.

1–3 Similar to steps $1-3$ of Algorithm 5.6.

4 Compute the upper bound $b_u \in \mathbb{N}$ and lower bound $b_l \in \mathbb{N}$ for the order of minimal telescopers for T w.r.t. k, respectively.

5 Set $N = \sigma_n(H)/H$ and $R = \ell_0 r_0$, where ℓ_0 is an indeterminate.
For $i = 1, 2, \ldots, b_u$ do

5.1 Similar to steps $4.1 - 4.3$ of Algorithm 5.6, compute $g_i, r_i \in \mathbb{K}(n,k)$ such that (5.4) holds, and $R + \ell_i r_i$ is a residual form w.r.t. K, where ℓ_i is an indeterminate.

5.2 Update R to $R + \ell_i r_i$. If $i > b_l$ then find $\ell_j \in \mathbb{F}$ such that $R = 0$ by solving a linear system in $\ell_0, \ldots, \ell_i$ over $\mathbb{F}$.
If there is a nontrivial solution, return $\left(\sum_{j=0}^{i} \ell_j S_n^j,\ \sum_{j=0}^{i} \ell_j g_j H\right)$.

6.4 Comparison of bounds

In 2005, upper and lower bounds for the order of telescopers for hypergeometric terms have been studied in [49] and [6], respectively. In this section, we are going to review these known bounds and also compare them to our bounds.

6.4.1 Apagodu-Zeilberger upper bound

Let T be a *proper* hypergeometric term over $\mathbb{K}(n,k)$, i.e., it can be written in the form

$$T = pw^n z^k \prod_{i=1}^{m} \frac{(\alpha_i n + \alpha_i' k + \alpha_i'' - 1)!(\beta_i n - \beta_i' k + \beta_i'' - 1)!}{(\mu_i n + \mu_i' k + \mu_i'' - 1)!(\nu_i n - \nu_i' k + \nu_i'' - 1)!}, \tag{6.10}$$

where $p \in \mathbb{K}[n,k]$, $w, z \in \mathbb{K}$, $m \in \mathbb{N}$ is fixed, $\alpha_i, \alpha_i', \beta_i, \beta_i', \mu_i, \mu_i', \nu_i, \nu_i'$ are nonnegative integers and $\alpha_i'', \beta_i'', \mu_i'', \nu_i'' \in \mathbb{K}$. Further assume that there exist no integers i and j with $1 \leq i, j \leq m$ such that

$$\begin{aligned} & \big(\alpha_i = \mu_j \quad \text{and} \quad \alpha_i' = \mu_j' \quad \text{and} \quad \alpha_i'' - \mu_j'' \in \mathbb{N}\big) \\ \text{or} \quad & \big(\beta_i = \nu_j \quad \text{and} \quad \beta_i' = \nu_j' \quad \text{and} \quad \beta_i'' - \nu_j'' \in \mathbb{N}\big). \end{aligned}$$

We refer to this as the *generic* situation. Then Apagodu and Zeilberger [49] stated that the order of a minimal telescoper for T w.r.t. k is bounded by

$$B_{AZ} = \max\left\{\sum_{i=1}^{m}(\alpha_i' + \nu_i'), \sum_{i=1}^{m}(\beta_i' + \mu_i')\right\},$$

and this bound is generically sharp.

We now show that B_{AZ} is at least the order bound given in Theorem 6.9. Reordering the factorial terms in (6.10) if necessary, let $\mathcal{S}$ be the maximal set of integers i with $1 \leq i \leq m$ satisfying

$$\begin{aligned} & \big(\alpha_i = \mu_i \quad \text{and} \quad \alpha_i' = \mu_i' \quad \text{and} \quad \mu_i'' - \alpha_i'' \in \mathbb{N}\big) \\ \text{or} \quad & \big(\beta_i = \nu_i \quad \text{and} \quad \beta_i' = \nu_i' \quad \text{and} \quad \nu_i'' - \beta_i'' \in \mathbb{N}\big). \end{aligned}$$

Rewrite T as

$$rw^n z^k \prod_{i=1,\, i\notin\mathcal{S}}^{m} \frac{(\alpha_i n + \alpha_i' k + \alpha_i'' - 1)!(\beta_i n - \beta_i' k + \beta_i'' - 1)!}{(\mu_i n + \mu_i' k + \mu_i'' - 1)!(\nu_i n - \nu_i' k + \nu_i'' - 1)!},$$

where $r \in \mathbb{K}(n,k)$. For $q \in \mathbb{K}[n,k]$ and $m \in \mathbb{N}$, let

$$q^{\overline{m}} = q(q+1)(q+2)\cdots(q+m-1)$$

with the convention $q^{\overline{0}} = 1$. By an easy calculation,

$$K = z \prod_i \frac{(\alpha_i n + \alpha_i' k + \alpha_i'')^{\overline{\alpha_i'}} (\nu_i n - \nu_i' k + \nu_i'' - \mu_i')^{\overline{\nu_i'}}}{(\mu_i n + \mu_i' k + \mu_i'')^{\overline{\mu_i'}} (\beta_i n - \beta_i' k + \beta_i'' - \beta_i')^{\overline{\beta_i'}}} \tag{6.11}$$

where the product runs over all i from 1 to m such that $i \notin \mathcal{S}$, $\alpha_i', \beta_i' > 0$ and $\mu_i', \nu_i' > 0$, is a kernel of T and $S = r$ is a corresponding shell. Let $K = u/v$ with $u, v \in \mathbb{K}(n)[k]$ and $\gcd(u, v) = 1$. Note that the right-hand side of (6.11) already has the reduced form, then a straightforward calculation yields

$$\deg_k(u) = \sum_{i=1, i\notin\mathcal{S}}^{m} (\alpha_i' + \nu_i') \quad \text{and} \quad \deg_k(v) = \sum_{i=1, i\notin\mathcal{S}}^{m} (\beta_i' + \mu_i').$$

Applying the modified Abramov-Petkovšek reduction to T w.r.t. k yields (6.3), in which b is integer-linear. Since b only comes from the shift-free part of the denominator of r, it factors into shift-inequivalent integer-linear polynomials of degree one which are separately shift-equivalent to either $(\mu_i n + \mu_i' k + \mu_i'')$ or $(\beta_i n - \beta_i' k + \beta_i'')$ w.r.t. n, k for some $i \in \mathcal{S}$. Note that each i in $\mathcal{S}$ corresponds to at most one integer-linear factor of b, and increases the multiplicity of the corresponding factor in b by at most 1. Hence, the bound given in Theorem 6.9 is no more than

$$\max\{\deg_k(u), \deg_k(v)\} - [\![\deg_k(v-u) \le \deg_k(u) - 1]\!] + \sum_{i=1, i\in\mathcal{S}}^{m} (\beta_i' + \mu_i'),$$

which is exactly equal to

$$B_{AZ} - [\![\deg_k(v-u) \le \deg_k(u) - 1]\!],$$

since $\sum_{i=1, i\in\mathcal{S}}^{m} (\alpha_i' + \nu_i') = \sum_{i=1, i\in\mathcal{S}}^{m} (\beta_i' + \mu_i')$.

In general, i.e., in the generic situation, the order bound in Theorem 6.9 is almost the same as B_{AZ}, which is not suprising since B_{AZ} is already generically sharp. However, our bound can be much better in some special examples.

Example 6.12. Consider a rational function

$$T = \frac{\alpha^2 k^2 + \alpha^2 k - \alpha\beta k + 2\alpha n k + n^2}{(n + \alpha k + \alpha)(n + \alpha k)(n + \beta k)},$$

where α, β are positive integers and $\alpha \ne \beta$. Rewriting T into the proper form (6.10) yields $B_{AZ} = \alpha + \beta$. On the other hand, 1 is the only kernel of T since T is a rational function. By the modified Abramov-Petkovšek reduction, $b = n + \beta k$ in (6.3). By Theorem 6.9, a minimal telescoper for T w.r.t. k has order no more than β, which is in fact the real order of minimal telescopers for T w.r.t. k.

Remark 6.13. Together with [4, Theorem 10], the upper order bound B_{AZ} on minimal telescopers derived in [49] can be also applied to non-proper hypergeometric terms. On the other hand, Theorem 6.9 can be directly applied to any hypergeometric term provided that its telescopers exist.

6.4.2 Abramov-Le lower bound

With Convention 6.2, further assume that T has the initial reduction (6.3), in which b is integer-linear. Let $H' = H/v$. A direct calculation leads to

$$\frac{\sigma_k(H')}{H'} = \frac{u}{\sigma_k(v)},$$

which can be easily checked to be shift-reduced w.r.t. k. Let $d' \in \mathbb{K}(n)[k]$ be the denominator of $\sigma_n(H')/H'$. Then the algorithm *LowerBound* in [6] asserts that the order of telescopers for T w.r.t. k is at least

$$B_{AL} = \max_{\substack{p \mid b \\ \text{irred. \& monic} \\ \deg_k(p) > 0}} \min \left\{ \rho \in \mathbb{N} \setminus \{0\} : \begin{array}{c} \sigma_k^\ell(p) \mid \sigma_n^\rho(b) \\ \text{or} \\ \sigma_k^\ell(p) \mid \sigma_n^{\rho-1}(d') \text{ for some } \ell \in \mathbb{Z} \end{array} \right\}$$

Comparing to B_{AL} from above, one easily sees that the lower bound given in Theorem 6.10 can be better but never worse than B_{AL}.

Example 6.14. Consider a hypergeometric term

$$T = \frac{1}{(n - \alpha k - \alpha)(n - \alpha k - 2)!},$$

where $\alpha \in \mathbb{N}$ and $\alpha > 1$. By the algorithm *LowerBound*, a telescoper for T w.r.t. k has order at least 2. On the other hand, a telescoper for T w.r.t. k has order at least α by Theorem 6.10. In fact, α is exactly the order of minimal telescopers for T w.r.t. k.

6.5 Implementation and timings

In Maple 18, we have implemented Algorithm 6.11 and embedded it into the package **ShiftReductionCT**, under the name of BoundReductionCT. For a detailed explanation, one may refer to Appendix A.

In this section, we focus on the two procedures – BoundReductionCT and ReductionCT in the package **ShiftReductionCT**, and their runtime is compared. All timings are measured in seconds on a Linux computer with 388Gb RAM and twelve 2.80GHz Dual core processors. No parallelism was used in this experiment. Moreover, a comparison of the memory requirements is given in Appendix B. For brevity, we denote

- RCT_{tc}: the procedure ReductionCT in **ShiftReductionCT**, which computes a minimal telescoper and a corresponding normalized certificate;

- RCT_t: the procedure ReductionCT in **ShiftReductionCT**, which computes a minimal telescoper without constructing a certificate.
- BRCT_{tc}: the procedure BoundReductionCT in **ShiftReductionCT**, which computes a minimal telescoper and a corresponding normalized certificate;
- BRCT_t: the procedure BoundReductionCT in **ShiftReductionCT**, which computes a minimal telescoper without constructing a certificate.
- LB: the lower bound for telescopers given in Theorem 6.10.
- order: the order of the resulting minimal telescoper.

Example 6.15. Consider the same hypergeometric term as in Example 6.14, i.e.,

$$T = \frac{1}{(n - \alpha k - \alpha)(n - \alpha k - 2)!},$$

where α is an integer greater than 1. For different choices of α, Table 6.1 shows the timings of the above procedures. Note that since the term T in this example is very simple, there is little difference in the timings for the two procedures with and without construction of a certificate.

α	RCT_t	RCT_{tc}	BRCT_t	BRCT_{tc}	LB	order
20	2.00	2.02	1.07	1.13	20	20
30	7.01	7.19	2.86	2.96	30	30
40	20.08	20.13	7.06	7.18	40	40
50	42.15	42.68	14.96	15.05	50	50
60	104.07	106.31	25.54	25.93	60	60
70	225.67	229.04	45.76	45.97	70	70

Table 6.1: Timing comparison of two reduction-basedcreative telescoping with and without construction of a certificate for Example 6.15 (in seconds)

Example 6.16 (Example 6 in [6])**.** Consider the hypergeometric term

$$T = \Delta_k (T_1) + T_2,$$

where

$$T_1 = \frac{1}{(nk-1)(n-\alpha k-2)^m(2n+k+3)!} \text{ and } T_2 = \frac{1}{(n-\alpha k-2)(2n+k+3)!}$$

for α, m positive integers. For different choices of α and m, we compare the timings of the procedures from above. Table 6.2 shows the final experimental results.

(m, α)	RCT_t	RCT_{tc}	BRCT_t	BRCT_{tc}	LB	order
(1,1)	0.20	0.24	0.20	0.23	1	2
(1,10)	5.25	9.56	4.60	8.74	10	11
(1,15)	57.06	76.01	37.73	58.69	15	16
(1,20)	538.59	656.99	264.04	324.09	20	21
(2,10)	5.29	9.11	4.43	8.36	10	11
(2,15)	79.34	96.48	40.26	54.85	15	16
(2,20)	574.00	658.20	282.54	377.84	20	21

Table 6.2: Timing comparison of two reduction-basedcreative telescoping with and without construction of a certificate for Example 6.16 (in seconds)

Remark 6.17. Compared to linear dependence, determining linear independence takes much less time because with high probability, independence can be recognized by a computation in a homomorphic image. For this reason, the procedure BoundReductionCT makes no big difference from the procedure ReductionCT if the lower bound is far away from the real order of minimal telescopers. In fact, their perform almost the same in this case.

Part II

Limits of P-recursive sequences

Chapter 7

D-finite Functions and P-recursive Sequences

In this chapter, we recall [34, 41] basic notions related to the class of D-finite functions and P-recursive sequences, and also present some useful properties.

7.1 Basic concepts

Recall [41] that a *formal power series* is an infinite series of the form

$$f(z) = \sum_{n=0}^{\infty} a_n z^n,$$

where z is a formal indeterminate. It generalizes the notions of polynomials and power series in some sense. A formal power series differs from a polynomial in that it allows an infinite number of terms, and it differs from power series by assuming a formal variable and ignoring analytic properties. One way to view a formal power series $f(z)$ is to take it as an infinite sequence $(a_n)_{n=0}^{\infty}$, where the powers indicate the order of terms. We will also call a formal power series $f(x)$ the *generating function* of its coefficient sequence $(a_n)_{n=0}^{\infty}$. Note that these three notions – formal power series, sequences, generating functions – all refer to the same object.

For a ring R, we denote by $R[[z]]$ the ring of formal power series endowed with termwise addition $(+)$ and *Cauchy product* $(\cdot)$:

$$\left(\sum_{n=0}^{\infty} a_n z^n\right) + \left(\sum_{n=0}^{\infty} b_n z^n\right) = \sum_{n=0}^{\infty} (a_n + b_n)\, z^n,$$

$$\left(\sum_{n=0}^{\infty} a_n z^n\right) \cdot \left(\sum_{n=0}^{\infty} b_n z^n\right) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k}\right) z^n,$$

and by $R^{\mathbb{N}}$ the ring of infinite sequences endowed with termwise addition (+) and *Hadamard product* ($\odot$):

$$(a_n)_{n=0}^{\infty} + (b_n)_{n=0}^{\infty} = (a_n + b_n)_{n=0}^{\infty},$$
$$(a_n)_{n=0}^{\infty} \odot (b_n)_{n=0}^{\infty} = (a_n b_n)_{n=0}^{\infty}.$$

Also recall [34] that a complex function $f(z)$ is called *analytic* at a point $\zeta \in \mathbb{C}$ if for any z in a neighborhood of ζ, it can be represented by a convergent power series over $\mathbb{C}$,

$$f(z) = \sum_{n=0}^{\infty} a_n (z-\zeta)^n, \quad \text{where } a_n \in \mathbb{C} \text{ for all } n \in \mathbb{N}.$$

A function is *analytic* in an open set if it is analytic at every point of the set.

Throughout the chapter, let R be a subring of $\mathbb{C}$ and $\mathbb{F}$ be a subfield of $\mathbb{C}$. We consider linear operators that act on sequences or power series and analytic functions. Recall from the previous chapters that we write σ_n for the shift operator w.r.t. n which maps a sequence $(a_n)_{n=0}^{\infty}$ to $(a_{n+1})_{n=0}^{\infty}$. Also we denote by $\mathbb{F}[n]\langle S_n\rangle$ the ring of linear recurrence operators of the form $L := p_0 + p_1 S_n + \cdots + p_\rho S_n^\rho$, with $p_0, \ldots, p_\rho \in \mathbb{F}[n]$, where $S_n r = \sigma_n(r) S_n$ for all $r \in \mathbb{F}[n]$. This ring forms an Ore algebra. Analogously, we write D_z for the derivation operator w.r.t. z which maps a power series or function $f(z)$ to its derivative $f'(z) = \frac{d}{dz} f(z)$. Also the set of linear operators of the form $L := p_0 + p_1 D_z + \cdots + p_\rho D_z^\rho$, with $p_0, \ldots, p_\rho \in \mathbb{F}[z]$, forms an Ore algebra; we denote it by $\mathbb{F}[z]\langle D_z\rangle$. For an introduction to Ore algebras and their actions, please refer to [17]. When $p_\rho \neq 0$, we call ρ the *order* of the operator and $\mathrm{lc}(L) := p_\rho$ its *leading coefficient*.

Definition 7.1.

1. *A sequence* $(a_n)_{n=0}^{\infty} \in R^{\mathbb{N}}$ *is called* P-recursive *or* D-finite *over* $\mathbb{F}$ *if there exists a nonzero operator* $L = \sum_{j=0}^{\rho} p_j(n) S_n^j \in \mathbb{F}[n]\langle S_n\rangle$ *such that*
$$L \cdot a_n = p_\rho(n) a_{n+\rho} + \cdots + p_1(n) a_{n+1} + p_0(n) a_n = 0$$
for all $n \in \mathbb{N}$*.*
2. *A formal power series* $f(z) \in R[[z]]$ *is called* D-finite *over* $\mathbb{F}$ *if there exists a nonzero operator* $L = \sum_{j=0}^{\rho} p_j(z) D_z^j \in \mathbb{F}[z]\langle D_z\rangle$ *such that*
$$L \cdot f(z) = p_\rho(z) D_z^\rho f(z) + \cdots + p_1(z) D_z f(z) + p_0(z) f(z) = 0.$$
3. *A formal power series* $f(z) \in \mathbb{F}[[z]]$ *is called* algebraic *over* $\mathbb{F}$ *if there exists a nonzero bivariate polynomial* $P(z,y) \in \mathbb{F}[z,y]$ *such that* $P(z, f(z)) = 0$*.*

In general, D-finite power series are called D-finite functions instead. A formal power series is D-finite if and only if its coefficient sequence is P-recursive. Many elementary functions like rational functions, exponentials, logarithms, sine, algebraic functions, etc., as well as many special functions, like hypergeometric series, the error function, Bessel functions, etc., are D-finite. Hence their respective coefficient sequences are P-recursive.

7.2 Useful properties

The class of D-finite functions (resp. P-recursive sequences) is closed under certain operations: addition, multiplication, derivative (resp. forward shift) and integration (resp. summation). In particular, the set of D-finite functions (resp. P-recursive sequences) forms a left-$\mathbb{F}[z]\langle D_z\rangle$-module (resp. a left-$\mathbb{F}[n]\langle S_n\rangle$-module). Also, if f is a D-finite function and g is an algebraic function, then the composition $f \circ g$ is D-finite. These and further closure properties are easily proved by linear algebra arguments, whose proofs can be found for instance in [59, 57, 41]. We will make free use of these facts.

We will be considering singularities of D-finite functions. Recall from the classical theory of linear differential equations [40] that a linear differential equation $p_0(z)f(z) + \cdots + p_\rho(z)f^{(\rho)}(z) = 0$ with polynomial coefficients $p_0, \ldots, p_\rho \in \mathbb{F}[z]$ and $p_\rho \neq 0$ has a basis of analytic solutions in a neighborhood of every point $\zeta \in \mathbb{C}$, except possibly at roots of p_ρ. The roots of p_ρ are therefore called the *singularities* of the equation (or the corresponding linear operator). If $\zeta \in \mathbb{C}$ is a singularity of the equation but the equation nevertheless admits a basis of analytic solutions at this point, we call it an *apparent singularity*. It is well-known [40, 25] that for any given linear differential equation with some apparent and some non-apparent singularities, we can always construct another linear differential equation (typically of higher order) whose solution space contains the solution space of the first equation and whose only singularities are the non-apparent singularities of the first equation. This process is known as desingularization.

For later use, we will give a proof of the composition closure property for D-finite functions which pays attention to the singularities.

Theorem 7.2. *Let $P(z, y) \in \mathbb{F}[z, y]$ be a square-free polynomial of degree d, and let $L \in \mathbb{F}[z]\langle D_z\rangle$ be nonzero. Let $\zeta \in \mathbb{C}$ be such that P defines d distinct analytic algebraic functions $g(z)$ with $P(z, g(z)) = 0$ in a neighborhood of ζ, and assume that for none of these functions, the value $g(\zeta) \in \mathbb{C}$ is a singularity of L. Fix a solution g of P and an analytic solution f of L defined in a neighborhood of $g(\zeta)$. Then there exists a nonzero operator $M \in \mathbb{F}[z]\langle D_z\rangle$ with $M \cdot (f \circ g) = 0$ which does not have ζ among its singularities.*

Proof. (borrowed from [42]) Consider the operator $\tilde{L} = L(g, (g')^{-1}D_z) \in \overline{\mathbb{F}(z)}\langle D_z\rangle$. It is easy to check that $L \cdot f = 0$ if and only if $\tilde{L} \cdot (f \circ g) = 0$ for every solution g of P near ζ. Therefore, if $f_1, \ldots, f_\rho$ is a basis of the solution space of L near $g(\zeta)$, then $f_1 \circ g, \ldots, f_\rho \circ g$ is a basis of the solution space of $\tilde{L}$ near ζ.

Let $g_1, \ldots, g_d$ be all the solutions of P near ζ, and let M be the least common left multiple of all the operators $L(g_j, (g_j')^{-1}D_z)$. Then the solution space of M near ζ is generated by all the functions $f_i \circ g_j$. Since the coefficients of M are symmetric w.r.t. the conjugates $g_1, \ldots, g_d$, they belong to the ground field $\mathbb{F}(z)$, and after clearing denominators (from the left) if necessary, we may assume that

M is an operator in $\mathbb{F}[z]\langle D_z\rangle$ whose solution space is generated by functions that are analytic at ζ. Therefore, by the remarks made about desingularization, it is possible to replace M by an operator (possibly of higher order) which does not have ζ among its singularities. □

By a similar argument, we see that algebraic extensions of the coefficient field of the recurrence operators are useless. Moreover, it is also not useful to make $\mathbb{F}$ bigger than the quotient field of R.

Lemma 7.3.

1. *If $\mathbb{E}$ is an algebraic extension field of $\mathbb{F}$ and $(a_n)_{n=0}^{\infty}$ is P-recursive over $\mathbb{E}$, then it is also P-recursive over $\mathbb{F}$.*
2. *If $R \subseteq \mathbb{F}$ and $(a_n)_{n=0}^{\infty} \in R^{\mathbb{N}}$ is P-recursive over $\mathbb{F}$, then it is also P-recursive over* $\mathrm{Quot}(R)$*, the quotient field of R.*
3. *If $\mathbb{F}$ is closed under complex conjugation and $(a_n)_{n=0}^{\infty}$ is P-recursive over $\mathbb{F}$, then so are $(\bar{a}_n)_{n=0}^{\infty}$,* $(\mathrm{Re}(a_n))_{n=0}^{\infty}$*, and* $(\mathrm{Im}(a_n))_{n=0}^{\infty}$.

Proof. 1. Let $L \in \mathbb{E}[n]\langle S_n\rangle$ be an annihilating operator of $(a_n)_{n=0}^{\infty}$. Then, since L has only finitely many coefficients, $L \in \mathbb{F}(\theta)[n]\langle S_n\rangle$ for some $\theta \in \mathbb{E}$. Let M be the least common left multiple of all the conjugates of L. Then M is an annihilating operator of $(a_n)_{n=0}^{\infty}$ which belongs to $\mathbb{F}[n]\langle S_n\rangle$. The claim follows.

2. Let us write $\mathbb{K} = \mathrm{Quot}(R)$. Let $L \in \mathbb{F}[n]\langle S_n\rangle$ be a nonzero annihilating operator of $(a_n)_{n=0}^{\infty}$. Since $\mathbb{F}$ is an extension field of $\mathbb{K}$, it is a vector space over $\mathbb{K}$. Write

$$L = \sum_{m=0}^{\rho}\sum_{j=0}^{d_m} p_{mj} n^j S_n^m,$$

where $r, d_m \in \mathbb{N}$ and $p_{mj} \in \mathbb{F}$ not all zero. Then the set of the coefficients p_{ij} belongs to a finite dimensional subspace of $\mathbb{F}$. Let $\{\alpha_1, \dots, \alpha_s\}$ be a basis of this subspace over $\mathbb{K}$. Then for each pair (m, j), there exists $c_{mj\ell} \in \mathbb{K}$ such that $p_{mj} = \sum_{\ell=1}^{s} c_{mj\ell}\alpha_\ell$, which gives

$$0 = L \cdot a_n = \sum_{\ell=1}^{s} \alpha_\ell \underbrace{\left(\sum_{m=0}^{\rho}\sum_{j=0}^{d_m} c_{mj\ell} n^j a_{n+m}\right)}_{=:b_n \in \mathbb{K}}.$$

For all $n \in \mathbb{N}$, it follows from the linear independence of $\{\alpha_1, \dots, \alpha_s\}$ over $\mathbb{K}$ that $b_n = 0$. Therefore

$$\sum_{m=0}^{\rho} \underbrace{\left(\sum_{j=0}^{d_m} c_{mj\ell} n^j\right)}_{\in \mathbb{K}[n]} S_n^m \cdot a_n = 0 \quad \text{for all } n \in \mathbb{N} \text{ and } \ell = 1, \dots, s.$$

Thus $(a_n)_{n=0}^{\infty}$ has a nonzero annihilating operator with coefficients in $\mathbb{K}[n]$.

3. Since $(a_n)_{n=0}^{\infty}$ is P-recursive over $\mathbb{F}$, there exists a nonzero operator L in $\mathbb{F}[n]\langle S_n\rangle$ such that $L \cdot a_n = 0$. Hence $\bar{L} \cdot \bar{a}_n = 0$ where $\bar{L}$ is the operator obtained from L by taking the complex conjugate of each coefficient. Since $\mathbb{F}$ is closed under complex conjugation by assumption, we see that $\bar{L}$ belongs to $\mathbb{F}[n]\langle S_n\rangle$, and hence $(\bar{a}_n)_{n=0}^{\infty}$ is P-recursive over $\mathbb{F}$.

   Because of $\mathrm{Re}(a_n) = \frac{1}{2}(a_n + \bar{a}_n)$ and $\mathrm{Im}(a_n) = \frac{1}{2i}(a_n - \bar{a}_n)$ with i the imaginary unit, the other two assertions follow by closure properties.

□

Of course, all the statements hold analogously for D-finite functions instead of P-recursive sequences.

If we consider a D-finite function as an analytic complex function defined in a neighborhood of zero, then this function can be extended by analytic continuation to any point in the complex plane except for finitely many ones, namely the singularities of the given function. In this sense, D-finite functions can be evaluated at any non-singular point by means of analytic continuation. Numerical evaluation algorithms for D-finite functions have been developed in [26, 62, 63, 64, 47, 48], where the last two references also provide a Maple implementation, namely the **NumGfun** package, for computing such evaluations. These algorithms perform arbitrary-precision evaluations with full error control.

Chapter 8

D-finite Numbers[1]

As mentioned in the introduction, the class of algebraic numbers and the class of algebraic functions are naturally connected to each other. For instance, evaluating an algebraic function over $\mathbb{Q}$ at an algebraic point gives an algebraic number. Also the values of compositional inverses of algebraic functions at algebraic points are algebraic. In particular, roots of an algebraic function over $\mathbb{Q}$ are all algebraic numbers. Moreover, we will see below that every algebraic number can appear as a limit of the coefficient sequence of an algebraic function. However, the class of algebraic numbers is quite small. Almost all real and complex numbers are not algebraic, including many important numbers like π and Euler's number e.

Motivated by the above relation, we aim to establish a similar correspondence between numbers and the class of D-finite functions. To this end, we introduce the following class of numbers.

Definition 8.1. *Let R be a subring of $\mathbb{C}$ and let $\mathbb{F}$ be a subfield of $\mathbb{C}$.*

1. *A number $\xi \in \mathbb{C}$ is called* D-finite *(with respect to R and $\mathbb{F}$) if there exists a convergent sequence $(a_n)_{n=0}^{\infty}$ in $R^{\mathbb{N}}$ with $\lim_{n\to\infty} a_n = \xi$ and some polynomials $p_0, \dots, p_\rho \in \mathbb{F}[n]$, $p_\rho \neq 0$, not all zero, such that*

$$p_0(n)a_n + p_1(n)a_{n+1} + \cdots + p_\rho(n)a_{n+\rho} = 0$$

   for all $n \in \mathbb{N}$.

2. *The set of all D-finite numbers with respect to R and $\mathbb{F}$ is denoted by $\mathcal{D}_{R,\mathbb{F}}$. If $R = \mathbb{F}$, we also write $\mathcal{D}_{\mathbb{F}} := \mathcal{D}_{\mathbb{F},\mathbb{F}}$ for short.*

It turns out that the class of D-finite numbers is closely related to the class of (regular or singular) holonomic constants [35], i.e., the set of all finite values of D-finite functions at (regular or singular) algebraic points.

In this chapter, we show that D-finite numbers are in fact holonomic constants, and conversely, the regular holonomic constants, i.e., the values D-finite

[1] The main results in this chapter are joint work with M. Kauers [39].

functions can assume at non-singular algebraic number arguments, are essentially D-finite numbers over the Gaussian rational field. Together with the work on arbitrary-precision evaluation of D-finite functions [26, 62, 63, 64, 47, 48], it follows that D-finite numbers are computable in the sense that for every D-finite number ξ there exists an algorithm which for any given $n \in \mathbb{N}$ computes a numeric approximation of ξ with a guaranteed precision of 10^{-n}. Consequently, all non-computable numbers have no chance to be D-finite. Besides these artificial examples, we do not know of any explicit real numbers which are not in $\mathcal{D}_{\mathbb{Q}}$, and we believe that it may be very difficult to find some.

We see from Definition 8.1 that the class $\mathcal{D}_{R,\mathbb{F}}$ depends on two subrings of $\mathbb{C}$: the ring R where the sequence lives, and the field $\mathbb{F}$ over which the difference equation is defined. Obviously, different choices of subrings may or may not lead to different classes of D-finite numbers. One goal for this chapter is to investigate what kind of choices of R and $\mathbb{F}$ can be made without changing the resulting class of D-finite numbers.

8.1 Examples of D-finite numbers

Throughout the chapter, R is a subring of $\mathbb{C}$ and $\mathbb{F}$ is a subfield of $\mathbb{C}$, as in Definition 8.1 above. Thanks to many mathematicians' work, we can easily recognize for many constants that they in fact belong to $\mathcal{D}_{\mathbb{Q}}$.

Example 8.2.

1. Archimedes' constant π. Let

$$f_n = \sum_{k=0}^{n} \frac{1}{16^k} \left(\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right).$$

   It is clear that $(f_n)_{n=0}^{\infty}$ is a P-recursive sequence in $\mathbb{Q}$. According to the Bailey-Borwein-Plouffe formula [13], $\lim_{n\to\infty} f_n = \pi$.

2. Euler's number e. By the Taylor series of the exponential function, we have

$$\lim_{n\to\infty} f_n = \mathrm{e} \quad \text{where } f_n = \sum_{k=0}^{n} \frac{1}{k!}.$$

   It is clear that the terms f_n form a P-recursive sequence over $\mathbb{Q}$.

3. Logarithmic value $\log 2$. By the Taylor series of the natural logarithm, we find a P-recursive sequence $(f_n)_{n=0}^{\infty} \in \mathbb{Q}^{\mathbb{N}}$ with

$$f_n = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k},$$

   such that $\lim_{n\to\infty} f_n = \log(2)$.

4. Pythagoras' constant $\sqrt{2}$. One easily finds a P-recursive sequence $(f_n)_{n=0}^{\infty}$ over $\mathbb{Q}$ with
$$f_n = \sum_{k=0}^{n} \binom{\frac{1}{2}}{k},$$
and we have $\lim_{n\to\infty} f_n = \sqrt{2}$ by the binomial theorem.

5. Apéry's constant $\zeta(3)$. By the definition, we see that
$$\lim_{n\to\infty} f_n = \zeta(3) \quad \text{with } f_n = \sum_{k=1}^{n} \frac{1}{k^3}.$$
It is readily seen that $(f_n)_{n=0}^{\infty} \in \mathbb{Q}^{\mathbb{N}}$ is D-finite.

6. The number $1/\pi$. Thanks to Ramanujan, we know that the terms
$$f_n = \sum_{k=0}^{n} \binom{2k}{k}^3 \frac{42k+5}{2^{12k+4}},$$
tend to $1/\pi$ as $n \to \infty$ and form a P-recursive sequence over $\mathbb{Q}$.

7. Euler's constant γ. A desired P-recursive sequence is found by Fischler and Rivoal at their work [31]. They showed that
$$\lim_{n\to\infty} f_n = \gamma \quad \text{with } f_n = \sum_{k=1}^{n} (-1)^k \binom{n}{k} \frac{1}{k} \left(1 - \frac{1}{k!}\right).$$

8. Any value of the Gamma function to a rational number $\Gamma(\alpha)$ with $\alpha < 1$ in $\mathbb{Q}$. Again, Fischler and Rivoal [31] proved that
$$\lim_{n\to\infty} f_n = \Gamma(\alpha) \quad \text{with } f_n = \sum_{k=0}^{n} \binom{n+\alpha}{k+\alpha} \frac{(-1)^k}{k!(k+\alpha)}.$$

8.2 Algebraic numbers

Before turning to general D-finite numbers, let us consider the subclass of algebraic functions. We will show that in this case, the possible limits are precisely the algebraic numbers. For the purpose of this chapter, let us say that a sequence $(a_n)_{n=0}^{\infty} \in \mathbb{F}^{\mathbb{N}}$ is *algebraic* over $\mathbb{F}$ if the corresponding power series $\sum_{n=0}^{\infty} a_n z^n \in \mathbb{F}[[z]]$ is algebraic in the sense of Definition 7.1. Since algebraic functions are D-finite (Abel's theorem), it is clear that algebraic sequences are P-recursive. We will write $\mathcal{A}_{\mathbb{F}}$ for the set of all numbers $\xi \in \mathbb{C}$ which are limits of convergent algebraic sequences over $\mathbb{F}$.

Recall [34] that two sequences $(a_n)_{n=0}^{\infty}$, $(b_n)_{n=0}^{\infty}$ are called *asymptotically equivalent*, written $a_n \sim b_n$ $(n \to \infty)$, if the quotient a_n/b_n converges to 1 as $n \to \infty$. Similarly, two complex functions $f(z)$ and $g(z)$ are called *asymptotically equivalent* at a point $\zeta \in \mathbb{C}$, written $f(z) \sim g(z)$ $(z \to \zeta)$, if the quotient $f(z)/g(z)$ converges to 1 as z approaches ζ. These notions are connected by the following classical theorem.

Theorem 8.3.

1. *(Transfer theorem [33, 34]) For every $\alpha \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$ we have*

$$[z^n]\frac{1}{(1-z)^{\alpha}} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} \quad (n \to \infty),$$

   where $\Gamma(z)$ stands for the Gamma function and the notation $[z^n]f(z)$ refers to the coefficient of z^n in the power series $f(z) \in \mathbb{F}[[z]]$.

2. *(Basic Abelian theorem [32]) Let $(a_n)_{n=0}^{\infty} \in \mathbb{F}^{\mathbb{N}}$ be a sequence that satisfies the asymptotic estimate*

$$a_n \sim n^{\alpha} \quad (n \to \infty),$$

   where $\alpha \geq 0$. Then the generating function $f(z) = \sum_{n=0}^{\infty} a_n z^n$ satisfies the asymptotic estimate

$$f(z) \sim \frac{\Gamma(\alpha+1)}{(1-z)^{\alpha+1}} \quad (z \to 1^-).$$

   This estimate remains valid when z tends to 1 in any sector with vertex at 1 symmetric about the horizontal axis, and with opening angle less than π.

To show that $\mathcal{A}_{\mathbb{F}}$ is in fact a field, we need the following lemma. It indicates that depending on whether $\mathbb{F}$ is a real field or not, every real algebraic number or every algebraic number can appear as a limit.

Lemma 8.4. *Let $p(z) \in \mathbb{F}[z]$ be an irreducible polynomial of degree d. Then there is a square-free polynomial $P(z,y) \in \mathbb{F}[z,y]$ of degree d in y and admitting d distinct analytic algebraic functions $f(z) \in \mathbb{F}[[z]]$ with $P(z, f(z)) = 0$ in a neighborhood of 0 such that 1 is the only dominant singularity of each f and*

1. *if $\mathbb{F} \subseteq \mathbb{R}$, then for each root $\xi \in \bar{\mathbb{F}} \cap \mathbb{R}$ of $p(z)$ there exists a solution $f(z)$ of $P(z,y)$ with $\lim_{n\to\infty}[z^n]f(z) = \xi$;*
2. *if $\mathbb{F} \setminus \mathbb{R} \neq \emptyset$, then for each root $\xi \in \bar{\mathbb{F}}$ of $p(z)$ there exists a solution $f(z)$ of $P(z,y)$ with $\lim_{n\to\infty}[z^n]f(z) = \xi$.*

Proof. The two assertions can be proved simultaneously as follows.

Let $\varepsilon > 0$ be such that any two (real or complex) roots of p have a distance of more than ε to each other. Such an ε exists because p is a polynomial, and polynomials have only finitely many roots. The roots of a polynomial depend continuously on its coefficients. Therefore there exists a real number $\delta > 0$ so that perturbing the coefficients by up to δ won't perturb the roots by more than $\varepsilon/2$. Any positive smaller number than δ will have the same property. By the choice of ε, any such perturbation of the polynomial will have exactly one (real or complex) root in each of the balls of radius $\varepsilon/2$ entered at the roots of p.

Let ξ be a root of p. If $\xi = 0$, then $p(z) = z$. Letting $P(z, y) = y$ yields the assertions. Now assume that $\xi \neq 0$. Let $m \in \mathbb{F}$ be the maximal modulus of coefficients of p. Then $m \neq 0$ since p is irreducible. Therefore, we always can find a number $a_0 \in \mathbb{F}$ such that $|a_0 - \xi| < \delta/m$, with the δ from above. Indeed, we have the following case distinction.

For part 1 where $\mathbb{F} \subseteq \mathbb{R}$, we only consider $\xi \in \bar{\mathbb{F}} \cap \mathbb{R}$. In this case, $\mathbb{F}$ is dense in $\mathbb{R}$ since $\mathbb{F} \supseteq \mathbb{Q}$. Hence such $a_0 \in \mathbb{F} \subseteq \mathbb{R}$ exists.

For part 2 where $\mathbb{F} \setminus \mathbb{R} \neq \emptyset$, there exists a non-real complex number α in $\mathbb{F}$. Therefore, $\mathbb{Q}(\alpha)$ is dense in $\mathbb{C}$. Since $\mathbb{Q}(\alpha) \subseteq \mathbb{F}$, such $a_0 \in \mathbb{F}$ is guaranteed by the density of $\mathbb{F}$ in $\mathbb{C}$.

After finding $a_0 \in \mathbb{F}$ with $|a_0 - \xi| < \delta/m$, for both cases, we have

$$|p(a_0)| = |p(a_0) - p(\xi)| \leq m|a_0 - \xi| < \delta.$$

Replace this δ by $|p(a_0)|$ for such a choice of a_0. The remaining argument works for both cases.

Consider the perturbation $\tilde{p}(y) = p(y) - p(a_0)(1 - z)$. For any $z \in [0, 1]$,

$$|-p(a_0)(1 - z)| < |p(a_0)| = \delta.$$

Therefore, as z moves from 0 to 1, each root of $p(y) - p(a_0)$ moves to the corresponding root of $p(y)$, which belongs to the same ball. In particular, the root a_0 of $\tilde{p}|_{z=0}$ will move to the root ξ of $\tilde{p}|_{z=1}$. Define

$$P(z, y) = p((1 - z)y) - p(a_0)(1 - z) \in \mathbb{F}[z, y].$$

We claim that $P(z, y)$ determines an analytic algebraic function $f(z)$ in $\mathbb{F}[[z]]$ with the dominant singularity 1 and whose coefficient sequence converges to ξ. To prove this, we make an ansatz

$$f(z) = \sum_{n=0}^{\infty} a_n z^n,$$

where the a_0 is from above and $(a_n)_{n=1}^{\infty}$ are to determined. Observe that for any $c(z) \in \mathbb{F}[z]$, $c(z)/(1 - z)$ is a root of $P(z, y)$ if and only if $c(z)$ is a root of $\tilde{p}(y)$, so $f(z)$ admits the following Laurent expansion at $z = 1$,

$$f(z) = \frac{\xi}{1 - z} + \sum_{n=0}^{\infty} b_n (1 - z)^n \quad \text{for } b_n \in \mathbb{C}.$$

Hence $z = 1$ is a singularity of $f(z)$ as $\xi \neq 0$.

The above argument also implies that $z = 1$ is the only dominant singularity of $f(z)$. Indeed, note that $z = 1$ is the only root of the leading coefficient of $P(z, y)$ w.r.t. y, so the other singularities of $f(z)$ could only be branch points, i.e., roots of discriminant of $P(z, y)$ w.r.t. y. However, the choices of ε and δ make it impossible for $f(z)$ to have branch points in the disk $|z| \leq 1$, because in order to have a branch point, two roots of the polynomial $P(z, y)$ w.r.t. y would need to touch each other, and we have ensured that they are always separated by more than ε. Consequently, $z = 1$ is the dominant singularity of $f(z)$, which gives $a_n \sim \xi$ as $n \to \infty$ by part 1 of Theorem 8.3. Therefore $\lim_{n\to\infty} a_n = \xi$ since $\xi \neq 0$.

To complete the proof, it remains to show that the coefficients of $f(z)$ are indeed in $\mathbb{F}$. This is observed by plugging the ansatz of $f(z)$ into $P(z, y)$ and comparing the coefficients of like powers of z to zero. Since $p(z)$ is irreducible and ξ is arbitrary, one sees that $P(z, y)$ admits d distinct analytic solutions in $\mathbb{F}[[z]]$ in a neighborhood of 0. □

The following theorem clarifies the converse direction for algebraic sequences. It turns out that every element in $\mathcal{A}_{\mathbb{F}}$ is algebraic over $\mathbb{F}$.

Theorem 8.5. *Let $\mathbb{F}$ be a subfield of $\mathbb{C}$.*

1. *If $\mathbb{F} \subseteq \mathbb{R}$, then $\mathcal{A}_{\mathbb{F}} = \bar{\mathbb{F}} \cap \mathbb{R}$.*
2. *If $\mathbb{F} \setminus \mathbb{R} \neq \emptyset$, then $\mathcal{A}_{\mathbb{F}} = \bar{\mathbb{F}}$.*

Proof.

1. Let $\xi \in \bar{\mathbb{F}} \cap \mathbb{R}$. Then there is an irreducible polynomial $p(z) \in \mathbb{F}[z]$ such that $p(\xi) = 0$. By part 1 of Lemma 8.4, ξ is in fact a limit of an algebraic sequence in $\mathbb{F}^{\mathbb{N}}$, which implies $\xi \in \mathcal{A}_{\mathbb{F}}$.

   To show the converse inclusion, we let $\xi \in \mathcal{A}_{\mathbb{F}}$. When $\xi = 0$, there is nothing to show. Assume that $\xi \neq 0$. Then there is an algebraic sequence $(a_n)_{n=0}^{\infty} \in \mathbb{F}^{\mathbb{N}}$ such that $\lim_{n\to\infty} a_n = \xi$. Since $\xi \neq 0$, $a_n \sim \xi$ $(n \to \infty)$.

   Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$. Clearly $f(z)$ is an algebraic function over $\mathbb{F}$. By part 2 of Theorem 8.3, $f(z) \sim \xi/(1 - z)$ $(z \to 1^-)$, implying that $z = 1$ is a simple pole of $f(z)$ and

   $$f(z) = \frac{\xi}{1 - z} + \sum_{n=0}^{\infty} b_n (1 - z)^n \quad \text{for } (b_n)_{n=0}^{\infty} \in \mathbb{C}^{\mathbb{N}}.$$

   Setting $g(z) = f(z)(1 - z)$ establishes that $g(z) = \xi + \sum_{n=0}^{\infty} b_n (1 - z)^{n+1}$, and then $g(z)$ is analytic at 1. Sending z to 1 gives $g(1) = \xi$. By closure properties, $g(z)$ is again an algebraic function over $\mathbb{F}$. Thus $\xi = g(1) \in \bar{\mathbb{F}} \cap \mathbb{R}$ as $\mathbb{F} \subseteq \mathbb{R}$.

2. By part 2 of Lemma 8.4 and a similar argument as above, we have $\mathcal{A}_{\mathbb{F}} = \bar{\mathbb{F}}$.

□

If we were to consider the class $\mathcal{C}_{\mathbb{F}}$ of limits of convergent sequences in $\mathbb{F}$ satisfying linear difference equations with constant coefficients over $\mathbb{F}$, sometimes called C-finite sequences, then an argument analogous to the above proof would imply that $\mathcal{C}_{\mathbb{F}} \subseteq \mathbb{F}$, because the power series corresponding to such sequences are rational functions, and the values of rational functions over $\mathbb{F}$ at points in $\mathbb{F}$ evidently gives values in $\mathbb{F}$. The converse direction $\mathbb{F} \subseteq \mathcal{C}_{\mathbb{F}}$ is trivial, so $\mathcal{C}_{\mathbb{F}} = \mathbb{F}$.

Corollary 8.6. *If $\mathbb{F} \subseteq \mathbb{R}$, then $\bar{\mathbb{F}} = \mathcal{A}_{\mathbb{F}(i)} = \mathcal{A}_{\mathbb{F}}[i] = \mathcal{A}_{\mathbb{F}} + i\mathcal{A}_{\mathbb{F}}$, where i is the imaginary unit.*

Proof. Since $\mathcal{A}_{\mathbb{F}}$ is a ring and $i^2 = -1 \in \mathbb{F} \subseteq \mathcal{A}_{\mathbb{F}}$, we have $\mathcal{A}_{\mathbb{F}}[i] = \mathcal{A}_{\mathbb{F}} + i\mathcal{A}_{\mathbb{F}}$. Since $i \in \bar{\mathbb{F}}$ and $\mathbb{F} \subseteq \mathbb{R}$, $\bar{\mathbb{F}}$ is closed under complex conjugation and then

$$\bar{\mathbb{F}} = (\bar{\mathbb{F}} \cap \mathbb{R}) + i(\bar{\mathbb{F}} \cap \mathbb{R}) = \mathcal{A}_{\mathbb{F}} + i\mathcal{A}_{\mathbb{F}},$$

by part 1 of Theorem 8.5. It follows from part 2 of Theorem 8.5 that $\mathcal{A}_{\mathbb{F}(i)} = \overline{\mathbb{F}(i)}$. Since $\mathcal{A}_{\mathbb{F}} \subseteq \mathcal{A}_{\mathbb{F}(i)}$ and $i \in \mathcal{A}_{\mathbb{F}(i)}$, we have

$$\bar{\mathbb{F}} = \mathcal{A}_{\mathbb{F}} + i\mathcal{A}_{\mathbb{F}} \subseteq \mathcal{A}_{\mathbb{F}(i)} = \overline{\mathbb{F}(i)} = \bar{\mathbb{F}}.$$

The assertion holds. □

The following lemma says that every element in $\bar{\mathbb{F}}$ can be represented as the value at 1 of an analytic algebraic function vanishing at zero, provided that $\mathbb{F}$ is dense in $\mathbb{C}$. This will be used in the next section to extend the evaluation domain.

Lemma 8.7. *Let $\mathbb{F}$ be a subfield of $\mathbb{C}$ with $\mathbb{F} \setminus \mathbb{R} \neq \emptyset$. Let $p(z) \in \mathbb{F}[z]$ be an irreducible polynomial of degree d. Assume that $\xi_1, \ldots, \xi_d$ are all the (distinct) roots of p in $\bar{\mathbb{F}}$. Then there is a square-free polynomial $P(z, y) \in \mathbb{F}[z, y]$ of degree d in y and admitting d distinct analytic algebraic functions $g_1(z), \ldots, g_d(z)$ with $P(z, g_j(z)) = 0$ in a neighborhood of 0 such that all g_j's are analytic in the disk $|z| \leq 1$ with $g_j(0) = 0$ and, after reordering (if necessary), $g_j(1) = \xi_j$.*

Proof. By part 2 of Lemma 8.4, there exists a bivariate square-free polynomial $\tilde{P}(z, y) \in \mathbb{F}[z, y]$ of degree d in y and admitting d distinct analytic algebraic functions $f_1(z), \ldots, f_d(z)$ with $P(z, f_j(z)) = 0$ in a neighborhood of 0 such that 1 is the only dominant singularity of each $f_j(z)$ and, after reordering (if necessary),

$$\lim_{n\to\infty} [z^n] f_j(z) = \xi_j, \quad j = 1, \ldots d.$$

If $\xi_j = 0$ for some j then $p(z) = z$. Letting $P(z, y) = y$ yields the assertion. Otherwise all roots $\xi_1, \ldots, \xi_d$ are nonzero, and thus $[z^n] f_j(z) \sim \xi_j$ $(n \to \infty)$ for each j. By part 2 of Theorem 8.3,

$$f_j(z) \sim \frac{\xi_j}{1 - z} \ (z \to 1^-),$$

which implies that $z = 1$ is a simple pole of each f_j. Let $g_j(z) = f_j(z)z(1-z)$. Then $g_1(z), \ldots, g_d(z)$ are distinct and each $g_j(z) \in \mathbb{F}[[z]]$ is analytic for any z in the disk $|z| \leq 1$. Moreover, $g_j(0) = 0$ and $g_j(1) = \xi_j$. By closure properties, $g_j(z)$ is again algebraic over $\mathbb{F}$. Define a square-free polynomial

$$P(z, y) = \prod_{j=1}^{d} (y - g_j(z)) = \prod_{j=1}^{d} \big(y - f_j(z)z(1-z)\big) \in \overline{\mathbb{F}(z)}[y].$$

Then $P \in \mathbb{F}[z, y]$ since P is symmetric in $f_1, \ldots, f_d$. The lemma follows. $\square$

8.3 Rings of D-finite numbers

Let us now return to the study of D-finite numbers. Let R be a subring of $\mathbb{C}$ and $\mathbb{F}$ be a subfield of $\mathbb{C}$. Recall that by Definition 8.1, the elements of $\mathcal{D}_{R,\mathbb{F}}$ are exactly limits of convergent sequences in $R^{\mathbb{N}}$ which are P-recursive over $\mathbb{F}$. Some facts about P-recursive sequences translate directly into facts about $\mathcal{D}_{R,\mathbb{F}}$.

Proposition 8.8.

1. *$R \subseteq \mathcal{D}_{R,\mathbb{F}}$ and $\mathcal{A}_{\mathbb{F}} \subseteq \mathcal{D}_{\mathbb{F}}$.*
2. *If $R_1 \subseteq R_2$ then $\mathcal{D}_{R_1,\mathbb{F}} \subseteq \mathcal{D}_{R_2,\mathbb{F}}$, and if $\mathbb{F} \subseteq \mathbb{E}$ then $\mathcal{D}_{R,\mathbb{F}} \subseteq \mathcal{D}_{R,\mathbb{E}}$.*
3. *$\mathcal{D}_{R,\mathbb{F}}$ is a subring of $\mathbb{C}$. Moreover, if R is an $\mathbb{F}$-algebra then so is $\mathcal{D}_{R,\mathbb{F}}$.*
4. *If $\mathbb{E}$ is an algebraic extension field of $\mathbb{F}$, then $\mathcal{D}_{R,\mathbb{F}} = \mathcal{D}_{R,\mathbb{E}}$.*
5. *If $R \subseteq \mathbb{F}$, then $\mathcal{D}_{R,\mathbb{F}} = \mathcal{D}_{R,\mathrm{Quot}(R)}$.*
6. *If R and $\mathbb{F}$ are closed under complex conjugation, then so is $\mathcal{D}_{R,\mathbb{F}}$.*
   In this case, we have $\mathcal{D}_{R,\mathbb{F}} \cap \mathbb{R} = \mathcal{D}_{R\cap\mathbb{R},\mathbb{F}}$.
   Moreover, if the imaginary unit $i \in \mathcal{D}_{R,\mathbb{F}}$ then $\mathcal{D}_{R,\mathbb{F}} = \mathcal{D}_{R\cap\mathbb{R},\mathbb{F}} + i\mathcal{D}_{R\cap\mathbb{R},\mathbb{F}}$.

Proof.

1. The first inclusion is clear because every element of R is the limit of a constant sequence, and every constant sequence is P-recursive. The second inclusion follows from the fact that algebraic functions are D-finite, and the coefficient sequences of D-finite functions are P-recursive.
2. Clear.
3. Follows directly from the corresponding closure properties for P-recursive sequences.
4. Follows directly from part 1 of Lemma 7.3.
5. Follows directly from part 2 of Lemma 7.3.

6. For any convergent sequence $(a_n)_{n=0}^{\infty} \in R^{\mathbb{N}}$, we have

$$\operatorname{Re}\left(\lim_{n\to\infty} a_n\right) = \lim_{n\to\infty} \operatorname{Re}(a_n), \quad \operatorname{Im}\left(\lim_{n\to\infty} a_n\right) = \lim_{n\to\infty} \operatorname{Im}(a_n),$$

and thus $\overline{\lim_{n\to\infty} a_n} = \lim_{n\to\infty} \bar{a}_n$. Hence the first assertion follows by $(\bar{a}_n)_{n=0}^{\infty} \in R^{\mathbb{N}}$ and part 3 of Lemma 7.3.

Since R is closed under complex conjugation, $(\operatorname{Re}(a_n))_{n=0}^{\infty} \in (R \cap \mathbb{R})^{\mathbb{N}}$. Then the inclusion $\mathcal{D}_{R,\mathbb{F}} \cap \mathbb{R} \subseteq \mathcal{D}_{R\cap\mathbb{R},\mathbb{F}}$ can be shown similarly as the first assertion. The converse direction holds by part 2. Thus $\mathcal{D}_{R,\mathbb{F}} \cap \mathbb{R} = \mathcal{D}_{R\cap\mathbb{R},\mathbb{F}}$.

If $i \in \mathcal{D}_{R,\mathbb{F}}$, then $\mathcal{D}_{R\cap\mathbb{R},\mathbb{F}} + i\mathcal{D}_{R\cap\mathbb{R},\mathbb{F}} \subseteq \mathcal{D}_{R,\mathbb{F}}$ since $\mathcal{D}_{R\cap\mathbb{R},\mathbb{F}} \subseteq \mathcal{D}_{R,\mathbb{F}}$. To show the converse inclusion, let $\xi \in \mathcal{D}_{R,\mathbb{F}}$. Then $\bar{\xi} \in \mathcal{D}_{R,\mathbb{F}}$ by the first assertion. Since $i \in \mathcal{D}_{R,\mathbb{F}}$ and R is closed under complex conjugation, $\operatorname{Re}(\xi), \operatorname{Im}(\xi)$ both belong to $\mathcal{D}_{R,\mathbb{F}} \cap \mathbb{R} = \mathcal{D}_{R\cap\mathbb{R},\mathbb{F}}$ by the second assertion. Therefore we have $\xi = \operatorname{Re}(\xi) + i\operatorname{Im}(\xi) \in \mathcal{D}_{R\cap\mathbb{R},\mathbb{F}} + i\mathcal{D}_{R\cap\mathbb{R},\mathbb{F}}$.

□

Example 8.9.

1. We have $\mathcal{D}_{\mathbb{Q}(\sqrt{2}),\mathbb{Q}(\pi,\sqrt{2})} = \mathcal{D}_{\mathbb{Q}(\sqrt{2}),\mathbb{Q}(\sqrt{2})} = \mathcal{D}_{\mathbb{Q}(\sqrt{2}),\mathbb{Q}}$. The first identity holds by part 5, the second by part 4 of the proposition.
2. We have $\mathcal{D}_{\bar{\mathbb{Q}},\mathbb{Q}} = \mathcal{D}_{\bar{\mathbb{Q}},\mathbb{R}}$. The inclusion "$\subseteq$" is clear by part 2. For the inclusion "$\supseteq$", let $\xi \in \mathcal{D}_{\bar{\mathbb{Q}},\mathbb{R}}$. Then $\xi = a + ib$ for some $a, b \in \mathbb{R}$, and there exists a sequence $(a_n + ib_n)_{n=0}^{\infty}$ in $\bar{\mathbb{Q}}^{\mathbb{N}}$ and a nonzero operator $L \in \mathbb{R}[n]\langle S_n\rangle$ such that $L \cdot (a_n + ib_n) = 0$ and $\lim_{n\to\infty}(a_n + ib_n) = a + ib$. Since the coefficients of L are real, we then have $L \cdot a_n = 0$ and $L \cdot b_n = 0$. Furthermore, we see that $\lim_{n\to\infty} a_n = a$ and $\lim_{n\to\infty} b_n = b$. Therefore,

$$a, b \in \mathcal{D}_{\bar{\mathbb{Q}}\cap\mathbb{R},\mathbb{R}} \overset{\text{part } 5}{=} \mathcal{D}_{\bar{\mathbb{Q}}\cap\mathbb{R},\bar{\mathbb{Q}}\cap\mathbb{R}} \overset{\text{part } 4}{=} \mathcal{D}_{\bar{\mathbb{Q}}\cap\mathbb{R},\mathbb{Q}},$$

which implies $a + ib \in \mathcal{D}_{\bar{\mathbb{Q}}\cap\mathbb{R},\mathbb{Q}} + i\mathcal{D}_{\bar{\mathbb{Q}}\cap\mathbb{R},\mathbb{Q}} \overset{\text{part } 6}{=} \mathcal{D}_{\bar{\mathbb{Q}},\mathbb{Q}}$, as claimed.

Lemma 8.7 motivates the following theorem, which says that every D-finite number is essentially the value at 1 of an analytic D-finite function, and thus a holonomic constant.

Theorem 8.10. *Let R be a subring of $\mathbb{C}$ and $\mathbb{F}$ be a subfield of $\mathbb{C}$. Then for any $\xi \in \mathcal{D}_{R,\mathbb{F}}$, there exists $g(z) \in R[[z]]$ D-finite over $\mathbb{F}$ and analytic at 1 such that $\xi = g(1)$.*

Proof. The statement is clear when $\xi = 0$. Assume that ξ is nonzero. Then there exists a sequence $(a_n)_{n=0}^{\infty} \in R^{\mathbb{N}}$, P-recursive over $\mathbb{F}$, such that $\lim_{n\to\infty} a_n = \xi$. Since ξ is nonzero, we have $a_n \sim \xi$ $(n \to \infty)$. Let $f(z) = \sum_{n=0}^{\infty} a_n z^n$. Then by Theorem 8.3, we see that

$$f(z) \sim \frac{\xi}{1-z} \ (z \to 1^-),$$

which implies that $z = 1$ is a simple pole of $f(z)$. Let $g(z) = f(z)(1 - z)$. Then $g(z)$ belongs to $R[[z]]$ and is analytic at $z = 1$. Write

$$f(z) = \frac{\xi}{1-z} + \sum_{n=0}^{\infty} b_n (1-z)^n \quad \text{with } b_n \in \mathbb{C}.$$

It follows that $g(z) = f(z)(1-z) = \xi + \sum_{n=0}^{\infty} b_n (1-z)^{n+1}$, which gives $\xi = g(1)$. The assertion follows by noticing that $g(z)$ is D-finite over $\mathbb{F}$ due to closure properties. $\square$

Example 8.11. We have $\zeta(3) = \sum_{n=1}^{\infty} \frac{1}{n^3} = \mathrm{Li}_3(1)$, where $\mathrm{Li}_3(z) = \sum_{n=1}^{\infty} \frac{1}{n^3} z^n$ is the polylogarithm function, D-finite over $\mathbb{Q}$ and analytic at 1.

Note that the above theorem implies that D-finite numbers are computable when the ring R and the field $\mathbb{F}$ consist of computable numbers. This allows the construction of (artificial) numbers that are not D-finite.

Some kind of converse of Theorem 8.10 can be proved for the case when $\mathbb{F}$ is not a subfield of $\mathbb{R}$, namely $\mathbb{F} \setminus \mathbb{R} \neq \emptyset$. Note that this condition is equivalent to saying that $\mathbb{F}$ is dense in $\mathbb{C}$. To this end, we first need to develop several lemmas.

The following lemma says that the value of a D-finite function at any non-singular point in $\bar{\mathbb{F}}$ can be represented by the value at 1 of another D-finite function.

Lemma 8.12. *Let $\mathbb{F}$ be a subfield of $\mathbb{C}$ with $\mathbb{F} \setminus \mathbb{R} \neq \emptyset$ and R be a subring of $\mathbb{C}$ containing $\mathbb{F}$. Assume that $f(z) \in \mathcal{D}_{R,\mathbb{F}}[[z]]$ is analytic and annihilated by a nonzero operator $L \in \mathbb{F}[z]\langle D_z \rangle$ with zero an ordinary point. Then for any non-singular point $\zeta \in \bar{\mathbb{F}}$ of L, there exists an analytic function $h(z) \in \mathcal{D}_{R,\mathbb{F}}[[z]]$ and a nonzero operator $M \in \mathbb{F}[z]\langle D_z \rangle$ with 0 and 1 ordinary points such that $M \cdot h(z) = 0$ and $f(\zeta) = h(1)$.*

Proof. Let $\zeta \in \bar{\mathbb{F}}$ be a non-singular point of L. Then there exists an irreducible polynomial $p(z) \in \mathbb{F}[z]$ such that $p(\zeta) = 0$. Let $\zeta_1 = \zeta, \ldots, \zeta_d$ be all the roots of p in $\bar{\mathbb{F}}$. By Lemma 8.7, there exists a square-free polynomial $P(z, y) \in \mathbb{F}[z, y]$ of degree d in y and admitting d distinct analytic algebraic functions $g_1(z), \ldots, g_d(z)$ with $P(z, g_j(z)) = 0$ in a neighborhood of 0. Moreover, $g_1(z), \ldots, g_d(z)$ are all analytic in the disk $|z| \leq 1$ with $g_j(1) = \zeta_j$ and $g_j(0) = 0$.

Since $g_1(1) = \zeta$ is not a singularity of L by assumption, none of $g_j(1) = \zeta_j$ is a singularity of L. Suppose otherwise that for some $2 \leq \ell \leq d$, the point $g_\ell(1) = \zeta_\ell$ is a root of $\mathrm{lc}(L)$. Since $\mathrm{lc}(L) \in \mathbb{F}[z]$ and p is the minimal polynomial of ζ_ℓ over $\mathbb{F}$, we know that p divides $\mathrm{lc}(L)$ over $\mathbb{F}$. Thus ζ is also a root of $\mathrm{lc}(L)$, a contradiction.

Note that $g_1(z), \ldots, g_d(z)$ are analytic in the disk $|z| \leq 1$ and $g_j(0) = 0$. By Theorem 7.2, there exists a nonzero operator $M \in \mathbb{F}[z]\langle D_z \rangle$ with $M \cdot (f \circ g_1) = 0$ which does not have 0 or 1 among its singularities. By part 1 of Proposition 8.8, $\mathbb{F} \subseteq R \subseteq \mathcal{D}_{R,\mathbb{F}}$. Since $f(z) \in \mathcal{D}_{R,\mathbb{F}}[[z]]$ and $g_1(z) \in \mathbb{F}[[z]]$ with $g_1(0) = 0$, we have $f(g_1(z)) \in \mathcal{D}_{R,\mathbb{F}}[[z]]$. Setting $h(z) = f(g_1(z))$ completes the proof. $\square$

With the above lemma, it suffices to consider the case when the evaluation point is in $R \cap \mathbb{F}$. This is exactly what the next two lemmas are concerned about.

Lemma 8.13. *Assume that* $f(z) = \sum_{n=0}^{\infty} a_n z^n \in R[[z]]$ *is D-finite over* $\mathbb{F}$ *and convergent in some neighborhood of* 0. *Let* $\zeta \in R \cap \mathbb{F}$ *be in the disk of convergence. Then* $f^{(k)}(\zeta) \in \mathcal{D}_{R,\mathbb{F}}$ *for all* $k \in \mathbb{N}$.

Proof. For $k \in \mathbb{N}$, it is well-known that $f^{(k)}(z) \in R[[z]]$ is also D-finite and has the same radius of convergence at zero as $f(z)$. Note that since $f(z)$ is D-finite over $\mathbb{F}$, so is $f^{(k)}(z)$. Thus to prove the lemma, it suffices to show the case when $k = 0$, i.e., $f(\zeta) \in \mathcal{D}_{R,\mathbb{F}}$.

Since $f(z)$ is D-finite over $\mathbb{F}$, the coefficient sequence $(a_n)_{n=0}^{\infty}$ is P-recursive over $\mathbb{F}$. Note that $\zeta \in R \cap \mathbb{F}$ is in the disk of convergence of $f(z)$ at zero, so

$$f(\zeta) = \sum_{n=0}^{\infty} a_n \zeta^n = \lim_{n \to \infty} \sum_{\ell=0}^{n} a_\ell \zeta^\ell.$$

Since $(\zeta^n)_{n=0}^{\infty}$ is P-recursive over $\mathbb{F}$, the assertion follows by noticing that the sequence $(\sum_{\ell=0}^{n} a_\ell \zeta^\ell)_{n=0}^{\infty} \in R^{\mathbb{N}}$ is P-recursive over $\mathbb{F}$ due to closure properties. □

Example 8.14. Since $\exp(z) = \sum_{n=0}^{\infty} \frac{1}{n!} z^n \in \mathbb{Q}[[z]]$ is D-finite over $\mathbb{Q}$, and converges everywhere, we get from the lemma that the numbers $\mathrm{e}, 1/\mathrm{e}, \sqrt{\mathrm{e}}$ belong to $\mathcal{D}_{\mathbb{Q},\mathbb{Q}}$. More precisely, since we are currently only considering non-real fields $\mathbb{F}$, we could say that the function $\exp(z)$ is D-finite over $\bar{\mathbb{Q}}$, therefore $\mathrm{e}, 1/\mathrm{e}, \sqrt{\mathrm{e}}$ all belong to $\mathcal{D}_{\mathbb{Q},\bar{\mathbb{Q}}}$, but by Proposition 8.8, $\mathcal{D}_{\mathbb{Q},\bar{\mathbb{Q}}} = \mathcal{D}_{\mathbb{Q},\mathbb{Q}}$.

Lemma 8.15. *Let* R *be a subring of* $\mathbb{C}$ *containing* $\mathbb{F}$. *Let* $f(z) = \sum_{n=0}^{\infty} a_n z^n$ *in* $\mathcal{D}_{R,\mathbb{F}}[[z]]$ *be an analytic function. Assume that there exists a nonzero operator* $L \in \mathbb{F}[z]\langle D_z \rangle$ *with zero an ordinary point such that* $L \cdot f(z) = 0$. *Let* $r > 0$ *be the smallest modulus of roots of* $\operatorname{lc}(L)$ *and let* $\zeta \in \mathbb{F}$ *with* $|\zeta| < r$. *Then* $f^{(k)}(\zeta) \in \mathcal{D}_{R,\mathbb{F}}$ *for all* $k \in \mathbb{N}$.

Proof. Let ρ be the order of L. Since zero is an ordinary point of L, there exist P-recursive sequences $(c_n^{(0)})_{n=0}^{\infty}, \dots, (c_n^{(\rho-1)})_{n=0}^{\infty}$ in $\mathbb{F}^{\mathbb{N}} \subseteq R^{\mathbb{N}}$ with $c_j^{(m)}$ equal to the Kronecker delta δ_{mj} for $m, j = 0, \dots, \rho - 1$, so that the set $\{\sum_{n=0}^{\infty} c_n^{(m)} z^n\}_{m=0}^{\rho-1}$ forms a basis of the solution space of L near zero. Note that the singularities of solutions of L can only be roots of $\operatorname{lc}(L)$. Hence the power series $f(z) = \sum_{n=0}^{\infty} a_n z^n$ as well as $\sum_{n=0}^{\infty} c_n^{(m)} z^n$ for $m = 0, \dots, \rho - 1$ are convergent in the disk $|z| < r$. It follows from $|\zeta| < r$ and Lemma 8.13 that the set $\{\sum_{n=0}^{\infty} c_n^{(m)} \zeta^n\}_{m=0}^{\rho-1}$ belongs to $\mathcal{D}_{R,\mathbb{F}}$. Since $a_0, \dots, a_{\rho-1} \in \mathcal{D}_{R,\mathbb{F}}$,

$$f(\zeta) = \sum_{n=0}^{\infty} a_n \zeta^n = a_0 \sum_{n=0}^{\infty} c_n^{(0)} \zeta^n + \cdots + a_{\rho-1} \sum_{n=0}^{\infty} c_n^{(\rho-1)} \zeta^n$$

is D-finite by closure properties. In the same vein, we find that for $k > 0$, the derivative $f^{(k)}(\zeta)$ also belongs to $\mathcal{D}_{R,\mathbb{F}}$. □

Example 8.16.

1. We know from Proposition 8.8 that $\sqrt{2} \in \mathcal{D}_{\mathbb{Q}}$. The series
$$(z+1)^{\sqrt{2}} = 1 + \sqrt{2}z + (1 - \frac{1}{\sqrt{2}})z^2 + \cdots \in \mathbb{Q}(\sqrt{2})[[z]] \subseteq \mathcal{D}_{\mathbb{Q}}[[z]]$$
is D-finite over $\mathbb{Q}$, an annihilating operator is $(z+1)^2 D_z^2 + (z+1)D_z - 2$. Here we have the radius $r = 1$. Taking $\zeta = \sqrt{2} - 1$, the lemma implies that $\sqrt{2}^{\sqrt{2}}$ belongs to $\mathcal{D}_{\mathbb{Q}}$.

2. Observe that the lemma refers to the singularities of the operator rather than to the singularities of the particular solution at hand. For example, it does not imply that $J_1(1) \in \mathcal{D}_{\mathbb{Q},\mathbb{Q}}$, where $J_1(z)$ is the first Bessel function, because its annihilating operator is $z^2 D_z^2 + zD_z + (z^2 - 1)$, which has a singularity at 0. It is not sufficient that the particular solution $J_1(z) \in \mathbb{Q}[[z]]$ is analytic at 0. Of course, in this particular example we see from the series representation $J_1(1) = \frac{1}{2}\sum_{n=0}^{\infty} \frac{(-1/4)^n}{(n+1)n!^2}$ that the value belongs to $\mathcal{D}_{\mathbb{Q},\mathbb{Q}}$.

3. The hypergeometric function $f(z) := {}_2F_1(\frac{1}{3}, \frac{1}{2}, 1, z + \frac{1}{2})$ can be viewed as an element of $\mathcal{D}_{\mathbb{Q},\mathbb{Q}}[[z]]$:
$$f(z) = \underbrace{\sqrt[3]{2}\sum_{n=0}^{\infty} \frac{(1/3)_n (1/2)_n}{n!^2}(-1)^n}_{\in \mathcal{D}_{\mathbb{Q}}} + \underbrace{\frac{\sqrt[3]{2}}{3}\sum_{n=0}^{\infty} \frac{(1/2)_n (4/3)_n}{(2)_n n!}(-1)^n}_{\in \mathcal{D}_{\mathbb{Q}}} z$$
$$+ \underbrace{\frac{2\sqrt[3]{2}}{3}\sum_{n=0}^{\infty} \frac{(1/2)_n (7/3)_n}{(3)_n n!}(-1)^n}_{\in \mathcal{D}_{\mathbb{Q}}} z^2 + \cdots .$$
The function f is annihilated by the operator
$$L = 3(2z-1)(2z+1)D_z^2 + (22z-1)D_z + 2.$$
This operator has a singularity at $z = 1/2$, and there is no annihilating operator of f which does not have a singularity there. Although
$$f(1/2) = \frac{\Gamma(1/6)}{\Gamma(1/2)\Gamma(2/3)}$$
is a finite and specific value, the lemma does not imply that this value is a D-finite number.

Theorem 8.17. *Let $\mathbb{F}$ be a subfield of $\mathbb{C}$ with $\mathbb{F} \setminus \mathbb{R} \neq \emptyset$ and let R be a subring of $\mathbb{C}$ containing $\mathbb{F}$. Assume that $f(z) \in \mathcal{D}_{R,\mathbb{F}}[[z]]$ is analytic and there exists a nonzero operator $L \in \mathbb{F}[z]\langle D_z\rangle$ with zero an ordinary point such that $L \cdot f(z) = 0$. Further assume that $\zeta \in \bar{\mathbb{F}}$ is not a singularity of L. Then $f^{(k)}(\zeta)$ belongs to $\mathcal{D}_{R,\mathbb{F}}$ for all $k \in \mathbb{N}$.*

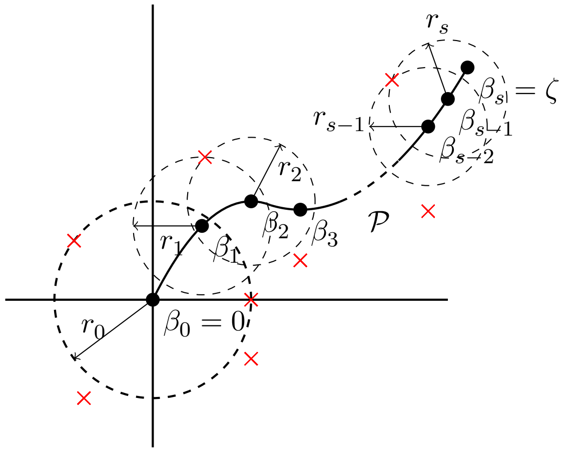

Figure 8.1: a simple path $\mathcal{P}$ with a finite cover $\bigcup_{j=0}^{s} \mathcal{B}_{r_j}(\beta_j)$ ($\times$ stands for the roots of $\operatorname{lc}(L)$)

Proof. By Lemma 8.12, it suffices to show the assertion holds for $\zeta = 1$ (or more generally $\zeta \in \mathbb{F}$). Now assume that $\zeta \in \mathbb{F}$. We apply the method of analytic continuation.

Let $\mathcal{P}$ be a simple path with a finite cover $\bigcup_{j=0}^{s} \mathcal{B}_{r_j}(\beta_j)$, where $s \in \mathbb{N}$, $\beta_0 = 0$, $\beta_s = \zeta$, $\beta_j \in \mathbb{F}$, $r_j > 0$ is the distance between β_j and the zero set of $\operatorname{lc}(L)$, and $\mathcal{B}_{r_j}(\beta_j)$ is the open circle centered at β_j and with radius r_j. Moreover, β_{j+1} is inside $\mathcal{B}_{r_j}(\beta_j)$ for each j (as illustrated by Figure 8.1). Such a path exists because $\mathbb{F}$ is dense in $\mathbb{C}$ and the zero set of $\operatorname{lc}(L)$ is finite. Since the path $\mathcal{P}$ avoids all roots of $\operatorname{lc}(L)$, the function $f(z)$ is analytic along $\mathcal{P}$. We next use induction on the index j to show that $f^{(k)}(\beta_j) \in \mathcal{D}_{R,\mathbb{F}}$ for all $k \in \mathbb{N}$.

It is trivial when $j = 0$ as $f^{(k)}(\beta_0) = f^{(k)}(0) \in \mathcal{D}_{R,\mathbb{F}}$ for $k \in \mathbb{N}$ by assumption. Assume now that $0 < j \leq s$ and $f^{(k)}(\beta_{j-1}) \in \mathcal{D}_{R,\mathbb{F}}$ for all $k \in \mathbb{N}$. We consider $f(\beta_j)$ and its derivatives.

Recall that $r_{j-1} > 0$ is the distance between β_{j-1} and the zero set of $\operatorname{lc}(L)$. Since $f(z)$ is analytic at β_{j-1}, it is representable by a convergent power series expansion

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(\beta_{j-1})}{n!}(z - \beta_{j-1})^n \quad \text{for all } |z - \beta_{j-1}| < r_{j-1}.$$

By the induction hypothesis, $f^{(n)}(\beta_{j-1})/n! \in \mathcal{D}_{R,\mathbb{F}}$ for all $n \in \mathbb{N}$ and thus $f(z)$ belongs to $\mathcal{D}_{R,\mathbb{F}}[[z - \beta_{j-1}]]$.

Let $Z = z - \beta_{j-1}$, i.e., $z = Z + \beta_{j-1}$. Define $g(Z) = f(Z + \beta_{j-1})$ and $\tilde{L}$ to be the operator obtained by replacing z in L by $Z + \beta_j$. Since $\beta_{j-1} \in \mathbb{F} \subseteq \mathcal{D}_{R,\mathbb{F}}$ and $D_z = D_Z$, we have $g(Z) \in \mathcal{D}_{R,\mathbb{F}}[[Z]]$ and $\tilde{L} \in \mathbb{F}[Z]\langle D_Z \rangle$. Note that $L \cdot f(z) = 0$

and β_{j-1} is an ordinary point of L as $r_{j-1} > 0$. It follows that $\tilde{L} \cdot g(Z) = 0$ and zero is an ordinary point of $\tilde{L}$. Moreover, we see that r_{j-1} is now the smallest modulus of roots of $\mathrm{lc}(\tilde{L})$. Since $|\beta_j - \beta_{j-1}| < r_{j-1}$, applying Lemma 8.15 to $g(Z)$ yields

$$f^{(k)}(\beta_j) = g^{(k)}(\beta_j - \beta_{j-1}) \in \mathcal{D}_{R,\mathbb{F}} \quad \text{for } k \in \mathbb{N}.$$

Thus the assertion holds for $j = s$. The theorem follows. □

Example 8.18. By the above theorem, $\exp(\sqrt{2})$ and $\log(1+\sqrt{3})$ both belong to $\mathcal{D}_{\mathbb{Q}}$. We also have $\mathrm{e}^{\pi} \in \mathcal{D}_{\mathbb{Q}}$. This is because $\mathrm{e}^{\pi} = (-1)^{-i}$ with i the imaginary unit, is equal to the value of the D-finite function $(z+1)^{-i} \in \mathbb{Q}(i)[[z]]$ at $z = -2$ (outside the radius of convergence; analytically continued in consistency with the usual branch cut conventions) and then $\mathrm{e}^{\pi} \in \mathcal{D}_{\mathbb{Q}(i)} \cap \mathbb{R} = \mathcal{D}_{\mathbb{Q}}$. Furthermore, as remarked in the introduction, the numbers obtained by evaluating a G-function at algebraic numbers which avoid the singularities of its annihilating operator are in $\mathcal{D}_{\mathbb{Q}(i)}$, because G-functions are D-finite.

8.4 Open questions

We have introduced the notion of D-finite numbers and made some first steps towards understanding their nature. We believe that, similarly as for D-finite functions, the class is interesting because it has good mathematical and computational properties and because it contains many special numbers that are of independent interest. We conclude this chapter with some possible directions of future research.

Evaluation at singularities. While every singularity of a D-finite function must also be a singularity of its annihilating operator, the converse is in general not true. We have seen above that evaluating a D-finite function at a point which is not a singularity of its annihilating operator yields a D-finite number. It would be natural to wonder about the values of a D-finite function at singularities of its annihilating operator, including those at which the given function is not analytic but its evaluation is finite. Also, we always consider zero as an ordinary point of the annihilating operator. If this is not the case, the method used in this chapter fails, as pointed out by part 2 of Example 8.16.

Quotients of D-finite numbers. The set of algebraic numbers forms a field, but we do not have a similar result for D-finite numbers. It is known that the set of D-finite functions does not form a field. Instead, Harris and Sibuya [37] showed that a D-finite function f admits a D-finite multiplicative inverse if and only if f'/f is algebraic. This explains for example why both e and $1/\mathrm{e}$ are D-finite, but it does not explain why both π and $1/\pi$ are D-finite. It would be interesting to know more precisely under which circumstances the multiplicative inverse of a D-finite number is D-finite. Is $1/\log(2)$ a D-finite number? Are there choices of R and $\mathbb{F}$ for which $\mathcal{D}_{R,\mathbb{F}}$ is a field?

Roots of D-finite functions. A similar pending analogy concerns compositional inverses. We know that if f is an algebraic function, then so is its compositional inverse f^{-1}. The analogous statement for D-finite functions is not true. Nevertheless, it could still be true that the values of compositional inverses of D-finite functions are D-finite numbers, although this seems somewhat unlikely. A particularly interesting special case is the question whether (or under which circumstances) the roots of a D-finite function are D-finite numbers.

Evaluation at D-finite number arguments. We see that the class $\mathcal{C}_{\mathbb{F}}$ of limits of convergent C-finite sequences is the same as the values of rational functions at points in $\mathbb{F}$, namely the field $\mathbb{F}$. Similarly, the class $\mathcal{A}_{\mathbb{F}}$ of limits of convergent algebraic sequences essentially consists of the values of algebraic functions at points in $\bar{\mathbb{F}}$. Continuing this pattern, is the value of a D-finite function at a D-finite number again a D-finite number? If so, this would imply that also numbers like $\mathrm{e}^{\mathrm{e}^{\mathrm{e}^{\mathrm{e}}}}$ are D-finite. Since $1/(1-z)$ is a D-finite function, it would also imply that D-finite numbers form a field.

Appendices

Appendix A

The **ShiftReductionCT** Package

In order to be able to experiment with the algorithms proposed in the first part of this thesis, we have implemented all of them and encapsulated the procedures as a Maple package, namely the **ShiftReductionCT** package. This package was developed for MAPLE 18 and it is available upon request from the author. Here is a description of the package.

```
> eval(ShiftReductionCT);
```

module( )
  option package;
  export *ReductionCT*, *BoundReductionCT*,
  ModifiedAbramovPetkovsekReduction, *ShiftMAPReduction*, *IsSummable*,
  ShellReduction, *KernelReductionCT*, *PolynomialReduction*,
  TranslateDRF, *VerifyMAPReduction*, *VerifyRCT*;
  description
  "Creative Telescoping for Bivariate Hypergeometric Terms via
  the Modified Abramov-Petkovsek Reduction";
end module

This appendix is intended to give a detailed instruction for the package. All export commands will be discussed in the order of their first appearance in the thesis, but only some of them will be emphasized particularly. By applying them to some concrete examples, we show the usage of the package as well as its applications. These examples are chosen to take virtually no computation time.

The appendix contains a whole Maple session. The inputs are given exactly in the way how the commands need to be used in Maple and displayed in the type of Maple notation, while the outputs are displayed in 2-D math notation. To start with, we load the package in Maple.

```
> read(ShiftReductionCT):
> with(ShiftReductionCT):
```

Commands related to Chapter 3

We first consider univariate hypergeometric terms. Let T be the hypergeometric term in Example 3.7 (or Example 3.19).

```
> T:=k^2*k!/(k+1);
```

$$T := \frac{k^2 k!}{k+1}$$

By commands from the built-in Maple package **SumTools[Hypergeometric]**, we find a kernel $K = k + 1$ and its corresponding shell $S = k^2/(k + 1)$ of T.

The command ShellReduction performs Algorithm 3.5 and returns a decomposition of the form (3.3) for the shell S with respect to its kernel K.

```
> res:=ShellReduction(numer(K),denom(K),numer(S),denom(S),k);
```

$$res := \left[\left[-\frac{1}{k+1}\right], -1, k+2, k\right]$$

Using the notations in (3.3), we check the correctness by

```
> S1:=add(res[1][i],i=1..nops(res[1])):
> a:=res[2]: b:=res[3]: p:=res[4]:
> normal(K*subs(y=y+1,S1)-S1+(a/b+p/denom(K))-S);
```

$$0$$

The command PolynomialReduction, namely Algorithm 3.16, projects a polynomial onto the image space of the map for polynomial reduction with respect to a shift-reduced rational function, and the standard complement of the image space.

```
> res:=PolynomialReduction(p,numer(K),denom(K),k);
```

$$res := [1], 0$$

Using the notations in Algorithm 3.16, we check the correctness by

```
> f:=add(res[1][i],i=1..nops(res[1])):
> q:=res[2]: normal(numer(K)*subs(k=k+1,f)-denom(K)*f+q-p);
```

$$0$$

The built-in Maple command SumDecomposition, which is in the package **SumTools[Hypergeometric]**, is implemented based on the Abramov-Petkovšek reduction. It computes a minimal additive decomposition described in Section 3.1 for a given hypergeometric term.

```
> SumTools[Hypergeometric][SumDecomposition](T,k);
```

$$\left[\frac{k \prod_{\_k=1}^{k-1} (\_k+1)}{k+1}, -\frac{\prod_{\_k=1}^{k-1} (\_k+1)}{k+2}\right]$$

To avoid solving any auxiliary recurrence equations explicitly, we present a modified version of the Abramov-Petkovšek reduction, namely Algorithm 3.17, and implement it as the command ModifiedAbramovPetkovsekReduction. This command can be used in the following (default) way.

```
> res:=ModifiedAbramovPetkovsekReduction(T,k);
```

$$res := \left[\left[\frac{k}{k+1}, -\frac{1}{k+2}\right], k!\right]$$

Using the notations in Algorithm 3.17, we have

```
> f:=res[1][1]: r:=res[1][2]: H:=res[2]:
```

The package also provides the command VerifyMAPReduction to verify the reduction. This command is used according to the presented form of the result. In the default case, we say

```
> VerifyMAPReduction(res,T,k);
```

$$true$$

Moreover, we can change the outputs of ModifiedAbramovPetkovsekReduction by specifying the third argument. For example, we would like to display the result in terms of hypergeometric terms,

```
> res:=ModifiedAbramovPetkovsekReduction(T,k,output=
> hypergeometric);
> VerifyMAPReduction(res,T,k,output=hypergeometric);
```

$$res := \left[\frac{kk!}{k+1}, -\frac{k!}{k+2}\right]$$

$$true$$

or we can also perform it as a list of functions, which specifies the standard form of the residual forms.

```
> res:=ModifiedAbramovPetkovsekReduction(T,k,output=list);
> VerifyMAPReduction(res,T,k,output=list);
```

$$res := \left[\left[\left[-\frac{1}{k+1}, 0, 1\right], [-1, k+2, 0]\right], k!\right]$$

$$true$$

As mentioned in Section 3.3, we also implement a procedure based on the modified Abramov-Petkovšek reduction, which is only used to determine hypergeometric summability and performs in a similar way as Gosper's algorithm, namely the command IsSummable.

```
> IsSummable(T,k);
```

false

The built-in Maple command for Gosper's algorithm is Gosper in the package **SumTools[Hypergeometric]**.

```
> SumTools[Hypergeometric][Gosper](T,k);
```

Error, (in SumTools:-Hypergeometric:-Gosper) no solution found

Commands related to Chapter 4

In Chapter 4, we showed that the sum of two residual forms is congruent to a residual form (see Theorem 4.19), which plays an important role in developing the reduction-based creative telescoping algorithm for hypergeometric terms (i.e., Algorithm 5.6).

To prove Theorem 4.19, we introduced two congruences in Lemma 4.15. These two congruences stand for two types of kernel reduction in the shift case, that is, denominator type and numerator type, respectively. We implemented them by the command KernelReduction. To call it in Maple, using the notations from Lemma 4.15, one just says

```
> KernelReduction(p1,numer(K),denom(K),m,k,type=denominator);
```

or

```
> KernelReduction(p2,numer(K),denom(K),m,k,type=numerator);
```

The key idea of Algorithm 4.20 is to move the significant denominator of a residual form to a required form according to a given residual form, so that the resulting sum is again a residual form. This process was implemented as the command TranslateDRF. We also provide a command named SignificantDenom to extract the significant denominator of a residual form.

Now let's consider Example 4.11. For $K = 1/k$ shift-reduced, we have two residuals form w.r.t. K: $r = 1/(2k+1)$ and $s = 1/(2k+3)$.

```
> K:=1/k: r:=1/(2*k+1): s:=1/(2*k+3):
```

One can compute a residual form of $r + s$ in terms of the significant denominator of r by

```
> res:=TranslateDRF(s, SignificantDenom(r,K,k), K, k);
> S1:=res[1]: a:=res[2][1]: b:=res[2][2]: q:=res[2][3]:
> new:=r+normal(a/b)+q/denom(K); # evaluate the sum
> normal(K*subs(k=k+1, S1)-S1+new-r-s); # check the result
```

$$b := k + \frac{1}{2}$$

$$res := \left[-\frac{3}{2(2k+1)}, \left[-\frac{3}{4}, k + \frac{1}{2}, \frac{1}{2}\right]\right]$$

$$new := -\frac{1}{2(2k+1)} + \frac{1}{2k}$$

$$0$$

This confirms the result given in Example 4.11. Of course, one can also compute a residual form of $r + s$ in terms of the significant denominator of s,

```
> b:=SignificantDenom(s,K,k);
> res:=TranslateDRF(r, b, K, k);
> new:=s+normal(res[2][1]/b)+res[2][2]/denom(K);
> normal(K*subs(k=k+1, res[1])-res[1]+new-r-s);
```

$$res := \left[-\frac{1}{2k+1}, \left[-\frac{1}{3}, k + \frac{3}{2}, \frac{1}{3}\right]\right]$$

$$new := \frac{1}{3(2k+3)} + \frac{1}{3k}$$

$$0$$

Commands related to Chapter 5

Now let's turn our attention to bivariate hypergeometric terms. Consider the following hypergeometric term from Example 5.10.

```
> T:=binomial(n,k)^3;
```

$$T := \text{binomial}(n, k)^3$$

Based on the modified Abramov-Petkovšek reduction, Algorithm 5.6 is implemented in the command ReductionCT, which (by default) returns the (monic) minimal telescoper for a given hypergeometric term.

```
> ReductionCT(T,n,k,Sn);
```

$$-\frac{8(n^2 + 2n + 1)}{n^2 + 4n + 4} - \frac{(7n^2 + 21n + 16)Sn}{n^2 + 4n + 4} + Sn^2$$

As illustrated by the following commands, if a fifth argument is specified then the command also returns a corresponding certificate, whose form depends on the specification. More precisely, we get a certificate as a list of a normalized rational function and a hypergeometric term by saying

```
> res:=ReductionCT(T,n,k,Sn,output=normalized);
```

$$res := \left[-\frac{8(n^2+2n+1)}{n^2+4n+4} - \frac{(7n^2+21n+16)Sn}{n^2+4n+4} + Sn^2, \right.$$

$$\left[\frac{1}{(-n-1+k)^3(n^2+4n+4)(-n-2+k)^3} \Big(k^3(4k^3n^2 - 18k^2n^3 + 27kn^4 \right.$$

$$- 14n^5 + 8k^3n - 66k^2n^2 + 147kn^3 - 102n^4 + 4k^3 - 78k^2n + 291kn^2$$

$$\left. \left. - 290n^3 - 30k^2 + 249kn - 402n^2 + 78k - 272n - 72) \Big), \text{binomial}(n,k)^3 \right] \right]$$

or get one as a list of a linear combination of several simple rational functions and a hypergeometric term by

```
> res:=ReductionCT(T,n,k,Sn,output=unnormalized);
```

$$res := \left[-\frac{8(n^2+2n+1)}{n^2+4n+4} - \frac{(7n^2+21n+16)Sn}{n^2+4n+4} + Sn^2, \left[\frac{4(n^2+2n+1)}{n^2+4n+4} \right. \right.$$

$$- \frac{(7n^2+21n+16)(n^3+3n^2+3n+1)}{(n^2+4n+4)(-n-1+k)^3}$$

$$- \frac{(n+1)^3(6k^2+3kn+n^2+6k+4n+4)}{(n^2+4n+4)(-n-1+k)^3}$$

$$+ \frac{1}{(n^2+4n+4)(-n-1+k)^3}(12k^2n^3 - 12kn^4 + 11n^5 + 36k^2n^2$$

$$- 48kn^3 + 62n^4 + 36k^2n - 72kn^2 + 140n^3 + 12k^2 - 48kn$$

$$+ 158n^2 - 12k + 89n + 20)$$

$$- \frac{((n+1)^3 + 3(n+1)^2 + 3n + 4)(n+1)^3}{(-n-2-k)^3(-n-1+k)^3},$$

$$\left. \left. \text{binomial}(n,k)^3 \right] \right]$$

The result returned by the command ReductionCT can be verified by the command VerifyRCT.

```
> VerifyRCT(res,T,n,k,Sn);
```

$$true$$

Maple's implementation for Zeilberger's algorithm is the command Zeilberger, which is also in the package **SumTools[Hypergeometric]**.

```
> SumTools[Hypergeometric][Zeilberger](T,n,k,Sn);
```

$$\Bigg[(n^2+4n+4)\,Sn^2+(-7n^2-21n-16)\,Sn-8n^2-16n-8,$$

$$\frac{1}{(-n-2+k)^3(-n-1+k)^3}\left(\left(k^3+\left(-\frac{9}{2}n-\frac{15}{2}\right)k^2\right.\right.$$

$$\left.+\left(\frac{27}{4}n^2+\frac{93}{4}n+\frac{39}{2}\right)k-\frac{7}{2}n^3-\frac{37}{2}n^2-32n-18\right)$$

$$\left.k^3\,\mathrm{binomial}(n,k)^3(4n^2+8n+4)\right)\Bigg]$$

In view of Remark 5.9, we introduce the command ShiftMAPReduction, which performs the same function as applying ModifiedAbramovPetkovsekReduction with respect to k to the m-th shift $\sigma_n^m(T)$ for a bivariate hypergeometric term $T(n,k)$ but in a faster way as pointed out by the remark. Moreover, this command always uses the same kernel independent of the value of m. Note that when $m=0$ the command is the same as the command ModifiedAbramovPetkovsekReduction.

To illustrate this command, we consider the same hypergeometric term T as before.

```
> T:=binomial(n,k)^3:
```

Then it has a minimal additive decomposition

```
> ModifiedAbramovPetkovsekReduction(T,k);
```

$$\left[\left[-\frac{1}{2},\frac{1}{2}\,\frac{3k^2n-3kn^2+n^3+3k^2+3k+1}{(k+1)^3}\right],\mathrm{binomial}(n,k)^3\right]$$

For the first shift of T w.r.t. n, we have

```
> ModifiedAbramovPetkovsekReduction(subs(n=n+1,T),k);
```

$$\left[\left[-\frac{1}{2},\frac{1}{2}\,\frac{3k^2n-3kn^2+n^3+6k^2-6kn+3n^2+3n+2}{(k+1)^3}\right],\mathrm{binomial}(n+1,k)^3\right]$$

On the other hand, applying the command ShiftMAPReduction gives

```
> ShiftMAPReduction(T,n,k,1);
```

$$\left[\left[\frac{n^3+3n^2+3n+1}{(-n-1+k)^3}, \frac{n^3+3n^2+3n+1}{(k+1)^3}\right], \mathrm{binomial}(n,k)^3\right]$$

Commands related to Chapter 6

Combining the bounds given in Chapter 6, we implemented Algorithm 6.11 as the command BoundReductionCT. The function of this command is illustrated as follows.

Consider Example 6.12 with $\alpha = 5$.

```
> alpha:=5: T:=1/((n-alpha*k-alpha)*(n-alpha*k-2)!);
```

$$T := \frac{1}{(-5k+n-5)(-5k+n-2)!}$$

In Maple, the built-in command for the algorithm *LowerBound* [6] is also named LowerBound in the package **SumTools**[**Hypergeometric**]. With only three arguments, it returns a lower order bound of the telescopers for a given hypergeometric term,

```
> SumTools[Hypergeometric][LowerBound](T,n,k);
```

$$2$$

Moreover, by specifying a fourth and a fifth argument, the command also gives information about telescopers as well as certificates.

```
> SumTools[Hypergeometric][LowerBound](T,n,k,Sn,'Zpair');
> Zpair;
```

$$2$$

$$\left[Sn^5-1, \left[\frac{1}{\Gamma(n+4)\,(5k-n)}\left(\prod_{\_k=0}^{k-1}(-(5\_k-n-3)(5\_k-n-2)\right.\right.\right.$$
$$\left.\left.\left.(5\_k-n-1)(5\_k-n+1)(5\_k-n)\right)\right]\right]$$

In the same fashion, our implementation for Algorithm 6.11, namely the command BoundReductionCT, with three arguments specified returns an upper bound as well as a lower bound for the order of minimal telescopers for a given hypergeometric term.

```
> BoundReductionCT(T,n,k);
```

$$[5,\ 10]$$

In addition, depending on the numbers of specified arguments and the specifications, the command performs in the same manner as the command ReductionCT introduced above. To be precise, we have the following commands.

```
> BoundReductionCT(T,n,k,Sn);
```

$$Sn^5 - 1$$

```
> res:=BoundReductionCT(T,n,k,Sn,output=normalized):
```

$$res := \left[Sn^5 - 1,\ \left[\frac{5}{(5k-n)^2(5k-3-n)(5k-n-2)(5k-1-n)(5k-n+1)},\right.\right.$$
$$\left.\left. -\frac{1}{5(-5k+n-2)!}\right]\right]$$

```
> res:=BoundReductionCT(T,n,k,Sn,output=unnormalized);
```

$$res := \left[Sn^5 - 1,\ \left[\frac{5}{5k-n+5} - \frac{5}{5k-n+1} + \frac{20}{(5k-n+5)(5k-n+1)}\right.\right.$$
$$+ \frac{5}{(5k-n)^2(5k-3-n)(5k-n-2)(5k-1-n)(5k-n+1)},$$
$$\left.\left. -\frac{1}{5(-5k+n-2)!}\right]\right]$$

Appendix B

Comparison of Memory Requirements

In this section, we collect all comparisons of memory requirements between our new procedures from the **ShiftReductionCT** package (see Appendix A) and Maple's implementations of known algorithms. All memory requirements are obtained by the Maple command

```
>  kernelopts("bytesused");
```

and measured in bytes on a Linux computer with 388Gb RAM and twelve 2.80GHz Dual core processors. Recall that

- G: the procedure Gosper in **SumTools**[**Hypergeometric**], which is based on Gosper's algorithm;
- AP: the procedure SumDecomposition in **SumTools**[**Hypergeometric**], which is based on the Abramov-Petkovšek reduction;
- Z: the procedure **SumTools**[**Hypergeometric**][Zeilberger], which is based on Zeilberger's algorithm;
- S: the procedure IsSummable in **ShiftReductionCT**, which determines hypergeometric summability in a similar way as Gosper's algorithm;
- MAP: the procedure ModifiedAbramovPetkovsekReduction in **ShiftReductionCT**, which is based on the modified reduction.
- RCT_{tc}: the procedure ReductionCT in **ShiftReductionCT**, which computes a minimal telescoper and a corresponding normalized certificate;
- RCT_{t}: the procedure ReductionCT in **ShiftReductionCT**, which computes a minimal telescoper without constructing a certificate.
- BRCT_{tc}: the procedure BoundReductionCT in **ShiftReductionCT**, which computes a minimal telescoper and a corresponding normalized certificate;

- BRCT_t: the procedure BoundReductionCT in **ShiftReductionCT**, which computes a minimal telescoper without constructing a certificate.
- LB: the lower bound for telescopers given in Theorem 6.10.
- order: the order of the resulting minimal telescoper.

Tables for Example 3.23 and Example 3.24.

(λ, μ)	G	AP	S	MAP
$(0, 0)$	1.80015e7	2.79579e7	2.19643e7	2.20057e7
$(5, 5)$	6.92148e7	5.45788e8	8.31337e7	1.00876e8
$(10, 10)$	1.06237e8	1.74321e9	1.63963e8	2.23078e8
$(10, 20)$	3.67295e8	4.22563e9	3.41155e8	7.14421e8
$(10, 30)$	9.08446e8	2.06166e10	5.73637e8	2.07008e9
$(10, 40)$	1.79107e9	3.74146e10	8.60492e8	5.01724e9
$(10, 50)$	3.19600e9	4.98811e10	1.16624e9	9.80644e9

Table B.1: Memory comparison of Gosper's algorithm, the Abramov-Petkovšek reduction and the modified version for random hypergeometric terms (in bytes)

(λ, μ)	G	AP	S	MAP
$(0, 0)$	1.49566e8	3.83358e8	1.96563e8	1.97086e8
$(5, 5)$	2.76453e8	9.42523e8	2.40684e8	2.40927e8
$(10, 10)$	3.15859e8	1.86511e9	2.50334e8	2.50661e8
$(10, 20)$	6.81883e8	4.15802e9	3.19633e8	3.20250e8
$(10, 30)$	1.48580e9	7.60674e9	3.61856e8	3.60798e8
$(10, 40)$	2.66329e9	1.24394e10	3.81800e8	3.82879e8
$(10, 50)$	4.96349e9	2.22568e10	4.15063e8	4.14124e8

Table B.2: Memory comparison of Gosper's algorithm, the Abramov-Petkovšek reduction and the modified version for summable hypergeometric terms (in bytes)

Tables for Example 5.11.

$(d_1, d_2, \alpha, \lambda, \mu)$	Z	RCT_{tc}	RCT_t	order
$(1, 0, 1, 5, 5)$	2.05992e9	5.36111e8	1.58646e8	4
$(1, 0, 2, 5, 5)$	6.13485e9	3.33929e9	9.01651e8	6
$(1, 0, 3, 5, 5)$	2.05569e10	1.12736e10	2.59005e9	7
$(1, 8, 3, 5, 5)$	2.84955e10	1.46063e10	3.24756e9	7
$(2, 0, 1, 5, 10)$	3.58374e10	6.87524e9	6.90891e8	4
$(2, 0, 2, 5, 10)$	3.03599e10	4.30070e10	7.44379e9	6
$(2, 0, 3, 5, 10)$	6.95166e10	1.29853e11	2.56292e10	7
$(2, 3, 3, 5, 10)$	7.63196e10	1.34622e11	2.78371e10	7
$(2, 0, 1, 10, 15)$	1.72175e11	2.44536e10	1.52217e9	4
$(2, 0, 2, 10, 15)$	8.27362e10	1.38827e11	2.09677e10	6
$(2, 0, 3, 10, 15)$	1.79564e11	4.57813e11	1.04973e11	7
$(2, 5, 3, 10, 15)$	2.01763e11	4.49569e11	1.06872e11	7
$(3, 0, 1, 5, 10)$	7.48174e11	4.17901e10	5.18114e9	6
$(3, 0, 2, 5, 10)$	3.63162e11	2.25463e11	5.19205e10	8
$(3, 0, 3, 5, 10)$	7.60572e11	6.16676e11	1.78310e11	9

Table B.3: Memory comparison of Zeilberger's algorithm to reduction-based creative telescoping with and without construction of a certificate (in bytes)

Tables for Example 6.15 and Example 6.16.

α	RCT_t	RCT_{tc}	BRCT_t	BRCT_{tc}	LB	order
20	2.53275e8	2.57797e8	1.42371e8	1.46826e8	20	20
30	1.04691e9	1.05593e9	4.73815e8	4.83413e8	30	30
40	3.16905e9	3.18565e9	1.31468e9	1.33395e9	40	40
50	7.69274e9	7.71999e9	3.12029e9	3.15161e9	50	50
60	1.62442e10	1.62819e10	6.24941e9	6.28674e9	60	60
70	3.15561e10	3.16084e10	1.19886e10	1.20418e10	70	70

Table B.4: Memory comparison of two reduction-based creative telescoping with and without construction of a certificate for Example 6.15 (in bytes)

(m, α)	RCT_t	RCT_{tc}	BRCT_t	BRCT_{tc}	LB	order
(1,1)	2.64768e7	3.12387e7	2.64914e7	3.12548e7	1	2
(1,10)	9.91388e8	1.62603e9	8.94051e8	1.50416e9	10	11
(1,15)	2.01112e10	2.32990e10	1.33834e10	1.75427e10	15	16
(1,20)	2.23859e11	2.43209e11	1.13767e11	1.29430e11	20	21
(2,10)	1.03547e9	1.65297e9	9.12084e8	1.52683e9	10	11
(2,15)	2.70850e10	3.02579e10	1.38753e10	1.64594e10	15	16
(2,20)	2.37348e11	2.48004e11	1.29174e11	1.41685e11	20	21

Table B.5: Memory comparison of two reduction-based creative telescoping with and without construction of a certificate for Example 6.16 (in bytes)

Bibliography

[1] Sergei A. Abramov. On the summation of rational functions. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 11:1071–1075, 1971.

[2] Sergei A. Abramov. The rational component of the solution of a first order linear recurrence relation with rational right hand side. *Ž. Vyčisl. Mat. i Mat. Fiz.*, 15(4):1035–1039, 1090, 1975.

[3] Sergei A. Abramov. Indefinite sums of rational functions. In *ISSAC 1995—Proceedings of the 20th International Symposium on Symbolic and Algebraic Computation*, pages 303–308. ACM, New York, 1995.

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

[5] Sergei A. Abramov, Jacques J. Carette, Keith O. Geddes, and Ha Q. Le. Telescoping in the context of symbolic summation in Maple. *J. Symbolic Comput.*, 38(4):1303–1326, 2004.

[6] Sergei A. Abramov and Ha Q. Le. On the order of the recurrence produced by the method of creative telescoping. *Discrete Math.*, 298(1-3):2–17, 2005.

[7] Sergei A. Abramov and Marko Petkovšek. Minimal decomposition of indefinite hypergeometric sums. In *ISSAC 2001—Proceedings of the 26th International Symposium on Symbolic and Algebraic Computation*, pages 7–14 (electronic). ACM, New York, 2001.

[8] Sergei A. Abramov and Marko Petkovšek. Proof of a conjecture of Wilf and Zeilberger, 2001. Preprints Series of the Institute of Mathematics, Physics and Mechanics 39(748)(2001), Ljubljana, March 9, 2001.

[9] Sergei A. Abramov and Marko Petkovšek. On the structure of multivariate hypergeometric terms. *Adv. in Appl. Math.*, 29(3):386–411, 2002.

[10] Sergei A. Abramov and Marko Petkovšek. Rational normal forms and minimal decompositions of hypergeometric terms. *J. Symbolic Comput.*, 33(5):521–543, 2002. Computer algebra (London, ON, 2001).

[11] Gert Almkvist and Doron Zeilberger. The method of differentiating under the integral sign. *J. Symbolic Comput.*, 10(6):571–591, 1990.

[12] Moa Apagodu and Doron Zeilberger. Multi-variable Zeilberger and Almkvist-Zeilberger algorithms and the sharpening of Wilf-Zeilberger theory. *Adv. in Appl. Math.*, 37(2):139–152, 2006.

[13] David Bailey, Peter Borwein, and Simon Plouffe. On the rapid computation of various polylogarithmic constants. *Math. Comp.*, 66(218):903–913, 1997.

[14] Alin Bostan, Shaoshi Chen, Frédéric Chyzak, and Ziming Li. Complexity of creative telescoping for bivariate rational functions. In *ISSAC 2010—Proceedings of the 35th International Symposium on Symbolic and Algebraic Computation*, pages 203–210. ACM, New York, 2010.

[15] Alin Bostan, Shaoshi Chen, Frédéric Chyzak, Ziming Li, and Guoce Xin. Hermite reduction and creative telescoping for hyperexponential functions. In *ISSAC 2013—Proceedings of the 38th International Symposium on Symbolic and Algebraic Computation*, pages 77–84. ACM, New York, 2013.

[16] Alin Bostan, Pierre Lairez, and Bruno Salvy. Creative telescoping for rational functions using the Griffiths-Dwork method. In *ISSAC 2013—Proceedings of the 38th International Symposium on Symbolic and Algebraic Computation*, pages 93–100. ACM, New York, 2013.

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

[18] Shaoshi Chen. *Some applications of differential-difference algebra to creative telescoping.* PhD thesis, February 2011.

[19] Shaoshi Chen, Hui Huang, Manuel Kauers, and Ziming Li. A modified Abramov-Petkovšek reduction and creative telescoping for hypergeometric terms. In *ISSAC 2015—Proceedings of the 40th International Symposium on Symbolic and Algebraic Computation*, pages 117–124. ACM, New York, 2015.

[20] Shaoshi Chen and Manuel Kauers. Order-degree curves for hypergeometric creative telescoping. In *ISSAC 2012—Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation*, pages 122–129. ACM, New York, 2012.

[21] Shaoshi Chen and Manuel Kauers. Trading order for degree in creative telescoping. *J. Symbolic Comput.*, 47(8):968–995, 2012.

[22] Shaoshi Chen, Manuel Kauers, and Christoph Koutschan. A generalized Apagodu-Zeilberger algorithm. In *ISSAC 2014—Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, pages 107–114. ACM, New York, 2014.

[23] Shaoshi Chen, Manuel Kauers, and Christoph Koutschan. Reduction-based creative telescoping for algebraic functions. In *ISSAC 2016—Proceedings of the 41st International Symposium on Symbolic and Algebraic Computation*, page 175–182. ACM, New York, 2016.

[24] Shaoshi Chen, Manuel Kauers, and Michael F. Singer. Telescopers for rational and algebraic functions via residues. In *ISSAC 2012—Proceedings of the 37th International Symposium on Symbolic and Algebraic Computation*, pages 130–137. ACM, New York, 2012.

[25] Shaoshi Chen, Manuel Kauers, and Michael F. Singer. Desingularization of Ore operators. *J. Symbolic Comput.*, 74:617–626, 2016.

[26] David V. Chudnovsky and Gregory V. Chudnovsky. Computer algebra in the service of mathematical physics and number theory. In *Computers in mathematics (Stanford, CA, 1986)*, volume 125 of *Lecture Notes in Pure and Appl. Math.*, pages 109–232. Dekker, New York, 1990.

[27] Frédéric Chyzak and Bruno Salvy. Non-commutative elimination in Ore algebras proves multivariate identities. *J. Symbolic Comput.*, 26(2):187–227, 1998.

[28] Richard M. Cohn. *Difference algebra*. Interscience Publishers John Wiley & Sons, New York-London-Sydeny, 1965.

[29] Mary Celine Fasenmyer. Some generalized hypergeometric polynomials. *Bull. Amer. Math. Soc.*, 53:806–812, 1947.

[30] Stéphane Fischler and Tanguy Rivoal. On the values of G-functions. *Comment. Math. Helv.*, 89(2):313–341, 2014.

[31] Stéphane Fischler and Tanguy Rivoal. Arithmetic theory of E-operators. *J. Éc. polytech. Math.*, 3:31–65, 2016.

[32] Philippe Flajolet, Stefan Gerhold, and Bruno Salvy. On the non-holonomic character of logarithms, powers, and the nth prime function. *Electron. J. Combin.*, 11(2):Article 2, 16 pp. (electronic), 2004/06.

[33] Philippe Flajolet and Andrew Odlyzko. Singularity analysis of generating functions. *SIAM J. Discrete Math.*, 3(2):216–240, 1990.

[34] Philippe Flajolet and Robert Sedgewick. *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.

[35] Philippe Flajolet and Brigitte Vallée. Continued fractions, comparison algorithms, and fine structure constants. In *Constructive, experimental, and nonlinear analysis (Limoges, 1999)*, volume 27 of *CMS Conf. Proc.*, pages 53–82. Amer. Math. Soc., Providence, RI, 2000.

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

[37] William A. Harris, Jr. and Yasutaka Sibuya. The reciprocals of solutions of linear ordinary differential equations. *Adv. in Math.*, 58(2):119–132, 1985.

[38] Hui Huang. New bounds for hypergeometric creative telescoping. In *ISSAC 2016—Proceedings of the 41st International Symposium on Symbolic and Algebraic Computation*, pages 279–286. ACM, New York, 2016.

[39] Hui Huang and Manuel Kauers. D-finite numbers, 2016. in preparation.

[40] Edward L. Ince. *Ordinary Differential Equations*. Dover Publications, New York, 1944.

[41] Manuel Kauers and Peter Paule. *The Concrete Tetrahedron*. Texts and Monographs in Symbolic Computation. SpringerWienNewYork, Vienna, 2011. Symbolic sums, recurrence equations, generating functions, asymptotic estimates.

[42] Manuel Kauers and Gleb Pogudin. in preparation, 2016.

[43] Maxim Kontsevich and Don Zagier. Periods. In *Mathematics unlimited—2001 and beyond*, pages 771–808. Springer, Berlin, 2001.

[44] Ha Q. Le. *Algorithms for the construction of the minimal telescopers*. PhD thesis, University of Waterloo, Waterloo, Ont., Canada, 2003. AAINQ83003.

[45] Leonard Lipshitz. *D*-finite power series. *J. Algebra*, 122(2):353–373, 1989.

[46] Christian Mallinger. Algorithmic manipulations and transformations of univariate holonomic functions and sequences. Master's thesis, RISC, J. Kepler University, August 1996.

[47] Marc Mezzarobba. NumGfun: a package for numerical and analytic computation and D-finite functions. In *ISSAC 2010—Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation*, pages 139–146. ACM, New York, 2010.

[48] Marc Mezzarobba and Bruno Salvy. Effective bounds for P-recursive sequences. *J. Symbolic Comput.*, 45(10):1075–1096, 2010.

[49] Mohamud Mohammed and Doron Zeilberger. Sharp upper bounds for the orders of the recurrences output by the Zeilberger and q-Zeilberger algorithms. *J. Symbolic Comput.*, 39(2):201–207, 2005.

[50] Oystein Ore. Sur la forme des fonctions hypergéométriques de plusieurs variables. *J. Math. Pures Appl. (9)*, 9(4):311–326, 1930.

[51] Peter Paule. Greatest factorial factorization and symbolic summation. *J. Symbolic Comput.*, 20(3):235–268, 1995.

[52] Peter Paule and Markus Schorn. A Mathematica version of Zeilberger's algorithm for proving binomial coefficient identities. *J. Symbolic Comput.*, 20(5-6):673–698, 1995. Symbolic computation in combinatorics Δ_1 (Ithaca, NY, 1993).

[53] Marko Petkovšek. Hypergeometric solutions of linear recurrences with polynomial coefficients. *J. Symbolic Comput.*, 14(2-3):243–264, 1992.

[54] Marko Petkovšek, Herbert S. Wilf, and Doron Zeilberger. $A = B$. A K Peters, Ltd., Wellesley, MA, 1996. With a foreword by Donald E. Knuth, With a separately available computer disk.

[55] Roberto Pirastu and Volker Strehl. Rational summation and Gosper-Petkovšek representation. *J. Symbolic Comput.*, 20(5-6):617–635, 1995. Symbolic computation in combinatorics Δ_1 (Ithaca, NY, 1993).

[56] John Riordan. *Combinatorial identities*. Robert E. Krieger Publishing Co., Huntington, N.Y., 1979. Reprint of the 1968 original.

[57] Bruno Salvy and Paul Zimmermann. Gfun: a maple package for the manipulation of generating and holonomic functions in one variable. *ACM Transactions on Mathematical Software*, 20(2):163–177, 1994.

[58] Carl L. Siegel. Über einige Anwendungen diophantischer Approximationen [reprint of Abhandlungen der Preußischen Akademie der Wissenschaften. Physikalisch-mathematische Klasse 1929, Nr. 1]. In *On some applications of Diophantine approximations*, volume 2 of *Quad./Monogr.*, pages 81–138. Ed. Norm., Pisa, 2014.

[59] Richard P. Stanley. Differentiably finite power series. *European J. Combin.*, 1(2):175–188, 1980.

[60] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*, volume 62 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. With a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin.

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

[62] Joris van der Hoeven. Fast evaluation of holonomic functions. *Theoret. Comput. Sci.*, 210(1):199–215, 1999.

[63] Joris van der Hoeven. Fast evaluation of holonomic functions near and in regular singularities. *J. Symbolic Comput.*, 31(6):717–743, 2001.

[64] Joris van der Hoeven. Efficient accelero-summation of holonomic functions. *J. Symbolic Comput.*, 42(4):389–428, 2007.

[65] Herbert S. Wilf and Doron Zeilberger. Rational functions certify combinatorial identities. *J. Amer. Math. Soc.*, 3(1):147–158, 1990.

[66] Herbert S. Wilf and Doron Zeilberger. Towards computerized proofs of identities. *Bull. Amer. Math. Soc. (N.S.)*, 23(1):77–83, 1990.

[67] Herbert S. Wilf and Doron Zeilberger. An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities. *Invent. Math.*, 108(3):575–633, 1992.

[68] Herbert S. Wilf and Doron Zeilberger. Rational function certification of multisum/integral/"q" identities. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):148–153, 1992.

[69] Doron Zeilberger. A fast algorithm for proving terminating hypergeometric identities. *Discrete Math.*, 80(2):207–211, 1990.

[70] Doron Zeilberger. A holonomic systems approach to special functions identities. *J. Comput. Appl. Math.*, 32(3):321–368, 1990.

[71] Doron Zeilberger. The method of creative telescoping. *J. Symbolic Comput.*, 11(3):195–204, 1991.

Notation

The following list describes the most important mathematical notations that have been used in this thesis. For each group, the order follows roughly the order of first appearance in the text.

Abbreviations

gcd	The greatest common divisor
min	The minimum
max	The maximum
$p \mid q$	A polynomial p divides a polynomial q over the domain where the polynomials live
$p \nmid q$	A polynomial p does not divide a polynomial q over the domain where the polynomials live
log	The natural logarithm
exp	The exponential function

Number Sets

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$	Sets of natural, integer, rational, real, complex numbers
$\mathbb{Q}(i)$	The Gaussian rational field
$\mathcal{D}_{R,\mathbb{F}}$	The set of D-finite numbers with respect to R and $\mathbb{F}$
$\mathcal{D}_{\mathbb{F}}$	The set $\mathcal{D}_{\mathbb{F},\mathbb{F}}$
$\mathcal{A}_{\mathbb{F}}$	The set of limits of convergent algebraic sequences over $\mathbb{F}$
$\emptyset$	The empty set
$\mathcal{C}_{\mathbb{F}}$	The set of limits of convergent C-finite sequences over $\mathbb{F}$

Operators

σ_k	The shift operator w.r.t. k which maps $r(k)$ to $r(k+1)$ for every rational function $r \in \mathbb{F}(k)$
Δ_k	The difference of σ_k and the identity map
S_n	The operator in the ring of linear recurrence operators over $\mathbb{F}$ which satisfies $S_n r = \sigma_n(r) S_n$ for all $r \in \mathbb{F}$.
$\sum_{j=0}^{\rho} p_j S_n^j$	A recurrence operator with polynomial coefficients p_j
D_z	The derivation operator w.r.t. z which maps a power series or function $f(z)$ to its derivative $f'(z) = \frac{d}{dz} f(z)$
$\sum_{j=0}^{\rho} p_j D_z^j$	A differential operator with polynomial coefficients p_j

Rings/Fields/Algebra

$\mathrm{Quot}(R)$	The quotient field of the ring R
$\mathbb{K}$	A field of characteristic zero
$\mathbb{F}$	A field of characteristic zero, or the field $\mathbb{K}(n)$ (Chapter 5), or a subfield of $\mathbb{C}$ (Chapter 8)
$\mathbb{F}(k)$	The field of univariate rational functions in k over $\mathbb{F}$
$\mathbb{F}[k]$	The ring of univariate polynomials in k over $\mathbb{F}$
$\mathbb{D}$	A difference ring extension of $\mathbb{F}(k)$
$\mathbb{K}(n,k)$	The field of bivariate rational functions in n, k over $\mathbb{K}$
$\mathbb{F}[n]\langle S_n \rangle$	The Ore algebra of linear recurrence operators with polynomial coefficients w.r.t. n
$\mathbb{K}[n,k]$	The ring of bivariate polynomials in n, k over $\mathbb{K}$
$\mathbb{K}(n)[k]$	The ring of polynomials in k over the field $\mathbb{K}(n)$
R	A subring of $\mathbb{C}$
$R[[z]]$	The ring of formal power series over R
$R^{\mathbb{N}}$	The ring of all sequences from $\mathbb{N}$ to R
$\mathbb{F}[z]\langle D_z \rangle$	The Ore algebra of linear differential operators with polynomial coefficients wr.t. z
$\bar{\mathbb{F}}$	The algebraic closure of the field $\mathbb{F}$

Other Symbols

$\sum_{j=a}^{b} f(k)$	The sum $f(a) + f(a+1) + \cdots + f(b)$
$\deg_k(p)$	Degree of a polynomial p w.r.t. k
$\mathrm{lc}_k(p)$	Leading coefficient of a polynomial p w.r.t. k
$A \setminus B$	The relative complement of a set B with respect to a set A
$k!$, $\binom{n}{k}$	Factorial $k! = 1 \cdot 2 \cdot 3 \ldots (k-1) \cdot k$ and binomial coefficient $\binom{n}{k} = n(n-1)\ldots(n-k+1)/k!$
$\mathbb{U}_T$	The union of $\{0\}$ and the set of summable hypergeometric terms that are similar to a hypergeometric term T
$\mathbb{V}_K$	The set $\{K\sigma_k(r) - r \mid r \in \mathbb{F}(k)\}$ where K is a shift-reduced rational function in $\mathbb{F}(k)$
$A \equiv_k B \mod C_k$	The expression $A - B$ belongs to a set C_k
ϕ_K	The map for polynomial reduction with respect to a shift-reduced rational function K
$\mathrm{im}(\phi_K)$	The image space of the map ϕ_K
$\mathbb{W}_K$	The standard complement of $\mathrm{im}(\phi_K)$
$A \oplus B$	The direct sum of two vector spaces A and B
$A \cap B$	The intersection of two sets A and B
$A \cup B$	The union of two sets A and B
$\lvert\mathcal{P}\rvert$	The number of elements of the set $\mathcal{P}$
$[\![\varphi]\!]$	The Iversion bracket, namely $[\![\varphi]\!]$ equals 1 if the expression φ is true, otherwise it is 0.
$\prod_{j=a}^{b} f(k)$	The product $f(a)f(a+1)\ldots f(b)$
$p \sim_k q$	A polynomial p is shift-equivalent to a polynomial q w.r.t. k
$p \approx_k q$	A shift-free polynomial p is shift-related to a shift-free polynomial q w.r.t. k
$L(T)$	The application of a recurrence operator L to a hypergeometric term T

$p \sim_{n,k} q$	A polynomial p is shift-equivalent to a polynomial q w.r.t. n and k
$\lvert\xi\rvert$	The modulus of a complex number ξ
$\dim_{\mathbb{K}(n)}(\mathbb{W}_K)$	The dimension of the vector space $\mathbb{W}_K$ over the field $\mathbb{K}(n)$
$\delta^{(\lambda,\mu)}$	The operator $\sigma_n^\alpha \sigma_k^\beta$ where λ, μ are coprime integers and $\alpha\lambda + \beta\mu = 1$ with $\lvert\alpha\rvert < \lvert\mu\rvert$ and $\lvert\beta\rvert < \lvert\lambda\rvert$
$\sum_{n=0}^{\infty} a_n z^n$	A power series with the coefficient sequence $(a_n)_{n=0}^{\infty}$
$(a_n)_{n=0}^{\infty}$	An infinite sequence $a_0, a_1, a_2, \ldots$
$f'(z)$	The first derivative of a power series or function $f(z)$ w.r.t. z
$\mathrm{lc}(L)$	The leading coefficient of an operator L
$L \cdot a_n$	The application of a recurrence operator L to an infinite sequence $(a_n)_{n=0}^{\infty}$
$L \cdot f(z)$	The application of a differential operator L to a power series f
$f \circ g$	The composition $f(g)$ of functions f and g
$A \subseteq B$	A set A is contained by a set B
$\bar{\xi}$	The complex conjugation of a complex number ξ
$\mathrm{Re}(\xi)$	The real part of a complex number ξ
$\mathrm{Im}(\xi)$	The imaginary part of a complex number ξ
$a_n \sim b_n \ (n \to \infty)$	The quotient a_n/b_n converges to 1 as $n \to \infty$
$f(z) \sim g(z) \ (z \to \zeta)$	The quotient $f(z)/g(z)$ converges to 1 as z approaches ζ
$[z^n]f(z)$	The coefficient of z^n in a power series $f(z) \in \mathbb{F}[[z]]$
$f^{(k)}(z)$	The kth derivative of a power series or function $f(z)$ w.r.t. z

Index

Curriculum Vitae

Personal data

Name	Hui HUANG
Gender	Female
Date of birth	August 28, 1989
Place of birth	Fujian province, China
Nationality	People's Republic of China

Contact

Address	Institute for Algebra Johannes Kepler University Linz Altenberger Straße 69 4040 Linz, Austria
Email	huanghui@amss.ac.cn

Education

2013–2017	Doctorate studies in symbolic computation at the FWF Doctoral Program "Computational Mathematics" (Johannes Kepler University, Linz, Austria) and Key Laboratory of Mathematics Mechanization (University of Chinese Academy of Sciences, Beijing, China) Co-supervisors: Manuel Kauers and Ziming Li
2011–2013	Studies in symbolic computation at Key Laboratory of Mathematics Mechanization, University of Chinese Academy of Sciences, Beijing, China Supervisor: Ziming Li
2007–2011	Studies in applied mathematics at Xiamen University, Fujian, China

Awards

2016	Distinguished Female Student Award at ISSAC 2016
2014	Distinguished Poster Award at ISSAC 2014 (together with S. Chen and Z. Li)

Scientific Work

Refereed Publications

1. Hui Huang. New bounds for hypergeometric creative telescoping. In *ISSAC 2016—Proceedings of the 41st International Symposium on Symbolic and Algebraic Computation*, pages 279–286. ACM, New York, 2016.
2. Shaoshi Chen, Hui Huang, Manuel Kauers, and Ziming Li. A modified Abramov-Petkovšek reduction and creative telescoping for hypergeometric terms. In *ISSAC 2015—Proceedings of the 40th International Symposium on Symbolic and Algebraic Computation*, pages 117–124. ACM, New York, 2015.

Posters

- Shaoshi Chen, Hui Huang, and Ziming Li. Improved Abramov-Petkovšek's Reduction and Creative Telescoping for Hypergeometric Terms (Poster at ISSAC 2014). In *ACM Commun. Comput. Algebra*, 48(3/4):106-108. ACM, New York, 2014.

Talks

July 2016	*New Bounds for Hypergeometric Creative Telescoping.* ISSAC 2016, Waterloo, Canada.
June 2016	*Reduction and Creative Telescoping for Hypergeometric Terms.* Center for Combinatorics Seminar, Tianjin, China.
November 2015	*Two Applications of the Modified Abramov-Petkovšek Reduction.* CM 2015, Hefei, China.
July 2015	*A Modified Abramov-Petkovšek Reduction and Creative Telescoping for Hypergeometric Terms.* ISSAC 2015, Bath, United Kingdom.
June 2015	*Abramov-Petkovšek Reduction and Creative Telescoping for Rational Functions.* CanaDAM 2015, Saskatoon, Canada.
August 2013	*An Improved Abramov-Petkovšek Reduction for Hypergeometric Terms.* CM 2013, Changchun, China.